

UNIVERSITÀ CA' FOSCARI - VENEZIA



FACOLTÀ DI LETTERE E FILOSOFIA

CORSO DI LAUREA SPECIALISTICA IN ARCHIVISTICA E BIBLIOTECONOMIA

La bibliometria: metodi e strumenti per la misurazione della ricerca scientifica e delle raccolte bibliotecarie

LAUREANDA
TAMARA ANZILIERO
MATRICOLA 820314

RELATORE
PROFESSOR RICCARDO RIDI

ANNO ACCADEMICO 2011-2012

Indice

Indice.....	ii
Introduzione.....	v
1.La bibliometria.....	1
1.1.Definizione della disciplina.....	1
1.2.Lineamenti storici: dalla bibliografia statistica alla bibliometria moderna.....	7
1.3.Basi matematiche	14
1.3.1.Legge di Lotka.....	15
1.3.2.Legge di Bradford.....	16
1.3.3.Legge di Zipf.....	18
1.4.Fondamenti teorici.....	19
1.4.1.John Bernal e la scienza come sistema materiale di relazioni sociali e simboliche	20
1.4.2.Robert Merton e la concezione normativa della scienza.....	21
1.4.3.Derek de Solla Price e la scienza come sistema "termodinamico"	24
1.4.4.Henry Small e le citazioni bibliografiche come simboli di concetti.....	28
1.5.Applicazioni.....	30
1.5.1.La cartografia della scienza.....	30
1.5.2.La valutazione della ricerca	33
1.5.3.La bibliometria come metrica del Web.....	39
2. L'impiego di metodi bibliometrici nella valutazione e nella gestione delle collezioni	
bibliotecarie.....	42
2.1.Leggi e modelli bibliometrici per la valutazione e gestione della collezione.....	42
2.1.1.La distribuzione statistica per la determinazione di un nucleo di documenti.....	42
2.1.2.L'obsolescenza della letteratura.....	50
2.2.L'analisi citazionale come strumento per la gestione e la valutazione della collezione.....	56
2.3.La bibliometria per la gestione e valutazione della biblioteca digitale.....	61
3.L'analisi citazionale: il più diffuso dei metodi bibliometrici.....	66
3.1.La citazione bibliografica.....	67
3.1.1.Il significato.....	67
3.1.2.Il comportamento citazionale.....	68
3.1.3.Il ruolo della citazione.....	72
3.2.I primi lavori di analisi delle citazioni.....	74

3.3.Le tecniche citazionali.....	76
3.3.1.L'analisi delle co-citazioni (cluster o co-citation analysis).....	77
3.3.2.L'analisi degli accoppiamenti bibliografici (bibliographic coupling analysis).....	78
4.Gli indici di citazioni.....	79
4.1.Citation indexes: dai manoscritti ebraici (XII sec.) allo Science Citation Index (SCI).....	80
4.2.Le potenzialità degli indici di citazioni nel campo delle strategie di information retrieval...	88
4.3.Indici di citazioni multidisciplinari online.....	93
4.3.1.ISI Web of Science.....	93
4.3.2.Scopus.....	105
4.3.3.Google Scholar.....	110
4.4.Altri indici citazionali specializzati in specifici settori disciplinari.....	115
4.5.Considerazioni sulla copertura degli indici di citazioni.....	117
5.Gli indicatori bibliometrici.....	121
5.1.Indicatori primari.....	122
5.1.1.Il conteggio del numero di pubblicazioni.....	123
5.1.2.Il conteggio del numero di citazioni.....	125
5.2.Indicatori secondari o derivati.....	127
5.2.1.L'Impact Factor (IF).....	127
5.2.2.Altri indicatori contenuti nel Journal Citation Reports.....	131
5.2.3.H-Index e indici derivati.....	132
5.2.4.Eigenfactor (EF) e Article Influence Score (AIS).....	139
5.2.5.SCIImago Journal Rank (SJR) e Source Normalised Impact per Paper (SNIP).....	140
6.Le metriche del Web.....	142
6.1.Dall'analisi citazionale all'analisi del link (PageRank).....	142
6.2.Web Impact Factor (WIF).....	143
6.3.Indicatori d'uso non basati sulle citazioni.....	147
6.3.1.Progetti internazionali per la definizione di standard comuni e di metodi condivisi di raccolta e analisi dei dati sull'utilizzo.....	147
6.3.1.1. COUNTER.....	147
6.3.1.2. MESUR.....	149
6.3.1.3. SUSHI.....	151
6.3.2.Usage Factor (UF).....	152
6.4.Aggregazione di metriche a livello dell'articolo (Article Level Metrics): il caso PLoS.....	154
7.Le nuove sfide per le biblioteche e i bibliotecari.....	156

8.Caso di studio: antropologia a Ca' Foscari.....	162
8.1.Introduzione.....	162
8.2.Gli strumenti utilizzati per la rilevazione dei dati bibliometrici.....	163
8.3.Metodologia d'analisi.....	164
8.4.I risultati.....	166
8.5.Conclusioni.....	179
Conclusioni.....	181
Bibliografia.....	183

Introduzione

La crescente importanza della bibliometria nella società attuale dovuta sia allo sviluppo delle moderne tecnologie dell'informazione sia ai significativi cambiamenti che stanno interessando la politica e l'organizzazione della ricerca scientifica, oltre a rappresentare certamente una grande opportunità, pone nuove sfide per le biblioteche specie sotto il profilo del ruolo da loro svolto all'interno del processo di comunicazione scientifica che risulta notevolmente ridefinito e ampliato.

L'approccio di tipo quantitativo all'analisi valutativa offerto dalla bibliometria, che rappresenta in qualche modo garanzia di oggettività dei risultati, ha fatto di questa metodologia, basata sul conteggio delle pubblicazioni e degli elementi estrapolati dalle unità bibliografiche, uno strumento efficace e in continua espansione nell'ambito sia della valutazione della ricerca scientifica sia in quello delle politiche documentarie in biblioteca. I dati provenienti dall'analisi bibliometrica sono utili infatti per tutte le operazioni che riguardano non solo la politica di selezione, di acquisizione, di accesso e di uso delle collezioni bibliografiche, ma anche per quelle di scarto e di conservazione del materiale. Le informazioni di queste analisi sono una base scientifica e oggettiva che aiuta i responsabili della biblioteca a pianificare, ottimizzare e razionalizzare le risorse strumentali, finanziarie e umane. Inoltre, con l'estensione dei metodi bibliometrici allo studio della comunicazione mediata dalla rete di *hyperlink* che popolano il World Wide Web, si è iniziato ad applicare l'analisi citazionale, il più noto dei metodi bibliometrici, al contesto delle biblioteche digitali allo scopo di determinare impatto e *ranking* di documenti e riviste ad uso dell'utenza della biblioteca.

Per le biblioteche accademiche e scientifiche si profila un impiego della bibliometria che va ben oltre gli ambiti tradizionali d'applicazione già ampiamente collaudati e consolidati, ed è sostanzialmente finalizzato all'elaborazione di dati d'alta qualità a sostegno della valutazione della ricerca svolta all'interno di università e istituti di ricerca, ma che può arrivare fino a concretizzarsi in un vero e proprio nuovo servizio di "*reference service* di tipo bibliometrico" a sostegno in questo caso del bisogno informativo di tipo bibliometrico dell'utenza composta soprattutto da ricercatori ma estendibile anche a studenti e dottorandi.

Si è scelto di utilizzare per questo lavoro una metodologia d'indagine capace di far emergere luci e ombre della bibliometria, dei metodi e degli strumenti bibliometrici; nel concreto, tale impostazione ha significato mettere in evidenza sia i punti di forza sia i punti di criticità già a partire dall'unità informativa di base dell'analisi citazionale: la citazione bibliografica.

La difficoltà principale nello scrivere questa tesi è stata sicuramente quella di saper guardare oltre i facili entusiasmi che accompagnano questa metodologia di valutazione, che la presentano quasi come una sorta di panacea capace di risolvere ogni problema di tipo valutativo, e di considerare invece la disciplina per quello che effettivamente è, cioè un metodo di valutazione che rispetto alla tradizionale revisione dei pari, è certamente più veloce, più economico e per certi aspetti più obiettivo, ma procede inevitabilmente, e non può essere altrimenti, ad una semplificazione di una realtà che invece è molto più complessa.

La tesi si apre con un ampio capitolo completamente dedicato alla definizione della disciplina anche in rapporto con quelle metrologie che pur presentando delle differenze metodologiche si distinguono per avere finalità e ambiti di ricerca contigui ma diversi rispetto alla bibliometria; all'evoluzione storica della disciplina, alle basi matematiche, ai fondamenti teorici e alle principali applicazioni della bibliometria che vanno dalla costruzione di mappe dettagliate delle specificità scientifiche (cartografia della scienza) all'analisi della tipologia di connessioni ipertestuali e dei modelli di utilizzo delle risorse della rete. Ampio spazio viene dato all'applicazione più importante ai fini di questo lavoro, vale a dire all'impiego dei metodi bibliometrici nella valutazione e nella gestione delle collezioni bibliotecarie sia cartacee che digitali. In tale contesto particolare importanza è stata data ai due concetti bibliometrici considerati fondamentali per efficacia valutativa e sui quali è stato possibile ricavare modelli generali di analisi: la distribuzione statistica e l'obsolescenza delle pubblicazioni.

La trattazione si è fatta più approfondita a partire dal capitolo 3 in cui viene presentato il metodo dell'analisi citazionale che rappresenta la base anche dei due capitoli successivi. Il metodo dell'analisi citazionale è stato al centro di un processo di perfezionamento che, pur mantenendone inalterata la metodologia, ha dilatato l'unità informativa su cui poggia

il metodo stesso. In altre parole, il metodo non si avvale solo ed esclusivamente della citazione diretta, vale a dire della citazione di un documento precedente da parte di uno successivo, ma nel tempo ha sviluppato delle tecniche più raffinate ed efficaci, capaci cioè, spingendo l'analisi ancora più in profondità, di far emergere e misurare un numero sempre maggiore di nessi significativi e rappresentativi del grado di affinità tra i documenti. In questo capitolo verranno analizzate proprio queste tecniche vale a dire l'analisi delle co-citazioni (*co-citation analysis*) e l'analisi degli accoppiamenti bibliografici (*bibliographic coupling analysis*), non prima però di avere esaminato il significato e il ruolo di quello che rappresenta il fulcro del metodo, nonché il fenomeno di base su cui sono costruiti gli indicatori bibliometrici più utilizzati: la citazione bibliografica.

Il capitolo successivo è dedicato invece a quello speciale repertorio bibliografico rappresentato dall'indice di citazioni, a cui la bibliometria deve parte del suo sviluppo e della sua fortuna, a quello strumento cioè che elenca tutti i riferimenti bibliografici contenuti in un *corpus* selezionato di documenti e abbina ad ogni singola pubblicazione citata il documento citante. Il capitolo si apre con una breve panoramica storica dello strumento bibliografico allo scopo di far emergere le sue ampie finalità applicative che nel corso dei secoli sono passate dall'ambito giuridico-religioso a quello biblioteconomico, del recupero e dell'organizzazione dell'informazione, storico e valutativo della ricerca. Con Garfield e il suo *SCI*, l'indice di citazioni diventerà quello strumento moderno conosciuto attualmente, dalle grandi potenzialità bibliometriche e rivoluzionario come sistema di *information retrieval*, che l'avvento del Web ha saputo ulteriormente sviluppare e che l'integrazione nella piattaforma *Web of Knowledge* ha trasformato in quel database multidisciplinare chiamato *WoS*, che rappresenta il più frequente punto di partenza per l'analisi citazionale. La fortuna di *WoS* ha inoltre contribuito alla nascita di altri database citazionali multidisciplinari come *Scopus*, di proprietà dell'editore *Elsevier*, e *Google Scholar*, quest'ultimo ancora più innovativo rispetto ai precedenti perché più alto è il suo livello di integrazione bibliografica anche con risorse non disponibili in formato digitale, ma soprattutto perché capace di sfruttare le enormi opportunità offerte dal mondo *open access*. In questo capitolo i tre database multidisciplinari *WoS*, *Scopus* e *Google Scholar* sono descritti in modo analitico, mentre ricevono una trattazione più sommaria quei database che, pur essendo identici per funzionalità e

struttura ai precedenti, sono specializzati in specifici ambiti disciplinari. L'ultimo paragrafo del capito sarà invece dedicato ai criteri di selezione dei materiali inseriti negli indici di citazioni, un aspetto questo per niente trascurabile poiché oltre a modificare direttamente il grado di copertura dello strumento, influenza indirettamente, fino anche ad arrivare a falsare, i valori degli indicatori bibliometrici, che proprio per il fatto di essere calcolati sulla base dei dati estratti da questi repertori citazionali, risentono delle scelte di inclusione o esclusione dei materiali da indicizzare effettuate dai curatori.

Nel capitolo 5 vengono trattati nel dettaglio i più noti e più diffusi indicatori bibliometrici di tipo citazionale suddivisi in primari (numero di pubblicazioni e di citazioni) e secondari o derivati (*IF*, *H-Index*, *Eigenfactor*, *Article Influence Score*, *SJR*, *SNIP*), di cui verranno discussi i punti di forza e le criticità. Prima però di procedere a tale discussione sono necessarie alcune considerazioni che riguardano indistintamente qualsiasi tipo di indicatore bibliometrico e che permettono di inquadrare tali strumenti nell'ottica di un contesto più ampio di valutazione al fine di evitare inutili semplificazioni, o peggio, di cedere alla tentazione di assolutizzare i risultati.

Nel capitolo 6 sono invece analizzate le innovative metriche basate sull'analisi dei *link* e sul *download* dei documenti elettronici, frutto dell'applicazione di gran parte del bagaglio teorico e pratico accumulato in molti decenni dalla bibliometria allo studio della struttura del Web. In questo capitolo sono inoltre trattati i progetti internazionali per la definizione di standard comuni e metodi condivisi di raccolta e analisi dei dati d'uso (*COUNTER*, *ME-SUR* e *SUSHI*), iniziative cioè accomunate da un'esigenza di normalizzazione che la crescita esponenziale della letteratura scientifica online ha reso una stringente necessità.

Il capitolo 7 è incentrato su alcune considerazioni riguardanti le sfide, ma anche le opportunità, che l'impiego della bibliometria pone a biblioteche e a bibliotecari specie di università e centri di ricerca nel più vasto processo di comunicazione e valutazione delle ricerca scientifica.

La tesi si conclude con un'esercitazione che ha avuto come obiettivo principale quello di dimostrare concretamente come possono essere applicati ad uno specifico *set* di dati bibliografici i metodi e gli strumenti bibliometrici più importanti analizzati nel corso della tesi. Sono stati utilizzati gli indicatori noti come primari (numero di pubblicazioni e di citazio-

ni) per tratteggiare la fisionomia generale (la proporzione tra libri e riviste, la suddivisione in lingue, l'età, la percentuale di presenza nei vari indici di citazioni analizzati) e il quadro citazionale del *set* di dati di riferimento, mentre sono stati impiegati gli indicatori più evoluti, quelli noti come secondari, ad un *set* di dati più ristretto (ricavato effettuando un *ranking* su base citazionale) composto da autori e riviste allo scopo di dimostrare soprattutto le discrepanze nei risultati forniti dagli indici di citazioni utilizzati (*Web of Science*, *Scopus*, *Google Scholar*) dovute essenzialmente al diverso grado di copertura di tali strumenti. Il capitolo si chiude con la rilevazione del numero, e la relativa percentuale, di documenti recuperati nel catalogo generale di Ca' Foscari in modo tale da non fare dell'esercitazione un lavoro di analisi relegato ad un piano astratto o un puro esercizio fine a se stesso bensì di ancorarla al contesto ben concreto di una specifica collezione bibliotecaria.

1. La bibliometria

1.1. Definizione della disciplina

La definizione della disciplina e l'introduzione del termine "bibliometria" si fa comunemente risalire ad Alan Pritchard che, in un articolo del 1969, propose di sostituire il termine poco utilizzato e alquanto ambiguo di "bibliografia statistica"¹ (*statistical bibliography*) con "bibliometria" (*bibliometrics*) per indicare l'applicazione della matematica e dei metodi statistici ai libri e alle altre forme di comunicazione scritta.² In realtà, l'equivalente francese del termine (*bibliometrie*) era già stato usato in campo biblioteconomico da Paul Otlet nel 1934, il quale dedicò una sezione del suo *Traité de documentation* all'argomento *Le livre et la mesure – Bibliometrie*,³ e da S. R. Ranganathan nel 1948 che, in occasione della conferenza ASLIB di Leamington Spa, preferì invece parlare di 'librametry', traendo chiaramente spunto dalla metrologie scientifiche esistenti in altre discipline (*biometry, econometry, psychometry* etc.), come di quella disciplina basata sull'impiego di metodi matematici per l'ottimizzazione dei flussi informativi nei servizi bibliotecari (Ranganathan 1949). Anche se molte altre definizioni della "bibliometria" si sono susseguite nel tempo, queste non hanno sostanzialmente modificato quella fornita da Pritchard che rimane tuttora quella comunemente accettata.⁴ Tali contributi hanno però il merito di aver precisato e approfondito gli scopi e l'ambito di studio (numero di pubblicazioni per autore, tipologia di pubblicazioni etc.) della disciplina, dilatando le sue possibilità di indagine fino all'analisi quantitativa della produttività scientifica su larga scala.⁵

1 L'ambiguità del termine nasce dalle due possibili interpretazioni di "bibliografia statistica" che può essere intesa sia come statistica della bibliografia sia come una bibliografia sulla statistica.

2 «BIBLIOMETRICS, i.e. the application of mathematics and statistical methods to books and other forms of written communication» (Pritchard 1969 p. 349).

3 La rivendicazione dell'origine francese del termine unita a una critica alla tendenza di molti autori di lingua inglese di ignorare i contributi nelle lingue romanze è discussa in Fonseca (1973 p. 5-6) e ripresa da Wilson (1995 p.133).

4 Nel 1976 anche *The British Standard Institution* propose una nuova definizione che sostanzialmente non modificava quella di Pritchard: «Bibliometrics as the application of mathematical and statistical methods in the study of the use of documents and publication patterns» (British Standard Institution 1976). Altre definizioni della disciplina si devono ad autori come Fairthorne (1969), Hawkins (1977), Khawaja (1987), Burton (1988), Egghe (1988), Khurshid – Sahai (1991a,b) and Tague – Sutcliffe (1992a).

5 Broadus (1987b p. 376) definì la bibliometria «... the quantitative study of physical published units, or of bibliographic units, or of surrogates of either ...» e White – McCain (1989 p. 119) «... is the quantitative study of

Visto il suo ampio spettro d'indagine, nel definire la disciplina diventa fondamentale stabilire i confini della stessa soprattutto in rapporto a quelle metrologie come la "scientometria", l'"informetrica" e la "webmetrica" che, pur non presentando delle differenze metodologiche, si distinguono per avere finalità e ambiti di ricerca contigui ma diversi rispetto alla bibliometria. I termini quindi impiegati per indicare le varie metrologie non possono venire usati come sinonimi, nonostante i confini sempre più vaghi e le concrete sovrapposizioni tra le varie discipline.⁶

La bibliometria ha come scopo principale quello di analizzare le collezioni bibliografiche e i servizi bibliografici estrapolando le relazioni quantitative tra i documenti e tra gli elementi che li compongono (parole, citazioni, autori, istituzioni etc.), a prescindere dall'ambito di applicazione o dalla specificità disciplinare.

La scientometria studia invece la scienza come prodotto sociale e culturale e, in particolare, il progresso scientifico e tecnologico attraverso la valutazione e la misurazione del contributo di scienziati, istituzioni, nazioni all'avanzamento delle conoscenze. La scientometria nel valutare la ricerca scientifica si basa sia su un approccio qualitativo avvalendosi di strumenti come la *peer review* (comitato dei pari), il *panel*, l'analisi costi-benefici, il grado di internazionalizzazione, sia su un approccio quantitativo attraverso calcoli sulle pubblicazioni e sulle citazioni. In questo secondo caso «le due aree della scientometria e della bibliometria diventano praticamente indistinguibili» (De Bellis 2005 p. 22).⁷

L'informetrica studia l'informazione in qualunque forma e contesto sociale essa venga espressa utilizzando un approccio di tipo quantitativo. Essa quindi è considerata «l'insieme degli insiemi di tutte le altre metriche poiché tutte contano un qualche tipo d'informazione» (De Bellis 2005 p. 23).

literatures as they are reflected in bibliographies. Its task, immodestly enough, is to provide evolutionary models of science, technology, and scholarship.»

6 «Though the techniques of scientometrics and bibliometrics are closely similar, their different roles are distinguished by their different contexts» (Brookes 1990 p.42).

7 «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» (De Bellis 2005 p. 39).

La webmetrica (o anche netmetrica, o cybermetrica) utilizza invece metodi e concetti bibliometrici e informetrici per lo studio quantitativo dei vari aspetti legati alla costruzione e all'uso delle risorse Web. «Tenta di misurare il World Wide Web al fine di ottenere conoscenze sul numero e sul tipo di connessioni ipertestuali, gli *hyperlinks*, sulle strutture del Web e sui modelli di utilizzo delle risorse che popolano il cyberspazio» (De Robbio 2007 p. 21).

Le varie metrologie quindi utilizzano metodologie praticamente uguali, quello che però è diverso per ognuna è l'oggetto dell'analisi, che cambia a seconda che nell'ambito di indagine si voglia porre l'accento sulle problematiche relative ai documenti (bibliometria), alla scienza (scientometria), all'informazione (informetrica) o al Web (webmetrica).

Pertanto, unità informative di base necessarie per una qualsiasi analisi bibliometrica sono i libri e "le altre forme di comunicazione scritta". Queste comprendono un insieme molto ampio di oggetti: monografie, rapporti, tesi, articoli di riviste e periodici, articoli pubblicati su collane seriali, come i *working papers* di un dipartimento, atti di convegni, libri elettronici (*e-books*), articoli pubblicati su riviste elettroniche (*e-journals*), articoli elettronici pubblicati in archivi disciplinari o istituzionali (*e-prints*), e i siti web. Tuttavia l'articolo scientifico *peer reviewed* è considerato l'unità bibliografica più adatta per l'analisi bibliometrica (Glänzel 2003 p. 12), questo perché «se uno scienziato ha qualcosa da dire, pubblica o cerca di pubblicare le sue scoperte su una rivista internazionale», esso cioè «rappresenta l'apice del lavoro creativo degli scienziati: ne legittima le rivendicazioni conoscitive e preannuncia la ricompensa istituzionale dei loro sforzi» (De Bellis 2005 p. 39). L'indagine bibliometrica parte dalle informazioni estrapolate dalle unità bibliografiche e, con l'impiego di specifici metodi di analisi, ha lo scopo sia di rintracciare le relazioni quantitative tra autori, pubblicazioni ed aree di ricerca sia di "misurare" la qualità (*performance*) dei ricercatori. I metodi bibliometrici generalmente usati dalla bibliometria sono l'analisi del contenuto e l'analisi citazionale. Della bibliometria, la quale ha vaste applicazioni in molti altri ambiti, sono proprio i metodi bibliometrici a rientrare nel campo della *LIS Library and Information Science* e discipline affini, questo perché fondamentalmente i loro meccanismi si basano sui documenti e in particolare sulle relazioni tra le citazioni dei periodici accademici.

L'analisi delle citazioni è il metodo più noto della bibliometria. Essa si basa «sul presupposto teorico che il numero di citazioni ricevute da una pubblicazione sia un indice della qualità della stessa e riesca a rivelare l'impatto di un lavoro di ricerca sulle comunità scientifiche di riferimento» (Cassella-Bozzarelli 2011 p.68): più volte (e per più tempo) un lavoro è citato più alto è il suo valore scientifico. La diffusione dell'uso delle citazioni come strumento valutativo, avvenuta grazie alla creazione di archivi come quelli dell'*ISI*, ha determinato un miglioramento delle tecniche di indagine. Non sono più bastati i semplici conteggi del numero di pubblicazioni o del numero di citazioni ricevute da un autore (ma anche da un gruppo di ricerca, da una rivista o da un'istituzione), cioè l'impiego di misure 'grezze' necessariamente approssimative specie se non poste in relazione con altri parametri, ma si sono ricercati indicatori bibliometrici dagli algoritmi sempre più raffinati. Con l'impiego poi delle potenzialità dell'analisi bibliometrica in ambiti sempre più vasti, anche le tecniche citazionali sono via via diventate sempre più complesse tanto che, in alcuni ambiti, alle singole citazioni si sono preferiti gli accoppiamenti di citazioni (*co-citation*) o gli accoppiamenti bibliografici (*bibliographic coupling*).

Volendo schematizzare si può affermare che attualmente gli studi bibliometrici si suddividono in due aree concettuali fondamentali: descrittiva e valutativa. La bibliometria descrittiva studia i modelli di comunicazione e informazione scientifica e loro evoluzione in relazione ad area geografica e settore disciplinare, mentre quella valutativa ha come scopo l'identificazione dei modelli d'uso delle citazioni bibliografiche ed è anche conosciuta come "analisi citazionale" (Di Cesare 2002 p. 135). La prima ha un approccio *top-down*, vale a dire dall'alto verso il basso, l'analisi quantitativa si basa cioè sul conteggio delle pubblicazioni di intere aree disciplinari (*productivity count*). La seconda invece ha un approccio *bottom-up*, vale a dire dal basso verso l'alto, l'analisi quantitativa si basa cioè sul conteggio delle citazioni ricevute dai ricercatori in un dato settore disciplinare (*Literature Usage Count*).⁸

I dati provenienti dall'analisi bibliometrica sono diventati per i bibliotecari uno strumento efficace e in continua espansione su almeno due grandi fronti: nella valutazione/gestione delle collezioni bibliografiche e nella valutazione/gestione dei servizi bibliotecari. Sono utili infatti per tutte le operazioni che riguardano non solo la politica

⁸ Per una estesa trattazione e un confronto tra le due aree di studi bibliometrici cfr. il saggio di Van Leeuwen *Descriptive versus evaluative bibliometrics* contenuto in Moed – Glänzel – Schmoch (2004 p. 373-388).

di selezione, di acquisizione, di accesso e di uso delle collezioni bibliografiche, ma anche per quelle di scarto e di conservazione del materiale. Le informazioni di queste analisi sono una base scientifica e oggettiva che aiuta i responsabili della biblioteca a pianificare, ottimizzare e razionalizzare le risorse strumentali, finanziarie e umane. «Aiutano i bibliotecari poi a progettare regole di catalogazione adeguate alle tipologie di *authorship*, a scegliere i periodici cruciali per una disciplina e a progettare, grazie soprattutto alla nascita e alla continua implementazione degli archivi citazionali, sistemi più efficienti di *information retrieval*» (De Bellis 2005 p. 38). Inoltre, con l'estensione dei metodi bibliometrici allo studio della comunicazione mediata dalla rete di *hyperlink* che popolano il World Wide Web, si è iniziato ad applicare l'analisi citazionale, il più noto dei metodi bibliometrici, al contesto delle biblioteche digitali allo scopo di determinare impatto e *ranking* di documenti e riviste ad uso dell'utenza della biblioteca.

Si avvalgono delle tecniche bibliometriche, oltre ai bibliotecari, molte altre categorie di professionisti. Sono impiegate infatti dai sociologi della comunicazione scientifica per analizzare la struttura della scienza, cioè per costruire mappe delle specialità disciplinari, delle tematiche ricorrenti, delle aree di ricerca attive e di quelle che si stanno formando. Dagli storici della scienza per la costruzione di bibliografie retrospettive al fine di tracciare i percorsi seguiti dalle idee e dalle teorie scientifiche nella loro evoluzione storica. Dai responsabili delle politiche pubbliche come strumento semplice ed economico per confrontare e valutare l'attività di ricerca e nelle istituzioni accademiche per selezionare i ricercatori e distribuire i fondi di ricerca. Dai redattori delle riviste come valido strumento di riferimento per individuare i *peer reviewers* delle pubblicazioni. E infine, da tutti gli specialisti di un determinato settore per l'analisi quantitativa della disciplina di appartenenza.

Guardando agli ambiti di applicazione della bibliometria emerge chiaramente la sua spiccata interdisciplinarietà che però ha anche significato scarsa standardizzazione dei metodi di misurazione impiegati e assenza di un «accordo su una teoria della scienza che faccia da guida nella costruzione degli indicatori» (De Bellis 2005 p. 46). Normalmente la procedura seguita prevede una prima fase dedicata alla raccolta ed elaborazione dei dati, accompagnata solo successivamente da una fase dedicata invece alla esplicazione dei risultati (Van Raan 2004 p. 21). «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» (De Bellis 2005 p. 46). Glänzel (2003 p. 9) ha individuato tre grandi componenti o macro-aree di ricerca della bibliometria contemporanea: «a) una bibliometria per bibliometri, fatta prevalentemente di riflessioni sullo statuto metodologico della disciplina; b) una bibliometria per le discipline scientifiche e storico-sociali, orientata alle diverse modalità di trattamento quantitativo dell'informazione specifica di ciascuna disciplina; c) 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» (De Bellis 2005 p.46).

Contrariamente quindi a quello che comunemente si pensa, la bibliometria non è, o non è soltanto, una disciplina per la costruzione di indicatori bibliometrici utili per la valutazione della ricerca o uno strumento ausiliario che non necessita di ulteriore ricerca metodologica. Questo appare chiaro soprattutto in ambito bibliotecario dove fin dall'inizio la disciplina viene applicata per l'analisi dei meccanismi di crescita ed obsolescenza della letteratura scientifica, le finalità pertanto sono sì valutative, ma riguardanti qualità e quantità delle collezioni bibliografiche in senso biblioteconomico.

Attualmente della disciplina esiste una letteratura vastissima prodotta soprattutto, ma non solo, nei paesi anglosassoni che comprende, oltre a una grande produzione di articoli scientifici e monografie sull'argomento, riviste internazionali, manuali, rassegne, dizionari, conferenze internazionali, liste di discussione.⁹ Ciononostante a tutt'oggi in Italia la bibliometria è ancora una «una disciplina poco studiata e poco applicata, tanto che la massiccia *Guida classificata* diretta da Mauro Guerrini non le dedica nessuna delle sue 138 voci né delle sue 299 intestazioni di soggetto, e che nell'ultima edizione di *BIB* solo 25 delle oltre

9 Delle riviste dove la bibliometria occupa uno spazio considerevole si ricorda: *Scientometrics* (1978-); *Journal of Informetrics* (2007-); *Cybermetrics* (1997-); *Information Visualization* (2002-); *Journal of the American Society for Information Science and Technology* (1970-); *Journal of Documentation* (1945-); *Journal of Information Science* (1979-); *Social Studies of Science* (1975-); *Science Studies* (1988-); *American Sociological Review* (1936-); *Research Evaluation* (1991-); *Research Policy* (1971-); *Information Processing & Management* (1975-); *Library Trends* (1952-); *Library Quarterly* (1931-); *Journal of Computer-Mediated Communication* (1995-); e *Webology* (2004-). Tra i manuali in ordine cronologico: Ravichandra Rao (1983); Van Raan (1988); Egghe – Rousseau (1990); Glänzel (2003); Moed – Glänzel – Schmoch (2004); Thelwall (2004); Moed (2005); Vinkler (2010). Tra le rassegne: Narin (1976); Narin – Moll (1977); White – McCain (1989); Wilson (1999); Nicolaisen (2007); Bar-Ilan (2008); De Bellis (2009). Il dizionario di Diodato (1994). Conferenze internazionali: *The International Conference on Scientometrics and Informetrics* (1987-); *Conferences on Science & Technology Indicators* (CWTS, 1988-); *International Conference on Webometrics, Informetrics and Scientometrics & COLLNET International Research Network on Collaboration in Science and Technology* (2000-). Lista di discussione: *SIGMETRICS Listserv* (1999-).

30.000 schede (relative a quasi 50.000 testi pubblicati dal 1971 al 2004) sono risultate direttamente riconducibili ad essa» (Ridi 2010).

1.2. Lineamenti storici: dalla bibliografia statistica alla bibliometria moderna

La storia della bibliometria¹⁰ può essere suddivisa in tre grandi periodi. Il primo periodo (seconda metà del XIX sec.) si riferisce a quella fase degli studi bibliometrici, ancora poco ambiziosi nello scopo e poco raffinati nella tecnica, incentrati sulla semplice analisi della quantità della letteratura prodotta analizzando un arco temporale prestabilito. Il secondo periodo (prima metà del XX sec.) vede la pubblicazione dei lavori di Lotka, Bradford e Zipf che costituiranno il nucleo fondamentale delle leggi bibliometriche, frutto dell'applicazione anche nel campo dell'informazione e della comunicazione scientifica di un approccio metodologico quantitativo. Il terzo (seconda metà del XX sec.) è il periodo moderno che inizia con le pubblicazioni di Garfield, Pritchard e Price.

Precursori degli studi bibliometrici possono essere persino considerati i semplici conteggi e le classificazioni dei libri a fini patrimoniali e amministrativi condotti fin dai tempi più antichi. Si conosce infatti il posseduto della Biblioteca di Alessandria, che ammontava a 490.000 rotoli, dato registrato già nel III secolo a.C..

Degni di nota anche altri esempi più recenti basati sul conteggio delle unità bibliografiche come la lista compilata nel 1837 sul posseduto di alcune importanti biblioteche. Dove si legge che la Biblioteca Reale di Parigi possedeva 620.000 volumi e nel 1850 circa le biblioteche pubbliche degli Stati Uniti ne possedevano circa 1.300.000. Nello stesso anno venne rapportato il numero dei volumi delle biblioteche di alcune città europee con la popolazione delle rispettive città. Monaco aveva 750 libri su 100 persone, Firenze 313, Parigi 143,

¹⁰ Una eccellente panoramica della storia della bibliometria è fornita da Hertzell (1987). L'autrice descrive le tappe fondamentali dello sviluppo della disciplina fin dalle sue origini nella bibliografia e nella statistica, prestando particolare attenzione alle leggi di Lotka, Bradford e Zipf. Altro contributo simile è quello di Broadus (1987a) che traccia la storia della disciplina fino al 1969 quando Pritchard introduce il termine "bibliometria", esamina lo sviluppo delle tre leggi bibliometriche, dell'analisi citazionale e dell'impiego della disciplina in ambito bibliotecario. Brookes (1990) oltre a tracciare il percorso storico della disciplina si sofferma anche sull'uso terminologico delle metriche associate alla bibliometria. Wilson (2001) dedica una sezione alla storia della terminologia delle tre metriche e della 'Librametry'. Tra gli altri contributi che affrontano l'argomento sotto l'aspetto storico si ricordano: Wittig (1978), Griffith (1979), Roy (1980), Schmidmaier (1984), Schrader (1984), Deogan (1987), White – McCain (1989), Pierce (1992), Roman (1994), Tague – Sutcliffe (1994), Buckland – Liu (1995), Portal (1995) and Chongde (1996). In lingua italiana De Bellis (2005 e 2009).

Bruxelles 100 e Londra solo 20 (Broadus 1987 p. 127). In un primo periodo quindi la “bibliometria” contava semplicemente i libri in circolazione, statistiche ufficiali erano a quel tempo pressoché inesistenti.

É solo alla fine del XIX secolo che si comincia ad assistere ad un generale affinamento delle tecniche bibliometriche grazie all'applicazione della statistica ai libri e ad altri prodotti della comunicazione scritta, allo scopo di trarre conclusioni generali sul progresso di una disciplina scientifica o di ricavare indicazioni utili alla politica delle acquisizioni in biblioteca (De Bellis 2005 p. 24). Questa attività costituiva la prima fase di “incubazione” di quella che in seguito sarebbe stata chiamata “bibliografia statistica” e più tardi “bibliometria” che, fondamentalmente, focalizzava l'attenzione sui meccanismi di crescita e obsolescenza della letteratura scientifica e che operativamente si basava sulle bibliografie disciplinari ampiamente disponibili già dal XVIII secolo grazie all'impulso dato alla specializzazione della ricerca scientifica.

Gli studi bibliometrici di prima generazione hanno la straordinaria importanza di aver gettato le basi della moderna bibliometria. Frank Campbell in una pubblicazione del 1896 impiegò il calcolo statistico per sintetizzare la ripartizione tematica delle pubblicazioni facenti parte delle più importanti bibliografie dell'epoca. F. J. Cole e Nellie Eales pubblicarono nel 1917 un lavoro sulla storia dell'anatomia comparata, nel quale venne applicata l'analisi statistica alla letteratura del settore, distinta per Paese, prodotta dal 1543 al 1860, misurando in questo modo il grado crescita e di obsolescenza di questa disciplina. E. Wyndham Hulme, bibliotecario presso il *Patent Office* di Londra, fu il primo ad utilizzare il termine di “bibliografia statistica” nel suo *pamphlet* del 1923, dove si occupò delle relazioni tra il numero di pubblicazioni scientifiche, il numero di brevetti registrati e l'ammontare delle esportazioni. Utilizzando più edizioni dello *International Catalogue of Scientific Literature* Hulme ricavò informazioni quali la ripartizione per settore disciplinare degli scienziati e la quantità totale di periodici indicizzati dall'intero servizio bibliografico britannico. «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 cronolo-

gicamente all'interno di ciascuna classe, si poteva costruire un'immagine bibliografica attendibile dei progressi della mente umana» (De Bellis 2005 p. 25).¹¹

Nel 1943, Charles Gosnell, basandosi sugli apparati bibliografici di alcune pubblicazioni accademiche, alle quali applicò il calcolo statistico, determinò l'indice di invecchiamento delle letterature suddivisa per ambito disciplinare. Gosnell, come Hulme, era convinto che il numero di libri, ma soprattutto il tasso di crescita e obsolescenza della letteratura nei diversi settori della conoscenza, fossero elementi utili su cui basarsi per poter trarre conclusioni sull'evoluzione generale della cultura. Si deve invece ai coniugi Gross, due chimici del piccolo college di Pomona, California del sud, il primo saggio di analisi delle citazioni pubblicato nel 1927. I Gross esaminarono un'annata del *Journal of the American Chemical Society* classificando i periodici in funzione del numero di citazioni ricevute allo scopo di fornire uno strumento valido alla politica delle acquisizioni della biblioteca del college in cui lavoravano. È qui che nasce l'equivalenza tra qualità dei periodici e numero di citazioni ricevute, assunto che ispirerà più tardi Eugene Garfield per l'ideazione dello *Science Citation Index*.¹²

In altri settori, come la psicologia, nel 1936 Hulsey Cason e Marcella Lubotsky analizzarono i riferimenti bibliografici di 28 periodici per stabilire le interrelazioni tra i diversi settori della disciplina e, allo stesso scopo, dieci anni più tardi Herman Henkle fece lo stesso per l'ambito biochimico.

In un articolo apparso nel 1944, Estelle Brodman andò oltre il lavoro dei Gross. Per l'individuazione dei *key journals*, decise di non limitare l'indagine ad un'unica fonte informativa, come invece avevano fatto i Gross in precedenza, ma di utilizzare più liste di periodici selezionate usando criteri diversi per ognuna. Considerando quindi la metodologia precedente sbagliata e riduttiva, anticipò alcuni problemi rivelatesi cruciali negli studi sulle citazioni dei decenni successivi.

11 «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» (Hulme 1923 p. 43)

12 Quattro anni dopo, Gross e Woodford (1931) pubblicarono uno studio simile per la geologia. Contrariamente però al precedente articolo, il lavoro risultò notevolmente più esteso e la lista delle pubblicazioni non più rispondente alle esigenze delle piccole biblioteche di un college come strumento di supporto nella politica delle acquisizioni.

Un altro punto di svolta è determinato da Herman Fussler, che in una dissertazione per la *Graduate Library School* dell'Università di Chicago del 1948, sviluppò ulteriormente il metodo di usare le referenze bibliografiche di una rivista per individuare i *core journals* di un settore disciplinare. A tal fine, Fussler si servì dei riferimenti bibliografici di due *key journals* delle discipline della fisica e chimica e cioè il *Physical Review* e il *Journal of the American Chemical Society*. Lo studio di Fussler viene considerato nell'ambito dell'analisi delle citazioni, quello che più di ogni altro aveva recepito quanto la bibliometria aveva fino ad allora elaborato come testimoniano le sue citazioni a Gross – Gross (1927), Brodman (1944) e a Hulme (1923) (Di Cesare 2002 p. 138).

Ma un vero e proprio spartiacque nella disciplina, quello che le conferirà una rinnovata consapevolezza nelle potenzialità delle tecniche di indagine, è rappresentato da tre lavori fondamentali apparsi negli anni '20 e '30 del '900 e risultato di quell'approccio metodologico quantitativo emerso nel campo dell'informazione e della comunicazione scientifica nella prima metà del XX secolo.¹³ Sono gli studi di Lotka (Lotka 1926) sulla distribuzione della produzione scientifica in rapporto al numero degli autori, di Zipf (Zipf 1930) sulla distribuzione delle parole in uno scritto, e di Bradford (Bradford 1934) sulla distribuzione del numero di articoli nei periodici. Grazie a queste pubblicazioni, l'intuizione si fa legge bibliometrica. Quella che era stata infatti fino ad allora una semplice regolarità, intuita analizzando empiricamente i dati, diventa ora un concetto espresso dalla matematica e dalla statistica.

La terza fase storica della disciplina si apre con la fine del secondo conflitto mondiale, nel nuovo assetto non solo politico, ma anche socio-economico e dell'organizzazione delle conoscenze soprattutto di carattere scientifico.

Nell'immediato dopoguerra tuttavia la valutazione della ricerca era condotta ancora senza l'impiego di indicatori bibliometrici, utilizzando invece o il collaudato sistema del *peer reviewing* o indicatori economici di macro-livello, impostati cioè all'analisi di dati provenienti da sistemi non meno ristretti di quello nazionale.

¹³ Di Cesare (2002 p. 133) sottolinea l'esistenza di precedenti in questo senso e menziona il tentativo di Francis Galton (1822-1917) di introdurre l'analisi nei più svariati ambiti dell'attività umana, compiendo studi importanti sulla "distribuzione della qualità tra gli scienziati" più in vista all'epoca.

Il mondo postbellico, diventato più complesso nell'ambito della produzione scientifica, recupera gli strumenti e l'idea di misurare la documentazione scientifica che la bibliografia statistica aveva già sviluppato a partire dalla fine del secolo XIX. Questa esigenza nasce dalla convinzione che l'oggetto dell'attività scientifica possa essere controllato, pianificato e indirizzato verso importanti obiettivi specifici, soprattutto da quando si è cominciato a vedere come leva della crescita economica lo sviluppo e l'innovazione messi in moto dalla scienza di base.

Il *Frascati Manual for the Measurement of Scientific and Technical Activities* (OECD 1963) può essere considerato una risposta alla crescente importanza economica di scienza e tecnologia diventata visibile nelle statistiche economiche negli anni '50 del '900. Del resto, proprio nel *Frascati Manual*, come già nel famoso report di Vannervar Bush *The Endless Frontier* (Bush 1945), viene sì sottolineata l'importanza della scienza di base per il progresso economico-sociale ma anche l'autonomia della stessa nei confronti del governo, interessato invece, a valutarla con misure di impatto e produttività. Il *Frascati Manual* come gli analoghi *Oslo Manual* e *Canberra Manual* non auspicavano comunque ancora l'impiego di indicatori bibliometrici per la misurazione della scienza sia perché considerati inadeguati a rappresentare una realtà sicuramente complessa e articolata come quella delle relazioni tra ricerca scientifica, sviluppo tecnologico e tessuto sociale ed economico, sia per la mancanza di linee guida internazionali a tal proposito.

Nella politica della ricerca scientifica, un vero e proprio cambiamento avviene nel 1957, con il lancio dello *Sputnik* da parte dell'Unione Sovietica. Una diretta conseguenza¹⁴ di quello che Leydesdorff (Leydesdorff 2005 p. 1510) definì «*Sputnik shock*» fu la riorganizzazione di grande portata del sistema di ricerca statunitense, allo scopo di colmare il *gap* scientifico nei confronti dei Sovietici, che l'importante traguardo da loro raggiunto aveva reso evidente.

L'OECD (*Organization for Economic Co-operation and Development*) venne creata nel 1961, nel solco del coordinamento militare della NATO, allo scopo di organizzare e coordinare le politiche in materia di scienza e tecnologia degli stati membri. L'organizzazione è però in-

¹⁴ Gli Stati Uniti risposero al lancio dello *Sputnik* e al pericolo di un eventuale conflitto militare soprattutto con la creazione della *NASA* (*National Aeronautics and Space Administration*) e dell'*ARPA* (*Advanced Research Project Agency*).

teressata più che ai risultati della ricerca, alle risorse umane e finanziarie investite, mentre la valutazione della ricerca rimane ancora fermamente ancorata nei meccanismi interni della comunicazione accademica e specialistica.

Ma l'attenzione per la misurazione della documentazione scientifica non si esaurisce nel solo interesse per la valutazione della ricerca. Nel corso degli anni '50 e '60 del XX secolo, la stessa comunità scientifica era diventata ben cosciente dell'aumento vertiginoso della letteratura scientifica che rendeva sempre più urgente l'ideazione di nuovi strumenti di controllo.¹⁵ Il volume della produzione scientifica, aggravato dai limiti insiti nei sistemi manuali di indicizzazione della letteratura, stava in effetti diventando tale da eccedere l'abilità di produttori e consumatori di scienza nell'identificare e utilizzare le informazioni più pertinenti per le loro ricerche.

A questo problema se ne andavano ad aggiungere altri. Gli anni del dopoguerra infatti avevano assistito anche alla marcata specializzazione delle aree di ricerca con la conseguente proliferazione settoriale. Tutto questo costituiva un ostacolo non solo per i soggetti più direttamente interessati, ma anche per il progresso scientifico generale che invece aveva bisogno di strumenti veloci di recupero e disseminazione dell'informazione.¹⁶

È a questo punto che entra nella storia della bibliometria la figura di Eugene Garfield che, operando presso il *John Hopkins Welch Medical Library Project*,¹⁷ si rese conto dell'inadeguatezza degli strumenti a disposizione degli scienziati per recuperare informazioni utili ai fini della loro attività di ricerca (Garfield 1955a).¹⁸ Allo scopo di superare i limiti imposti

15 Uno degli autori principali che ha trattato il problema dell'incremento esponenziale della letteratura scientifica durante il '900 è Derek J. De Solla Price, storico della scienza. In particolare: De Solla Price (1963) e (1965).

16 L'impossibilità di avere una visione completa della letteratura prodotta in un determinato ambito disciplinare poneva anche il problema della duplicazione delle scoperte scientifiche, sulla questa questione De Bellis (2005 p. 43) scrive: «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».

17 Scopo di tale progetto era lo studio dei problemi legati all'indicizzazione della letteratura medica e l'esplorazione di metodi basati, in particolare, sull'impiego di macchine capaci di costruire indici automatici della bibliografia medica. Larkey (1953 p. 1) scrive a proposito dello scopo del progetto: «To study the problems of indexing medical literature. To explore the theory and practice of subject heading (nomenclature) and classification (coding) as they concern medical literature. To explore existing and projected methods, emphasizing machine methods, applicable to medical bibliography, operating such pilot projects as may be necessary, and to report on the suitability of machine methods in the bibliographic operations of the Army Medical Library.»

18 In occasione del suo 75° compleanno, Garfield ha reso disponibile in un archivio aperto il testo completo dei propri contributi che comprendono 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). In questo open archive inoltre trovano spazio molti articoli e saggi di autori che si sono occupati di tematiche riconducibili agli indici di citazioni. L'archivio è liberamente consultabile all'indirizzo web: < <http://www.garfield.library.upenn.edu/pub.html> >.

connaturati ai tradizionali indici per soggetto, Garfield presentò l'idea di un progetto di indice interdisciplinare basato sulle citazioni per la prima volta sulla rivista *Science* nel 1955. Grazie all'interessamento del genetista Joshua Lederberg il progetto di Garfield trovò piena concretizzazione solo nel 1963 con la pubblicazione dello *Science Citation Index* (SCI).

Garfield era quindi riuscito a sfruttare, concretizzandolo in un sistema, il meccanismo interno alla citazione bibliografica che da sempre aveva legato documenti e autori. Per effetto quindi dell'applicazione della proprietà transitiva, le citazioni formano una rete di connessioni tra documenti che, strutturata in un indice, diventa, a sua volta, proprio quello che Garfield aveva definito un *association-of-ideas index*¹⁹ capace di collegare assieme documenti difficilmente catalogati sotto lo stesso *heading* all'interno dei tradizionali indici per soggetto.

Lo SCI ideato da Garfield svilupperà ulteriormente la corrispondenza tra valore qualitativo dei periodici e numero di citazioni ricevute, affermando l'uso delle citazioni come strumento valutativo e, indirettamente, degli indicatori bibliometrici nella valutazione della ricerca scientifica che, per tale scopo, si affermarono, a partire dagli anni '60-'70, prima negli Stati Uniti e poi anche in diversi paesi europei.

Il fermento innovativo contribuì inoltre a far emergere una riflessione sulla natura, gli ambiti e gli scopi della disciplina e, cosa non meno importante, sul termine più adatto per designarla. È qui che si colloca il contributo, già ricordato, di Pritchard (Pritchard 1969).

A partire dagli anni '70 vennero pubblicati numerosi repertori basati su dati ricavati dallo SCI di Garfield. Il primo esempio è rappresentato dai *Science Indicators Reports* pubblicati dal 1973, a cura del *National Science Board*, allo scopo di misurare la ricerca scientifica statunitense. Repertori analoghi seguiranno anche in Europa, Canada e Australia.²⁰

Nel 1976 Francis Narin pubblicò il suo studio pionieristico intitolato *Evaluative Bibliometrics* (Narin 1976) incontrando il favore della *National Science Foundation*. Se Henry Small (Small 1973) aveva precedentemente proposto un metodo per mappare la struttura discipli-

19 Garfield (1955b p. 108) scrive: «a citation [...] offers a new approach to subject control of the literature of science. By virtue of its different construction, it tends to bring together material that would never be collated by the usual subject indexing. It is best described as an association-of-ideas index, and it gives the reader as much leeway as he requires. Suggestiveness through association-of-ideas is offered by conventional subject indexes but only within the limits of a particular subject heading.»

20 Per citare alcuni esempi si ricorda: l'*Indicateurs de sciences et de technologies* francese, il *Wetenschaps- en Technologie-Indicatoren* olandese, il *Vlaams Indicatorenboek* belga e l'*European Report on S&T Indicators*.

nare della scienza basato sulla tecnica delle co-citazioni, Narin si concentrò invece sul concetto di impatto citazionale come misura della qualità delle pubblicazioni scientifiche a livello sovranazionale.

Dello stesso anno del contributo di Narin è anche quello di Derek de Solla Price (Price 1976) apparso su *Journal of the American Society for Information Science* che propose un modello concettuale unificato delle leggi bibliometriche, dove i singoli elementi alla base delle leggi di Lotka, Bradford, Price vengono considerati in maniera cumulativa.

Molti furono anche i governi e le istituzioni interessate al perfezionamento degli indici bibliometrici come valido strumento di supporto nelle decisioni in merito alla politica di distribuzione delle risorse da investire in ricerca. Mentre quindi negli anni '70 gli studi americani si incentrarono soprattutto sull'organizzazione della letteratura scientifica, negli anni '80 gli studi in Europa cominciarono ad applicare gli indicatori bibliometrici su base istituzionale, per la valutazione cioè delle performance di università, istituzioni e gruppi di ricerca. A questo proposito merita menzionare il lavoro dell'Università di Leida che, influenzato a sua volta dallo studio di Ben Martin e John Irvine (Martin – Irvine 1983) sulla misurazione dell'efficienza di quattro osservatori radioastronomici, a partire dai primi anni '90 venne preso a modello dai *Research Assessment Exercises* (REA), il programma relativo alla distribuzione dei fondi per la ricerca universitaria del Regno Unito.

Per la storia della bibliometria è stato importante anche l'ultimo decennio, il quale ha visto importanti innovazioni come la creazione di nuovi indici di citazioni anche di tipo *open access*, che hanno definitivamente decretato la fine del monopolio dettato dallo *SCI*; lo sviluppo di indicatori bibliometrici sempre più sofisticati; e l'applicazione della metodologia in ambiti sempre più vasti come il World Wide Web.

1.3. Basi matematiche

Le basi matematiche della bibliometria si sostanziano nelle tre leggi bibliometriche frutto degli studi di Lotka, Bradford e Zipf. Esse però non sono delle vere e proprie "leggi" in senso matematico o fisico, non sono cioè dei teoremi dal valore assoluto o delle espressioni di regolarità constatate in fenomeni naturali per i quali è possibile applicare metodi sperimentali rigorosi, ma dei concetti assertivi di regolarità riscontrabili in ambito documenta-

rio, espresse per mezzo di formule empiriche che non consentono previsioni accurate e che hanno un valore puramente probabilistico. La regolarità descritta dalle leggi bibliometriche è in realtà una sostanziale asimmetria: esiste un ristretto nucleo di cause che produce una quantità molto più consistente di effetti.

Tale asimmetria è comunque constatabile anche in molti altri settori naturali o sociali ed è comunemente tradotta in termini matematici mediante delle "distribuzioni iperboliche", dette anche "leggi di potenza" o "distribuzioni paretiane".

Le distribuzioni bibliometriche sono pertanto una tipologia specifica di distribuzioni iperboliche o paretiane, utilizzate per descrivere regolarità empiriche di fenomeni, processi ed eventi bibliografici.

1.3.1. Legge di Lotka

Nel 1926 Alfred Lotka (1880-1949), matematico e chimico, presidente dell'*American Statistical Society*, pubblicò uno studio pionieristico (Lotka 1926) sulla distribuzione della produzione scientifica in una comunità ben definita di scienziati.

Analizzando i dati relativi a un certo numero di autori e di pubblicazioni schedati nei *Chemical Abstracts* (dal 1907 al 1916) e nelle *Geschichtstafeln der Physik di Auerbach* (dalle origini al 1900), Lotka riscontrò che gli scienziati contribuiscono in misura diversa al progresso delle conoscenze, vale a dire che a fronte di pochi autori con un alto tasso di produttività, ne esiste un numero molto più alto di poco produttivi. Lotka sintetizzò le sue osservazioni in una legge empirica secondo cui il numero di autori che producono n pubblicazioni è circa pari a $1/n^2$ del numero di autori che pubblicano un solo contributo; e che la proporzione di tutti gli autori che producono un singolo contributo è all'incirca il 60 per cento dell'intera produzione (De Bellis 2005 p. 30).

Questa la formula ricavata:

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

dove p sta per il numero di autori che producono n pubblicazioni e K indica una costante specifica dell'ambito disciplinare di riferimento. Di conseguenza, se più della metà (pari al 60%) degli autori di un ambito disciplinare pubblica soltanto un lavoro, a scalare il 15% pubblicherà due contributi ($1/22 * 60$), il 7% tre contributi ($1/32 * 60$) etc. Il valore quindi

della legge di Lotka è strettamente collegato alla produttività media (K) di una specifica comunità scientifica. Essa dipende, a sua volta, dall'area disciplinare considerata, dalla comunità di scienziati studiata, e perfino, quando le precedenti variabili si mantengono costanti, dallo specifico momento storico (Bailón – Moreno et al. 2005 p. 217).

A partire dagli anni '70 la legge è andata incontro a tutta «una serie di applicazioni, reinterpretazioni, smentite e presunte riconferme» (De Bellis 2005 p. 29).²¹ Del resto, Lotka stesso (Lotka 1926 p. 323) ammise che le «distribuzioni di questo tipo 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».

In generale si può affermare che la legge di Lotka presenta una buona aderenza ai valori osservati in un'area dove il livello di produttività degli autori è basso, mentre quando il livello aumenta, l'aderenza della legge ai dati empirici peggiora notevolmente fino anche ad arrivare a perdere validità.²²

1.3.2. Legge di Bradford

Dopo otto anni dalla pubblicazione di Lotka, vide la luce quella che è considerata la più importante delle leggi bibliometriche, elaborata da Samuel C. Bradford (1878-1948), bibliotecario presso il Museo della Scienza di Londra, il quale pubblicò sulla rivista *Engineering* i risultati dei suoi studi (Bradford 1934), che ricevettero però piena attenzione sola alla fine degli anni '40, in particolare dopo la pubblicazione di un suo lavoro intitolato *Documentation* (Bradford 1950).

La legge elaborata è il frutto dell'analisi statistica che egli condusse su due bibliografie di geofisica: la *Current Bibliography of Applied Geophysics* e la *Quarterly Bibliography of Lubrication*. La legge afferma che, se si considera un elenco di riviste stilato tenendo conto della produttività decrescente in rapporto al numero di articoli nell'ambito di un determinato

21 Ad esempio, come scrive De Bellis (2005 p. 29): «ne è stata mostrata 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*', mentre uno studio recente sullo sviluppo di software *open source* da parte dei programmatori disseminati nel cyberspazio ne ha reclamato la sostanziale validità». Una vasta bibliografia di lavori su Lotka di 437 titoli è contenuta nel primo articolo di Vlachy pubblicato nella rivista *Scientometrics*, pubblicato nel 1978. Più recentemente una rassegna sui contributi inerenti le leggi informetriche e in particolare quelle di Lotka è contenuta in un articolo di Bar-Ilan, pubblicato nel *Journal of Informetrics* (Bar-Ilan 2008).

22 Cfr. Bailón – Moreno et al. (2005 p. 217), Egghe (1995) e Kretschmer – Rousseau (2001).

settore scientifico, il risultato è un gruppo di riviste che concentra la maggior parte degli articoli sull'argomento, e per questo è considerato il nucleo di quell'ambito disciplinare; ed altri gruppi formati da un numero di riviste che, per poter contenere lo stesso numero di articoli di quelle che formano il nucleo, cresce in modo esponenziale. Il numero di riviste che formano ogni gruppo è espresso dalla proporzione:

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

Questo vuol dire che se le riviste in un dato settore disciplinare sono classificate in base al numero di articoli in tre gruppi, un terzo degli articoli viene pubblicato da un limitato numero di periodici specializzati nel settore disciplinare in questione, un altro terzo da altri periodici dedicati ad argomento simile ma non coincidente col gruppo precedente, ed il terzo restante da molti periodici generalisti e quindi meno pertinenti.

Appare chiaro quindi dalla proporzione che per raggiungere il 100% della copertura bibliografica sull'argomento occorre aggiungere al nucleo formato da poche riviste specializzate un numero di riviste molto più ampio che tra l'altro cresce in modo esponenziale.

La legge di Bradford si dimostrò estremamente utile in campo biblioteconomico. Con l'aumento esponenziale della letteratura scientifica divennero infatti più incalzanti molte delle problematiche, legate all'obsolescenza, all'organizzazione e alla gestione delle collezioni bibliografiche, con anche quelle relative al recupero mirato delle informazioni in esse contenute. A queste si aggiunse la crisi del 1929 che portò ad un assottigliamento delle risorse finanziarie a disposizione delle biblioteche e, di conseguenza, ad un drastico ridimensionamento delle nuove acquisizioni di materiale bibliografico. Visto che necessariamente le acquisizioni di materiale bibliografico nella logica del "*just in case*" ("nel caso che" l'utenza ne avesse fatto richiesta) dovevano essere abbandonate senza portare ad un totale stravolgimento della completezza dei nuclei disciplinari della collezione (Guédon 2004 p. 33) e minimizzando l'impatto sull'utenza, divenne vitale un'ottimizzazione della spesa ricorrendo a studi di valutazione quantitativa per la selezione degli acquisti. La legge, individuando un nucleo di pubblicazioni considerate fondamentali all'interno di ogni disciplina scientifica, dava ai bibliotecari la possibilità di togliere tutte quelle che si allontanavano da tale nucleo. La distribuzione statistica espressa dalla legge non solo divenne la spiegazione "scientifica" che legittimava molti tagli nelle collezioni ma anche un caposal-

do nella politica delle acquisizioni (e anche dello scarto) in biblioteca: non tutte le pubblicazioni sono necessarie, o almeno oltre una certa soglia i benefici derivanti da ulteriori acquisizioni sarebbero stati nulli, o talmente limitati rispetto ai costi sostenuti, da ritenere inutile il loro acquisto.

Ottimizzando inoltre la selezione del materiale attorno ad un nucleo, anche la ricerca bibliografica ne traeva benefici con l'aumento della precisione e il conseguente abbattimento del rumore.

La legge di Bradford, come quella di Lotka e di Zipf, ha subito molte «analisi, estensioni, contestazioni, comparazioni e fusioni con altre leggi» (De Bellis 2005 p.32), Garfield, per esempio, ampliarà le sue potenzialità applicative estendendone la validità anche alla scienza concepita come organismo ben strutturato, ed utilizzandola come base per la scelta dei periodici che formeranno il nucleo dello *Science Citation Index*.

1.3.3. Legge di Zipf

George K. Zipf (1902-1950), professore di linguistica tedesca ad Harvard, formulò una legge interessante sia per la bibliometria che per la linguistica quantitativa, che egli derivò dallo studio della frequenza delle parole in un testo.

La legge di Zipf afferma che, se si prende un testo, che può essere più o meno lungo, e si procede a stilare una classifica delle parole sulla base della frequenza decrescente di apparizione, è possibile ricavare una costante moltiplicando il numero di posizione della parola nella classifica per il numero di volte in cui concretamente appare nel testo (Zipf 1949; De Bellis 2005 p. 34).

Si ottiene la formula:

$$r * f = k$$

dove r rappresenta il numero d'ordine della parola nella lista, f il numero di volte in cui la parola appare nel testo, k la costante. In altre parole, «la legge misura la significatività di una parola in funzione della frequenza con la quale appare nelle riviste di settore» (Di Cesare 2002 p. 134).

Per illustrare il suo lavoro, Zipf si servì di un indice delle parole dell'*Ulisse* di James Joyce. Dall'analisi del testo emerse che la decima parola nella classifica di frequenza di appari-

zione nel testo compariva 2653 volte, la parola in posizione 100 compariva 265 volte, quella in posizione 200 compariva 133 volte etc.

Questo significava che esiste un nucleo di parole molto frequenti e nuclei di parole molto meno frequenti. Zipf spiegò la sua legge con il principio del minimo sforzo con il quale affermò che le risorse si organizzano in modo da minimizzare il lavoro e che il 20-30% delle risorse produce il 70-80% dell'attività.

In un testo quindi l'autore tende a minimizzare il numero e la lunghezza delle parole e sfrutta invece la stessa parola per diversi significati. Il lettore del testo viceversa, tende alla diversificazione delle parole, assegnando significati diversi a parole diverse. Affinché la comunicazione possa essere efficace è necessario che queste due spinte divergenti trovino un equilibrio, dando origine così alla legge sulla distribuzione delle parole in un testo.

Nel corso degli anni, anche la legge di Zipf ha subito generalizzazioni, riformulazioni, reinterpretazioni e numerosissime applicazioni. Quello che sorprende però è che ancora non è stato spiegato perché essa esista e perché sia così diffusa.

1.4. Fondamenti teorici

La costruzione dei fondamenti teorici della disciplina si deve ai contributi di importanti studiosi come John Bernal, Robert Merton, Derek de Solla Price, Henry Small e Eugene Garfield.

Ad aprire la strada verso l'analisi quantitativa della scienza è John Bernal, il quale è convinto che la scienza sia una attività sociale controllabile mediante l'impiego di metodi matematici e statistici ai prodotti dell'attività di ricerca.

A questa premessa teorica si aggancia il pensiero di Robert Merton che, spingendosi oltre, arriva ad individuare le norme universali che governano l'operato degli scienziati e che fanno della ricerca scientifica un'attività lineare, perfettamente coerente e omogenea in tutte le sue fasi.

Lo schema normativo dato da Merton alla scienza è il punto di partenza per la teoria di Henry Small sulla citazione come simbolo del contributo di un documento. Small capisce che, se gli autori appartenenti ad uno stesso campo disciplinare condividono regole di condotta, linguaggio e schemi metodologici di ricerca, è piuttosto probabile che essi citino an-

che lo stesso documento per riferirsi allo stesso concetto. In questo modo, egli approfondisce la valenza cognitiva delle citazioni bibliografiche e, di riflesso, anche la funzione degli indici di citazioni sempre in senso cognitivo. Essi infatti saranno equiparati a indici di concetti scientifici utili per ricerche bibliografiche e per la mappatura delle relazioni sociali tra autori.

A Derek de Solla Price si deve il concetto di "riduzionismo bibliometrico": «la scienza si identifica con la letteratura scientifica prodotta» (Polanco 1995; De Bellis 2005 p. 49). Da ciò deriverà una sorta di legittimazione all'uso di strumenti bibliometrici per l'analisi quantitativa della letteratura pubblicata e, in particolare, della citazione, vista come segno dell'integrazione di un determinato soggetto nella comunità di riferimento.

E infine, ma non meno importante, il contributo dato da Eugene Garfield²³ con la creazione dello *Science Citation Index*, dello strumento cioè da cui è possibile estrarre concretamente i dati per la costruzione degli indicatori bibliometrici, primo fra tutti dell'*Impact Factor*.

1.4.1. John Bernal e la scienza come sistema materiale di relazioni sociali e simboliche

John D. Bernal (1901-1970) era un cristallografo di fama internazionale, docente di fisica e cristallografia presso l'Università di Londra e membro della *Royal Society*. Lo scienziato irlandese condusse importanti studi, nella stragrande maggioranza di carattere pionieristico, sulla diffrazione dei raggi X e sulla biologia molecolare, ma fu anche fondatore di una nuova disciplina, la 'scienza della scienza' e, ben presto, il suo libro, *The Social Function of Science* (Bernal 1939), divenne un classico in questo campo. La filosofia della scienza di Bernal si inserisce nel solco del materialismo dialettico di Engels applicato allo studio della natura e della società; la scienza era concepita da Bernal come un'attività direttamente collegata a tutte le altre attività umane (De Bellis 2005 p. 47). Per lo scienziato la filosofia, come anche la teoria sociale o, più in generale, la conoscenza, non può essere indipendente dalla scienza. Essa diventa la base di tutto. Pur essendo quindi il fondamento di ogni conoscenza, è comunque per Bernal un'attività sociale controllabile. È possibile pertanto con-

23 Alla figura e al contributo di Garfield sarà dato ampio spazio nella parte dedicata agli indici di citazioni (capitolo 4).

trollare l'avanzamento delle conoscenze scientifiche attraverso un'analisi di tipo quantitativo delle pubblicazioni scientifiche ed una programmazione dettagliata degli obiettivi che si intendono raggiungere. Nel capitolo *Scientific communication* (Bernal 1939), Bernal sintetizzò queste idee proponendo un meccanismo rivoluzionario di comunicazione scientifica basato sul principio della pianificazione centralizzata. Il suo sistema comprendeva la pubblicazione, il controllo bibliografico, la disseminazione, l'archiviazione, il coordinamento e la distribuzione dell'informazione scientifica. Propose inoltre la sostituzione delle riviste scientifiche individuali secondo i bisogni espressi dai fruitori della scienza e il superamento del regime monopolistico esercitato dalle società scientifiche sulla letteratura periodica (Bernal 1939 p. 300).

L'obiettivo di Bernal era creare un sistema efficiente dove tutta l'informazione scientifica avesse potuto essere disponibile per ogni ricercatore con un'ampiezza proporzionale al grado di rilevanza (Bernal 1939 p. 294; De Bellis 2005 p. 48), risolvendo così sia parte dei problemi connessi all'inadeguatezza dei sistemi tradizionali di indicizzazione sia di quelli legati alla crescita esponenziale della produzione scientifica.

Quando teorizzato da Bernal sarà in parte concretizzato da Garfield con la creazione dei *citation indexes* dell'ISI, che recupereranno proprio la sua idea di disseminazione flessibile dell'informazione legata dai servizi di *abstracting* e *indexing* convenzionali.

1.4.2. Robert Merton e la concezione normativa della scienza

La bibliometria e, più in generale, la scientometria devono parte del loro sviluppo alla sinergia con l'indirizzo metodologico avviato dalla sociologia della scienza già a partire dai primi studi degli anni '30 dello scorso secolo.

È comunque solo dopo il secondo dopoguerra che la sociologia si occupa di scienza come specifico oggetto d'indagine, questo è dovuto al ruolo cruciale svolto da alcuni scienziati e gruppi di ricerca proprio durante il conflitto mondiale che ha contribuito a rafforzare la convinzione che il potere politico e lo sviluppo economico dipendano dal contributo della scienza e della tecnologia.

Robert Merton (1910-2003) è considerato il fondatore della sociologia della scienza, perché è stato il primo sociologo a mostrare come la scienza sia un'istituzione sociale con una

sua propria struttura culturale e normativa. Egli indaga dunque l'organizzazione sociale della scienza, i valori e le norme che possono favorire od ostacolare il sorgere o lo stabilizzarsi della scienza, il ruolo sociale degli scienziati, lo sviluppo delle istituzioni scientifiche etc.

Nella sua tesi di dottorato, dal titolo *Science, Technology and Society in Seventeenth Century England* (1938), egli affronta in modo del tutto nuovo il rapporto tra attività scientifica e sviluppo del capitalismo. «La novità è rappresentata dal fatto che Merton si interessa allo sviluppo istituzionale della scienza in relazione alla diffusione di particolari valori religiosi e non all'impatto delle scoperte scientifiche e delle innovazioni tecnologiche sullo sviluppo economico e industriale» (Bucchi 2004 p. 578). Basandosi su dati storici concreti, Merton mette in luce come un numero sempre maggiore di individui provenienti dall'*élite* britannica si fosse dedicato alla scienza, e come una quota significativa dei loro lavori non fosse orientata ad alcuna finalità pratica (Merton 1938).

Le motivazioni all'attività scientifica dovevano quindi essere ricercate anche vagliando altri tipi di stimolo. Il tentativo di Merton è, dunque, evidenziare che l'attività scientifica e il ruolo dello scienziato non solo il prodotto di un insieme di attività e metodi condivisi da un gruppo di persone, bensì presuppongono l'esistenza di un insieme di valori e norme che caratterizzano tutta la comunità scientifica.

L'autore si sofferma sui meccanismi attraverso cui sono assegnate e distribuite all'interno della comunità scientifica risorse e ricompense, quali la possibilità di pubblicare e il prestigio, nonché gli aspetti organizzativi e funzionali della scienza in quanto istituzione, che definiscono nel loro complesso '*the social stability*' e '*the institutional integrity*' della scienza (Merton 1938; Bucchi 2004 p. 579).

Ma è in un suo articolo²⁴ del 1942 che Merton descrive il complesso normativo della scienza, ovvero i valori e le norme di condotta che garantiscono il funzionamento della produzione di conoscenza. Gli 'imperativi istituzionali' che costituiscono la base dell'*ethos of science* sono quattro: universalismo, comunitarismo, disinteresse e scetticismo organizzato (Bucchi 2004 p. 580; De Bellis 2005 p. 61-62). L'universalismo consiste nel fatto che i risultati scientifici vengono giudicati indipendentemente dalle caratteristiche del soggetto

²⁴ L'articolo, apparso per la prima volta nel *Journal of Legal and Political Sociology*, venne incluso nella raccolta di saggi *The Sociology of Science* (Merton 1973) e ristampato in traduzione italiana solo nel 1981 (Merton 1981).

che li ha formulati e dal luogo della scoperta. La scienza, infatti, rappresenta la ragione impersonale nella sua forma più compiuta: in altre parole, le verità della scienza non dipendono dalle caratteristiche personali o sociali dello scienziato, cioè dalla nazione di origine, dalla religione o dalla razza di appartenenza.²⁵

Il comunitarismo attiene al fatto che i risultati del singolo sono patrimonio della comunità scientifica e della società nella sua totalità, perché la scienza è un'attività sociale basata sugli sforzi pregressi e influenzata da quelli futuri, lo scienziato quindi può ottenere riconoscimenti e ricompense per la propria attività solo rendendola pubblica e mettendola a disposizione di tutta la società.²⁶

Il disinteresse delle ricerca scientifica risiede nel fatto che l'obiettivo primario del ricercatore è l'avanzamento della conoscenza, dal quale deriva indirettamente anche il proprio riconoscimento personale.

Infine lo scetticismo organizzato impone al ricercatore il dovere deontologico di valutare in modo critico qualunque risultato, inclusi i propri, formando il giudizio definitivo solo dopo aver raccolto tutte le prove necessarie.

Dalle norme di condotta individuate da Merton emerge quindi chiaramente perché gli scienziati hanno urgenza di pubblicare i propri contributi il più presto possibile. Essa risiede nel fatto che gli scienziati, comunicando al resto della comunità un avanzamento della conoscenza, stabiliscono pubblicamente la propria priorità nella scoperta. L'idea è che ogni scienziato aspiri ad essere riconosciuto dalla comunità dei pari come il primo ad aver raggiunto un certo risultato. Tale riconoscimento si concretizza in tre modi: l'eponimia, cioè l'associazione del nome dello scienziato al risultato da lui raggiunto; l'attribuzione di premi e onorificenza; e la citazione esplicita del lavoro di uno scienziato da parte di altri scienziati nelle loro pubblicazioni.

«La citazione bibliografica si configura quindi come un atomo di *peer recognition*» (De Bellis 2005 p. 62).

Il sistema del riconoscimento non è perfetto, tanto che in certe circostanze il meccanismo attraverso cui sono assegnate e distribuite risorse e ricompense all'interno della comu-

25 «L'imperativo dell'universalismo è profondamente radicato nel carattere impersonale della scienza» (Merton 1981 p. 352).

26 «...le scoperte sostanziali della scienza sono un prodotto della collaborazione sociale e sono assegnate alla comunità. Esse costituiscono una eredità comune in cui il diritto del produttore individuale è severamente limitato» (Merton 1981 p. 355).

nità scientifica spesso si inceppa e si trasforma in un circolo vizioso.

A tale cortocircuito Merton ha dato il nome di “effetto San Matteo”. L'espressione trae origine dal Vangelo secondo Matteo, dove si legge: «poiché a chi ha verrà dato, e sarà nell'abbondanza; ma a chi non ha, verrà tolto anche quello che ha» (Matteo 25, 29). Gli scienziati ricompensati una volta godono, rispetto ai colleghi anonimi, di un vantaggio che garantisce loro ulteriori ritorni in futuro in termini sia di accesso ai mezzi di produzione scientifica, sia di maggiore visibilità all'interno della comunità scientifica.²⁷

Le norme di Merton sono state importanti per lo sviluppo della bibliometria per il fatto che hanno contribuito a costruire una base di riferimento per la comprensione dei processi e dei comportamenti citazionali. Grazie a Merton infatti, le citazioni bibliografiche diventano un elemento controllabile e quindi in una certa misura prevedibile, proprio perché regolate da un insieme di norme che, per la produzione scientifica, ruota attorno al riconoscimento dei meriti individuali all'interno della comunità dei pari (De Bellis 2005 p. 65).

1.4.3. Derek de Solla Price e la scienza come sistema "termodinamico"

Derek J. De Solla Price (1922-1983), professore presso la Yale University oltre che specialista e storico della scienza dell'informazione, è considerato il padre della scientometria. I suoi studi sulla crescita esponenziale della scienza, degli scienziati e degli articoli scientifici furono importanti per lo sviluppo degli indicatori bibliometrici e, più in generale, per la creazione di un modello di analisi dell'attività scientifica. Price utilizzò le citazioni bibliografiche per analizzare e mappare non solo gli ambiti scientifici con le loro strutture e la loro evoluzione nel tempo, ma anche per definire l'organizzazione e la distribuzione delle fonti scientifiche di un determinato settore disciplinare. Esse vennero impiegate inoltre per formulare le teorie sul vantaggio cumulativo e sulla collaborazione informale all'interno dei "collegi invisibili" (*invisible colleges*).

Il occasione del VI Congresso Internazionale di Storia della Scienza (1950), Price formulò la tesi che il numero di articoli pubblicati in una rivista scientifica in un determinato arco di tempo costituisce un buon indicatore del livello di produttività scientifica del setto-

²⁷ «Un contributo scientifico avrà ulteriore visibilità nella comunità degli scienziati quando è introdotto da uno scienziato di alto profilo rispetto a quando è presentato da uno scienziato che non ha ancora lasciato il segno» (Merton 1973 p. 447).

re disciplinare oggetto della rivista esaminata.²⁸ Price capì che la scienza era in una fase di trasformazione e espansione tale da rendere necessari metodi di analisi nuovi e più efficaci rispetto ai tradizionali, capaci di dare risposte più convincenti ai problemi di organizzazione, amministrazione e controllo politico della scienza. Il titolo *Little Science, Big Science* di un suo libro del 1963 è eloquente da questo punto di vista: «egli definisce "piccola scienza" quella dei tempi di Galileo, che vedeva un uomo solo, chiuso nel suo laboratorio, con qualche sovvenzione di un mecenate o di una organizzazione, che inventava, produceva, scopriva, cioè "creava" la scienza. La "grande scienza" è invece quella dei *teams* di ricerca, dei gruppi collettivi, finanziati in base a stanziamenti enormi, che lavora per creare un cambiamento delle condizioni planetarie» (Price 1967 p. 36 e segg.). Price riteneva che fosse giunto il momento di applicare alla "grande scienza" le stesse discipline d'indagine che essa applicava alla natura, di utilizzare cioè metodologie come la statistica, allo scopo di far emergere i meccanismi interni che ne regolavano l'avanzamento e il funzionamento complessivo.

Per spiegare la sua tesi, lo scienziato utilizzò la metafora del comportamento del gas nella termodinamica come modello di riferimento per l'applicazione della scienza alla scienza. La termodinamica studia il comportamento del gas sotto determinate condizioni di temperatura e pressione, ma il suo interesse non si focalizza sulla traiettoria di una specifica molecola in un dato istante. Al contrario, essa analizza i fenomeni nel loro complesso. Price suggerì quindi di studiare la scienza in un modo simile: non focalizzare l'attenzione sulla singola "molecola", cioè sul singolo scienziato o sulla singola scoperta, ma utilizzare invece le statistiche su scienziati, articoli e citazioni per analizzare la scienza nel suo complesso, per capire cioè il "volume" della scienza, la traiettoria delle "molecole" nella scienza, il modo in cui queste interagiscono tra di loro, fino ad arrivare alle proprietà politiche e sociali di questo "gas"-scienza (Price 1963 prefazione, pp. xv-xvi).

La sua teoria sulla crescita esponenziale della scienza venne formulata dopo un attento lavoro di classificazione e conteggio di un numero molto elevato di articoli tratti dalle *Philosophical Transactions of the Royal Society of London* che vanno dal 1665 al 1850. Secondo Price la scienza non cresce in modo lineare, ma esponenziale, vale a dire che si moltiplica in

28 «Since the usual manner of recording a contribution to scientific knowledge is through the medium of the scientific paper published in some learned journal, one might expect that the number of papers appearing each year would be a usual barometer indicating the amount of activity, during that year, and over the range of subject-matter from which a count had been made» (Price 1951 p. 86).

eguali intervalli temporali, di un fattore costante. Uno dei dati forniti da Price riguardava il fatto «che circa l'87% degli scienziati di tutti i tempi risultava in attività, passando da circa 50.000 di fine '800 a oltre un milione. Analogamente, il numero delle riviste scientifiche era passato da circa 100 nel 1830 a diverse decine di migliaia; la quota del PIL destinata alla ricerca scientifica negli USA era passata dallo 0,2% del 1929 al 3% dei primi anni '60... [la scienza] insomma era cresciuta in misura esponenziale raddoppiando le proprie dimensioni approssimativamente ogni quindici anni. Per dare un'idea della rapidità di questa crescita, Price la metteva a confronto con altri fenomeni quali l'aumento della popolazione terrestre, il cui tempo di raddoppiamento si aggirava attorno ai cinquant'anni. "Ad ogni raddoppio della popolazione sono corrisposti almeno tre raddoppi della popolazione degli scienziati" (Price 1963 p. 12)... Da questi dati, Price traeva due ordini di considerazioni. Il primo ridimensionava il ruolo, spesso enfatizzato, della seconda guerra mondiale nello sviluppo dell'attività scientifica. Il tasso di crescita, infatti, era rimasto sostanzialmente identico negli anni precedenti e in quelli immediatamente successivi alla guerra... Il secondo ordine di considerazioni era in realtà una previsione. A meno di una drastica riorganizzazione dei suoi fondamenti, il processo di crescita esponenziale della scienza sarebbe inevitabilmente andato incontro a un limite massimo» (Bucchi 2002).

Sempre traendo spunto da queste considerazioni, Price formulò una legge che può essere considerata una variante della legge di Lotka. Secondo lo studioso, il numero di scienziati produttivi in un determinato settore disciplinare, che producono cioè circa la metà degli articoli pubblicati, può essere ricavato calcolando la radice quadrata della somma di tutti gli scienziati. La ricerca scientifica quindi è in mano a pochi scienziati, nell'ordine di qualche centinaio, e a poche riviste scientifiche, nell'ordine di qualche migliaio, la parte rimanente di quest'ultime costituisce semplicemente puro "rumore di sfondo" (Price 1965 p. 515). Il fatto che il fronte attivo della scienza si concentri in un nucleo ristretto è per Price riconducibile ad un fenomeno molto più ampio, che già Merton aveva chiamato "effetto San Matteo", mediante il quale il "successo genera successo", vale a dire che la ricchezza tende ad aumentare polarizzandosi su pochi soggetti.²⁹

²⁹ «It is common in bibliometric matters and in many diverse social phenomena, that success seems to breed success. A paper which has been cited many times is more likely to be cited again than one which has been little cited. An author of many papers is more likely to publish again than one who has been less prolific. A journal which has been frequently consulted for some purpose is more likely to be turned to again than one of previously infrequent use» (Price 1976 p. 292).

Nel 1976, sulla scorta di tale assunto, dal quale discendono le stesse leggi di Lotka, Bradford e Zipf, egli propose la teoria probabilistica generale dei vantaggi cumulativi (*Cumulative Advantage Distribution*). Tale teoria rappresenta un modello concettuale unificato delle leggi bibliometriche in cui i singoli elementi (distribuzione degli articoli, frequenza delle citazioni per articolo e delle parole in un testo) sono considerati in modo cumulativo, «rafforzando in tal modo la consuetudine che si era già affermata con la legge di Bradford e quella di Zipf, le quali venivano già di frequente associate, tanto che si parla oggi di *Bradford-Zipf distribution*» (Di Cesare 2002 p. 134).

La tesi di Price trovò ulteriore sviluppo grazie ai *citation indexes* di Garfield. Price infatti era convinto che fosse possibile, analizzando la rete delle citazioni, ricreare la struttura socio-cognitiva di un settore disciplinare fino ad arrivare, attraverso la rappresentazione grafica, addirittura alla creazione di vere e proprie mappe disciplinari (De Bellis 2005 p. 52). In sostanza Price anticipò teoricamente quello che concretamente sarà negli anni '70 il progetto che Small e Griffith avvieranno utilizzando l'analisi delle co-citazioni.

Le analisi quantitative dello *SCI* furono importanti per Price anche perché gli consentirono di trarre conclusioni che si dimostrarono importanti per lo sviluppo degli studi successivi sulle citazioni bibliografiche. Price notò che in un dato anno il 50% dei riferimenti bibliografici di un corpo selezionato di letteratura si riferisce ad un numero ristretto di pubblicazioni, mentre l'altro 50% si distribuisce su una quantità molto maggiore di letteratura prodotta negli anni precedenti. Lo *SCI* poteva quindi essere utilizzato anche per stilare una lista delle pubblicazioni più importanti ("*classics*" e "*superclassics*") di un determinato settore disciplinare (Price 1965 p. 512). Egli osservò inoltre che se le citazioni indicano la natura di un campo scientifico, su un campione selezionato il 35% degli articoli non veniva mai citato, il 49% aveva ricevuto una sola citazione, il 16% aveva circa tre citazioni, mentre solo l'1% ne aveva ricevute più di sei (Price 1965 p. 511). Infine scoprì che solo il 30% delle citazioni si riferiscono alla letteratura recente formata da articoli con età compresa tra 1 e 6 anni (Price 1965 p. 513). Quest'ultima considerazione, cioè la tendenza a citare soprattutto documenti recenti, è una caratteristica propria delle "scienze dure" rispetto alle altre discipline e conferma, a sua volta, l'esistenza di un nucleo ristretto che costituisce il fronte più impegnato e produttivo di una area disciplinare formato da quelli che Price definisce "col-

leggi invisibili" (Price 1963 p. 56-81 e p. 119-134), cioè le reti permanenti di comunicazione che si formano in seno a una disciplina scientifica e coinvolgono un gruppo ristretto di studiosi. Tale rete è anch'essa una diretta conseguenza dell'effetto della moltiplicazione delle pubblicazioni e della saturazione dei canali di informazione nelle istituzioni scientifiche: gli studiosi tendono a reagire creando proprio questi rapporti interni a una disciplina scientifica.

Nonostante le molte critiche avanzate al modello termodinamico e alle leggi di Price (Holton 1979; Tague-Sutcliffe – Beheshti – Reespotter 1981; Glänzel – Schubert 1985; Nicholls 1988), il contributo dello studioso è considerato determinante per lo sviluppo degli indici di citazioni, ma anche per il loro impiego nel campo della sociologia della scienza e della valutazione della ricerca scientifica.

1.4.4. Henry Small e le citazioni bibliografiche come simboli di concetti

Henry Small, membro dello staff dell'*Institute for Scientific Information* (ISI) dal 1972, ha condotto studi pionieristici nel campo dell'analisi delle citazioni, in particolare ha sviluppato ulteriormente le possibilità offerte dalle stesse fino ad arrivare ad impiegarle come strumento nella mappatura della scienza.

In un articolo del 1978 (Small 1978) ha sostenuto la funzione cognitiva della citazione, vale a dire il ruolo che essa gioca come simbolo di concetti o metodi. In un contributo scientifico, quando un autore decide di inglobare un riferimento bibliografico, non fa altro che citare un lavoro precedente che costituisce né più né meno la rappresentazione di un'idea trattata dall'autore stesso. Più in generale, i documenti citati diventano "simboli" di questa idea.

Small nell'illustrare la sua concezione del documento-simbolo fa un chiaro riferimento all'opera *Culture and Communication* (1976) dell'antropologo inglese Edmund Leach (1910-1989), in cui l'autore applica l'analisi strutturalista all'antropologia sociale. Per Leach la cultura è, come il linguaggio, un sistema di segni che deve essere decodificato all'interno del proprio contesto per essere realmente compreso, per capirlo cioè bisogna conoscerne il codice fatto di segnali, segni e simboli. Nel segno, la correlazione tra significante (forma visiva o espressiva) e significato (contenuto, oggetto) è contigua, perché esso ha la funzione di

indicare, cioè sta al posto di ciò che indica, o meglio: lo rappresenta. Nel simbolo invece, la correlazione tra significante e significato non è contigua, perché esso non indica una realtà distinta da se stesso, ma la contiene nella sua completezza, nell'esprimere il significato esso stesso diventa il significante. Esiste poi una distinzione tra il simbolo "occasionale" e quello "standardizzato". Il primo ha un valore individuale, cioè comunica informazione soltanto a chi lo esprime. Il secondo invece ha un valore universale, cioè comunica informazioni di pubblico dominio, comprensibili a tutti.

Applicando lo schema e la terminologia di Leach alla citazione bibliografica, Small la considera contemporaneamente sia un segno che un simbolo. È un segno che sta per un documento con cui condivide alcune caratteristiche; ed è anche un simbolo dei concetti, delle idee, delle teorie espresse all'interno del documento.

Quando si considerano poi le citazioni come simboli, anch'esse possono essere di tipo "occasionale", cioè utilizzate dall'autore come simboli personali per esprimere i propri concetti, o "standardizzate" quando invece rimandano a documenti, il cui contenuto, e quindi il loro significato, è conosciuto e condiviso da più autori. Naturalmente, i simboli "standardizzati" sono il frutto di un processo di selezione da parte di molti individui lungo un periodo di tempo. Questa dimensione sovraindividuale però non esclude la possibilità data al singolo autore di possedere un personale repertorio di tali concetti collettivi e dei corrispondenti documenti-simbolo (Small 1978 p. 329).

Citando un documento lo scienziato lo etichetta e gli attribuisce un proprio significato, che può anche discostarsi del tutto o in parte da quello conferitogli dall'autore, adattandolo al contesto in cui la citazione viene inserita.³⁰

Per lo scienziato tuttavia il caso più frequente di citazione è ancorato a una sorta di consenso normativo, quando cioè il significato espresso dall'autore citante è coincidente con quello dell'autore citato o con l'uso fatto dalla comunità scientifica di appartenenza (Small 2004 p. 76).

La corrispondenza tra concetti e citazioni di Small ha avuto una grande importanza per lo sviluppo dello *Science Citation Index*. Essa ha rappresentato la legittimazione all'uso di questo strumento per ricavare «informazioni utili a valutare in termini di impatto cogniti-

³⁰ «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» (Small 2004 p.75).

vo il numero di citazioni ricevute da un documento» (De Bellis 2005 p. 60). Guardando infatti lo *SCI* sotto l'ottica dell'equivalenza stabilita da Small, esso acquisisce una duplice valenza: quella di indice di citazioni e quella di «indice di concetti scientifici equiparabile ad un tesoro disciplinare» (De Bellis 2005 p. 60).

1.5. Applicazioni

1.5.1. La cartografia della scienza

La costruzione di mappe dettagliate della scienza, delle specialità scientifiche, dei "fronti attivi" di ricerca e di sviluppo, e dei temi ricorrenti è oramai considerata un affermato e fecondo campo di ricerca e applicazione delle bibliometria.

La mappatura della scienza, avviata da Garfield negli anni '60 dello scorso secolo e successivamente sviluppata da Small e Griffith «sulla scorta del riduzionismo bibliometrico di Price e della fiducia mertoniana nel carattere normativo delle pratiche citazionali» (De Bellis 2005 p. 96), ha trovato pieno sviluppo soltanto a partire dai primi anni '90, in coincidenza cioè, della messa a punto di comode interfacce grafiche a disposizione dell'utente. È stata infatti l'invitante possibilità di utilizzare una mappa come interfaccia, per reperire informazioni o investigare le strutture della scienza così generate, che ha contribuito in modo considerevole alla riscoperta di questo campo d'applicazione della bibliometria.

Come già ricordato, il primo a sviluppare concretamente l'idea della mappatura cronologica della scienza fu Eugene Garfield. La sua idea era semplice e poteva essere concretizzata in due fasi distinte ma complementari. Innanzitutto era necessario costruire una mappa con nodi e linee (grafo) rappresentativi delle fasi cruciali di un importante episodio della storia della scienza, formate nel concreto dagli autori e dai documenti "chiave" estratti, in precedenza, da una ricostruzione esauriente di tale episodio. La seconda fase era la sovrapposizione della mappa così creata ad un'altra, costruita, questa volta, sulle connessioni bibliografiche tra i documenti "chiave" identificati, estratti dallo *Science Citation Index* (Garfield 1964). L'esperimento diede buoni risultati e successivamente la procedura venne automatizzata con la messa a punto del software *HistCite*, che ha permesso di visualizzare graficamente la distribuzione spaziale dei nodi e dei *link* corrispondenti ai documenti più

citati e ai legami bibliografici che li connettono. Negli anni '70 del XX secolo, utilizzando la tecnica delle co-citazioni, saranno Small e Griffith a sviluppare ulteriormente quest'applicazione bibliometrica. Lo scopo era quello di disegnare una mappa della scienza capace di indicare visivamente non solo i documenti fondamentali di un settore disciplinare, ma anche le variazioni diacroniche degli stessi in concomitanza di eventuali trasformazioni concettuali interne a una disciplina scientifica (Small – Griffith 1974).

Una mappa della scienza è una rappresentazione in due e tre dimensioni di un campo di ricerca, un "*landscape of science*" (Noyons 2004 p. 238), i cui elementi non sono nient'altro che gli argomenti o i temi del campo mappato, come lo sono le città su una cartina geografica. In queste mappe gli elementi sono posizionati in relazione agli altri: la vicinanza o la lontananza tra gli elementi nella cartina corrisponde al grado di "prossimità" o "similitudine" concettuale degli elementi oggetto della mappa. Il fondamento teorico su cui poggia la costruzione delle mappe della scienza è quello concepito da Price (Price 1963): la scienza si identifica con la letteratura scientifica pubblicata, e quindi una mappa basata sui dati delle pubblicazioni scientifiche appartenenti a un determinato campo di ricerca è capace di rappresentare la struttura stessa di quel settore disciplinare. Naturalmente, molto dipende da quali informazioni, cioè da quali elementi di un record bibliografico si utilizzano per la costruzione della mappa, e dal tipo di struttura che si vuole generare.

Per la costruzione di tali mappe la bibliometria utilizza il principio della co-occorrenza di elementi bibliografici: «più due elementi compaiono assieme nello stesso documento, maggiore la probabilità che essi abbiano una correlazione strutturale con i concetti caratteristici del campo e che la loro posizione sulla superficie della mappa sia rappresentata da punti vicini» (De Bellis 2005 p. 92-93).

Mappe di questo tipo si possono costruire utilizzando diversi elementi di un record bibliografico. Poiché ogni elemento coglie un aspetto specifico della pubblicazione, esso sarà la fonte su cui si svilupperà uno specifico tipo di struttura che, pur essendo unica, è comunque strettamente collegata alle strutture generate da altri elementi, e questo perché tutti i dati si estrapolano da un'unica fonte: la pubblicazione scientifica. Una mappa costruita estraendo da un corpo di letteratura selezionata le co-occorrenze degli autori dei documenti è molto probabile restituisca molto della conformazione sociale di un settore di-

sciplinare (Peters – Van Raan 1991), in quanto la frequenza di tali co-occorrenze riesce a comunicare molto delle relazioni che si instaurano tra gli scienziati appartenenti ad uno stesso ambito di ricerca, fino a ricostruire la struttura di quelli che Price definì "collegi invisibili".

Una mappa costruita invece sulle co-occorrenze di termini (*co-word analysis*), ricavati per esempio dal titolo o dall'*abstract*, è capace di rappresentare meglio di altri tipi di mappa il contenuto dei documenti e, di conseguenza, anche la struttura cognitiva di un campo scientifico (Courtial – Callon 1984). Essa fornisce informazioni su: gli argomenti "caldi" (*core topics*), cioè i fronti attivi di ricerca in un determinato ambito scientifico; gli argomenti "emergenti; la rete di correlazioni tra argomenti "caldi", "secondari" ed "emergenti" che si nasconde sotto la superficie retorica del testo.

Ma per costruire mappe bibliometriche sono le co-occorrenze di citazioni ad essere le più utilizzate. Infatti, se la citazione può essere considerata simbolo di concetti o metodi (Small 1972) e regolata da un'etica professionale (Merton 1973), «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» (De Bellis 2005 p. 93). Tali mappe infatti consentono di individuare concetti, metodi e autori "chiave" più d'interesse per gli scienziati di un dato settore disciplinare e quindi di tracciare i confini concettuali tra discipline, specialità e aree di ricerca.

Mappe di questo tipo non sono comunque immuni da critiche,³¹ le quali ruotano soprattutto attorno al diverso peso che le citazioni possono avere a seconda della disciplina oggetto d'indagine, all'esistenza di documenti troppo citati o co-citati, all'intrinseca interdisciplinarietà di alcune discipline, o infine all'instabilità di molti fronti di ricerca che sembrano variare continuamente.

Dagli anni '90 nuove tecniche di mappatura e visualizzazione³² si sono affiancate a quelle tradizionali, spesso hanno aggiunto caratteristiche quali la tridimensionalità e l'interattività alle mappe che così sono diventate più facili da navigare e più adatte al recupero mirato di specifiche informazioni (De Bellis 2005 p. 110).

31 A questo proposito cfr. Edge (1977) e Hicks (1987).

32 Per una rassegna di tali tecniche vedi Börner – Chen – Boyack (2003).

Degli ultimi anni, tuttavia, la novità sicuramente più rilevante è stata l'integrazione delle tecniche di mappatura della scienza con i metodi messi a punto all'interno del processo di valutazione della ricerca (Noyons 2004 p. 245).

1.5.2. La valutazione della ricerca

La bibliometria essendo una disciplina basata sull'analisi quantitativa delle pubblicazioni è diventata, soprattutto a seguito dello sviluppo e dell'implementazione delle banche dati citazionali, un efficace strumento nella valutazione della ricerca. È noto che la pubblicazione scientifica non rappresenta solo un contributo all'avanzamento della conoscenza in un determinato campo, ma sia anche uno strumento capace di attivare il meccanismo premiale all'interno di una comunità scientifica. È attraverso di essa infatti che l'autore può stabilire pubblicamente la propria priorità in una scoperta e, di conseguenza, ottenere il riconoscimento e la legittimazione dei risultati raggiunti da parte della comunità dei pari e, indirettamente, anche dal decisore pubblico che ha il compito di allocare le risorse disponibili alle ricerche più promettenti dal punto di vista sociale.

Per secoli, la valutazione della qualità della letteratura scientifica è stata ritenuta di competenza esclusiva della comunità dei pari, cioè di un gruppo di esperti nella disciplina in cui rientra la pubblicazione, dominio quindi di quello che viene anche definito sistema di *peer reviewing*. Come per ogni procedura che prevede un giudizio finale, tale metodo poggia sul presupposto della totale imparzialità dei revisori e sulla necessità di una procedura obiettiva e non gravata da giudizi personali. Proprio perché l'attendibilità del metodo dipende da queste premesse, esso non è, come è facile immaginare, immune da critiche, anzi tra gli stessi scienziati si è fatta strada la consapevolezza che il «sistema dei *referees* è viziato da distorsioni, pregiudizi, conflitti di interesse, errori di valutazione ampiamente documentati nella letteratura critica» (De Bellis 2005 p. 113).

La difficoltà di trovare l'equilibrio tra l'esigenza di imparzialità del metodo e il comportamento opportunistico o i problemi di soggettività che a volte si riscontrano nei revisori, ha spinto il mondo della scienza a ricercare metodi di valutazione alternativi al *peer reviewing*. Inoltre, a partire dalla seconda metà del XX secolo, la necessità sempre più urgente per il decisore pubblico di avere informazioni sulla qualità della ricerca, ha spinto allo

sviluppo sistematico di metodi formali di misurazione e valutazione della qualità.³³ Come ben spiegato da Baccini (2010 p. 34) «l'obiettivo della politica della ricerca contemporanea può essere indicato nel disegno di norme e meccanismi istituzionali in grado di favorire lo sviluppo della ricerca di qualità. Il fine ultimo di questi meccanismi è l'allocatione delle risorse alle ricerche più promettenti... [che] dovrebbe avvenire attraverso meccanismi in grado di spingere il ricercatore e i gruppi di ricerca a condurre la ricerca al meglio delle loro possibilità, cioè in modo efficiente». Per gli scopi perseguiti dalla politica della ricerca diventa così fondamentale poter acquisire dalla comunità scientifica informazioni il più possibile corrette, frutto di una valutazione obiettiva ed estranee a interessi di parte.³⁴ È proprio in questo contesto che l'impiego di strumenti bibliometrici basati sull'analisi delle citazioni ha cominciato a ricevere grande attenzione da politici e manager della scienza. Rispetto alla valutazione basata sul giudizio dei pari, quella condotta con indicatori bibliometrici³⁵ è più veloce e meno costosa, i dati poi su cui poggia sono facilmente estraibili dagli indici di citazioni, gratuiti o a pagamento, disponibili online.

In Italia, l'interesse per la valutazione della ricerca, cresciuto in modo esponenziale negli ultimi dieci anni sulla scia degli decisioni prese a livello europeo con la *Bologna Declaration* del 1999 (per la didattica) e con la strategia di Lisbona del 2000 (per la ricerca), si è tradotto in un massiccio intervento normativo.

Nel 2004, istituito con Dlg. 204/98 presso il Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), diventa operativo il Comitato di indirizzo per la valutazione della ricerca (CIVR) che inaugura il primo esercizio triennale di valutazione della ricerca per gli anni 2001-2003 (VTR).³⁶ Istituita con D.L. n. 262/2006, ma in attesa di attivazione è la nuova Agenzia Nazionale di Valutazione dell'Università e della Ricerca (ANVUR). Nel marzo 2010 il CIVR ha avviato il secondo esercizio di valutazione quinquennale della ricerca (VQR) per gli anni 2004-2008.³⁷ Tale organismo è composto da un gruppo di «esperti con elevata competenza in campo scientifico, sociale e produttivo, a cui è affidato il compito

33 Sulle origini dei metodi di valutazione della ricerca basata su indicatori quantitativi si rimanda a Cole (2000) e alla bibliografia ivi citata.

34 La valutazione della ricerca sfrutta anche altri criteri come: l'analisi del beneficio economico, basata sulla convenienza finanziaria ed economica dei progetti di ricerca; e il grado di internazionalizzazione delle pubblicazioni, cioè la percentuale di contributi scientifici pubblicati su riviste internazionali. A questo proposito cfr. De Robbio (2008 p. 7-8).

35 Gli indicatori bibliometrici saranno oggetto di un'ampia e approfondita trattazione nel capitolo 5.

36 Sull'esperienza del VTR vedi Reale (2008).

37 Sull'iter normativo dell'ANVUR si rimanda a De Robbio (2009 p. 105-107).

fondamentale di promuovere l'attività di valutazione della ricerca attraverso l'abrogazione di omogenei criteri di valutazione delle attività e dei risultati». ³⁸ Il decreto del MIUR del 2009 sulla "Valutazione dei titoli e delle pubblicazioni scientifiche" ³⁹ dei candidati nell'ambito delle procedure di valutazione comparativa per il reclutamento dei ricercatori universitari prevede sia una valutazione di tipo *peer reviewing* (art. 3, comma 2), basata su criteri puntualmente elencati, sia una valutazione basata su indicatori bibliometrici (art. 3, comma 4), come il numero totale di citazioni, il numero medio di citazioni per pubblicazione, "impact factor" totale, "impact factor" medio per pubblicazione, e combinazioni dei precedenti parametri atte a valorizzare l'impatto della produzione scientifica del candidato (indice di *Hirsch* o simili). La valutazione introdotta dal MIUR per le pubblicazioni scientifiche è quindi di tipo "misto", cioè basata su un approccio sia qualitativo (giudizio dei pari) che quantitativo (impiego di indicatori bibliometrici), il cui utilizzo, nella misurazione del produzione del sapere scientifico, è largamente auspicato e incoraggiato.

Gli strumenti bibliometrici basati sull'analisi citazionale possono essere utilizzati sia per la valutazione dei documenti, che per la valutazione di periodici, che per la valutazione di persone, che per la valutazione di gruppi e istituzioni, ed infine per la valutazione di nazioni. Generalmente, si usa distinguere tre livelli a cui può essere applicato l'esercizio valutativo: al livello "macro" dei sistemi nazionali; al livello "meso" delle istituzioni, incluse le singole università; e infine al livello "micro" dei gruppi di ricerca o del singolo ricercatore (De Robbio 2009 p. 105-106).

Quanto appena illustrato però è applicabile alle discipline appartenenti al mondo delle cosiddette "scienze dure". L'applicazione di strumenti bibliometrici per la valutazione della ricerca al mondo delle scienze sociali e umane può essere problematica o produrre distorsioni nei risultati. In queste ultime infatti la forma di comunicazione più diffusa è la monografia specialistica, gli autori tendono ad adottare forme di citazione diverse a seconda della natura della pubblicazione, tanto che questa tendenza si spinge fino a creare due mondi separati: i libri citano prevalentemente altri libri e le riviste citano prevalentemente le riviste. Ne consegue che gli indicatori citazionali costruiti a partire dai libri e dalle riviste danno luogo a risultati diversi. Inoltre, nelle scienze sociali e umane i concetti espressi

³⁸ <http://www.ricercaitaliana.it/valutazionericerca.htm>

³⁹ Decreto Ministeriale 28 luglio 2009 prot. n. 89/2009 Valutazione dei titoli e delle pubblicazioni scientifiche <www.miur.it/0006Menu_C/0012Docume/0015Atti_M/7921Valuta_cf2.htm>.

non riescono ad essere comunicati in un linguaggio unificante come quello della matematica, per cui gli studiosi preferiscono esprimersi nella loro lingua madre. L'esistenza di una letteratura nazionale dà luogo anche a citazioni riferite alla letteratura nazionale (Hicks 2004). Viste queste problematiche anche i dati ricavati da indici di citazioni come quello dell'*ISI* si rivelano inadatti, poiché sono specializzati nell'indicizzazione di citazioni provenienti da riviste e non da monografie. Allo scopo di superare questi ostacoli, per le scienze sociali e umane, Moed e Torres-Salinas (2009) hanno proposto l'uso dell'analisi dei cataloghi online delle biblioteche (*LCA Library Catalog Analysis*) in alternativa all'analisi citazionale che meglio si adatta alle "scienze dure". La LCA ha lo scopo di descrivere quantitativamente un settore disciplinare partendo però dai titoli dei libri presenti nei cataloghi. Come ben spiegato da De Robbio (2009 p. 115) il lavoro di Moed e Torres-Salinas «propone un modello analogo alla tradizionale analisi citazionale effettuata per gli articoli dai periodici applicata ai libri, agli OPAC e illustra come la tecnica di mappatura tematica possa essere messa a frutto quale potente strumento per la valutazione delle monografie come produzioni intellettuali di ricerca a livello di singolo ricercatore, di dipartimento o come intera produzione di un Paese o di un editore».

Partendo quindi da queste considerazioni, appare chiaro che la scelta degli obiettivi e della metodologia valutativa varia in relazione al livello valutativo adottato e alla disciplina di ricerca indagata. L'applicazione della bibliometria alla valutazione della ricerca comporta quindi la conoscenza di molti aspetti tecnici necessari per un uso corretto delle stesse tecniche bibliometriche. Purtroppo però, essi sono spesso trascurati nel dibattito accademico e di politica della ricerca, e forse, indirettamente, tale assenza è anche all'origine di molti degli abusi degli strumenti bibliometrici riscontrati appunto in questo campo d'applicazione.

A tal proposito, da più parti si stanno sollevando molte obiezioni all'impiego stesso della bibliometria a fini valutativi,⁴⁰ soprattutto se ad essere valutata è la produttività scientifica, qualitativa e quantitativa, di ciascun studioso. Se l'impiego di indici bibliometrici, come il semplice conteggio, è considerato un efficace mezzo per l'analisi dell'impatto della produzione scientifica di università, istituzioni o gruppi di ricerca, esso non è tale, o comun-

⁴⁰ Molti dei punti critici che animano questo dibattito sono qui soltanto accennati, saranno sviluppati parte nella sezione relativa agli indicatori bibliometrici, di cui si analizzeranno punti di forza e criticità (capitolo 5) e parte nella sezione dedicata agli indici di citazioni (capitolo 4).

que necessita di «ponderate compensazioni, sul piano granulare della performance individuale» (Van Raan 2005; Cassella-Bozzarelli 2011 p. 69). Le critiche poi non sono rivolte soltanto agli indicatori bibliometrici, ma si spingono oltre, fino ad investire l'analisi citazionale nel suo complesso. Il comportamento citazionale è influenzato da molti fattori, primo fra tutti, il tempo. Capita così, e studi empirici lo hanno dimostrato, che sono maggiori per un articolo le possibilità di venire citato se la sua data di pubblicazione è vicina a quella del documento che effettua la citazione; oppure, come già dimostrato da Merton con il principio "il successo genera successo", un articolo che in passato ha già ricevuto citazioni ha maggiori probabilità di riceverne altre; a parità di qualità poi, l'articolo di un autore con più anni di carriera alle spalle ha maggiori possibilità di ricevere un maggior numero di citazioni rispetto a quello di un autore "entrante" o in formazione. Inoltre, ad influenzare i comportamenti citazionali ci sono, oltre all'anzianità, il prestigio di un autore, l'appartenenza o meno ad una importante istituzione scientifica, la lingua in cui si esprime, e perfino una caratteristica personale come il genere sembra in questo caso incidere. Anche alcune caratteristiche della rivista, come il grado di internazionalizzazione, l'accessibilità, la periodicità regolare e frequente e l'ordine editoriale attribuito ad un articolo all'interno della pubblicazione, possono influire sul comportamento citazionale. A questi fattori si deve inoltre aggiungere l'impiego fraudolento dell'autocitazione. È questo il caso di un ricercatore che nello scrivere un articolo cita in modo sproporzionato i suoi articoli precedenti, creando così una lunga coda di autocitazioni che possono gonfiare artificialmente i dati citazionali.

Ma il dibattito critico non si ferma qui e, soprattutto negli ultimi anni, un'accesa polemica si è sviluppata attorno all'impiego del più noto tra gli indicatori citazionali: l'*Impact Factor*.⁴¹ Il limite più grande dell'*IF* è da più parti considerato proprio il suo impiego nella valutazione della ricerca. L'indicatore infatti nasce per fini diversi da quelli valutativi (è stato creato per realizzare un *ranking* di riviste), ma invece viene applicato, senza correttivi, proprio per valutare l'attività di ricerca ed erogare finanziamenti.⁴² Anche il fatto che l'*IF* venga applicato senza tener conto del contesto disciplinare di riferimento suscita molte criti-

41 Per una visione chiara ed esaustiva dei limiti inerenti all'*IF* si segnalano in lingua italiana: Figà Talamanca (2000); Tammaro (2001); Valente – Luzi (2004); De Robbio (2007 p. 13-14); Morriello (2007); Baccini (2010); Cassella – Bozzarelli (2011).

42 Lo stesso creatore dell'*IF* critica questo abuso dell'indicatore cfr. Garfield (1996 e 2000).

che: riviste afferenti a settori disciplinari generici presentano fattori d'impatto più elevati rispetto agli specialistici. Lo stesso vale per le riviste caratterizzate da un alto grado di internazionalizzazione rispetto a quelle limitate al contesto nazionale, o comunque, con un scarso grado di internazionalizzazione. E poi c'è il fattore temporale. L'*IF* viene calcolato su un arco temporale di due anni e questo risulta particolarmente penalizzante almeno in due contesti: in quelle discipline, come la matematica e le scienze sociali, la cui durata delle citazioni si prolunga nel tempo; e in quelle riviste dotate di severe e lunghe procedure di *refereeing*.

Nel panorama italiano, in prima fila in questo dibattito troviamo il matematico italiano Alessandro Figà Talamanca che critica non solo l'impiego dell'*IF* per scopi valutativi, ma anche la politica di "marketing" dell'*ISI* (oggi *Thomson-Reuters*). A questo proposito è famoso il suo discorso al convegno SINM del 2000 (Figà Talamanca 2000), nel quale contesta all'*ISI* la volontà di «favorire i grandi editori commerciali, a scapito delle pubblicazioni legate ad istituzioni scientifiche, specialmente quelle dei paesi europei non di lingua inglese, di far lievitare i costi delle pubblicazioni». Figà Talamanca denuncia inoltre l'opera di convincimento perpetrata dall'*ISI* nei confronti delle biblioteche scientifiche di università e di centri di ricerca, che presenta l'*IF* «come uno strumento per decidere se vale la pena di acquisire o mantenere l'abbonamento ad una rivista». Per il matematico proprio qui scatterebbe un perverso circolo vizioso a danno prima di tutto della libertà di comunicare la conoscenza dell'intera comunità scientifica, ma anche a scapito delle biblioteche scientifiche che, pur con un bilancio ridotto, sono obbligate ad acquistare quelle riviste che hanno maggior *IF* e che quindi sono anche le più costose. Figà Talamanca scrive: «Se le principali biblioteche scientifiche americane si convincono che non vale la pena acquistare riviste che hanno un *IF* basso o nullo, e se, come conseguenza, le riviste che non appartengono alla banca dati *ISI* scompaiono dalle biblioteche, la banca dati dell'*ISI*, apparirà automaticamente completa ed affidabile. Essa rappresenterà l'universo delle pubblicazioni che vale la pena di acquistare o consultare. In effetti, la scomparsa di una rivista dalle maggiori biblioteche scientifiche americane ne decreta, in qualche modo, l'estinzione o la totale emarginazione. Il risultato finale è quindi quello di modificare la realtà esterna per renderla conforme alle esigenze di contenimento dei costi e massimizzazione del profitti dell'azienda *ISI*».

Anche sotto il profilo più strettamente biblioteconomico non mancano le critiche alle banche dati citazionali e non solo a quelle della *Thomson-Reuters*. È chiaro che la mancanza di un sistema «di *authority control* al loro interno e di forme di controllo sugli indici nominali identificativi degli autori contribuiscono alla proliferazione di errori ed ambiguità a cascata, basti pensare al problema degli omonimi e ai sinonimi, con ricadute apprezzabili sull'attendibilità dei dati citazionali» (Cassella – Bozzarelli 2011 p. 69).

Le critiche tuttavia hanno anche un lato positivo. Per risolvere infatti molti dei problemi che esse hanno evidenziato, la ricerca bibliometrica è andata avanti ed è stata capace di costruire nuovi indicatori bibliometrici, capaci di superare molte delle criticità insite nell'*IF*. La nascita poi di nuove banche dati citazionali, una delle quali, rifacendosi proprio al mondo *open access*, è completamente gratuita (è questo il caso di *Google Scholar*), ha rotto definitivamente il regime monopolistico imposto prima dall'*ISI* e poi dalla *Thomson-Reuters*.

1.5.3. La bibliometria come metrica del Web

Un fronte attivo di ricerca nello studio della struttura del Web è quello che utilizza le tecniche quantitative e i modelli matematici mutuati dalla bibliometria per indagare la tipologia e misurare il numero di connessioni ipertestuali (*hyprelink*), e per analizzare i modelli di utilizzo delle risorse della rete.⁴³

Grazie all'impetuoso sviluppo delle metriche per il Web, questo campo di ricerca, che ha potuto beneficiare di oltre mezzo secolo di studi bibliometrici, può oggi qualificarsi come una disciplina autonoma, ben distinta dalla bibliometria, indicata, come già ricordato, con il termine "webmetrica".

L'applicazione dei metodi bibliometrici allo studio della rete di *hyperlink* è dovuta alla sostanziale analogia tra gli stessi *hyperlink* e la citazione bibliografica: un sito o una pagina web che rinvia a un'altra unità informativa per mezzo di un *link* è né più né meno identico ad un documento che richiama nel proprio testo, per mezzo di un riferimento bibliografico, un altro documento. Per De Bellis (2005 p. 154) in un link ipertestuale la funzione di collegamento risulterebbe addirittura potenziata, rispetto alla citazione, nel senso che «in un ipertesto la citazione non è una componente tipograficamente e concettualmente margi-

43 Un'interessante rassegna sui diversi metodi di analisi si trova in Park – Thelwall (2003).

nale del documento, bensì la norma principale che regola la costruzione e l'organizzazione del testo». Un link uscente (*outlink*) da una pagina web è simile ad un riferimento bibliografico; un link entrante (*inlink*) è analogo ad una citazione; due link che puntano alla stessa risorsa web (*co-linking*) sono simili ad una co-citazione; un link che punta alla struttura interna di una pagina o di un sito (*selflink*) corrisponde ad una autocitazione (De Bellis 2005 p. 157).

L'analisi citazionale, tecnica portante della bibliometria, è così diventata analisi dei link in webmetrica.⁴⁴ Tale tecnica ha portato anche all'elaborazione di un indicatore bibliometrico modellato sull'*IF* di Garfield, il *Web Impact Factor* (WIF) utile per la valutazione comparativa dei siti web. Anche lo *Science Citation Index* è stato declinato al mondo del Web, con la creazione da parte della Thomson-Reuters del *Web Citation Index*, un indice di citazioni multidisciplinare delle risorse accademiche accessibili via web. L'indice, pur collocandosi all'interno della piattaforma *ISI Web of Knowledge*, che ospita anche gli indici tradizionali dai quali si calcola l'*IF*, utilizza la tecnologia messa a punto per la costruzione di *CiteSeer* (oggi *CiteSeerx*),⁴⁵ riuscendo in questo modo a conciliare la politica commerciale della Thomson-Reuters con le esigenze legate al mondo dell'*Open Access*, su cui poggia per la gran parte *CiteSeer*.

Inoltre, il sistema di *ranking* di motori di ricerca come *Google* si sviluppa proprio da un principio classico dell'analisi citazionale: la probabile affinità concettuale esistente tra documenti collegati da una citazione. In questo caso però la citazione diventa *link* ipertestuale e l'affinità concettuale è tra la pagina web che effettua il *link* e la pagina linkata.

Sempre sfruttando le tecniche bibliometriche, la webmetrica ha provato a percorrere strade nuove, non limitandosi quindi a trasportare semplicemente al Web applicazioni e ambiti di ricerca già sperimentati dalla bibliometria. Un primo esempio è lo studio del contenuto e del contesto delle pagine web in cui compaiono i nomi di singoli studiosi, come indice della loro notorietà (Cronin – Snyder – Rosenbaum et al. 1998 e Bar-Ilan 1998). Un'altra tecnica è quella indicata con il termine *co-site analysis*, che impiega il principio, già ricordato, della co-occorrenza di informazioni bibliografiche alle pagine web. In altre parole, lo

⁴⁴ Le applicazioni pratiche basate sull'analisi dei link sono in questo paragrafo soltanto accennate, riceveranno invece una più ampia trattazione nel capitolo 6.

⁴⁵ <http://citeseerx.ist.psu.edu/>

studio si focalizza su gruppi di siti web co-citati, formati da coppie di siti presenti contemporaneamente come *link* uscenti (*outlinks*) di altri siti, allo scopo di visualizzare la rete di correlazioni tra istituzioni accademiche (Polanco – Boudourides – Besagni et al. 2001), di creare gruppi di siti di interesse disciplinare (Larson 1996) o identificare comunità virtuali (*web communities*), cioè gruppi di individui che condividono un interesse comune attraverso le pagine web di siti dedicati (Kumar – Raghauam – Rajagopalan et al. 1999). E infine, la *co-word analysis* di parole chiave utilizzate dall'utenza nei motori di ricerca, utile per la creazione di sempre più sofisticati algoritmi di *ranking* (Ross – Wolfram 2000), e di termini usati in siti di rilevanza nazionale allo scopo di analizzare le relazioni esistenti tra governi, università e importanti soggetti privati (Leydesdorff – Curran 2000).

2. L'impiego di metodi bibliometrici nella valutazione e nella gestione delle collezioni bibliotecarie

2.1. Leggi e modelli bibliometrici per la valutazione e gestione della collezione

Come già ricordato nel capitolo 1 gli studi bibliometrici comprendono attualmente due aree concettuali fondamentali, quella descrittiva che cerca di contare la produzione di un corpo di letteratura in relazione ad un'area geografica, a un settore disciplinare o a un arco temporale, e quella valutativa che ha come obiettivo l'identificazione dei modelli d'uso delle citazioni bibliografiche.

Concretamente, molto dello sforzo fatto negli studi bibliometrici descrittivi si concentra nel ricavare dati da un'area disciplinare specifica con cui sviluppare modelli generali o distribuzioni che possono essere applicati o testati in altri campi come per esempio nelle politiche documentarie in biblioteca.

Un tema che attraversa infatti tutta la letteratura bibliometrica fin dalle sue origini è quello dell'impiego dei dati prodotti dagli studi bibliometrici condotti sulle collezioni bibliotecarie per fornire una base scientificamente più solida alle decisioni riguardanti le operazioni di selezione, acquisizione e revisione dei documenti in biblioteca.

Nella bibliometria descrittiva sono due in particolare i concetti bibliometrici, applicati alla valutazione della collezione bibliotecaria, considerati fondamentali per efficacia valutativa e sui quali è stato possibile ricavare modelli generali di analisi: la distribuzione statistica e l'obsolescenza delle pubblicazioni.

2.1.1. La distribuzione statistica per la determinazione di un nucleo di documenti

Il concetto di distribuzione delle pubblicazioni si basa sull'osservazione che l'uso di una

qualsiasi raccolta di elementi generalmente non è distribuito equamente tra le voci: alcuni elementi vengono utilizzati molto, altri ricevono un uso moderato ed infine altri ancora non vengono mai utilizzati. Si è inoltre scoperto che esiste una sorprendente regolarità per i modelli di distribuzione dell'uso dei documenti che compongono una raccolta bibliotecaria. Esempi di tale modello di distribuzione si riscontrano nei tassi di circolazione annuale relativi a una monografia o a una rivista, nel numero di articoli pubblicati in un determinato periodico di un determinato ambito disciplinare, e infine nel numero di riferimenti bibliografici fatti a un determinato articolo contenuti nei periodici successivi alla sua pubblicazione (Morse 1979 p. 187).

Il modello più importante per la distribuzione statistica delle pubblicazioni è considerata la legge di Bradford. Essa, che è anche valutata come la più importante delle leggi bibliometriche, fin dalla sua prima applicazione è diventata uno strumento indispensabile per la gestione delle acquisizioni, la legge infatti nasce come tentativo di rispondere ad un problema che per i bibliotecari era diventato pressante a seguito della crisi del 1929: abbassare il costo dell'uso di ogni documento (Guédon 2004 p. 33).

La legge consente di analizzare i criteri di distribuzione degli articoli più significativi fra i periodici di una certa disciplina e dimostra che essi tendono a concentrarsi in un numero limitato di riviste, sono le cosiddette *core publications*, che di quel settore costituiscono il nucleo. Al di fuori di esso, il numero di riviste che contengono tanti articoli rilevanti quanti sono quelli contenuti nelle riviste del nucleo, cresce in modo esponenziale.

Per una biblioteca poter individuare il nucleo di riviste considerate centrali all'interno di ogni disciplina scientifica ha determinato una considerevole diminuzione dei costi di consultazione¹. Il bibliotecario ha infatti la possibilità di disdire o comunque di non sottoscri-

1 Serrai (1984 p. 199-229) ha trattato molti problemi legati alla bibliometria. In particolare, l'autore definisce "una chimera" la convinzione secondo cui «con l'ausilio della legge di Bradford, ogni biblioteca venisse messa in grado di decidere quali fossero i periodici più utili da acquistare e quelli che si potevano trascurare nei limiti di un certo bilancio finanziario. Il guaio era che la legge di Bradford poteva venire applicata solo dopo che i periodici del settore interessato erano stati ordinati in una graduatoria di produttività, e quindi solo dopo che la biblioteca aveva compiuto l'indagine empirica per sapere quali fossero i periodici più fertili, e perciò i più convenienti nel rapporto produttività-costi» (Serrai 1984 p. 201-202). In altre parole, Serrai critica la volontà di applicare il potere predittivo delle leggi matematiche della fisica alla politica delle acquisizioni in biblioteca, ad un ambito cioè «che non presenta affatto vincoli e relazioni funzionali altrettanto semplici e rigorose quanto quelle della fisica e della astronomia» (Serrai 1984 p. 202). L'impiego della legge di Bradford al mondo delle biblioteche ha anche un altro risvolto negativo. Il suo uso costante può determinare una sorta di fissità nella scelta dei documenti e portare indirettamente all'estinzione di tutti quei periodici considerati meno pertinenti. La legge utilizzata in modo massiccio si trasforma così in una sorta di profezia autoavverante.

vere gli abbonamenti per tutte quelle riviste che ruotano lontano dal nucleo costituito da quelle considerate principali in una disciplina, contenendo contemporaneamente l'eventuale danno arrecato al bisogno informativo dell'utenza.

Il consolidamento all'interno della biblioteca di quei nuclei di riviste ritenute centrali per una disciplina permette non solo la formazione di una raccolta capace di rispondere in modo più efficace ai bisogni dell'utenza, ma anche una gestione più efficiente delle risorse impiegate per il loro soddisfacimento. Oltre all'evidente risparmio di risorse finanziarie, la compressione mirata delle acquisizioni vuol dire naturalmente anche risparmio di spazio. Si pensi soltanto, per esempio, a tutte quelle riviste disponibili unicamente nel formato cartaceo che non vengono acquistate proprio perché non appartenenti al nucleo, e al conseguente notevole risparmio di spazio per la biblioteca.

La collezione organizzata in base al concetto di pertinenza determina poi un aumento del grado di precisione della ricerca bibliografica e una sostanziale diminuzione del rumore dovuto all'assenza di riviste di secondaria o terziaria rilevanza, tutti vantaggi che si traducono in un notevole risparmio di tempo sia per l'utente che per il personale della biblioteca.

La prima specifica applicazione della bibliometria alla biblioteconomia si deve al Donald J. Urquhart, fondatore dell'odierno *British Library Document Supply Centre (BLDSC)*, il quale attribuendo alla probabilità un ruolo fondamentale nella gestione delle collezioni bibliotecarie², è il primo ad analizzare scientificamente le statistiche d'uso dei periodici in biblioteca e applicare loro un modello probabilistico (Bensman 2005a). Lo studioso introdusse per primo la distribuzione di Poisson³ come modello di analisi dell'utilizzo dei periodici scientifici in biblioteca, dimostrando non solo la sua importanza a livello delle singole biblioteche ma anche come essa giochi un ruolo importante all'interno di un sistema bibliotecario più complesso.

2 A questo proposito la sua assistente Bunn scrive: «Dr. Urquhart is beginning to regard an understanding of the use of probability mathematics in a library as a more essential requirement than a knowledge of cataloguing and classification schemes» (Bensman 2005a p. 198).

3 «La legge (o distribuzione) di Poisson, detta anche “legge degli eventi rari”, è una distribuzione di probabilità discreta che esprime la probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo, sapendo che mediamente se ne verifica un numero λ » (fonte Wikipedia: <http://it.wikipedia.org/wiki/Distribuzione_di_Poisson>)

Lo studio condotto conteggiando i prestiti di materiale periodico effettuati nel 1956 dalla biblioteca dello *Science Museum (SML)* permise a Urquhart di osservare almeno due regolarità che si rivelarono successivamente fondamentali per la formulazione di quella che più tardi egli stesso chiamerà “Legge di Urquhart”. La prima fu la relazione positiva tra il numero di prestiti interbibliotecari di un periodico scientifico e il numero totale di volte quello stesso periodico rientrava nel posseduto delle principali biblioteche britanniche come riportato nel *British Union Catalogue of Periodicals (BUCOP)*: più alto era il numero di prestiti di un periodico effettuati dal SML e più alto era il numero di volte il periodico rientrava nel posseduto delle biblioteche (Urquhart – Bunn 1959 p. 21; Bensman 2005a p. 199). La seconda regolarità osservata da Urquhart fu che soltanto una piccola porzione (meno del 10%) di tutto il materiale periodico posseduto dalla SML era sufficiente a soddisfare il grosso (circa l'80%) delle richieste di materiale proveniente da altre biblioteche (Bensman 2005a p. 200) e che un terzo dei periodici correnti era scarsamente usato tanto da suggerire che l'acquisizione di un'unica copia in una sola biblioteca del paese poteva ritenersi sufficiente a coprire l'intera domanda nazionale per quel periodico (Urquhart 1959 p. 293).

L'osservazione di queste regolarità lo portarono a formulare una legge (detta anche “Law of Superlibrary Use”) sull'uso di materiale periodico fornito dalla biblioteca all'utenza unicamente attraverso il servizio di prestito interbibliotecario o attraverso il servizio di *document delivery* (fornitura documenti) organizzato in modo centralizzato. In base a questa legge il numero di prestiti interbibliotecari (e di utilizzo del servizio di *document delivery*) di un periodico diventa una misura attendibile del suo utilizzo globale⁴.

La base statistica su cui poggia la legge di Urquhart è rappresentata, come accennato in precedenza, dalla distribuzione di probabilità di Poisson. Essa è utilizzata per lo studio di eventi rari sia sotto l'aspetto temporale che spaziale, per fare alcuni esempi: il numero di biciclette che percorrono una strada secondaria e poco trafficata in un dato intervallo temporale o il numero di volte un periodico viene preso in prestito in una grande biblioteca in un intervallo di tempo di una settimana. «Perché sia però possibile, fissato un ambito di ri-

4 Urquhart (1977 p. 149) definì la legge da lui formulata nel modo seguente: «in its more pedantic form the law states that the inter-library loan demand for a periodical is as a rule a measure of its total use. As far as I am aware the existence of such a law was first indicated in my report of a survey of the use of journals in the Science Museum Library in 1956. For this reason perhaps the law should be called Urquhart's Law».

ferimento (spazio o tempo), calcolare tale probabilità, e quindi sia possibile conteggiare eventi o realizzazioni di un processo, è necessario che tale evento soddisfi determinate condizioni e che esse rimangano le stesse ad ogni ripetizione dell'evento stesso. Le condizioni che descrivono il processo di Poisson sono: indipendenza degli eventi per quanto concerne la loro realizzazione, vale a dire che il verificarsi di un evento in un intervallo di tempo o di spazio non ha alcun effetto sulla probabilità che esso si verifichi una seconda volta nello stesso o in un altro intervallo; la probabilità di una singola realizzazione dell'evento in un dato intervallo è proporzionale alla lunghezza dell'intervallo; in ogni parte arbitrariamente piccola dell'intervallo, la probabilità che l'evento si verifichi più di una volta è trascurabile» (Garretto 2002 p. 134).

Urquhart applicò la distribuzione di Poisson all'uso del materiale periodico in biblioteca proprio perché esso rappresenta una ripetizione di eventi che soddisfano pienamente le condizioni appena elencate. La distribuzione inoltre si basa sulla media degli eventi (chiamata tecnicamente λ) che può essere stimata semplicemente contando il numero degli utilizzi in un determinato intervallo di tempo e dividendo per il numero complessivo di unità bibliografiche soggette a consultazione. Lo spazio, elemento necessario nella distribuzione di Poisson, è rappresentato in ambito bibliotecario dai singoli periodici o da le varie classi di soggetto a cui possono venire ricondotti questi periodici (Bensman 2005c p. 7).

Per Urquhart l'obiettivo principale dell'impiego del modello poissoniano era identificare i periodici maggiormente utilizzati per assicurare una loro maggiore disponibilità attraverso la duplicazione delle copie o attraverso procedure speciali di rilegatura dei fascicoli, o per garantire una loro sistemazione più efficace del punto di vista dei costi.

Uno studio successivo sullo spostamento a magazzino del materiale soggetto a perdita parziale di valore informativo condotto presso l'università di Chicago da Fussler e Simon (1969) non solo confermò il concetto di uso interbibliotecario come indicatore di utilizzo globale di un periodico ma anche ampliò e contribuì a rafforzare la legge di Urquhart come strumento bibliometrico per la gestione e lo sviluppo della collezione bibliotecaria. Confrontando infatti i dati d'uso riguardanti lo stesso e identico materiale bibliografico posseduto contemporaneamente dalle università di Chicago, Northwestern, Yale e Berkeley, i

due studiosi osservarono che i bisogni informativi dei ricercatori nelle diverse università erano sostanzialmente simili e che i dati d'uso delle monografie in una singola biblioteca universitaria potevano essere utilizzati per prevedere l'uso delle monografie in un'altra biblioteca universitaria.

Ma l'analisi non si limitò solamente alle monografie, Fussler e Simon scoprirono che perfino l'uso passato di un periodico poteva essere utilizzato per prevedere il suo uso futuro. Per i due studiosi la più importante caratteristica dei periodici è da ricercarsi nella loro stessa natura, essi infatti possono essere visti come gruppi compatti di volumi i cui modelli d'utilizzo risultano collegati tra loro. Sotto l'aspetto dell'uso dei documenti, per Fussler e Simon esiste un maggiore rapporto di interdipendenza tra i volumi che compongono una stessa serie di un periodico rispetto ai volumi scelti casualmente da altri periodici (Bensman 2005c p. 14-15). Si tratta di una sorta di "effetto contagio": il maggiore o minore uso di uno o più elementi si trasmette anche agli altri elementi che fanno parte dello stesso insieme. Ai fini della gestione dei periodici in biblioteca, questa osservazione rende possibile l'impiego dei dati relativi all'utilizzo passato dei volumi di un periodico per prevedere l'uso futuro dei volumi appartenenti allo stesso. In altre parole, l'approccio quantitativo proposto dai due autori si basa sul concetto di predittività dell'uso attuale di un documento rispetto all'uso futuro, ovvero sulla possibilità di prevedere, grazie all'analisi statistica dei dati di circolazione, la frequenza d'uso futura di un gruppo di documenti con caratteristiche definite. Il criterio è semplice e banale: se un documento non è mai stato prestato negli ultimi n anni, difficilmente avrà una probabilità maggiore di circolare nei successivi n anni.

Urquhart che già considera un periodico come un insieme di volumi strettamente collegati tra loro, da cui a sua volta è possibile ricavare una media d'uso modellata sulla distribuzione di Poisson, introdusse due concetti fondamentali nelle politiche inerenti lo sviluppo e la gestione della collezione bibliotecaria: la stabilità nel tempo dell'uso del materiale periodico e la regolarità osservata nello studio dell'università di Chicago, vale a dire la valenza predittiva dell'uso dei periodici nel passato. Egli applicò entrambi questi concetti ai periodici maggiormente utilizzati, cioè a quel 10% posseduto dalla biblioteca sufficiente a soddisfare l'80% della richiesta di materiale proveniente dall'esterno.

Sebbene la legge e gli studi di Urquhart siano ancora poco conosciuti dai bibliotecari, essi rappresentano un valido strumento di riferimento per la singola biblioteca e, ad un livello più alto, per un sistema bibliotecario che ingloba più biblioteche. A livello della singola biblioteca sapere che soltanto una piccola porzione di materiale periodico è sufficiente a soddisfare quasi l'intero bisogno informativo dell'utenza aiuta in fase di acquisizione di nuovo materiale a non disperdere risorse finanziarie per l'acquisto di materiale scarsamente o per nulla utilizzato; permette di organizzare gli spazi in funzione proprio di questo maggiore utilizzo; e favorisce l'avvio di interventi di conservazione mirata come per esempio la digitalizzazione, se si tratta di materiale cartaceo, dei periodici maggiormente utilizzato da parte dell'utenza di una biblioteca. A livello di sistema bibliotecario sapere che l'uso di un periodico è sostanzialmente assimilabile all'uso che di esso si fa in un'altra biblioteca dello stesso tipo, naturalmente al netto delle comprensibili minime differenze legate alla corrispondente utenza di riferimento, diventa utile per la progettazione e la gestione di un efficace ed efficiente servizio di prestito interbibliotecario e di fornitura documenti.

La legge di Urquhart poi diventa ancora più utile nei sistemi bibliotecari di tipo consortile dove la politica delle acquisizioni, lo stoccaggio e la conservazione delle collezioni avvengono a livello centralizzato. In un recente studio (Tonta – Ünal 2010) la validità della legge di Urquhart è stata testata relativamente all'uso di *e-journals* da parte dell'utenza di un consorzio bibliotecario. I risultati dello studio hanno mostrato che più alto risulta l'uso dei periodici in loco o attraverso il servizio di *document delivery* di un periodico cartaceo, più alto risulta anche l'uso della versione digitale dello stesso periodico da parte dell'utenza di un consorzio bibliotecario. Inoltre, benché il set di periodici maggiormente utilizzati cambi in base all'utente, il modello complessivo d'utilizzo dei periodici rimane simile, anche se non uguale, a quello rilevato da Urquhart: circa il 35% degli *e-journals* inclusi in un pacchetto soddisfa l'80% di tutti gli utilizzi di *e-journals* del consorzio bibliotecario, mentre nel restante 65% del pacchetto molti *e-journals* risultano scarsamente o mai utilizzati (Tonta – Ünal 2010 p. 806-807). I risultati di questo studio che sostanzialmente confermano quanto dimostrato da Urquhart possono diventare la base per la costruzione di una politica documentaria anche per le risorse elettroniche all'interno del consorzio. Gli *e-journals* maggiormente utilizzati possono venire confermati, mentre quelli scarsamente consultati dovreb-

bero essere esclusi dalla collezione, ottimizzando così la spesa con un impatto sull'utenza pressoché nullo o comunque alquanto limitato. Senza dimenticare poi che questi risultati possono venire impiegati per negoziare più vantaggiose condizioni contrattuali con gli editori per l'acquisto di licenze elettroniche.

Da più parti questa legge viene considerata l'equivalente, basato però sull'analisi dei dati d'uso del materiale bibliografico in biblioteca, della legge di Garfield che invece si basa sull'analisi dei dati citazionali (Bensman 2000 p. 826).

Lo studio sui dati d'uso condotto da Urquhart, che può essere considerato il primo del suo genere, ha anche il merito di aver aperto la strada a altri contributi sull'argomento da parte di studiosi che hanno ulteriormente indagato le potenzialità dell'impiego dei dati d'utilizzo per la valutazione e la gestione della collezione come il già ricordato studio sulla predittività di Fussler e Simon (1969) e quello sulla "regola 80/20" di Trueswell (1969). Richard Trueswell, docente di ingegneria industriale all'università del Massachusetts osservò, in base ad accurate ricerche sul campo, che l'80% delle richieste degli utenti delle biblioteche si concentra sul 20% dei documenti immagazzinati, analogamente a quanto era stato da tempo osservato nei magazzini per lo stoccaggio delle merci nei settori dell'industria e del commercio. Partendo da queste considerazioni e riprendendo il principio di predittività di Fussler e Simon, Trueswell propone un metodo fondato sull'analisi dell'uso pregresso (*Last Circulation Date, LCD*) che permette di individuare la parte della raccolta in grado di garantire una percentuale prefissata (ad esempio il 95%) di soddisfazione della domanda futura. Trueswell definì *core collection* la parte circolante della raccolta, da conservare, distinguendola dalla *non core collection*, la parte improduttiva della dotazione, da sottoporre a revisione (Parise 2008 p. 219).

Stanley Slote ha ulteriormente approfondito questa direzione di ricerca e introdotto il concetto di *shelf time period* per indicare il tempo di immobilizzazione sullo scaffale fra due prestiti, cioè il tempo medio di permanenza a scaffale di un documento, sulla base del quale individuare una data limite di non utilizzo (Slote 1989). Il metodo è molto semplice: se all'interno di un determinato arco temporale si considerano tutti i documenti in prestito raggruppandoli per anno in base al tempo di permanenza a scaffale fra due prestiti, è possibile calcolare per ciascun anno la percentuale di prestiti realizzata e individuare la soglia

temporale che corrisponde alla percentuale di soddisfazione della domanda desiderata. Ad esempio, se si vuole garantire una percentuale d'uso futuro pari al 95%, con questo metodo è possibile eliminare dallo scaffale tutti i documenti che eccedono il periodo necessario per raggiungere il 95% dei prestiti. Per i bibliotecari il metodo permette di collocare a deposito quella quota di documenti a basso indice d'uso che rappresenta, se si considera la regola del 80/20 di Trueswell, una porzione molto rilevante della raccolta (l'80%). Bisogna considerare poi che il tasso d'uso è soggetto a un declino lineare nel corso del tempo. Uno studio condotto nel 1979 all'università di Pittsburg ha dimostrato che esiste soltanto il 50% di probabilità che un nuovo acquisto sia prestato, e che questa si dimezza ogni due anni fino a scendere al 2% se nel frattempo il documento non è mai stato prestato (Kent et al. 1979 p. 10). A questi risultati tuttavia è riconducibile il secondo grande concetto bibliometrico, l'obsolescenza, che sarà oggetto del paragrafo seguente.

2.1.2. L'obsolescenza della letteratura

Il concetto di obsolescenza⁵ della letteratura si può riferire sia all'invecchiamento dei contenuti che alla diminuzione dell'uso di un documento. In sostanza si tratta del declino nel tempo della validità e dell'utilità delle informazioni nella prima accezione e della valutazione della probabilità di un uso futuro nella seconda (Parise 2008 p. 217).

Anche se sembra chiaramente sbagliato riferire esclusivamente il concetto di obsolescenza ai tassi di circolazione del documento, considerando quindi obsoleti i documenti poco utilizzati, non è comunque errato ammettere una correlazione fra i tempi di obsolescenza e i tassi di circolazione, poiché l'uso di opere non aggiornate tende generalmente a calare. Determinare quando un documento è diventato obsoleto, cioè non più capace di ricevere ulteriore uso, o comunque soggetto a una perdita parziale del valore informativo, rappresenta una delle informazioni più importanti nella gestione della collezione, perché aiuta il bibliotecario a stabilire, nelle operazioni di revisione, quando sia giunto il momento di inviare al macero un documento o semplicemente di allontanarlo dalla sezione o scaffale aperto, e in quelle di gestione delle acquisizioni, a capire quali segmenti delle collezioni

5 Il termine "obsolescenza" venne impiegato per la prima volta nel lavoro di Gross – Gross (1927), nel quale gli autori analizzarono le citazioni bibliografiche della rivista scientifica *Chemical Literature* del 1926, osservando che il numero di citazioni si dimezza in arco di tempo di 15 anni.

non corrispondono più agli interessi dell'utenza, evitando in tal modo di disperdere inutilmente risorse finanziarie per l'acquisto di materiale incapace di soddisfare il bisogno informativo della comunità in cui la biblioteca opera.

Lo sviluppo quindi di un metodo scientifico credibile e facilmente applicabile per la misurazione dell'obsolescenza dei documenti ha rappresentato, e rappresenta, un tema centrale nell'ambito di tutte le decisioni che riguardano le operazioni di revisione e acquisizione di materiale bibliografico.

È proprio per agevolare fasi così delicate della politica documentaria che si è cercato, e si cerca, di sviluppare metodi scientifici sempre più sofisticati in termini di affidabilità e facilità d'applicazione per la misurazione dell'obsolescenza dei documenti.

Un punto importante da tenere presente quando si procede all'analisi della raccolta sfruttando il dato dell'età dei documenti è che i tempi di obsolescenza variano in base alla disciplina e chiaramente sono più rapidi per le pubblicazioni scientifiche. Inoltre, per alcune biblioteche, ad esempio per quelle di conservazione, il concetto di obsolescenza potrebbe risultare del tutto estraneo, o essere parzialmente irrilevante se per *mission* la biblioteca si è data il compito di conservare alcune parti delle raccolte, come nel caso della sezione di storia locale.

Gli studi sull'obsolescenza si concentrano attorno a due grandi aree che si differenziano tra loro non tanto per la metodologia impiegata, bensì per le diverse informazioni estrapolate dalle unità bibliografiche su cui poggiano tali studi: quelli basati sulla valutazione dei dati d'uso dei documenti e quelli basati sull'analisi delle citazioni. Entrambi gli approcci producono risultati che oltre ad essere simili tendono anche ad essere molto regolari: se si stila una classifica di documenti tenendo conto della loro età in rapporto al numero di volte in cui vengono usati si ottiene che i documenti molto recenti costituiscono una porzione molto ampia di tutto il materiale utilizzato, mentre i documenti molto vecchi registrano uno scarsissimo utilizzo.

L'obsolescenza è frequentemente espressa in termini di "vita media" ("half-life") o "semivita" della letteratura, concetto mutuato dalla fisica nucleare per indicare un processo che indica un decadimento qualitativo per il quale, per esempio, la semivita di una sostanza radioattiva è il tempo necessario affinché la metà di essa decada (Burton – Kebler 1960

p. 18-19). Per la letteratura, il concetto di vita media è utilizzato per misurare il periodo di tempo nel quale è stata pubblicata la metà del totale della letteratura corrente. La vita media poi può essere più o meno lunga in base alla disciplina oggetto di analisi. Se la vita media relativa ad una disciplina è considerata lunga significa che la letteratura pubblicata in quest'ambito si manterrà corrente per un numero maggiore di anni rispetto, per esempio, ad una disciplina caratterizzata da una vita media più corta.

A partire da questo concetto è anche possibile distinguere all'interno dello stesso ambito disciplinare due o più tipi di letteratura periodica, ciascuno caratterizzato da una propria vita media. In molti campi esiste infatti un corpo di letteratura a cui universalmente si fa riferimento come la parte "classica", vale a dire quella parte costituita da tutte le pubblicazioni che per importanza teorica o storica non possono essere trascurate nello studio di una disciplina, caratterizzata da una vita media molto più lunga rispetto a quella parte più "effimera", di secondaria importanza, costituita per esempio dalle riviste generaliste che trattano una specifica disciplina in maniera più superficiale rispetto a quelle specializzate in un settore ben preciso (Burton – Kebler 1960 p. 20).

Sfruttando il dato citazionale delle riviste, il concetto di vita media ha dato origine a due indicatori frequentemente usati soprattutto dalle biblioteche americane nell'ambito delle politiche documentarie: il *Cited Half-Life* e l'*Immediacy Index*.

Il *Cited Half-Life* è l'indice, facilmente estraibile dalle basi dati di ISI, che misura il numero di anni, a partire dall'anno corrente e andando indietro nel tempo, degli articoli cui si riferisce la metà del totale delle citazioni ricevute nell'anno in corso dalla rivista considerata. Per esempio, se una rivista nel 2010 ha un indice *Cited Half-Life* pari a 3, significa che metà delle citazioni che ha ricevuto nel 2010 si riferisce ad articoli pubblicati negli ultimi 3 anni, mentre l'altra metà delle citazioni registrate si riferisce ad articoli pubblicati in precedenza. Tale indice, analizzando la distribuzione nel tempo delle citazioni, misura quindi la vita media degli articoli citati. Lo stesso meccanismo è valido per pesare le citazioni in uscita dalla rivista, quindi il *citing half-life* (Morriello 2007 p. 27).

L'*immediacy index* misura invece il numero medio di volte in cui un articolo pubblicato nell'anno considerato all'interno di una determinata rivista è stato citato nel corso dello stesso anno di uscita. Esso indica, in sostanza, quanto spesso in media gli articoli pubblica-

ti su una rivista sono citati durante lo stesso anno, vale a dire il loro impatto nell'anno in cui sono stati pubblicati. È calcolato dividendo il numero corrente di citazioni di una rivista in un dato anno per il numero di articoli pubblicati nell'anno corrente. Per esempio, se una rivista nel 2010 ha pubblicato 800 articoli che sempre nello stesso anno sono stati citati 4000 volte, il suo *immediacy index* sarà pari a 5.

Come sottolineato da Morriello (2007 p. 27) questi strumenti sono viziati da difetti come «la penalizzazione delle riviste i cui fascicoli escono in ritardo, l'incidenza del numero di articoli pubblicati dalla rivista, le diverse abitudini citazionali tra le varie discipline e tra specifiche comunità o individui all'interno della stessa disciplina, anche rispetto alla retrospettività, la tipologia di periodico e la sua diversa "immediatezza" nel complesso». Senza contare poi la penalizzazione del cartaceo rispetto alla versione elettronica, quest'ultima consultabile, e quindi citabile, con largo anticipo rispetto all'arrivo del periodico cartaceo in biblioteca.

Come già ricordato, il concetto di vita media è calcolabile anche partendo dai dati d'uso raccolti all'interno di una specifica biblioteca dagli stessi bibliotecari. Il metodo suddivide il numero di prestiti effettuati in un determinato periodo per la data di pubblicazione dei documenti prestati. L'anno di pubblicazione nel quale cade il valore corrispondente alla metà dei prestiti effettuati è considerato "mediana d'uso" (Lancaster 1988 p. 73), che può essere assunta come un indicatore dei tempi di obsolescenza della raccolta: più questa data è ravvicinata, più i tempi di obsolescenza sono rapidi. Parise (2008 p. 220) propone di ponderare questo dato rispetto al posseduto, in quanto considera essenziale che la distribuzione dei documenti prestati per anno di edizione tenga conto della consistenza delle raccolte nei vari anni. Praticamente egli propone di fare riferimento all'indice di circolazione (I_c)⁶ piuttosto che al numero assoluto dei prestiti. In simboli:

$$\text{Mediana d'uso} = [I_c(a^1) + I_c(a^2) + \dots + I_c(a^n)]/2$$

Utilizzando quindi il dato sulla circolazione, l'anno di pubblicazione nel quale cade il valore corrispondente alla metà della somma degli indici di circolazione è considerato "soglia di declino d'uso". Tradotto in termini pratici, se il bisogno informativo dell'utenza per

⁶ La Norma ISO 11620: B.2.4.1 definisce l'indice di circolazione come il numero totale dei prestiti in un determinato periodo, diviso per il numero totale dei documenti posseduti.

un determinato ambito disciplinare viene soddisfatto per il 95% dei prestiti con documenti di età inferiore a 5 anni, è auspicabile che la parte di materiale più vecchia venga rimossa dallo scaffale e sottoposta a revisione.

È da dire comunque che il metodo della soglia, pur essendo un valido strumento di riferimento soprattutto nella stesura dei protocolli di revisione, deve essere calibrato su quelle che sono le specificità della singola biblioteca, adattandolo quindi allo specifico quadro di riferimento di politica documentaria e alle diverse aree disciplinari. Per esempio, nei casi in cui sia disponibile un deposito cooperativo, il limite di obsolescenza potrà essere più basso, mentre nel caso di una biblioteca che non abbia altra soluzione che l'eliminazione dovrà essere più elevato. Sarà quindi il bibliotecario a decidere il limite di obsolescenza, cioè la soglia percentuale di prestiti oltre la quale si procede all'accantonamento o allo scarto.

Quest'ultima considerazione introduce quello che rappresenta il maggiore fattore di criticità dell'impiego in biblioteca di metodi di analisi quantitativi specie se basati sulle statistiche d'uso, vale a dire la drastica riduzione dei margini di discrezionalità del bibliotecario. Un approccio puramente oggettivo non tiene conto infatti di tutte quelle grandi o piccole variazioni che possono verificarsi nell'uso della raccolta a causa, per esempio, di modifiche nelle politiche d'offerta documentaria, nella composizione e nella manifestazione dei bisogni documentari dell'utenza. Anche assolutizzare il criterio del valore d'uso, dimenticando invece che un'opera può rivestire valore al di là dei tassi di circolazione registrati, può causare uno sviluppo della raccolta troppo appiattito sul presente.

Una possibile soluzione ai problemi legati a quest'approccio è sicuramente l'uso combinato delle statistiche d'uso con altri generi di valutazione come l'analisi citazionale, capace di correggere eventuali appiattimenti e far emergere il valore anche qualitativo del risultato, soprattutto quando è necessario prendere delle decisioni sui tagli all'offerta documentaria o quando si deve scegliere se mantenere la versione elettronica o cartacea di un periodico.

Anche il modello matematico sviluppato da Bertram C. Brookes, docente di scienza dell'informazione presso l'università di Londra, è considerato un valido strumento bibliometrico per lo sviluppo e la gestione della collezione. Brookes, oltre ad aver contribuito in ma-

niera determinante all'avanzamento delle conoscenze nel campo dell'obsolescenza della letteratura scientifica, ha anche il merito di aver definito ed introdotto il concetto fondamentale di "utilità" dei documenti sviluppando indici come l'*Utility Factor* (UF) e l'*Annual Aging Factor* (AAF) (Brookes 1970).⁷ Per Brookes l'"utilità totale" o "iniziale", $U(0)$, di un articolo di periodico è da intendersi il numero totale di citazioni che si ipotizza esso possa ricevere in futuro partendo dalla data di pubblicazione (Brookes 1970 p. 324). In termini statistici, questa utilità rappresenta il 100% del totale. Con il passare del tempo, l'utilità iniziale di un documento si riduce e si trasforma in utilità residua $U(t)$ secondo una funzione esponenziale decrescente fino ad arrivare a essere nulla entro un periodo di tempo che, non essendo determinabile visto che a priori non si può determinare quando un periodico ha esaurito la propria valenza informativa, è infinito (Jiménez-Contreras – Bailòn Moreno – Ruiz-Baños et al. 2006 p. 199).

Per calcolare l'utilità residua $U(t)$ Brookes propone di moltiplicare l'utilità iniziale $U(0)$ per il fattore d'invecchiamento annuale, l'*Annual Aging Factor* (AAF), a^t , dove a rappresenta il valore percentuale di riduzione dell'utilità per ogni anno che passa, mentre t è l'età, cioè la differenza di anni tra un dato anno e la data di pubblicazione dei documenti, in simboli:

$$U(t) = U(0) * a^t$$

L'*Annual Aging Factor*, a , può assumere valori tra zero e uno: più il valore si avvicina a 1 più grande è il grado di utilità del documento.

Altro indice introdotto da Brookes è l'*Utility Factor*, u , che dipende dall'*Annual Aging Factor*, infatti più alto è il valore del fattore d'invecchiamento annuale più alto è il valore del fattore d'utilità totale: per esempio, quando $a = 0,80$, $u = 5$; quando $a = 0,90$, $u = 10$, e quindi più lento è l'invecchiamento del documento, più verrà utilizzato dall'utenza (Brookes 1970 p. 323).

L'*Utility Factor*, u , si calcola a partire dal numero totale (assoluto) di citazioni e dal numero totale di citazioni dell'anno che ne ha ricevute di più. In simboli:

$$u = 1 / (1 - a) [1]$$

⁷ Molti studiosi hanno applicato il modello di Brookes, o un suo derivato, a diversi corpi di letteratura scientifica. Degno di nota particolare è il lavoro di Griffith et al. (1979) che ha determinato l'obsolescenza di tutta la letteratura scientifica citata nello SCI fino al 1979 e stabilito l'*Annual Aging Factor* di molte delle riviste presenti nello stesso indice di citazioni.

e

$$R = u * C \quad [2]$$

dove R sta per il totale (assoluto) delle citazioni e C sta per il totale delle citazioni dell'anno che ne riporta una maggiore quantità. Sostituendo l'equazione [1] con l'equazione [2] si ottiene:

$$R = C / (1 - a)$$

Applicando la formula inversa si ottiene il fattore d'invecchiamento a :

$$a = 1 - C / R$$

Per esempio se le citazioni totali, R , ricevute da un periodico sono 14000 e quelle dell'anno in cui ne riporta di più, C , sono 1000, si otterrà un fattore d'invecchiamento, a , di 0,929:

$$a = 1 - 1000 / 14000 \quad (a = 0,929)$$

2.2. L'analisi citazionale come strumento per la gestione e la valutazione della collezione

La seconda grande area concettuale che compone gli studi bibliometrici è quella valutativa, conosciuta anche come analisi citazionale, che ha lo scopo pratico di studiare come un corpo selezionato di letteratura viene utilizzato mediante l'identificazione dei modelli d'uso delle citazioni bibliografiche.

Nel quadro delle politiche per la gestione e la valutazione delle raccolte, l'analisi citazionale⁸, assieme a tecniche come l'osservazione e la valutazione a scaffale condotte dal bibliotecario, da specialisti o da commissioni, e a procedure comparative con bibliografie o biblioteche di riferimento, rientra tra i metodi per la valutazione qualitativa che sfruttano un approccio centrato sul posseduto. Queste misurazioni qualitative permettono, a differenza di una analisi puramente quantitativa, di raccogliere informazioni che, se viste nell'ottica

⁸ Una panoramica esauriente sull'uso dell'analisi citazionale come metodo applicato alla gestione delle raccolte si trova nei lavori di Smith (1981) e Broadus (1977). Nisonger (1998) fornisce una dettagliata analisi di questo metodo nell'ottica di una sua applicazione nella gestione della collezione di periodici nelle biblioteche universitarie. Degni di nota sono anche i lavori di Line (1978), Bensman (1985), MacRoberts – MacRoberts (1989) e Kelland Yonge (1994). Altri autori si sono invece concentrati sull'applicazione dell'analisi citazionale a documenti non coperti dai più importanti indici di citazioni: Dilevko – Atkinson (2002); Kelland – Buckley (1997) sulle monografie; Magrill – St. Clair (1990), Leiding (2005), Edwards (1999), Kushkowski – Parson – Wiese (2003), Beile – Boote – Killingsworth (2004) sulla letteratura grigia; e infine Cui (1999) e Bar-Ilan – Peritz (2002) sulle risorse elettroniche presenti nel Web.

di un approccio centrato sui bisogni dell'utenza, appaiono determinanti come lo possono essere quelle relative all'uso che viene fatto della raccolta, alla sua adeguatezza rispetto alle esigenze dell'utenza, alla copertura degli argomenti che compongono un'area disciplinare, e al livello di aggiornamento della raccolta al progresso delle conoscenze e delle pratiche nella disciplina in esame (Parise 2008 p. 124).

L'analisi citazionale, metodo portante della bibliometria, è l'esame della frequenza e del modello di citazioni in articoli o in testi in generale.

Essa in sostanza è una procedura che, partendo dalla raccolta e dal conteggio delle citazioni, arriva al *ranking* (cioè a stilare una graduatoria) di riviste di una certa area disciplinare tenendo conto del numero di volte un documento viene citato nei riferimenti bibliografici di un altro documento, nelle bibliografie o negli indici di citazioni. In sostanza, questo metodo misura il valore di una pubblicazione in base alle citazioni che ha ricevuto: maggiore è il numero di citazioni di un documento, maggiore sarà la sua qualità e il suo prestigio.

L'utilizzo di questo metodo per la gestione della collezione parte proprio dal presupposto che i documenti che ricevono un alto numero di citazioni generano un forte impatto o influenza sulla comunità scientifica, e quindi sono sufficientemente importanti da essere inclusi dai bibliotecari in un nucleo (*core*) di pubblicazioni che, a sua volta, può essere preso come punto di riferimento per valutare i punti di forza e di debolezza delle raccolte.

L'analisi citazionale permette quindi la creazione di bibliografie calibrate sui bisogni dell'utenza di riferimento della biblioteca, superando così le difficoltà che l'utilizzo di liste preconfezionate spesso pongono in termini di aggiornamento, esaustività e appropriatezza rispetto al livello di approfondimento richiesto.

Inoltre, a differenza delle bibliografie preconfezionate sviluppate a partire dal giudizio di qualcuno che definisce cosa e quanto ha valore in un determinato ambito disciplinare, le liste di "*core publications*" sono costruite sfruttando innanzitutto un approccio oggettivo (il conteggio di dati) ma anche utilizzando proprio le fonti alla quali gli autori si sono rivolti nel loro lavoro (le citazioni). Quest'ultimo punto dilata ulteriormente le potenzialità d'analisi offerte da uno studio condotto utilizzando le citazioni, lo fa diventare cioè un metodo

che permette alla letteratura di una determinata area disciplinare di definire le caratteristiche e i confini che le sono propri.

L'analisi delle citazioni, oltre ad essere impiegata per costruire liste di "core publications", importanti per definire il grado di approfondimento di una disciplina o, più in generale, per valutare la raccolta nel suo complesso, diventa uno strumento di supporto utilissimo in tutte le fasi che compongono la politica documentaria della biblioteca, dalla selezione e acquisizione all'accantonamento e scarto dei documenti. Basandosi sui dati provenienti dall'analisi bibliometrica, in fase di selezione e acquisizione dei documenti, il bibliotecario sceglierà di acquistare le pubblicazioni che riportano un alto numero di citazioni e di sottoscrivere o confermare l'abbonamento a periodici i cui articoli sono molto citati, viceversa in fase di revisione della raccolta, il bibliotecario deciderà di spostare a magazzino o scartare le pubblicazioni che riportano un basso numero di citazioni e di cancellare l'abbonamento a riviste i cui articoli sono poco citati.

Mediante la costruzione di un indicatore di efficienza, il dato citazionale può anche essere sfruttato in biblioteca per la pianificazione del budget, permettendo così al bibliotecario di sviluppare un metodo di valutazione in cui la citazione è considerata in rapporto alle risorse necessarie al conseguimento degli obiettivi fissati dalla politica documentaria, misurate in termini monetari e, precisamente, in termini di costo (Altmann 1998).

In simboli:

$$\text{Citazioni per articolo} \times \text{Totale articoli per anno} = \text{TOTALE CITAZIONI PER ANNO}$$
$$\text{Costo per anno} \times \text{periodico} / \text{Totale citazioni per anno} = \text{COSTO PER CITAZIONE}$$

La stessa formula può essere utilizzata sostituendo il "Totale delle citazioni" per anno con l'*Impact Factor* per rivista (Christensen 1992; Stankus 1992; Stankus – Mills 1992).

I bibliotecari possono condurre indagini di tipo citazionale partendo da più livelli di complessità. Come già ricordato, il primo indicatore di performance citazionale è costituito contando semplicemente il numero di citazioni contenute nel profilo citazionale di una pubblicazione (conosciuto come *Citation Index*). Il dato può essere ricavato agevolmente, ol-

tre che da bibliografie e riferimenti bibliografici, anche da almeno tre banche dati: *ISI Web of Science*, *Scopus* e *Google Scholar*.

ISI Web e *Scopus* riportano come risultato delle interrogazioni, oltre al numero totale di citazioni ricevute da un singolo prodotto di ricerca nell'intervallo temporale compreso tra la data di pubblicazione e il momento dell'interrogazione, anche il numero medio anno di citazioni.

Molti altri database bibliografici indicano, per ogni pubblicazione, il numero totale di citazioni ricevuto nello stesso database, o su Web. *JSTOR*⁹, per esempio, probabilmente il più grande archivio interdisciplinare di riviste scientifiche, indica per ogni articolo indicizzato il numero complessivo delle citazioni ricevute all'interno del database, e permette di risalire immediatamente agli articoli citati attraverso *link* ipertestuali. In modo del tutto analogo anche altri database bibliografici e tutti gli *open archives* tipo *arXiv*¹⁰ e *RePEC*¹¹ indicano per ogni prodotto di ricerca il numero di citazioni ricevute da altri articoli contenuti nello stesso archivio.

Quello che è bene sottolineare è che il numero di citazioni, non essendo un marchio registrato come l'*Impact Factor*, non può essere considerato un dato univoco in quanto il suo valore può mutare a seconda della base di dati utilizzata. Anche se, in linea teorica, il numero di citazioni ricevute da un articolo, non può che essere unico, vari sono in realtà i fattori che contribuiscono alla sua variazione: ad esempio, la copertura delle singole banche dati sia come numero di riviste indicizzate, sia come anni di pubblicazione o di sottoscrizione; oppure i criteri di spoglio e gli algoritmi di ricerca, per cui alcune tipologie di articoli possono non essere incluse e quindi non essere conteggiate (Piazzini 2010 p. 76).

I bibliotecari possono anche condurre le loro indagini utilizzando indicatori citazionali caratterizzati da un livello di analiticità più elevato rispetto al *Citation Index*, sfruttando cioè le potenzialità offerte da indicatori bibliometrici citazionali come l'*Impact Factor*, considerato l'indicatore bibliometrico per eccellenza e sinonimo di "valore di un periodico", e da altri prodotti simili, frutto dell'evoluzione dei processi editoriali e della nascita di para-

9 [Http://www.jstor.org/](http://www.jstor.org/)

10 <http://arxiv.org/>

11 <http://repec.org/>

metri alternativi, come per esempio *SCImago* e l'*Eigenfactor*.¹² In sostanza, in questi indicatori, il dato relativo al conteggio del numero di citazioni ricevute, che costituisce ancora la base per la loro costruzione, è affiancata da un'ampia gamma di parametri aggiuntivi che contribuiscono a fare di queste metriche strumenti di valutazione ancora più precisi e quindi più affidabili.

L'analisi citazionale applicata alla valutazione delle raccolte non è tuttavia esente da criticità. Benché il metodo sia annoverato nei più importanti manuali di biblioteconomia tra le tecniche di valutazione qualitativa della collezione, esso sfrutta essenzialmente un approccio quantitativo di analisi.¹³ Gli indicatori bibliometrici non misurano necessariamente la qualità o l'importanza di un'opera, essi piuttosto sono costruiti sull'idea che il numero di citazioni ricevute dal lavoro di uno scienziato sia una buona approssimazione del suo impatto (popolarità) all'interno della comunità scientifica di riferimento. Non è invece sempre automatico il ragionamento secondo cui il lavoro di alta qualità prodotto da uno scienziato inneschi un numero di citazioni da parte dei colleghi più elevato rispetto a un prodotto di più bassa qualità scientifica. Molto dipende dal fatto che spesso la pratica citazionale non riflette solamente il contenuto dell'opera citata, ma è inevitabilmente influenzata da altri fattori, in parte non scientifici o dipendenti da caratteristiche dell'autore, dell'istituzione, o dalle regole che determinano il gioco accademico (Baccini 2010 p. 148).

I bibliotecari dovrebbero poi considerare che le pratiche citazionali variano nelle diverse comunità scientifiche, che in alcuni campi gli autori tendono a citare di più la letteratura recente (Peters – Van Raan 1994), che i settori di dimensione ridotta tendono a ricevere un numero di citazioni molto più basso rispetto a quelli più grandi (Moed et al. 1985), e che le probabilità di essere citati dipende dalle caratteristiche della rivista in cui l'articolo è pubblicato, o delle riviste che contengono articoli che possono citarlo.

12 Gli indicatori bibliometrici saranno oggetto di una più ampia e dettagliata trattazione nel capitolo 5.

13 A tal proposito, una considerazione deve essere fatta sulla natura ambivalente della citazione. Da una parte essa può essere trattata in base al suo valore qualitativo indiretto: riflette il contenuto dell'opera citata, cioè di un lavoro che è stato sottoposto per la pubblicazione ad una valutazione qualitativa espressa da un comitato dei pari (*peer-review*). Dall'altra, se considerata cumulativamente alle altre citazioni ricevute dalla medesima opera, essa permette di calcolarne l'impatto (cioè la popolarità) sulla comunità scientifica di riferimento. Risiede in questo il suo valore quantitativo.

Per i bibliotecari poi può essere difficile selezionare e raccogliere le fonti di citazioni che riflettono i bisogni informativi dell'utenza di riferimento della biblioteca (Sylvia 1998 p. 20).

È fondamentale inoltre analizzare il grado di copertura degli indici di citazioni, tenendo sempre presente che la forma di comunicazione più diffusa nelle scienze sociali e umane è la monografia, mentre nelle scienze naturali e ingegneristiche è la rivista scientifica. Ne consegue che gli indicatori citazionali costruiti a partire dai libri e dalle riviste danno luogo a risultati anche molto diversi.

Un'altra questione delicata riguarda gli errori nelle citazioni quelli che possono provenire da problemi tecnici connessi alla gestione e interrogazione dei database citazionali, in primo luogo le omonimie, sinonimie e omografia (Moed 2005; Baccini 2010 p. 147) che, uniti all'assenza di un processo di *authority control* all'interno delle banche dati citazionali più accreditate (*WoS* e *Scopus*), si ripercuotono pesantemente sull'attendibilità dei dati citazionali.

Visti tutti questi punti critici, è ovvio per i bibliotecari che si accingono a gestire e valutare le raccolte utilizzando strumenti di analisi citazionale considerare il loro impiego in combinazione con metodologie di rilevazione ed analisi dei dati diverse e complementari.

Una puntualizzazione finale sembra essere a questo punto obbligatoria: se è vero che l'analisi delle citazioni trova un'ampia applicazione come supporto alle decisioni dei bibliotecari nel campo delle politiche documentarie, anzi si può affermare che il metodo sia nato proprio a questo scopo, è anche vero che molte delle sue potenzialità potrebbero essere sfruttate per lo sviluppo e il potenziamento di comparti altrettanto importanti per la biblioteca, specie universitaria, e cioè del servizio di *reference* e quello di disseminazione selettiva dell'informazione e dei documenti.¹⁴

2.3. La bibliometria per la gestione e valutazione della biblioteca digitale

Inserendosi a pieno titolo tra i modelli di valutazione che da circa un decennio la biblio-

¹⁴ Per le sfide offerte dalla bibliometria che attendono le biblioteche si rinvia al capitolo 7.

teconomia internazionale discute e propone per le biblioteche digitali, l'analisi bibliometrica sta diventando un approccio molto promettente anche per la determinazione dell'impatto e per il *ranking* dei documenti che compongono l'offerta informativa a disposizione dell'utenza remota di questo nuovo, ma sempre più diffuso, tipo di biblioteca.¹⁵ Le biblioteche digitali infatti hanno aggiunto un valore enorme ai servizi e alle collezioni offerte dalle biblioteche ai loro utenti, ma hanno anche elevato enormemente i costi di gestione di un sistema bibliotecario.

Nel corso dell'ultimo decennio sono stati realizzati e pubblicati nell'ambito delle scienze dell'informazione una serie di studi sulla valutazione delle biblioteche digitali considerate nel loro complesso¹⁶ o in relazione alla singole componenti (ad esempio l'OPAC o i depositi istituzionali)¹⁷, ai servizi (ad esempio il *reference* digitale) o alle collezioni¹⁸.

Come accade per le biblioteche tradizionali, i dati raccolti utilizzando metodi bibliometrici possono essere utilizzati dai bibliotecari per valutare la collezione e i servizi offerti della biblioteca digitale. Più nel dettaglio, la raccolta di dati e l'impiego di indicatori di rendimento per la valutazione delle risorse elettroniche sono, come ben spiegato in Tammaro (2000 p. 66), utili per: aiutare chi deve prendere decisioni per la distribuzione delle risorse per pianificare lo sviluppo futuro; identificare gli aspetti positivi e quelli negativi dei servizi per migliorarli; fornire dati di tendenza per sviluppare i servizi nel tempo; rendere conto delle spese e dei risultati; servire come prima fase nel *benchmarking*; e determinare la soddisfazione dell'utenza per la rete e i servizi.

In generale, sono sostanzialmente due gli approcci che possono essere adottati per condurre un'analisi bibliometrica della biblioteca digitale.

Il primo approccio è quello che vede l'impiego delle metodologie di analisi citazionali normalmente impiegate per la valutazione delle collezioni della biblioteca tradizionale. Ri-

15 Tra gli studi specificatamente dedicati alla bibliometria nel contesto delle biblioteche digitali si ricorda: Borgman (1990), che rappresenta uno dei pochi contributi a stampa sull'argomento; Bollen – Luce (2002); e Bollen – Luce – Vemulapalli (2003).

16 Diversi i modelli di valutazione delle biblioteche digitali proposti in questi anni nella vasta letteratura professionale sull'argomento. Tra gli altri si segnala: Marchionini (2000); Saracevic (2000); Fuhr et al. (2001); DeLone – McLean (2003); Zhang (2010); Cassella (2010).

17 Uno studio di caso sulla valutazione economica del catalogo è l'indagine condotta nel maggio 2002 dalla Biblioteca Nazionale neozelandese. Cfr. National Library of New Zealand, National Bibliographic database and National Union Catalogue: economic valuation, Research report, October 2002.

18 Sulla valutazione del *reference* digitale segnalò a titolo esemplificativo l'articolo di Pomerantz – Mon – McLure (2008).

spetto però a quanto spiegato nel paragrafo 2.2., l'analisi citazionale nel contesto digitale non si limita a determinare l'impatto di autori e documenti all'interno della collezione come normalmente avviene per la biblioteca tradizionale, essa trasforma di fatto la stessa biblioteca digitale in un indicatore. Se infatti si utilizza il metodo per determinare per quante volte e dove la stessa biblioteca è stata citata, si crea automaticamente un indicatore del suo impatto che, a sua volta, interessa una comunità ben più ampia dell'utenza di riferimento della biblioteca.

Il secondo approccio è invece centrato sull'utente, o meglio, sui dati d'utilizzo generati automaticamente dagli utenti. I dati d'uso possono essere analizzati alla stregua delle citazioni, quantitativamente o strutturalmente. Come spiegato da De Robbio (2008 p. 22), se gli indicatori generati dall'autore sono le citazioni, quelli generati dai lettori sono i dati sull'utilizzo. La fonte più nuova di raccolta di statistiche di rete è rappresentata dai *log files* o *transaction logs*. I *log files* contengono la registrazione delle interazioni che intervengono nel momento in cui un utente sottopone una ricerca al sistema. Analizzando i *log files* è possibile raccogliere molte informazioni utili ad un'indagine bibliometrica come (Galluzzi 2001 p. 6): il numero di accessi¹⁹ e delle visite²⁰; il numero di byte trasferiti; quali risorse sono state ricercate; da quali *workstation*; quando; e i numeri di indirizzi IP da cui sono state effettuate le connessioni. È necessario però normalizzare il lavoro di raccolta dei dati di *log* al fine della condivisione di metodi per ottenere analisi significative (De Robbio 2008 p. 6).

Un esempio innovativo di impiego della bibliometria nel contesto delle biblioteche digitali si deve ad uno studio di Bollen – Luce (2002). Nel tentativo di valutare quantitativamente l'impatto della collezione e dei servizi di una biblioteca digitale, in rapporto soprattutto alla loro capacità di soddisfare il bisogno informativo dell'utenza, Bollen e Luce hanno usato i dati raccolti analizzando i *log files* per esaminare le relazioni esistenti tra i documenti nel contesto digitale di una biblioteca. Partendo dall'analisi dei modelli di ricerca dell'utenza, i due autori scoprirono che i dati raccolti potevano essere utilizzati per costruire un indicatore capace di misurare l'impatto di un documento su una specifica comunità virtuale.

19 Gli accessi si identificano con i *download* effettuati su una pagina html (ogni volta che si scarica una pagina, si scaricano tutte le immagini contenute in quella pagina).

20 Le visite identificano il numero di *download* di un'intera pagina.

Nello specifico, Bollen e Luce sono stati capaci di determinare quali documenti in una biblioteca digitale sono considerati dall'utenza di pari valore informativo e quali sono più frequentemente ricercati e scaricati. Bollen e Luce propongono una nuova applicazione al contesto della biblioteca digitale della classica tecnica della co-citazione. Se le co-citazioni sono coppie di citazioni condivise da due o più documenti, nel contesto digitale i *co-retrieved documents* sono due o più documenti ricercati in stretta prossimità temporale dallo stesso utente. Se un articolo che cita in bibliografia due documenti effettua un qualche tipo di associazione tra essi, allora l'utente che ricerca sulla stessa piattaforma due documenti in un breve lasso di tempo crea una qualche associazione tra essi. Se altri utenti operano la stessa associazione ricercando contemporaneamente gli stessi documenti, allora è ipotizzabile che esista una reale correlazione tra di essi. La frequenza di documenti *co-retrieved* misura perciò il grado di affinità (concettuale o metodologica) tra i documenti così come è percepita dalla comunità degli utenti. Come accade per le co-citazioni, anche i *co-retrieved documents* possono essere utilizzati per definire e visualizzare graficamente la struttura di una disciplina, mediante la costruzione di una mappa che evidenzia il reticolo di correlazioni tra i documenti. Dall'analisi di queste reti Bollen e Luce hanno creato un indicatore di impatto del documento, il *Journal Consultation Frequency (JCF)*, che non è altro che la somma del numero di connessioni effettuate e ricevute da uno specifico documento rispetto agli altri documenti che compongono la biblioteca digitale.

L'indicatore proposto da Bollen e Luce non è tuttavia l'unico esempio di indicatore d'uso non basato sulle citazioni. Il contesto digitale e la massiccia transizione delle riviste scientifiche al formato elettronico hanno infatti reso disponibile online una vasta mole di letteratura che, a sua volta, ha contribuito a rendere più agevole rispetto al passato l'analisi statistica dell'uso dei documenti (*download*, visualizzazione online) attraverso il monitoraggio dei *log files* generati dai *server web* e favorito l'esplorazione di nuovi indicatori quantitativi come ad esempio lo *Usage Factor (UF)*, che utilizza il *download* della versione digitale di un articolo per misurare l'impatto delle riviste, o il *Web Impact Factor (WIF)* che misura la frequenza con cui la pagina web "media" di un sito è linkata in un certo momento dall'utente.

La mancanza poi di un vero e proprio standard per la misurazione dell'uso delle collezioni digitali e, nel contempo, l'esigenza di normalizzare l'elaborazione di report statistici ha determinato il proliferare di iniziative e progetti internazionali come *COUNTER*, *SUSHI* e *MESUR* allo scopo di definire un linguaggio comune e metodi condivisi di raccolta e analisi dei dati.²¹

²¹ Agli indicatori quantitativi non citazionali e ai progetti internazionali per la definizione di standard comuni e di metodi condivisi di raccolta e analisi dei dati d'utilizzo sarà dedicato l'intero capitolo 6, dove riceveranno una trattazione più ampia e dettagliata.

3. L'analisi citazionale: il più diffuso dei metodi bibliometrici

Come visto nei capitoli 1 e 2, l'analisi citazionale rappresenta lo strumento portante e più noto della bibliometria.

È bene tuttavia ricordare che anche l'analisi di contenuto, o analisi testuale, che si basa sulla tecnica delle co-occorrenze di termini (*co-word-analysis*), è di fatto un metodo bibliometrico frequentemente usato, con cui si studia la relazione esistente tra due o più parole tenendo conto della loro co-occorrenza in documenti (He 1999). In sostanza, partendo da un corpo di letteratura specializzata, si estraggono identici elementi del documento da specifici campi dei record bibliografici o da sezioni nel testo dei documenti per misurarne la frequenza. La tecnica di *co-word* viene spesso confusa con quella nota come *term co-occurrence analysis*, dalla quale invece si distingue non tanto per la procedura impiegata, che rimane pressoché identica, ma perché non utilizza come unità di analisi i termini espressi nel linguaggio naturale distribuiti nel testo (Schneider 2004 p. 135). La co-occorrenza nella tecnica di *co-word* infatti può riguardare «le parole del titolo o dell'*abstract*, le parole chiave o i codici di classificazione oppure i termini scelti convenzionalmente (da autori o bibliotecari) per rappresentare il contenuto dei documenti» (De Bellis 2005 p. 93).

Entrambe le tecniche comunque sono importanti strumenti bibliometrici per la creazione di tesauri, la misurazione della frequenza di termini e l'esplorazione delle strutture grammaticali e sintattiche dei testi.

In questo capitolo verranno analizzate le tecniche che sono invece specifiche dell'analisi citazionale, vale a dire l'analisi delle co-citazioni (*co-citation analysis*) e l'analisi degli accoppiamenti bibliografici (*bibliographic coupling analysis*), non prima però di avere esaminato il significato e il ruolo di quello che rappresenta il fulcro del metodo, nonché il fenomeno di base su cui sono costruiti gli indicatori bibliometrici più utilizzati: la citazione bibliografica.

3.1. La citazione bibliografica

3.1.1. Il significato

La citazione bibliografica è una registrazione convenzionale degli elementi bibliografici di un documento che, collocandosi all'interno di un testo, permette l'individuazione univoca di una fonte. Il rimando alla fonte di qualsiasi citazione, fatto normalmente per supportare o rinforzare il punto di vista di un autore, o semplicemente per fornire un termine di confronto o elementi informativi aggiuntivi, è diventato una pratica quasi necessaria per qualsiasi testo che abbia l'obiettivo di essere considerato come scientifico o accademico.

Da un punto di vista puramente funzionale, la citazione bibliografica, riportando soltanto le caratteristiche essenziali dei documenti, non ha la pretesa di descrivere i documenti in modo completo e dettagliato. Essa perciò presenta un formato essenziale e scarno, completamente diverso da quello delle descrizioni bibliografiche che si trovano nei cataloghi o nelle bibliografie nazionali per ordine di disposizione dei vari elementi (autore e data non si trovano mai all'inizio nelle descrizioni bibliografiche) e per l'assenza di alcune informazioni riguardanti in particolare l'edizione, la collana e le dimensioni fisiche del documento. Per quanto concerne poi il reperimento materiale dei documenti, la citazione, pur rappresentando un importante punto di inizio, non è sufficiente, perché non fornisce alcuna informazione sulla disponibilità di copie di documenti. A tal fine sarà quindi sempre necessario l'uso complementare di cataloghi di biblioteche o di case editrici.¹

Per definire meglio il significato della citazione è necessario soffermarsi brevemente su «una differenza che le consuetudini linguistiche tendono appiattare» (De Bellis 2005 p. 14) e che a prima vista può sembrare un aspetto puramente terminologico, ma che si rivela invece molto importante nel contesto bibliometrico, e cioè la distinzione tra “riferimento bibliografico” e “citazione”. Usarli in modo improprio, visto che spesso vengono impiegati in modo intercambiabile e definiti in modo analogo, pur avendo un ruolo diverso e specifico in un documento, può infatti comportare una riduzione del loro contenuto informativo-comunicativo (Osareh 1996 p. 152-153).

¹ Tra i contributi in lingua italiana che trattano la citazione bibliografica si ricorda: Eco (2001); Corsi – Manzi (2006); Dell'Orso (2000); Gnoli (2000); Revelli (2010); Ridi (1995 e 2006); Serrai (2010); Usberti (2002 e 2010); e Vianello (1970).

Il riferimento bibliografico è un elemento del testo citante, con cui l'autore crea una sorta di collegamento con il documento citato allo scopo di comunicare al lettore l'affinità di teorie e concetti esistente tra documento citante e citato. La citazione, al contrario, è un elemento del testo citato, è l'inversione speculare del riferimento bibliografico. «Il riferimento bibliografico è orientato al passato, guarda all'indietro, mentre la citazione implica un movimento in avanti, verso il futuro» (De Bellis 2005 p. 15). In breve, come spiegato da De Bellis (2005 p. 14): «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, 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». Questa distinzione appare evidente se si utilizza la banca dati multidisciplinare prodotta dall'*Institute of Scientific Information (ISI): Web of Science (WoS)*. All'interno della banca dati infatti è previsto l'aggiornamento settimanale di due indici che ricalcano la distinzione tra citazione e riferimento bibliografico: il *Citation Index* e il *Source Index*. Con il *Citation Index* è possibile la ricerca di un autore a partire dalle sue citazioni, mentre con il *Source Index* è possibile la ricerca di un autore presente come "fonte informativa" (referenza) e non come autore citato. Un autore è presente nel *Citation Index* solo quando un suo lavoro viene citato, mentre è presente nel *Source Index* se un suo lavoro è stato pubblicato su una delle riviste indicizzate e inserite nella base dati *WoS*, a prescindere dall'eventuale citazione. Dunque la presenza nel *Source Index* diventa un prerequisito necessario e fondamentale perché l'eventuale citazione risulti rilevante ai fini degli strumenti bibliografici dell'*ISI* (Di Cesare 2002 p. 140).

3.1.2. Il comportamento citazionale

Al di là delle definizioni più o meno complete di citazione, il suo significato, cosa siano in grado concretamente di misurare e la complessità del processo che contraddistingue il "comportamento citazionale" sono al centro di un fitto dibattito teorico e di un non trascurabile lavoro empirico teso a sostenere o respingere proposte teoriche alternative.

Si possono distinguere due approcci teorici molto diversi, entrambi concentrati nel dare una spiegazione, o anche soltanto nel descrivere le linee principali, del comportamento ci-

tazionale, che oltre ad essere un processo complicato da capire è anche condizionato da variabili esterne non sempre riportabili sul piano scientifico: la teoria normativa delle citazioni e la teoria socio-costruttivista del comportamento citazionale. Per i sostenitori della teoria normativa il conteggio delle citazioni rende conto dell'impatto e del prestigio di un lavoro scientifico. In sostanza la teoria normativa si basa sul pensiero di Merton, già ricordato nel paragrafo 1.4.2., secondo cui il ricercatore riconosce il proprio credito nei confronti dei colleghi citando lavori che lo hanno influenzato o su cui si è basato per dare vita al proprio lavoro (Merton 1988 p. 622). Più in generale, la teoria si basa sull'idea che la citazione rientri nel sistema premiale di riconoscimento dei meriti, adottato dalle comunità scientifiche. La citazione bibliografica cioè è lo strumento principale attraverso cui avviene il riconoscimento dei pari.

L'approccio teorico dei socio-costruttivisti, all'opposto, ritiene che il conteggio delle citazioni non rifletta l'impatto dell'attività scientifica, poiché il comportamento citazionale è condizionato da molti fattori che devono essere attentamente considerati, e che di fatto, distorcono il significato di ogni indicatore citazionale. La visione costruttivista in sostanza vede le citazioni come strumenti retorici di persuasione (Gilbert 1977), passibili, in aggiunta, di manipolazioni. Per i socio-costruttivisti gli scienziati sceglierebbero le citazioni non tanto sulla base del contenuto del lavoro oggetto di citazione, ma soprattutto in relazione ad alcune sue caratteristiche estrinseche come l'eminenza dell'autore, la sua collocazione editoriale o il grado di autorevolezza già associato dalla comunità scientifica a quel lavoro (Moed – Garfield 2004; Egghe – Ravichandra Rao – Bhusan Sahoo 2006).²

La contrapposizione tra le due teorie ha indotto l'avvio di molto lavoro empirico. Un primo gruppo di contributi si è focalizzato sulla ricerca di evidenze in grado di confermare la teoria normativa. Degni di nota, per esempio, i lavori di Cole – Cole (1967) e Garfield (1986) che hanno dimostrato la relazione positiva tra le misure delle citazioni, molto più numerose rispetto alla media, e il conseguimento di premi e onorificenze (il premio Nobel su tutti). Altri lavori più vicini alla visione costruttivista invece hanno mostrato come le citazioni difficilmente riflettano il solo impatto della ricerca scientifica. Il primo di questi contributi è quello di Garfield (1965) che elenca 15 funzioni diverse della citazione, in ag-

² Per Latour (1987 p. 37-38) la scelta di citare sarebbe politica, dettata dal rispetto di alcune regole: «indebolisci i tuoi nemici, paralizza quelli che non puoi indebolire [...], aiuta i tuoi alleati quando sono attaccati, assicurati la comunicazione con quelli che ti forniscono strumenti inattaccabili [...], obbliga i tuoi nemici a combattersi tra loro».

giunta a quello generico di documentare un'affermazione fatta nel testo: 1) tributare un omaggio ai pionieri; 2) attribuire crediti a lavori 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 di altri; 7) criticare lavori precedenti; 8) dimostrare i diritti di proprietà; 9) indicare lavori in corso di pubblicazione; 10) fornire indicazioni su lavori poco diffusi, non indicizzati o non citati; 11) autenticare dati e classi di fatti (costanti fisiche ecc...); 12) identificare le pubblicazioni originali nelle quali un'idea o un concetto è stato discusso; 13) identificare pubblicazioni originali o altri lavori che descrivono un concetto o un termine eponimo (morbo di Parkinson; bosone di Higgs); 14) negare il lavoro o le idee di altri (indicazioni negativa); 15) disputare le rivendicazioni di priorità di altri (omaggio negativo). Questa lista di Garfield comprende ragioni riconducibili sia alla teoria normativa che costruttivista, a cui in anni recenti se ne sono aggiunte molte altre dando luogo a numerose "categorie citazionali".

L'analisi empirica su questi temi ha seguito tre filoni diversi. Il primo si pone l'obiettivo di testare la correttezza delle teorie citazionali. Baldi (1998), studiando le citazioni nel campo dell'astrofisica, propende per la teoria normativa, poiché le caratteristiche non contenutistiche degli articoli non influenzano significativamente la probabilità di essere inseriti nella lista delle *references*. White (2004), studiando le distribuzioni delle citazioni nel database *ISI*, giunge alle stesse conclusioni formulate da Baldi (1998). Collins (1999) invece in uno studio sulle citazioni relative ad alcuni contributi di teoria fisica giunge a conclusioni diametralmente opposte. Per l'autore, il numero di citazioni, ma soprattutto l'accettazione di idee innovative nella comunità dei fisici, dipenderebbero da una serie di variabili esterne, come il contesto istituzionale e l'esistenza di un pensiero ortodosso che controlla le riviste disciplinari più prestigiose al fine di convogliare i finanziamenti per la ricerca.

Il secondo filone di ricerca empirica prende atto della pluralità di ragioni che spiegano le citazioni e tenta di misurare il peso relativo delle varie categorie citazionali nei vari settori disciplinari, con risultati però talmente frammentari che è impossibile ricondurre a degli schemi generali (tra i molti contributi si ricorda: Bornmann – Daniel 2008; Cole – Cole 1971).

Il terzo filone di ricerca empirica si basa sull'idea che le citazioni a una pubblicazione non registrano solo la sua importanza, la sua qualità e il suo impatto (Peters – Van Raan 1994). Il numero di citazioni dipenderebbe da una pluralità di fattori (Baccini 2005 p. 144-146):

- 1) il tempo. La probabilità di essere citati è tanto più alta quanto più vicina è la data di pubblicazione dell'articolo citante a quella di pubblicazione dell'articolo citato. A questo si aggiunge l'effetto San Matteo teorizzato da Merton, secondo cui chi ha ricevuto citazioni in passato ha più possibilità di riceverne in futuro (vedi paragrafo 1.4.2);
- 2) le pratiche citazionali variano nelle diverse comunità scientifiche e sono radicalmente diverse nelle scienze naturali e ingegneristiche rispetto a quelle nelle scienze sociali e umane;
- 3) la probabilità di essere citati può dipendere da fattori riconducibili alle caratteristiche della rivista in cui l'articolo è pubblicato come la frequenza di pubblicazione della rivista su cui sono pubblicati gli articoli citanti, l'ordine in cui l'articolo è inserito in una rivista (gli articoli pubblicati per primi o nella parte iniziale della rivista ricevono in media più citazioni rispetto agli altri), l'accessibilità, la visibilità, il grado di internazionalizzazione, il prestigio della rivista e il suo Impact Factor;
- 4) il numero di citazioni può dipendere da alcune caratteristiche strutturali della pubblicazione. Così gli articoli che contengono la descrizione di una nuova tecnica di indagine, gli articoli metodologici, le rassegne, gli articoli di indagine applicata, le lettere, i commenti sono molto diversi tra loro. Anche la lunghezza e il numero dei co-autori dell'articolo mostrano una correlazione positiva con il numero di citazioni ricevute;
- 5) la probabilità di essere citati può dipendere da fattori legati alle caratteristiche personali dell'autore o del lettore di un certo articolo, come la lingua dell'articolo o il genere dell'autore;
- 6) il numero di citazioni ricevute può dipendere dall'accessibilità delle pubblicazioni. Alcuni studi hanno evidenziato la correlazione positiva tra numero di citazioni e accessibilità gratuita in rete degli articoli (Harnad – Brody 2004; Harnad et al. 2008);

- 7) gli errori nelle citazioni possono modificare anche in modo sostanziale l'affidabilità dei dati. Essi sono spesso dovuti agli autori che non controllano i riferimenti bibliografici o che li copiano dalle liste dei riferimenti usati in altri articoli (Simkin – Roychowdhury 2003; 2006). Un'altra possibile fonte di errore è l'uso degli *abstracts* per operare citazioni, che spesso riportano dati errati, inconsistenti o assenti rispetto a quelli contenuti nel corpo degli articoli stessi (Pitkin – Branagan – Burmeister 1999);
- 8) altri errori possono provenire da problemi tecnici connessi alla gestione e interrogazione dei database citazionali (Moed 2005; Bornmann – Daniel 2008). Basta sperimentare una ricerca per nome su *WoS* o *Scopus* per accorgersi che lo sforzo più grande per ottenere un risultato soddisfacente è legato alla necessità di “ripulire” i dati elementari dalle voci incongrue derivanti da omonimie o simili;
- 9) l'autocitazione, che può verificarsi quando un autore cita nel suo lavoro altre pubblicazioni di cui lui stesso è autore o co-autore, oppure quando un articolo pubblicato su una rivista è citato da articoli pubblicati sulla stessa rivista. In questi casi la pratica autocitazionale può rivelarsi sospetta, esiste cioè il dubbio che un ricercatore, un gruppo di ricercatori, o una rivista possono farne uso allo scopo di modificare, o meglio “gonfiare” i valori assunti dagli indicatori (Frey 2009).

3.1.3. Il ruolo della citazione

Da quanto fin qui esposto, appare evidente la complessità dell'intero processo di citazione poiché esso non riflette puramente il contenuto dell'opera citata.

In questo processo entrano in gioco altri e diversi fattori che possono dipendere dalle caratteristiche dell'autore, della rivista su cui si pubblica, della comunità scientifica, dell'istituzione di appartenenza che spingono a specifici comportamenti citazionali, o da altre variabili come il tempo, gli errori nelle citazioni, o il fenomeno delle autocitazioni che possono incidere positivamente o negativamente sul numero di citazioni ricevute da una pubblicazione.

Fatte queste opportune considerazioni è comunque ragionevole ritenere che la citazione non sia casuale o non sia «un orpello inutile o un artificio retorico del testo» (De Bellis 2005 p. 93). Essa invece può essere considerata una misura dell'impatto della ricerca, uno stru-

mento cioè capace di quantificare il grado di riconoscimento attribuito dalla comunità scientifica al contributo pubblicato e citato. Pur non essendo in generale la misura ideale della performance scientifica, essa è tuttavia considerata, specialmente a livelli di aggregazione relativamente elevati, un buon indicatore della rilevanza o dell'impatto del lavoro di ricerca di gruppi di ricerca o istituzioni, si rivela invece più deficitaria e bisognosa di ponderate compensazioni sul piano granulare della performance individuale (Van Raan 2005). Una procedura di valutazione che prevede l'uso delle sole citazioni ricevute da un prodotto di ricerca, senza un'attenta considerazione del significato delle citazioni e del comportamento citazionale, espone alla probabilità di trarre considerazioni sull'impatto del singolo contributo falsate da interferenze inappropriate.

Da più parti si sottolinea infatti le criticità che emergono dall'applicazione di indicatori citazionali al livello individuale della performance scientifica. Tra gli argomenti più forti a questo proposito vi è l'applicazione di indicatori bibliometrici citazionali medi, riferiti, per esempio, come l'*Impact Factor* alla rivista su cui è pubblicato un articolo, incapaci di rendere conto dell'impatto effettivo del singolo prodotto scientifico, oppure il fenomeno noto come "belle addormentate" (*sleeping beauties*), espressione con cui si indicano pubblicazioni che passano inosservate, e quindi non sono citate, per un lungo periodo di tempo, e che a un certo momento, quasi improvvisamente, attraggono l'attenzione della comunità scientifica (Van Raan 2004). La spiegazione di questo fenomeno è riconducibile al fatto che alcuni lavori precorrono i tempi, sono cioè troppo avanzati o innovativi per essere compresi a pieno subito dalla comunità scientifica. Ne consegue che quando il valutatore si trova di fronte a un articolo senza o con un bassissimo numero di citazioni in un ristretto arco temporale, non può considerarlo a priori un prodotto di bassa qualità, perché di scarso impatto, in quanto non può escludere che si tratti invece di un articolo che ha il potenziale per produrre buoni risultati soltanto in un momento successivo. Questo ragionamento quindi impone grande cautela ai valutatori non soltanto in fase di scelta dello strumento bibliometrico da impiegare, ma anche nella valutazione di tutti quei fattori esterni che potrebbero influire sul comportamento citazionale.

3.2. I primi lavori di analisi delle citazioni

Il primo studio in assoluto sulle citazioni bibliografiche venne condotto in ambito biblioteconomico da Charles Coffin Jewett attorno alla metà del XIX secolo. Jewett, bibliotecario e vicesegretario del *Smithsonian Institute*, utilizzò il dato citazionale allo scopo di redigere una bibliografia che potesse essere usata come strumento comparativo di valutazione del posseduto della biblioteca presso cui lavorava. Jewett, analizzando le 139 citazioni presenti in *History of International Law* di Wheaton, i 251 riferimenti bibliografici della *History of Commerce* di Hofer, le 204 voci dello studio di etnologia di J. A. Bartlett e le 38 citazioni del rapporto di chimica di Berzelius, constatò che la raccolta del *Smithsonian Institute* era del tutto inadeguata rispetto al posseduto delle più importanti biblioteche di ricerca americane ed europee (Jewett 1849).

È del 1927 il già ricordato saggio dei coniugi Gross, due chimici dell'università di Pomona, California del Sud, in cui per la prima volta venne applicato il metodo citazionale per identificare i periodici più rilevanti di un determinato ambito disciplinare e vennero applicate le tecniche bibliometriche alla gestione di una collezione bibliografica (Gross – Gross 1927). Essi ipotizzarono che la qualità di un periodico riconducibile a un determinato settore disciplinare fosse direttamente proporzionale al numero di volte in cui veniva citato nella letteratura di quella disciplina. Per il loro studio i Gross si servirono di un'annata della rivista *Journal of the American Chemical Society*, considerata particolarmente rappresentativa dell'ambito disciplinare, ne esaminarono i 3633 riferimenti bibliografici e, sulla base del numero di citazioni ricevute, classificarono i periodici più importanti (*key journals*) della disciplina. Il loro metodo fu in seguito esteso a settori disciplinari diversi dalla chimica tanto da diventare un modello per molte indagini condotte da autori successivi, tra cui anche Garfield, che riprese la metodologia operativa dei Gross individuando attraverso l'analisi delle citazioni i *core journals*, il nucleo di riviste di base per ogni settore o ambito disciplinare.

Del 1943 è invece il saggio finale per il conseguimento del Master alla *School of Library Service* della *Columbia University* di Estelle Brodman che nel 1944, nelle grandi linee, riprese in un breve articolo pubblicato sul *Bullettin of the Medical Library Association* (Brodman

1944). L'autrice, al contrario dei coniugi Gross, decise di non limitare l'indagine ad un'unica fonte informativa, anche se considerata una delle più rilevanti per un dato settore, come era stato fatto nello studio del 1927, ma di utilizzare invece più liste di periodici selezionate usando criteri diversi per ciascuna. A tal proposito, la Brodman comparò «una lista di periodici selezionata dai membri del dipartimento di fisiologia della *Columbia University* con altre due liste: una, ottenuta contando le citazioni bibliografiche dalla rivista *Annual Review of Physiology*, l'altra ottenuta contando quelle di altri tre periodici di fisiologi. I risultati da lei ottenuti non coincisero con quelli dei Gross, e questo permise alla Brodman di arrivare a delle importanti conclusioni. Per i Gross infatti qualsiasi rivista selezionata avrebbe dovuto essere rappresentativa dell'intera disciplina, e dunque le diverse riviste incluse nel campione avrebbero dovuto comportarsi, in relazione alla frequenza di citazioni, in modo analogo» (Di Cesare 2002 p. 137). Comparando tra loro le tre liste, la Brodman scoprì, al contrario, che ciascuna lista si comportava in modo diverso rispetto alle altre due, sovvertendo così l'ipotesi formulata dai Gross.

Con il suo lavoro la Brodman ebbe il merito di scoprire in anticipo alcuni punti di criticità rivelatesi in seguito fondamentali negli studi citazionali, come ad esempio, tra gli altri, il differente peso delle riviste, il quale dipende anche dalle caratteristiche strutturali della pubblicazione stessa, classificate come più rilevanti per una data disciplina.

Altro importante contributo è quello di Herman Fussler del 1949 (Fussler 1949), con il quale egli sviluppò ulteriormente l'applicazione della citazione all'individuazione di nuclei di riviste considerate cruciali per un ambito disciplinare. Fussler analizzò i riferimenti bibliografici di due periodici pubblicati dalle principali associazioni nazionali professionali e, proprio per questo, ritenute dall'autore sufficientemente rappresentative dell'intera letteratura nel campo della fisica e della chimica. Egli esaminò i riferimenti estrapolati ad intervalli temporali regolari da annate diverse (1899, 1919, 1939) di ciascuna delle due riviste di riferimento. Le riviste citate vennero poi raggruppate sulla base della frequenza di citazione ottenuta e la lista venne confrontata con una lista di controllo, organizzata in conformità con la classificazione della *Library of Congress*. Da questo confronto Fussler ottenne un elenco di riviste che identificò come facente parte dei nuclei di riviste più rilevanti per le due discipline. Le riviste maggiormente citate vennero inoltre esaminate in relazione a di-

versi indicatori, come per esempio: la data di pubblicazione e quella di citazione, gli autori secondo il paese di origine, la lingua o la tipologia del materiale citato (seriale e non).

Il lavoro di Fussler viene considerato, nell'ambito dell'analisi delle citazioni, la sintesi di tutto quello che gli studi bibliometrici avevano elaborato fino ad allora, e in un certo modo anche il contributo che chiude la fase esplorativa del metodo, la fase cioè da cui prenderà avvio una nuova stagione di studi citazionali, incentrati, o comunque stimolati, dalla figura di Eugene Garfield e dalla creazione degli indici di citazioni dell'*ISI*.

La proliferazione della letteratura scientifica e il conseguente fenomeno noto come *information overload* sono alla base della proposta di Eugene Garfield che, cosciente della stringente necessità di un controllo bibliografico, propose, in un articolo apparso su *Science* nel 1955 (Garfield 1955b), di estendere l'uso di indici di citazioni alla letteratura scientifica, analogamente a quanto avveniva fin dal 1873 nelle professioni legali con lo *Shepard's Citation*. In sostanza nel suo articolo Garfield presentò il progetto di un sistema bibliografico semi-automatizzato di indicizzazione delle citazioni come strumento insieme di verifica della bontà delle citazioni, capace quindi di «eliminare la citazione acritica di dati fraudolenti, incompleti o obsoleti consentendo allo studioso diligente di essere al corrente delle critiche indirizzate ai lavori precedenti» (Garfield 1955b p. 468), e di controllo bibliografico. Il sistema di Garfield, che concretamente ha significato la creazione dello *Science Citation Index*³, è coinciso con l'avvio della fase moderna degli studi citazionali i quali, focalizzandosi soprattutto sull'elaborazione e l'applicazione di indicatori bibliometrici capaci di misurare l'attività scientifica, hanno contribuito alla nascita di una letteratura vastissima, di taglio soprattutto empirico, che ormai riguarda ogni paese e ogni settore o ambito disciplinare.

3.3. Le tecniche citazionali

Il metodo dell'analisi citazionale è stato al centro di un processo di perfezionamento che, pur mantenendone inalterata la metodologia, ha dilatato l'unità informativa su cui poggia il metodo stesso. In altre parole, il metodo non si avvale solo ed esclusivamente del-

3 L'ideazione e la realizzazione dello *SCI* saranno discusse nel dettaglio nel paragrafo 4.1.

la citazione diretta, vale a dire della citazione di un documento precedente da parte di uno successivo, ma nel tempo ha sviluppato delle tecniche più raffinate ed efficaci, capaci cioè, spingendo l'analisi ancora più in profondità, di far emergere e misurare un numero sempre maggiore di nessi significativi e rappresentativi del grado di affinità tra i documenti.

Queste due misure, utilizzate per determinare la similarità tra due o più documenti e importantissime soprattutto alla luce delle recenti applicazioni al contesto digitale, sono: la tecnica dell'analisi delle co-citazioni e quella dell'analisi degli accoppiamenti bibliografici.

3.3.1. L'analisi delle co-citazioni (*cluster o co-citation analysis*)

La co-citazione è la condizione di due o più documenti che si trovano citati insieme da altri documenti. In altre parole, pur non citandosi direttamente, se i documenti A e B sono entrambi citati dal documento C, quest'ultimo effettua un qualche tipo di associazione tra di essi. Se poi i documenti A e B vengono citati contemporaneamente molte volte da altri documenti, allora la loro associazione non rappresenta un fatto accidentale, bensì è la conferma di una loro reale correlazione concettuale o metodologica.

L'analisi delle co-citazioni, misurando il numero di volte in cui due o più documenti sono stati citati insieme, rappresenta una tecnica quantitativa chiara e immediata per raggruppare o costruire dei *cluster* di documenti citati.

L'impiego della tecnica, teorizzata contemporaneamente, e indipendentemente, da Henry Small (1973) a Philadelphia e da Irina Marshakova (1973) a Mosca, parte dal presupposto che «se due documenti sono citati da un terzo, esiste un vincolo cognitivo tra i due documenti e che l'intensità di questo vincolo dipende dalla frequenza con cui i due documenti sono citati insieme nelle letterature specialistiche» (De Bellis 2005 p. 103).

La frequenza di co-citazione inoltre identifica le relazioni tra documenti che sono considerati come cruciali dagli autori della disciplina specifica, estendendo ulteriormente le possibilità offerte da questa tecnica citazionale.

Accettato l'assunto che i documenti molto citati corrispondano a concetti portanti, procedure o esperimenti caratterizzanti uno specifico ambito disciplinare, è possibile sfruttare questa tecnica, che per l'appunto riesce a estrapolare queste relazioni tra documenti o concetti chiave, 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 sottoree e nuovi fronti di ricerca» (De Bellis 2005 p. 103).

3.3.2. L'analisi degli accoppiamenti bibliografici (*bibliographic coupling analysis*)

L'accoppiamento bibliografico invece, pur operando sullo stesso principio, può essere considerato l'immagine speculare della co-citazione. Esso infatti è il rapporto tra due o più documenti che ne citano un terzo. L'accoppiamento bibliografico parte dal presupposto che, pur non citandosi direttamente tra loro, esista una qualche correlazione significativa tra documenti che condividono almeno un riferimento bibliografico.

Il primo a teorizzare il valore concettuale dell'accoppiamento bibliografico fu nel 1963 M. M. Kessler che provò direttamente la sua teoria analizzando gli articoli pubblicati in 35 volumi del *Physical Review* (Kessler 1963).

L'accoppiamento bibliografico è tuttavia considerato una tecnica molto statica poiché in definitiva è limitato ai soli riferimenti bibliografici che formano l'apparato bibliografico dei documenti accoppiati (Small – Griffith 1974 p. 20; De Bellis 2005 p. 102).

Proprio per tale caratteristica, giudicata negativamente da Garfield e dai suoi collaboratori, in fase di implementazione dello *SCI* all'accoppiamento bibliografico venne preferita la co-citazione proprio perché capace di rappresentare in modo più preciso tutte quelle relazioni in costante movimento che intercorrono tra i documenti.

4. Gli indici di citazioni

La bibliometria deve parte del suo sviluppo e della sua fortuna alla messa a punto e all'implementazione di quello speciale repertorio bibliografico rappresentato dall'indice di citazioni, quello strumento cioè che elenca tutti i riferimenti bibliografici contenuti in un *corpus* selezionato di documenti e abbina ad ogni singola pubblicazione citata il documento citante.

Il capitolo si apre con una breve panoramica storica dello strumento bibliografico allo scopo di far emergere le sue ampie finalità applicative che nel corso dei secoli sono passate dall'ambito giuridico-religioso a quello biblioteconomico, del recupero e dell'organizzazione dell'informazione, storico e valutativo della ricerca.

Con Garfield e il suo *SCI*, l'indice di citazioni diventerà quello strumento moderno conosciuto attualmente, dalle grandi potenzialità bibliometriche e rivoluzionario come sistema di *information retrieval*, che l'avvento del Web ha saputo ulteriormente sviluppare e che l'integrazione nella piattaforma *Web of Knowledge* ha trasformato in quel database multidisciplinare chiamato *WoS*, che rappresenta il più frequente punto di partenza per l'analisi citazionale. La fortuna di *WoS* ha inoltre contribuito alla nascita di altri database citazionali multidisciplinari come *Scopus*, di proprietà dell'editore *Elsevier*, e *Google Scholar*, quest'ultimo ancora più innovativo rispetto ai precedenti perché più alto è il suo livello di integrazione bibliografica anche con risorse non disponibili in formato digitale, ma soprattutto perché capace di sfruttare le enormi opportunità offerte dal mondo *open access*.

In questo capitolo i tre database multidisciplinari *WoS*, *Scopus* e *Google Scholar* saranno descritti in modo analitico, mentre riceveranno una trattazione più sommaria quei database che, pur essendo identici per funzionalità e struttura ai precedenti, sono specializzati in specifici ambiti disciplinari.

L'ultimo paragrafo sarà invece dedicato ai criteri di selezione dei materiali inseriti negli indici di citazioni, un aspetto questo per niente trascurabile poiché oltre a modificare direttamente il grado di copertura dello strumento, influenza indirettamente, fino anche ad arrivare a falsare, i valori degli indicatori bibliometrici, che proprio per il fatto di essere cal-

colati sulla base dei dati estratti da questi repertori citazionali, risentono delle scelte di inclusione o esclusione dei materiali da indicizzare effettuate dai curatori.

4.1. *Citation indexes: dai manoscritti ebraici (XII sec.) allo Science Citation Index (SCI)*

Sebbene i *citation indexes* siano considerati un prodotto del XX secolo, è in realtà solo la creazione della loro versione automatica o semi-automatica che può essere ricondotta, grazie al lavoro di Garfield, al secolo scorso. L'impiego di questo particolare tipo di strumento bibliografico ha infatti nella tradizione ebraica una storia ormai secolare alla spalle che poi è proseguita e si è ulteriormente rafforzata nel sistema giuridico anglosassone dei secoli XVIII e XIX.

Il primo indice di citazioni documentato in letteratura è contenuto nel manoscritto ebraico *Mafteah ha-Derashot* ("indice dell'omiletica") ed è organizzato in ordine alfabetico per frase biblica per permettere di trovare il punto esatto dove la frase stessa è citata all'interno di un ampio *corpus* di fonti rabbiniche. Il manoscritto è attribuito a Moshe ben Maimon, conosciuto anche come Maimonide (1135-1204), uno dei più eminenti filosofi medievali di tradizione ebraica, nonché importante rabbino e medico spagnolo. Nella tradizione ebraica gli indici di citazioni bibliche assolvevano ad una doppia finalità: erano utili per la preparazione di sermoni e indispensabili per questioni di natura legale (Weinberg 1997 p. 318).¹ Anche *'En Mishpat* ("fonte di giustizia"), il primo indice di citazioni ebraico compilato dopo l'invenzione della stampa,² venne inserito in un codice giuridico: il *Talmud* babilonese contenente la legge orale del giudaismo prodotta dal I al IV secolo d. C. *'En Mishpat*, indicando dove un dato argomento talmudico viene trattato nei codici giuridici successivi, rappresenta uno strumento bibliografico che guarda avanti nel tempo. Pur essendo impiegati quindi per scopi religiosi e, se si vuole, in minima parte anche a fini di ricerca alla stregua del loro utilizzo moderno, i *citation indexes* talmudici hanno però principalmente una funzione di natura giuridica, essi cioè sono stati ideati allo scopo di mostrare dove un pas-

1 Gaster (1929) descrive l'applicazione omiletica da parte dei predicatori degli indici di citazioni bibliche.

2 In realtà, *'En Mishpat* (1546) viene considerato soltanto il primo indice di citazioni ebraico compilato dopo l'invenzione della stampa. Per Hyman (1979 p. xii) e Gaster (1929) infatti il primissimo indice ebraico ad essere stato stampato, nel 1511, è *Hagadot-ha-Talmud* ("leggende del *Talmud*") che però, oltre a non riferirsi all'intero *Talmud*, ma soltanto a una sua parte, venne compilato prima dell'invenzione della stampa.

so viene citato nei codici della legge ebraica. Questa loro funzione è, a sua volta, direttamente riconducibile al meccanismo di funzionamento base su cui poggia la stessa legge ebraica, riassumibile nel concetto di «legittimazione verticale» del diritto (Weinreich 1980 p. 207): i giudizi rabbinici generalmente citano un'autorità precedente. «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» (De Bellis 2005 p. 67). Per chiarire questo meccanismo di stratificazione della legge ebraica e la conseguente funzione dei *citation indexes* diventa necessario un esempio. In un'edizione dell'indice di citazioni talmudico *'En Mishpat* pubblicato nel 1714 venne inserito un simbolo (una stella) per indicare tutte quelle decisioni inserite nel *Talmud* annullate da una norma di legge successiva (Weinberg 1997 p. 230; Rabinowitz 1952 p. 103). L'idea di inserire un simbolo per segnalare l'abrogazione di norme per effetto di decisioni successive sarà ripresa e applicata come modello nello *Shepard's Citation*, di cui si parlerà tra breve.

A ben guardare la componente giuridica dei *citation indexes* ebraici si è conservata anche in quelli moderni. Quest'ultimi infatti sono diventati strumenti capaci di misurare il grado di riconoscimento accordato ad una nuova idea dalla comunità dei pari legittimando al contempo il diritto di proprietà intellettuale del suo autore.

Sullo stesso concetto del “ragionare per precedenti” degli indici di citazioni ebraici poggia il sistema del *Common Law*, l'ordinamento giuridico diffuso in molti paesi anglosassoni, basato sui precedenti giurisprudenziali che, proprio per questa sua caratteristica, ha iniziato a produrre già dalla seconda metà del XVIII secolo tavole di “casi precedenti” inserite nelle rassegne giuridiche, che diventeranno, nel corso del XIX secolo, dei veri e propri indici di citazioni strutturati. Il primo esempio di tavole di “casi precedenti” è il *Raymond's Report* del 1743. Ma il primo indice di citazioni come lo si può concepire attualmente è *A Table of Cases in California as Affirmed, Overruled, Modified, Commented upon, or Altered by Statutory Enactment* compilato da Henry J. Labatt nel 1860. L'indice di Labatt venne organizzato in ordine alfabetico per nome del caso a cui si aggiunsero brevi annotazioni riguardanti le modificazioni apportate da casi successivi.

Un altro indice molto simile al precedente venne pubblicato nel 1872 per lo stato di New York da William Wait e a questo ne seguirono altri dedicati ad altri stati e giurisdizioni federali.

Ma il più importante successore di Labatt e Wait è stato Frank Shepard che già nel 1873 cominciò a stampare, su carta gommata, le citazioni ai casi giudiziari della Corte Suprema dell'Illinois, che poi venivano incollate dagli abbonati a questo servizio nelle proprie rassegne giudiziarie (Shepard 1948 p. 3). Successivamente attorno al 1875 lo *Shepard's Citation* si arricchì sia a livello di copertura geografica, estendendosi all'intero sistema giudiziario americano, sia a livello di contenuto, con l'introduzione di citazioni alle decisioni giudiziarie, agli statuti e ad altre fonti giurisprudenziali (Shapiro 1992 p. 337). L'indice di Shepard, che comunque rimane nella sua versione online uno strumento di consultazione ancora ampiamente utilizzato nel campo della ricerca giuridica,³ è strutturato in modo tale che, partendo da un caso discusso in uno dei tribunali dei 48 stati americani o nei tribunali federali, identificati in modo univoco da un codice formato dall'indicazione del volume e del numero di pagina del documento originario in cui è riportato, indica sotto di esso i riferimenti bibliografici di pubblicazioni successive che lo hanno trattato unitamente a tutte le istanze dei tribunali che vi si sono richiamati per confermare, mettere in discussione o capovolgere la sentenza originaria.

A partire dal XIX secolo si acutizza un fenomeno, evidente già da alcuni secoli soprattutto a scienziati e bibliotecari, noto come *information overload*, il sovraccarico informativo, dovuto alla crescita esponenziale delle pubblicazioni tecnico-scientifiche⁴ a cui si accompa-

3 Nel 1996 lo *Shepard's Citation* è stato acquistato dalla *LexisNexis*, una controllata dal 1994 della *Reed Elsevier*. Sebbene sempre meno consultata, la versione cartacea dell'indice è ancora in uso e dal marzo del 1999 è stata affiancata dalla versione online disponibile a pagamento sul sito di *LexisNexis*: <http://law.lexisnexis.com/shepards>.

4 Baldazzi (2002 p. 101) riporta il contributo di svariati autori che nel corso del '900 hanno descritto o cercato di quantificare il fenomeno: Schultz (1968) ipotizza che verosimilmente negli ultimi 500 anni l'umanità avesse prodotto circa 12 milioni di volumi, ma che ben 10 milioni fossero apparsi negli ultimi 100 anni; Rider (1944), ponendo problemi di gestione fisica, sul finire della seconda guerra mondiale, si preoccupava per la crescita esagerata di alcune biblioteche universitarie americane e più ancora di quelle di alcuni college femminili, che eludendo la legge di accrescimento cumulativo previsionale di un raddoppiamento dei volumi ogni 16 anni avevano invece raddoppiato nel giro di soli nove anni ed anche meno il loro patrimonio, che avrebbe finito per sommergere il bibliotecario e l'utente; Armstrong H. E., parlando della Società chimica di Londra, già nel 1894 poneva le sue preoccupazioni sul piano professionale, di utilizzo e di qualità di accesso all'informazione: «la letteratura chimica comincia ad essere inutilizzabile e incontrollabile per la sua vastità. Non solo per la quantità crescente di volumi di anno in anno, ma nuovi periodici escono costantemente. Si deve pur fare qualcosa per assistere i chimici e restare aggiornati nei loro campi» (Meadows 1974); e infine Thomas Young (Young 1845) che riferendosi al 1807 scriveva: «mentre contempliamo la stupefacente magnitudine della letteratura di ogni dipartimento scientifico esiste una più che buona ragione di apprensione per il fatto che, dalla continua moltiplicazione di nuovi saggi che sono mera

gna il timore di non riuscire ad approntare un controllo bibliografico adeguato. Ma è solo a partire dal secondo dopoguerra che la risoluzione di tale problema diventa una questione stringente. Con lo sviluppo dell'industria militare ereditata dalla seconda guerra mondiale, la corsa agli armamenti e la conquista dello spazio da parte delle due superpotenze USA-URSS unitamente alla necessità di sviluppare sistemi d'*intelligence* sempre più raffinati in cui l'informazione giocava e gioca un ruolo fondamentale poiché assicura alle aziende, alla ricerca e ai governi un vantaggio non trascurabile soprattutto se fruita nel minor tempo possibile, sono alla base di un'attenzione sempre più marcata per le innovazioni *machine-based-information*, che cominciano a farsi strada proprio in questo periodo, capaci di rendere più veloci e snelle le operazioni di ricerca delle informazioni da parte dei ricercatori e quindi di contribuire al progresso della scienza. Sono gli stessi scienziati e *information scientist* in enti di ricerca o imprese commerciali, insoddisfatti dei sistemi tradizionali di circolazione dell'informazione, a spingere per nuove forme di gestione bibliografica e a proporre modelli più efficaci di organizzazione delle conoscenze (Baldazzi 2002 p. 102).

È questo il caso di Eugene Garfield e dell'attività dell'istituto da lui fondato nel 1958, l'*ISI*, l'*Institute for Scientific Information*, che si configura come servizio d'informazione secondaria a sostegno della ricerca scientifica e della gestione d'impresa. In questo contesto storico e tecnologico Garfield svilupperà una nuova metodologia di recupero ed organizzazione dell'informazione, lo *SCI*, lo *Science Citation Index*, rifacendosi proprio al metodo utilizzato dallo *Shepard's Citation*.

Al di là del modello concreto rappresentato dallo *Shepard's* e dei contributi di autori importanti come Bernal, Bliss, Watson Davis e Wells, per l'ideazione dello *SCI* saranno determinanti i lavori di Shaw e Shera, due autori che negli anni '50 affrontano il problema del controllo bibliografico attraverso l'impiego di strumentazione elettronica, a cui Garfield direttamente rinvierà in un suo articolo del 2001 (Garfield 2001). Ralph Shaw per il suo "selettore d'informazione", uno strumento di indicizzazione automatica sviluppato negli anni '40 del XX secolo in collaborazione con Bush e gli ingegneri del *MIT* (Shaw 1951). Shera per aver utilizzato, a sostegno del principio di organizzazione bibliografica che a sua volta rappresenta uno dei fondamenti della sua filosofia sociale, proprio lo strumento messo a pun-

ripetizione di altri che sono stati dimenticati, la scienza sarà in breve inghiottita dal suo stesso crescente volume».

to da Shaw allo scopo di generalizzare i risultati della sperimentazione e di adottarne le applicazioni, stipulando a tal fine nel 1959 un contratto con la *Western Reserve's University* (Shera – Egan 1951).

Nell'ottica di sviluppo del suo progetto, i contributi di questi due autori sono importanti per Garfield perché sono i primi ad affrontare il problema del controllo bibliografico non in maniera teorica o astratta, bensì concretamente, prevedendo l'impiego di strumenti di indicizzazione automatica.

Garfield iniziò la sua carriera nel 1951, presso la *John Hopkins University* lavorando al *Welch Medical Library Indexing Project* per il settore chimico dove, occupandosi di indicizzazione per soggetto, venne a stretto contatto con i problemi inerenti il controllo bibliografico (Garfield 2000). Nella prospettiva di organizzare saperi facilmente accessibili, Garfield si rese conto che sia le classificazioni tradizionali sia le tecniche di soggettazione e indicizzazione risultavano inadeguate a trattare contenuti complessi e multidisciplinari, la scienza infatti non si manifesta in settori a sé stanti ma è concretamente multidisciplinare. Per questa ragione, anche gli indici della letteratura scientifica corrente (gli Indici di *Poole* o l'*Index Medicus*) presentavano a suo modo di vedere limiti importanti: un'accentuata duplicazione e sovrapposizione di *item*, una relativa dispersione dei soggetti, una terminologia spesso non significativa e troppo rigida, o comunque, non abbastanza flessibile per il ricercatore, e costi del trattamento semantico spesso superiori ai benefici (Garfield 2001; Baldazzi 2002 p. 105). Una possibile soluzione a questi problemi fu per Garfield quella di organizzare i dati dell'indicizzazione in un formato standardizzato e coerente che potesse essere utilizzato come un indice, idea questa che si inseriva perfettamente nel *Welch Project*, incentrato sulla *Current List of Medical Literature*, il cui scopo principale era quello di capire se i calcolatori elettronici dell'epoca potessero essere impiegati «per rendere più efficiente l'indicizzazione e la ricerca della letteratura biomedica» (De Bellis 2005 p. 68). In seno a questo progetto e spesso in anticipo sulla distribuzione a stampa dei fascicoli, nacquero i *Current Contents* (CC) che, grazie alla messa a punto di un sistema semi-automatico di indicizzazione basato sulla fotoriproduzione, resero disponibili a una vasta utenza, con periodicità settimanale, gli spogli di riviste di biblioteconomia e scienze dell'informazione (1953), di ma-

nagement e scienze sociali (1955), di farmacia e scienze biomediche (1957) e di scienze fisiche (1960).

Nel 1960 i *Currents Contents* vennero inseriti in un servizio di *document delivery* a pagamento e collegati a un repertorio di autori creato per facilitare lo scambio di informazioni e *reprint* tra i membri della comunità scientifica stessa. I *Currents Contents* fornirono i presupposti tecnici di un *citation index* per le scienze.

Intanto, nel marzo 1953 William Adair, vice presidente della società produttrice dello *Shepard's Citation* suggerì al *Welch Project*, che già si era dimostrato una valida soluzione al problema del controllo bibliografico, di considerare il metodo utilizzato dallo *Shepard's* come possibile tecnica di indicizzazione (Garfield 1979 p. 6-18). Dopo aver esaminato l'indice di Shepard, Garfield ebbe un'intuizione fondamentale: il criterio delle citazioni forniva una metodologia di indicizzazione delle recensioni che poteva essere estesa alla letteratura scientifica in generale. Egli capì in sostanza che era diventato opportuno spostare l'unità di analisi dalle parole ai riferimenti bibliografici: le prime infatti sono di per sé ambigue perché debbono molto del loro significato al contesto in cui si trovano; i secondi invece, riferendosi al contenuto del documento citato specificano il loro significato proprio attraverso l'intreccio di relazioni che si instaura con altri documenti.

I *citation indexes*, presentati da Garfield in un articolo apparso su *Science* nel 1955, costituiscono un vero e proprio «sistema bibliografico per la letteratura scientifica che può eliminare le citazioni non critiche di dati disonesti, incompleti, obsoleti [...] che offre un nuovo approccio al controllo dei soggetti con cui è indicizzata la letteratura scientifica. In virtù della sua diversa costruzione, esso tende a mettere insieme materiale che non potrebbe mai essere raccolto con l'usuale indicizzazione per soggetto. Può essere meglio descritto come un indice per associazione di idee e richiede al lettore solo il tempo necessario. Suggestioni per associazione sono offerte anche dagli indici convenzionali per soggetto ma soltanto all'interno dei limiti di un'intestazione per soggetto particolare» (Garfield 1955).

Il *citation index* di Garfield all'interno di un determinato ambito disciplinare è a ben guardare uno strumento bibliografico estremamente semplice. Si parte da un *corpus* prelezionato di riviste scientifiche e si catalogano tutti i riferimenti bibliografici contenuti negli articoli. L'elenco così ottenuto viene invertito e il riferimento bibliografico diventa fun-

zionale all'individuazione di tutti i documenti citanti. Partendo dal testo citato, l'indice di citazioni di Garfield avrebbe permesso in primo luogo di individuare l'insieme dei testi citanti, ma anche di misurarne il numero e la distribuzione nello spazio e nel tempo, informazioni cioè che avrebbero consentito di trarre anche delle considerazioni sull'impatto cognitivo del testo citato. In fase di realizzazione dell'indice, una delle questioni più delicate, a cui Garfield dovette trovare una soluzione adeguata, fu la copertura disciplinare, cioè la quantità e la qualità delle riviste da indicizzare. Convinto, come lo fu Bradford, che un numero ristretto di riviste produce la maggior parte dei risultati scientifici rilevanti, Garfield non puntò sulla copertura totale dell'indice, bensì sui criteri quali-quantitativi per estendere la copertura a di là del nucleo di riviste considerate fondamentali in un ambito disciplinare. A tal proposito, innovativo fu l'impiego del metodo di conteggio delle citazioni e, in particolare, dell'indicatore bibliometrico noto come fattore d'impatto (*Impact Factor*) delle riviste, vale a dire il numero medio di citazioni ricevute dagli articoli di una rivista in un determinato arco temporale.

Il primo vero banco di prova per l'indice di Garfield, o comunque, il passaggio dalla sua teorizzazione alla sua concretizzazione avvenne grazie all'interessamento dei due genetisti Gordon Allen e Joshua Lederberg alla ricerca di un sistema, per conto del *National Institutes of Health*, capace di rendere conto dell'impatto sulla comunità di studiosi esercitato dalle proprie ricerche. Nel 1963, grazie ai fondi assegnati dall'istituto, fu pubblicato il *Genetics Citation Index*; il repertorio si caratterizzò subito per il suo taglio multidisciplinare a cui Garfield giunse proprio lavorando agli articoli di interesse genetico. I contributi a questa disciplina erano contenuti solo per una parte molto ristretta in riviste specialistiche, la fetta maggiore si distribuiva infatti su un numero molto elevato di riviste di taglio generalista. La soluzione fu la messa a punto di una base di dati onnicomprensiva derivante dallo spoglio di 600 riviste coperte dai *Currents Contents* e lo sviluppo di un insieme di tecniche per estrarre in modo automatico, dal database iniziale, tutto il materiale ipoteticamente rilevante per la genetica. Soluzione questa che si rivelò a lungo andare vincente perché estendeva le potenzialità dell'indice «[allo] studio dei confini tra settori di ricerca e discipline, [alla] mappatura della rete di interrelazioni tra scienziati, gruppi di ricerca, [alla] valuta-

zione dell'impatto del lavoro individuale sulla comunità dei ricercatori» (De Bellis 2005 p. 75).

A seguito della cessazione dei finanziamenti erogati dal *National Institutes of Health*, l'ISI si trasformò in un'organizzazione di tipo commerciale anche perché oramai era diventato evidente che, vendendo l'accesso agli studiosi, agli amministratori di biblioteche e a tutti coloro che ne erano interessati, l'implementazione e il mantenimento dell'indice di citazioni si configurava come un vero e proprio *business*.

Nel 1964 lo *Science Citation Index* venne finalmente pubblicato e negli anni successivi lo seguirono altri due indici di citazioni specialistici: il *Social Sciences Citation Index* (1972) e l'*Arts & Humanities Citation Index* (1978).

Lo *Science Citation Index* consente attualmente di effettuare ricerche in più di 3700 riviste scientifiche e tecniche considerate tra le più rilevanti a livello internazionale che coprono oltre 100 discipline diverse. Esso comprende tre indici diversi pubblicati con periodicità trimestrale: il *Source Index*, il *Citation Index* e il *Permuterm Subject Index*.

Il *Source Index*, l'indice delle fonti, elenca, in ordine alfabetico per nome del primo autore, tutte le registrazioni bibliografiche dei documenti pubblicati nelle riviste indicizzate dell'anno in cui si è effettuato lo spoglio. Ogni registrazione fornisce il nome del primo autore, titolo completo dell'articolo (o una traduzione inglese del titolo in altra lingua), titolo della rivista, volume, fascicolo, paginazione, numero di riferimenti bibliografici e tipologia di fonte informativa.

Il *Citation Index*, l'indice di citazioni propriamente detto, elenca, in ordine alfabetico per nome del primo autore, anno di pubblicazione e titolo della rivista, ogni documento, edito o inedito, citato negli articoli delle fonti. L'indice elenca inoltre per ogni citazione gli estremi bibliografici necessari ad identificare gli articoli presenti nel *Source Index*, in cui proprio i documenti presenti nel *Citation Index* sono stati citati.

Appare quindi chiaro che il *Source Index*, essendo costruito sullo spoglio degli articoli appartenenti all'annata corrente delle riviste che l'ISI ha scelto di indicizzare, risulta rigidamente delimitato sia cronologicamente che per tipologia documentaria. Il *Citation Index*, al contrario, non presenta alcuna limitazione. A patto che essi siano citati negli articoli del

Source Index, al suo interno si possono trovare documenti di qualsiasi tipo e di autori vissuti in qualsiasi epoca, dunque non solo articoli e non solo autori contemporanei.

Il terzo indice di citazioni, il *Permuterm Subject Index*,⁵ l'indice permutato delle parole chiave, è costruito con le parole dei titoli degli articoli presenti nel *Source Index* (Garfield 1976). Impiegando quindi tali parole come voci di soggetto in corrispondenza dei nomi dei primi autori degli articoli di provenienza, l'indice si fonda sul linguaggio naturale, cioè la terminologia corrente della scienza e della tecnologia, e non su liste precompilate di termini o su sistemi di classificazione spesso modellati troppo rigidamente.

Alla struttura di base dello *SCI* si devono inoltre aggiungere due indici separati: il *Corporate Index*, l'indice delle organizzazioni di appartenenza degli autori; e il *Patent Citation Index*, l'indice dei brevetti citati nelle riviste indicizzate.

Nel 1992 l'*ISI* è stato venduto dal suo fondatore alla *Thomson Business Information*, una sezione della *Thomson Corporation* di Toronto che, a sua volta, nel 2008 ha acquisito l'agenzia di stampa *Reuters*, creando il gruppo *Thomson-Reuters*⁶, uno dei gruppi più importanti a livello mondiale nella creazione e distribuzione di informazioni di tutte le tipologie, e traghettando così definitivamente l'indice di Garfield nel contesto digitale.

Pur venendo ancora pubblicato nella versione cartacea, lo *SCI* è oggi integrato nella banca dati online *Web of Science (WoS)*⁷ fornita dalla *Thomson-Reuters* e accessibile dal portale *ISI Web of Knowledge*.

4.2. Le potenzialità degli indici di citazioni nel campo delle strategie di *information retrieval*

Se si considerano gli indici di citazioni sotto il profilo delle potenzialità offerte nel campo delle strategie di *information retrieval*, appare subito chiaro quanto innovativa sia la loro metodologia non solo rispetto ai tradizionali metodi di ricerca e recupero dell'informazione, ma anche come applicazione nel contesto digitale.

5 Come ben spiegato da De Bellis (De Bellis 2005 p. 78) «la 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».

6 http://thomsonreuters.com/content/press_room/corporate/212378

7 *WoS* sarà trattato dettagliatamente nel paragrafo 4.3.1.

L'indice di citazioni non fa altro che amplificare ulteriormente una delle caratteristiche che è alla base dell'unità informativa di cui è composto, vale a dire la capacità connettiva della citazione bibliografica. Nell'intento di rendere più accessibile l'informazione, Garfield pensò di raccogliere in un unico repertorio quell'insieme di collegamenti che si attivano tra i documenti quando uno di essi ne cita un altro.

Nei repertori bibliografici di questo tipo, la citazione non è più ancorata al documento di cui ha sempre fatto parte ma comincia ad acquisire un proprio valore informativo, diventa in sostanza uno strumento flessibile che, riflettendo il contenuto del documento citato, crea collegamenti spontanei tra concetti. L'indice di citazioni si configura quindi come una sorta di indice per associazione di idee che «non ha bisogno di *subject headings* e non richiede la mediazione concettuale dei bibliotecari» (De Bellis 2005 p. 14). Andando oltre si può tranquillamente affermare che questo repertorio bibliografico non subisce ingerenze esterne, non ha cioè bisogno per la sua costruzione e il suo funzionamento di trattamenti che puntano ad organizzare le informazioni, al massimo si può parlare di un'operazione di raccolta di dati, dal momento che la rete citazionale è creata spontaneamente e inconsapevolmente già dagli stessi scienziati. È proprio qui che si innesta l'aspetto innovativo della ricerca per citazioni reso possibile grazie allo sviluppo dei *citation indexes* rispetto alle tradizionali tecniche di *information retrieval*.

In un apparato bibliografico “senza soggetto” la tecnica è quella di recuperare, partendo da un documento che riteniamo rilevante, altri documenti in cui proprio quel contributo è stato inserito tra i riferimenti bibliografici dagli autori che, al pari nostro, lo hanno ritenuto rilevante per la loro ricerca. Se si compara infatti questa tecnica di ricerca bibliografica con metodologie quali la ricerca basata sui metadati (autore, titolo, data, soggetto, fonti etc.) effettuata nei cataloghi e nelle bibliografie o quella a partire dai riferimenti bibliografici di un documento considerato rilevante, nota anche come “ricerca a palla di neve” (Zani 2005; Ridi 2007 p. 206), si scopre che la citazione come strumento di *information retrieval* è più intuitiva e flessibile. Recuperare l'informazione mediante l'esplorazione della rete citazionale che si crea tra documenti, cioè lavorando in quell'apparato bibliografico costruito già di per sé dagli stessi scienziati, senza alcuna intermediazione esterna, ricalca perfettamente lo stile cognitivo di ricerca dell'informazione dello studioso. Per spiegare questo concetto bi-

sogna partire da un confronto tra il comportamento dello scienziato rispetto a quello del bibliotecario nell'accesso all'informazione. Il primo ha un approccio più mirato nel senso che quando avvia una ricerca ha sempre un elemento di partenza specifico in mente che può essere una citazione bibliografica, un determinato documento, un'idea, un fatto concreto, il secondo ha invece un approccio più generale all'informazione, il suo punto di partenza è infatti quasi sempre un concetto generale, una rubrica, un soggetto «che per quanto specifico risulta sempre generico per il ricercatore» (Baldazzi 2002 p. 106).

L'indice di citazioni inoltre non richiedendo un lavoro esterno di catalogazione semantica dei contenuti permette di effettuare una ricerca e un recupero dell'informazione non vincolati dai limiti dei sistemi di classificazione e soggettazione a cui invece rimane ancorata una ricerca basata sui metadati. Un accesso di tipo semantico all'informazione è gravato infatti da tutti quei punti di criticità insiti negli stessi processi di rappresentazione del contenuto concettuale del documento. Sebbene si tratti di un linguaggio controllato, una tecnica di indicizzazione come la catalogazione per soggetto risente in ogni caso della dinamicità del linguaggio naturale. Non è sempre facile per l'indicizzatore restituire il contenuto di un documento mediante dei termini normalizzati perché la natura del linguaggio è così dinamica che di continuo vengono introdotte parole nuove mentre altre scompaiono o vengono utilizzate con significati nuovi. Senza contare poi che l'uso delle parole varia da persona a persona e che una troppo rigida standardizzazione dei termini, allo scopo di garantire coerenza all'interno del sistema di indicizzazione, riduce in proporzione la ricchezza e la varietà descrittiva degli strumenti a disposizione degli indicizzatori. L'impiego invece della citazione bibliografica come elemento di indicizzazione fornisce una metodologia di ricerca che restituisce documenti dal contenuto semanticamente più preciso e più stabile. È più precisa innanzitutto perché supera lo scoglio di una terminologia spesso poco significativa o troppo rigida insita nei tradizionali sistemi di indicizzazione, ma anche perché la rete citazionale apporta quantitativamente più elementi, riesce cioè a specificare meglio il contenuto di un documento. È chiaro infatti che più grande è la quantità di informazioni impiegate per specificare un documento, più dettagliata è anche la descrizione dello stesso. In altri termini, più profondo è il livello di indicizzazione, più probabilità ha l'utente di trovare le informazioni di cui ha bisogno. Se si considera infatti che un artico-

lo contiene in media circa 15 citazioni, appare chiaro quanto più precisa sia una metodologia di ricerca condotta su un indice di citazioni che può contare in media su 15 chiavi d'accesso diverse ad un record bibliografico. Il notevole grado di profondità dell'indicizzazione che si ritrova nei repertori bibliografici di questo tipo rende la ricerca per citazione molto efficiente in quanto ottimizza il rapporto tra richiamo (recupero del maggior numero di documenti rilevanti rispetto all'obiettivo della ricerca) e precisione (recupero soltanto dei documenti rilevanti rispetto all'obiettivo della ricerca), riducendo di conseguenza il rumore (risultati non pertinenti recuperati). L'accesso all'informazione è più duraturo perché il contenuto dei documenti, o meglio, la sua rappresentazione è affidata alla stessa rete citazionale che oltre a renderla stabile ha la possibilità di specificarla gradualmente nel corso del tempo. Rispetto poi alla staticità di una ricerca condotta a partire da una bibliografia di un documento ritenuto rilevante, questa metodologia è più flessibile perché permette di recuperare le citazioni da e verso un determinato documento, vale a dire sia i riferimenti bibliografici contenuti nei documenti in questione sia le citazioni a quello stesso documento contenute negli articoli successivi la sua pubblicazione. Tutto ciò ha permesso di individuare oltre ai testi più rilevanti e a quelli che hanno segnato la storia o che rappresentano i filoni innovativi di un determinato settore, anche le relazioni che intercorrono tra aree disciplinari diverse consentendo così una ricerca multidisciplinare difficilmente attuabile attraverso le tradizionali tipologie di ricerca bibliografica (Baldazzi 2002 p. 107; De Bellis 2005 p. 84; Ridi 2007 p. 206-207).

I molti aspetti innovativi della ricerca per citazione non possono e non devono tuttavia far dimenticare i punti di criticità insiti nel metodo e, giustamente, i punti di forza di una tipologia di ricerca basata sui metadati. L'accesso di tipo semantico all'informazione garantito dal lavoro degli indicizzatori è basato in gran parte sull'analisi concettuale del documento che, pur presentando i punti deboli appena descritti, rimane comunque un metodo insostituibile quando si vuole ottenere la più esatta rappresentazione possibile del contenuto di un oggetto informativo, garanzia quest'ultima che non sempre la citazione e la rete citazionale riescono ad assicurare soprattutto se si considera che spesso gli autori citano altri documenti senza mai averli letti, ma copiando meccanicamente dalle liste di riferimenti bibliografici di altri documenti, o che la conoscenza dei documenti da citare è limitata alla

lettura dell'*abstract* che frequentemente riporta dati errati, inconsistenti o assenti rispetto a quelli contenuti nel corpo dei documenti stessi (vedi paragrafi 2.2. e 3.1.2.). Altro punto da tenere in mente è che con una ricerca basata sui metadati ogni documento ha la stessa identica probabilità di essere recuperato rispetto ad un altro, mentre con una ricerca per citazione la maggiore o minore visibilità di un documento all'interno della rete citazionale, e di conseguenza la maggiore o minore facilità di un suo recupero, dipende da se e quanto il documento stesso viene citato da altri documenti: se un documento viene poco citato sarà meno visibile all'interno della rete citazionale, dunque sarà più difficile recuperarlo; e nel caso in cui un documento, pur esistendo nel docuverso, non venga mai citato, automaticamente non potrà rappresentare un nodo della rete citazionale, dunque sarà impossibile un suo recupero. Rispetto poi alla ricerca per citazione, che può produrre nell'utente un senso di spaesamento per la complessità dei rinvii, la ricerca per metadati è più semplice e intuitiva, per cui l'utente, anche quello meno esperto, si sente molto più guidato nella ricerca.

Ritornando alle potenzialità di ricerca degli indici di citazioni è da dire, inoltre, che queste si sono ulteriormente ampliate in ambiente digitale. Nel Web, la separazione tra indici di citazioni, indici delle fonti e indici delle parole chiave di fatto scompare, permettendo una navigazione ipertestuale che «ha reso tangibili e fisicamente trasparenti i legami concettuali tra documenti incorporati nelle citazioni bibliografiche» (De Bellis 2005 p. 87). Se si utilizzano questi strumenti è possibile individuare i riferimenti bibliografici di un dato documento, gli articoli che lo hanno citato il grado di correlazione e di affinità con altri documenti mediante rispettivamente la tecnica dell'accoppiamento bibliografico e quella della co-citazione, ma è anche possibile modificare l'ordinamento dei risultati specificando il punto di accesso come avviene nei repertori bibliografici tradizionali (cataloghi e bibliografie). In sostanza ci troviamo di fronte ad uno speciale repertorio bibliografico che nella sua versione online riesce ad integrare tutte le tipologie di ricerca e recupero dell'informazione esistenti: quella per metadati, quella a partire dalla bibliografia di un documento rilevante, quella per citazione e, se si vuole, anche quella per *full-text* (De Bellis 2005 p. 87-88; Zani 2005; Ridi 2007 p. 207).

A questa tipologia di repertorio bibliografico appartengono le basi di dati citazionali come, per citare le più importanti, *WoS*, *Scopus* e *Google Scholar*, che saranno oggetto di trattazione nel paragrafo seguente.

4.3. Indici di citazioni multidisciplinari online

4.3.1. ISI Web of Science

Erede degli indici di citazioni inventati da Garfield negli anni '60 è *Web of Science*⁸, la grande banca dati bibliografica citazionale e multidisciplinare online dal 2002, accessibile a pagamento attraverso la piattaforma *Web of Knowledge*⁹ di proprietà della *Thomson-Reuters*.

La banca dati *WoS* consente di effettuare ricerche in oltre 12000 riviste, 148000 atti di convegni e, grazie al recente inserimento del *Book Citation Index*, in 28000 libri. *WoS* comprende nove indici di citazioni¹⁰ che possono essere raggruppati in quattro grandi aree:

1. il primo gruppo comprende, interrogabili sia simultaneamente che singolarmente, i tre database citazionali che sono la versione online aggiornata e ampliata dei tre indici citazionali di Garfield; complessivamente censiscono 12000 riviste internazionali suddivise in tre macro ambiti disciplinari:
 - a) *Science Citation Index Expanded (SCI-Expanded)*, con una copertura dal 1899 ad oggi, indicizza completamente oltre 7100 riviste in 150 discipline scientifiche;
 - b) *Social Sciences Citation Index (SSCI)*, con una copertura dal 1899 ad oggi, indicizza completamente oltre 2100 riviste in 50 discipline di scienze sociali e 3500 delle principali riviste scientifiche e tecniche del mondo;
 - c) *Arts & Humanities Citation Index (A&HCI)*, con una copertura dal 1975 ad oggi, indicizza completamente oltre 1200 riviste di arte e discipline umanistiche e anche voci selezionate da più di 6000 riviste scientifiche e di scienze sociali;
2. il secondo gruppo comprende due indici citazionali di atti di convegni, cioè la letteratura scientifica emersa (e in seguito pubblicata) durante i lavori di importanti con-

8 http://wokinfo.com/products_tools/multidisciplinary/webofscience/

9 <http://wokinfo.com/wok/>

10 Per maggiori informazioni sulle funzionalità, nonché suggerimenti di ricerca e esempi cfr. *Thomson-Reuters, Quick Reference Guide*, 2011, disponibile online:

http://thomsonreuters.com/content/science/pdf/ssr/training/wok5_wos_qrc_it.pdf

vegni, simposi, seminari, *workshop*, che indicizzano complessivamente oltre 148000 atti rientranti in 256 discipline e basati su riviste e libri sia di ambito scientifico che umanistico, con una copertura retrospettiva dal 1990:

- a) *Conference Proceedings Citation Index – Science (CPCI – S)*, per le discipline scientifiche;
 - b) *Conference Proceedings Citation Index – Social Science & Humanities (CPCI – SSH)*, per le scienze sociali e le discipline umanistiche;
3. il terzo gruppo comprende due indici di citazioni che censiscono complessivamente oltre 28000 libri e capitoli di libri selezionati a partire dall'anno di pubblicazione 2005:
- a) *Book Citation Index – Science (BKCI – S)*, per le discipline scientifiche;
 - b) *Book Citation Index – Social Sciences & Humanities (BKCI – SSH)*, per le scienze sociali e le discipline umanistiche;
4. il quarto gruppo comprende due indici settoriali per la chimica:
- a) *Index Chemicus (IC)*, con copertura retrospettiva dal 1993, indicizza le strutture e i dati critici di supporto di oltre 2,6 milioni di composti organici riportati in importanti riviste internazionali;
 - b) *Current Chemical Reactions (CCR - Expanded)*, con copertura retrospettiva dal 1986, indicizza i più recenti metodi di sintesi pubblicati nei più importanti brevetti e periodici internazionali specializzati nel campo della chimica organica, assicurando così l'accesso a circa un milione di reazioni.

Oltre a queste basi di dati che formano *WoS*, il portale *Web of Knowledge* ne ospita altre, tra cui *MEDLINE*¹¹ e *Inspec*¹², interrogabili sia singolarmente (cliccando sul pulsante “Select a database”) sia simultaneamente (cliccando sul pulsante “All databases”).

11 *MEDLINE (Medical Literature Analysis and Retrieval System Online)* <<http://wokinfo.com/media/pdf/qrc/medlineqrc.pdf>>, è il database bibliografico prodotto dalla *National Library of Medicine (NLM)* degli Stati Uniti. Esso ha una copertura retrospettiva dal 1950 e contiene circa 16 milioni di record tratti da oltre 5200 riviste scientifiche specializzate pubblicate in 37 lingue e rientranti negli ambiti disciplinari della medicina, dell'infermieristica, della farmacologia, dell'odontoiatria, della medicina veterinaria e dell'assistenza sanitaria in generale, ma anche della biologia e della biochimica.

12 *Inspec* <<http://wokinfo.com/media/pdf/qrc/inspecqrc.pdf>> è una base di dati bibliografica prodotta dall'*Institution of Engineering and Technology (IET)* che indicizza oltre 4000 riviste tecnico-scientifiche tra le più rappresentative a livello mondiale nei campi della fisica, dell'elettronica, dall'informatica, dell'ingegneria e dell'*information technology*. Aggiornata settimanalmente, *Inspec* contiene oltre 10,8 milioni di record, a cui se ne aggiungono circa 400000 di nuovi all'anno, e ha una copertura retrospettiva dal 1898.

Cliccando sul pulsante “Web of Science” si accede invece al database multidisciplinare che, come già ricordato, permette di effettuare una tipologia di ricerca sia per metadati sia per citazioni.

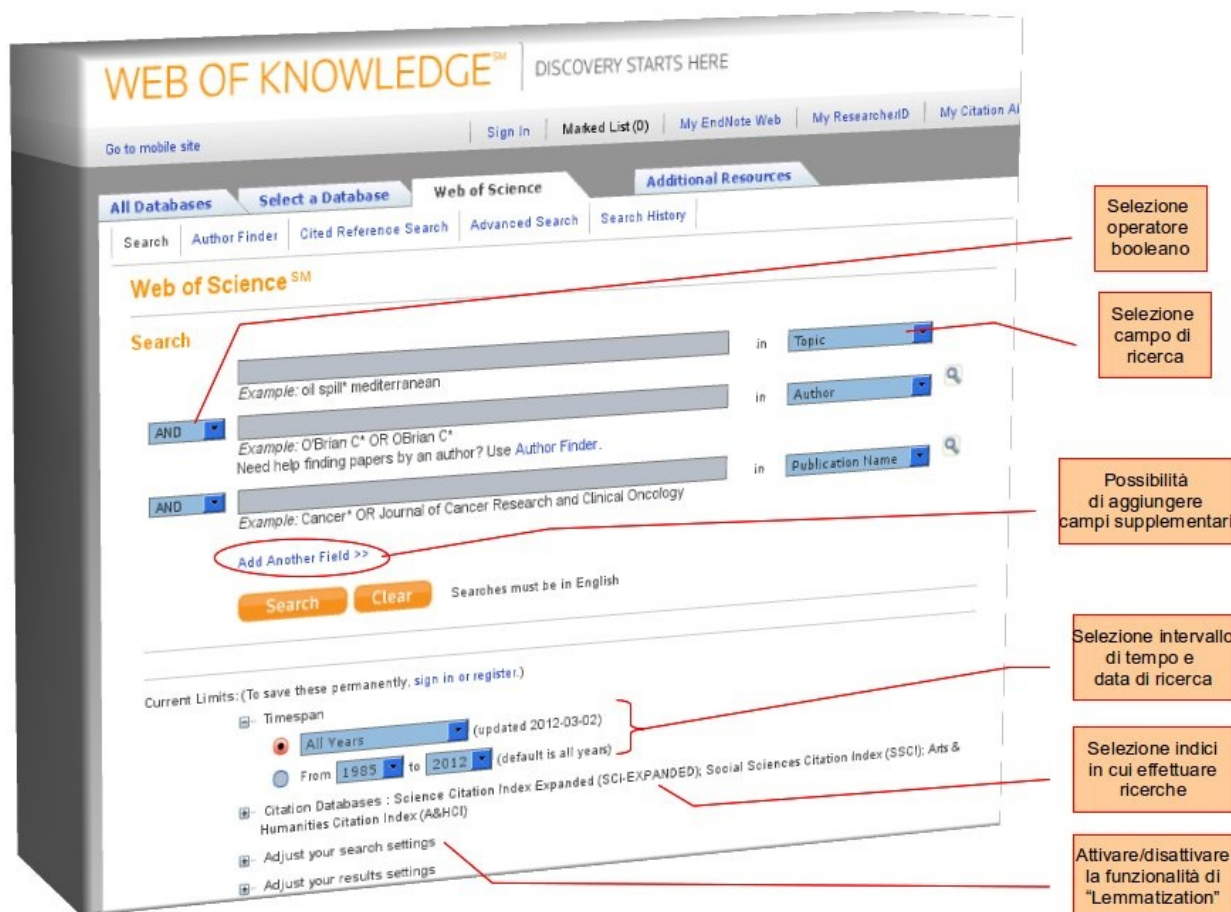


Figura 1: Schermata iniziale di Web of Science (rielaborazione da fonte WoS)

Per descrivere al meglio WoS è necessario, essendo una base di dati molto complessa sotto il profilo della varietà di azioni, spesso integrate, consentite all'utente, suddividere le sue funzionalità nelle seguenti quattro macro-aree: ricerca generale; ricerca di riferimenti citati; report e mappe di citazioni; e personalizzazione e gestione dei risultati.

Per ricerca generale si intende la possibilità di combinare parole e frasi per effettuare ricerche tra i record di WoS. Cliccando su “Search” si apre un form di interrogazione per metadati integrato con diversi strumenti che consentono all'utente di aumentare il grado di precisione della ricerca. L'utente può:

- a) utilizzando il menu a discesa posto in coincidenza di ogni casella di interrogazione, selezionare l'area di ricerca desiderata (argomento, titolo, autore, anno di pubblicazione etc.) o limitare la ricerca per lingua di pubblicazione o per tipologia documentaria;
- b) utilizzando il menu a discesa, selezionare gli operatori booleani AND, OR, NOT, per cambiare le relazioni tra i campi di ricerca;
- c) aggiungere campi di ricerca allo scopo di aumentare il grado di precisione dell'indagine;
- d) modificare i limiti cronologici e la data di ricerca;
- e) selezionare gli indici in cui effettuare la ricerca (il sistema effettua la ricerca solo nei prodotti per cui si è sottoscritto l'abbonamento);
- f) attivare o disattivare le funzionalità di "Lemmatization" (se attivata consente di trovare varianti mediante la riduzione al tema di parole flesse, di ricercare tempi verbali e livelli di comparazione aggettivali diversi);
- g) altre opzioni riguardano: l'uso degli operatori di prossimità quali NEAR/n e SAME; delle parentesi per raggruppare istruzioni booleane complesse, di virgolette per la ricerca di frasi esatte per i campi "Topic" e "Title"; e dei simboli di troncamento, l'asterisco (*) e il punto interrogativo (?), per ottenere forme plurali e varianti ortografiche con funzionalità "Lemmatization" disattivata.

Cliccando sul pulsante "Author Finder" l'utente ha la possibilità di fare ricerche per nome dell'autore, mentre selezionando "Advanced Search" il sistema permette di effettuare una ricerca avanzata con l'impiego di indicatori di campo ("*field tags*") e degli operatori logici in abbinamento non solo a parole o frasi ma anche a *set* di risultati.

In "Search History" sono visualizzate le ricerche effettuate in una sessione di lavoro. Il sistema permette di salvare una ricerca, di creare un messaggio di allerta (*Citation Alerts*) e di impostare *feed RSS*.

Una ricerca restituisce un record di questo tipo:

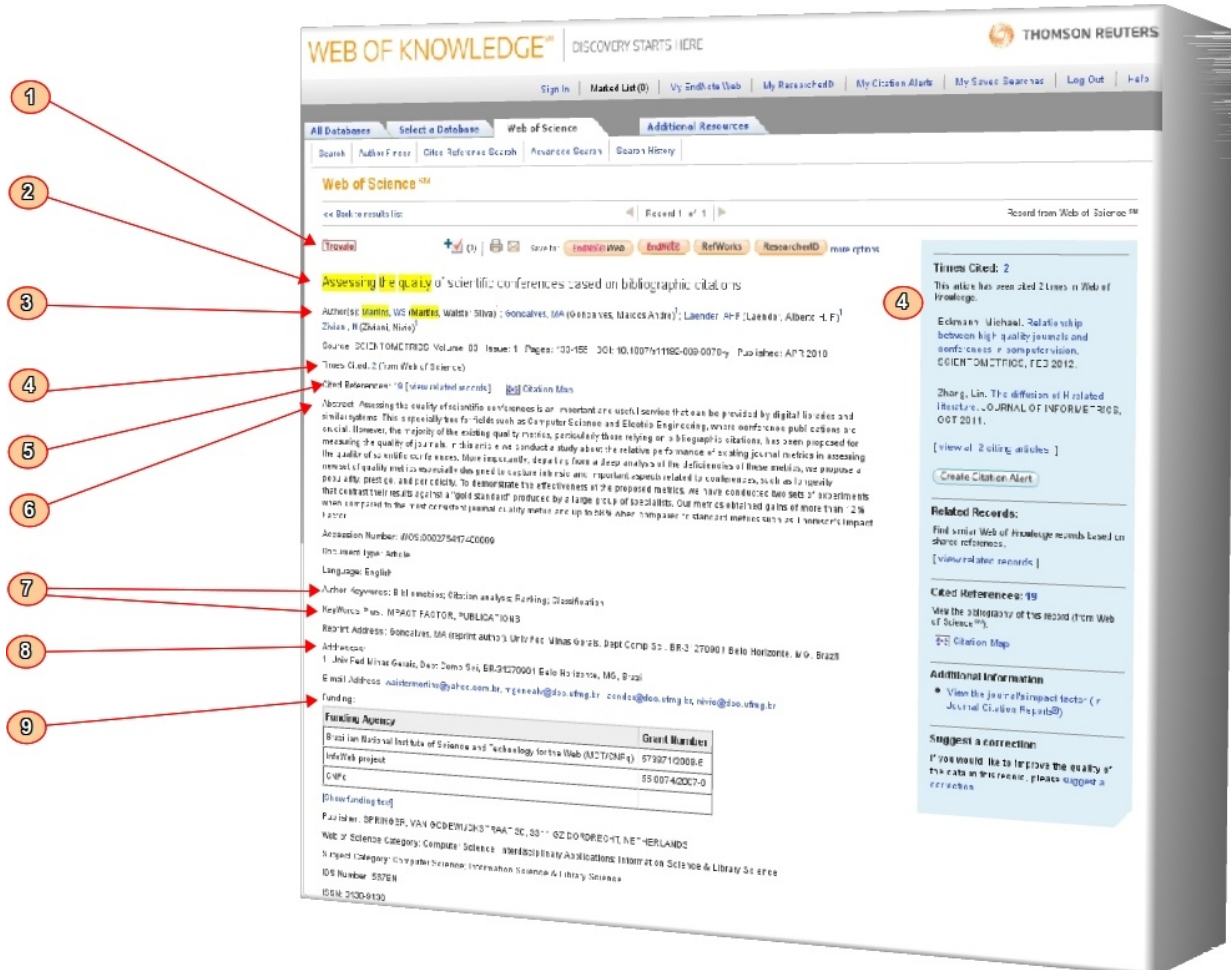


Figura 2: Schermata dei risultati di ricerca (rielaborazione da fonte WoS)

1	Link alle informazioni relative agli articoli completi e/o biblioteche
2	I TITOLI completi sono indicizzati e ricercabili
3	AUTHOR(S): indicizzati e ricercabili
4	TIMES CITED (conteggi delle citazioni ricevute dall'articolo): visualizzati in dettaglio nel riquadro di destra
5	CITED REFERENCES (riferimenti citati): link all'elenco completo dei relativi record
6	ABSTRACT: indicizzati nell'ordine di pubblicazione nella rivista (dal 1991 ad oggi)
7	AUTHOR KEYWORDS (parole chiave autore): indicizzate e ricercabili KEYWORDS PLUS: parole o frasi derivate dai titoli degli articoli citati
8	ADDRESSES (indirizzi degli autori): indicizzati e ricercabili
9	FUNDING (finanziamenti): -informazioni che includono anche l'ente finanziatore, numeri delle grant e testo relativo all'accettazione dei finanziamenti (dal 2008 ad oggi)- indicizzati e ricercabili

Tabella 1: Descrizione schermata dei risultati di ricerca

Come già indicato in figura 2, cliccando sul numero delle “cited references” è possibile passare alla relativa visualizzazione del record.

Cliccando sul pulsante “Cited Reference Search”, il sistema permette di effettuare una ricerca per ri-

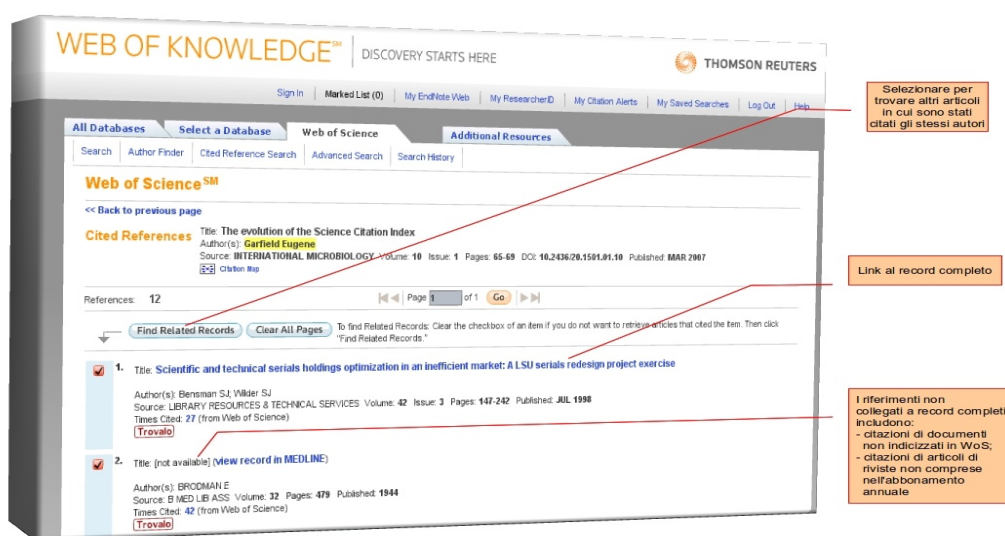


Figura 3: Schermata elenco record dei riferimenti citati (rielaborazione da fonte WoS)

fermenti citati. Anche in questo caso si apre un *form* di interrogazione per metadati in cui l'utente può:

- utilizzare operatori booleani, di prossimità, simboli di troncamento, parentesi e virgolette all'interno dei campi di ricerca;
- aggiungere campi supplementari;
- modificare i limiti cronologici e la data di ricerca;
- selezionare gli indici in cui effettuare la ricerca;
- attuare e disattivare la funzionalità di “Lemmatization”.

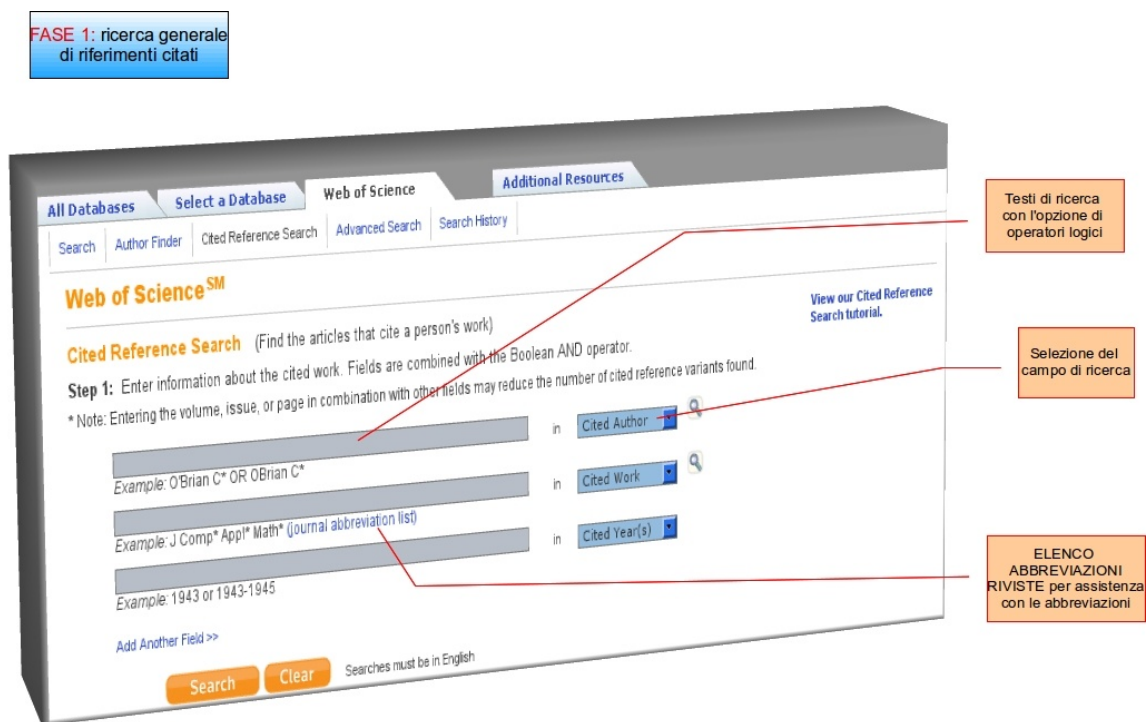


Figura 4: Schermata ricerca per riferimenti citati (rielaborazione da fonte WoS)

Con WoS è possibile visualizzare il riepilogo dei risultati di ricerca e procedere ad un loro raffinamento utilizzando “Refine” che permette di recuperare i primi 100 risultati nelle categorie di ricerca più significative.

FASE 2: selezione dei riferimenti da includere nella ricerca



Figura 5: Schermata selezione riferimenti di ricerca (rielaborazione da fonte WoS)

FASE 3: riepilogo dei risultati

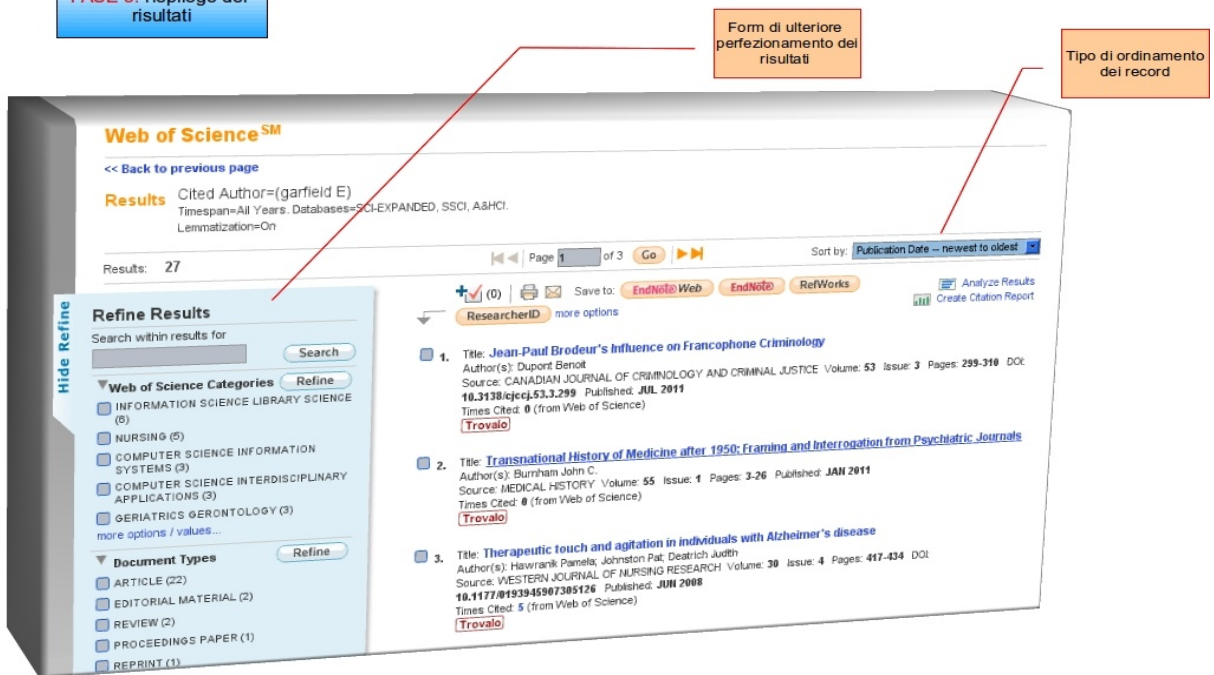


Figura 6: Schermata di riepilogo risultati di ricerca (rielaborazione da fonte WoS)

L'utente ha anche la possibilità di creare *report* e mappe di citazioni. Cliccando sul collegamento nell'angolo superiore destro di una pagina di riepilogo dei risultati è possibile generare un *report* di citazioni per l'insieme dei risultati di ricerca.

I *report* di citazioni possono offrire utili parametri di misurazione della produttività e delle prestazioni per un *set* di risultati che può arrivare a contare fino a 10000 record. I parametri sono: pubblicazioni in un anno; citazioni in un anno; H-index (n record con almeno n citazioni); citazioni totali per tutti i documenti; citazioni medie per elemento; numero di articoli di riferimento univoci.



Figura 7: Schermate creazione *report* di citazioni (rielaborazione da fonte WoS)

Cliccando invece sul pulsante "Citation map" collocato in corrispondenza del numero di riferimenti bibliografici "Cited references" è possibile creare una mappa di citazioni che consente di visualizzare fino a due generazioni di citazioni all'indietro e in avanti.



Figura 8: Schermate creazione mappa di citazioni (rielaborazione da fonte WoS)

L'utente di WoS può inoltre creare un profilo *Web of Knowledge* personale per sfruttare i vantaggi di efficaci opzioni di personalizzazione. La creazione di un profilo WoK consente di:

- salvare un numero illimitato di ricerche;
- creare *Search Alerts* (avvisi di ricerca);
- creare *Citation Alerts* (avvisi relativi a citazioni);
- creare una libreria *EndNote Web*;

- creare un profilo *ResearcherID*, uno spazio gratuito pubblico e online che può essere utilizzato per includere informazioni relative alle pubblicazioni istituzionali, agli interessi di ricerca e alle pubblicazioni personali dell'utente.

I risultati della ricerca possono essere gestiti attraverso *EndNote Web*¹³ che consente di salvare fino a un massimo di 10000 referenze bibliografiche. Lo strumento accessibile da *WoS* cliccando sul pulsante “My EndNote Web” permette di aggiungere e formattare i riferimenti bibliografici in un documento di testo e di effettuare ricerche in altre banche dati e cataloghi di biblioteche online.

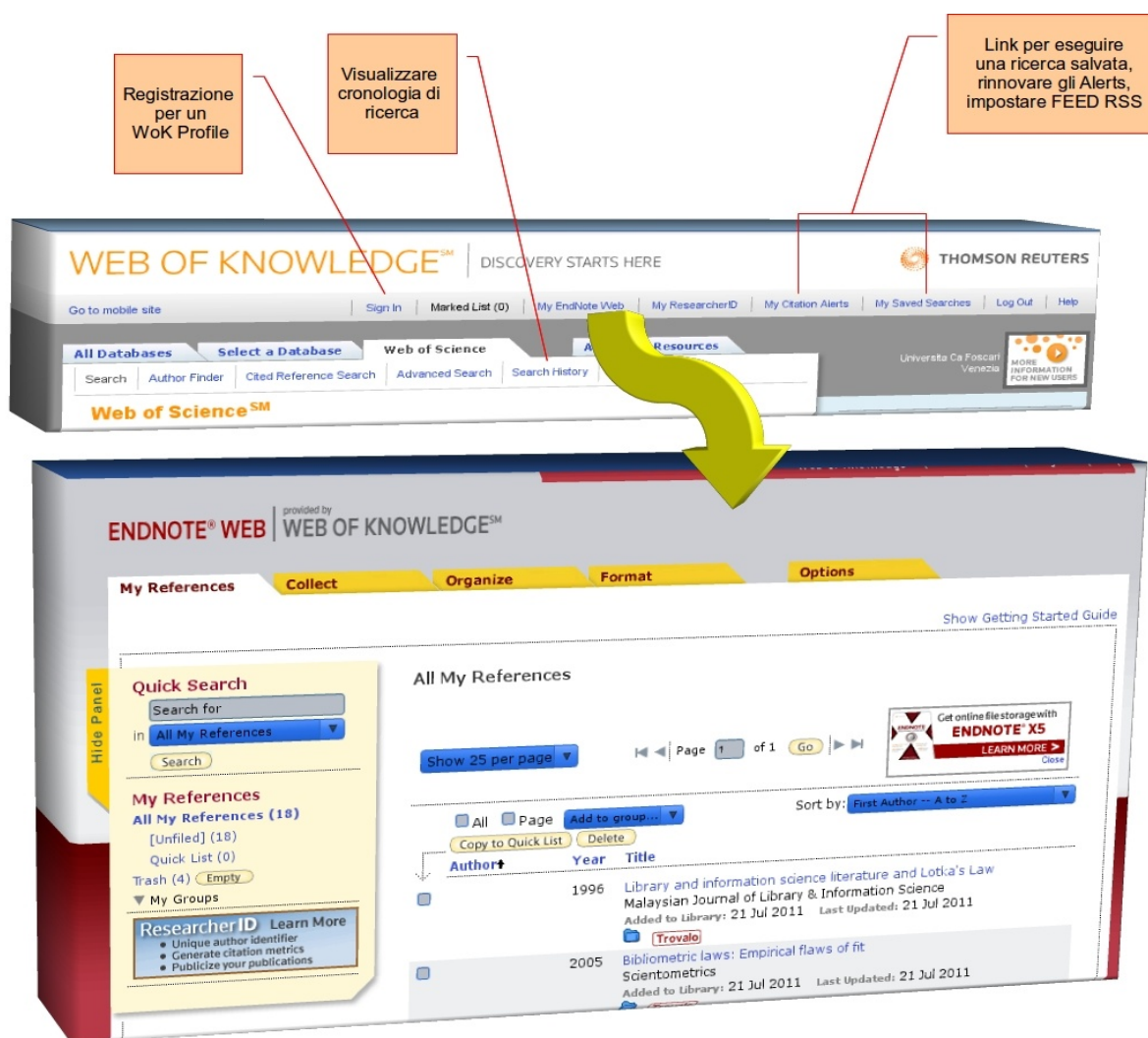


Figura 9: Schermata gestione dei risultati attraverso *EndNote Web* (rielaborazione da fonte *WoS*)

13 Raggiungile anche visitando il sito: <www.myendnoteweb.com>

Altro strumento integrato dentro la piattaforma WoK, ma distinto da WoS, è il *Journal Citation Reports (JCR)*,¹⁴ raggiungibile cliccando il pulsante “Additional Resources” visibile nella schermata iniziale di WoK. È disponibile in due edizioni:

- *JCR Science Edition*: contiene i dati di oltre 5900 riviste in 171 discipline;
- *JCR Social Science Edition*: contiene i dati di oltre 1700 riviste in 55 discipline.

Nella *homepage* è possibile selezionare l'anno di edizione di *JCR* e una delle seguenti opzioni di ricerca: categoria di argomento, editore, paese, ricerca di una specifica rivista, visualizzazione di tutte le riviste.

I risultati corrispondenti ai criteri di ricerca impostati vengono visualizzati nella pagina “Journal Summary List”.

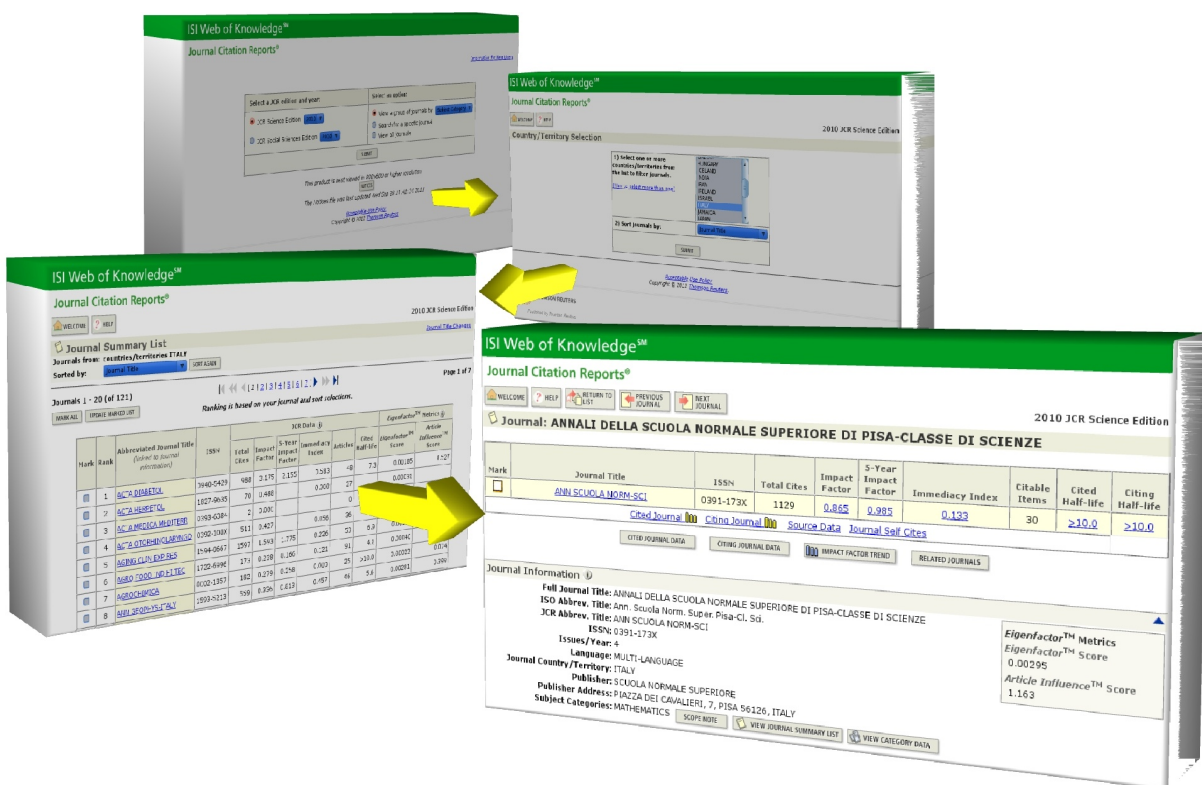


Figura 10: Schermate del *Journal Citation Reports* (rielaborazione da fonte WoK)

Facendo clic sul titolo abbreviato di una rivista si accede alla pagina del record completo in cui si trovano informazioni dettagliate su ogni rivista come il titolo completo e il recapito dell'editore. Disposti in una tabella sotto il titolo sono riportati i valori dei seguenti indicatori bibliometrici relativi alla rivista visualizzata comprensivi di link che rimandano a

14 <<http://ip-science.thomsonreuters.com/m/pdfs/mgr/jcr-1005-q-it.pdf>>

maggiori dettagli sulla pagina: totale citazioni; *Impact Factor*; indice di immediatezza (*Immediacy Index*); numero di articoli; *Cited Half-Life*; *Citing Half-Life*.¹⁵

4.3.2. Scopus

Lanciata nel 2004, *Scopus*¹⁶ è la banca dati citazionale a pagamento di ambito scientifico, tecnologico, biomedico e delle scienze sociali accessibile attraverso la piattaforma *SciVerse* dell'editore *Elsevier*. *Scopus* indicizza circa 18500 riviste *peer-reviewed* di oltre 5000 editori, 1800 riviste *open access*, 250 atti di convegni (per un totale di 4,8 milioni di documenti), 425 pubblicazioni commerciali, 350 serie di libri, *Medline* (copertura completa), 375 milioni di pagine web di qualità attraverso *Scirus*¹⁷, 24,8 milioni di registrazioni di brevetti. Complessivamente contiene 46 milioni di registrazioni, di cui 25 milioni con riferimenti bibliografici fino al 1996 e 21 milioni anteriori al 1996 che risalgono fino al 1823.

Il database si apre sulla schermata di ricerca generale ("Document Search"), da cui è possibile passare anche a una modalità diversa di ricerca cliccando sui pulsanti in alto ("Author Search", "Affiliation Search", "Advanced Search"). Sono dunque complessivamente possibili quattro funzionalità di consultazione e analisi.¹⁸

1. "Document Search": ricerca generale per campi selezionabili dal menu a tendina collocato in corrispondenza dell'area di interrogazione; per aggiungere campi supplementari e modificare le relazioni tra di essi mediante gli operatori booleani cliccare sul link "Add search field"; la ricerca può essere limitata cronologicamente per anno di pubblicazione o per data di inserimento nel database, per tipologia documentaria e per area disciplinare; i risultati della ricerca sono disposti in tabella e possono essere ordinati in base all'anno, al numero di citazioni ricevute alla pertinenza del documento all'autore e al titolo della rivista/fonte. I risultati forniti da *Scopus* sono tratti da cinque tipi di fonti:

- a) "Scopus": risultati tratti da periodici *peer-reviewed*;

15 Parte degli indicatori bibliometrici qui elencati saranno trattati nel dettaglio nel capitolo 5.

16 <<http://www.info.sciverse.com/scopus/scopus-in-detail/facts>>

17 *Scirus* <<http://www.info.scirus.com/>> è un motore di ricerca specializzato in fonti scientifiche, ricerca sia nel Web in generale sia in banche dati specializzate (*ScienceDirect*, *Medline*, *BioMedNet*, *ChemWeb* etc.).

18 Per maggiori informazioni collegarsi al servizio di *Elsevier* dedicato a *Scopus* <<http://www.info.sciverse.com/scopus>> o consultare *Elsevier, SciVerse Scopus, Guida Utente, 2010* disponibile online all'indirizzo: <http://www.unica.it/UserFiles/File/Utenti/scocumel/SCOPUS_Guida_Utenti_2.pdf>

- b) "More": risultati tratti da fonti selezionate e da fonti non presenti in Scopus;
- c) "Web": risultati provenienti dal Web trovati attraverso Scirus;
- d) "Patents": risultati provenienti da uffici brevetti indicizzati attraverso Scirus;
- e) "Selected Sources": risultati personalizzati tratti da open archives istituzionali di Scirus e raccolte di argomenti specialistici.

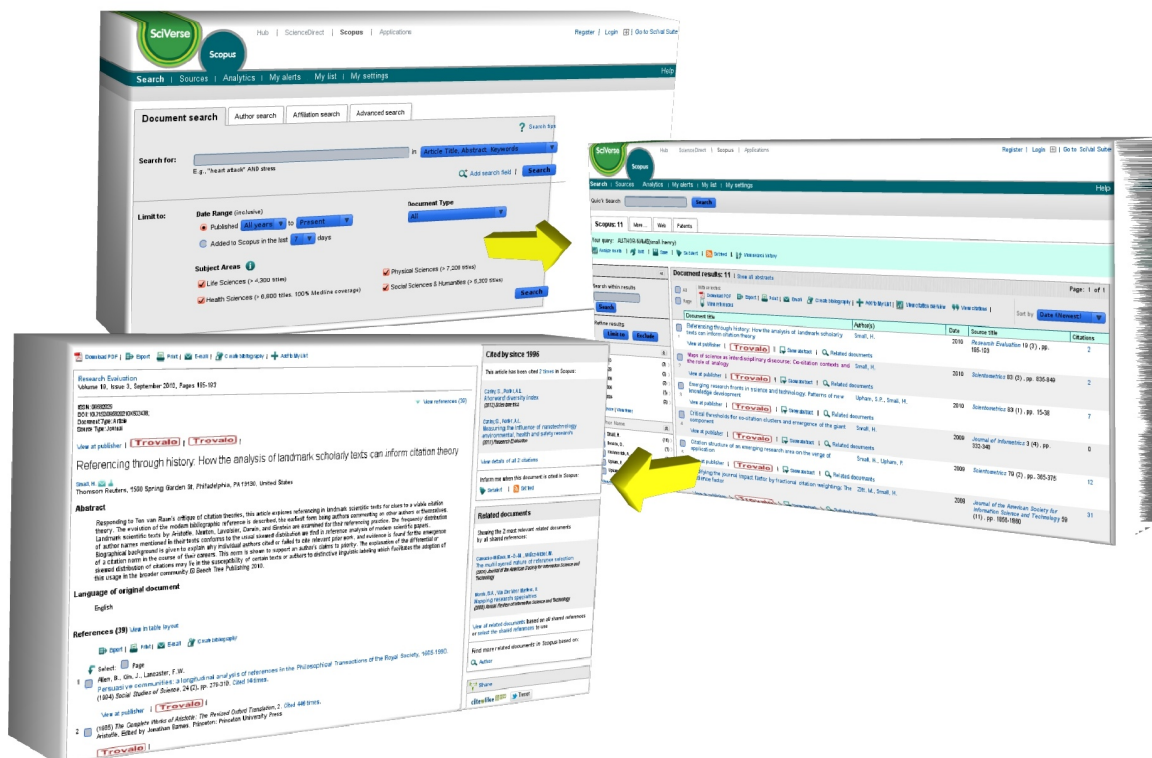


Figura 11: Schermate di ricerca generale (rielaborazione da fonte Scopus)

Per ogni risultato di ricerca è possibile visualizzare: il *full-text* del documento se si possiedono i privilegi d'accesso; l'*abstract* (clic su "Show Abstract"); i documenti correlati (documenti cioè che condividono uno o più riferimenti bibliografici con quello visualizzato); il documento alla pagina web originale dell'editore (clic su "View at Publisher"). Dalla stessa schermata è possibile ridefinire la ricerca utilizzando le opzioni della casella "Refine Results", che consente di perfezionare i risultati per le seguenti categorie: titolo della fonte/rivista; nome dell'autore; anno; tipo di documento; e area tematica. Il nome dell'autore è un link che con-

sente di visualizzare informazioni dettagliate relative all'autore stesso ("Author details"). Molte di queste informazioni sono anche riportate nel record completo dei risultati, il quale visualizza inoltre: la bibliografia dell'articolo che dispone di collegamenti ipertestuali all'*abstract* e al testo completo, ove disponibile, e del numero di citazioni relative a ogni riferimento bibliografico; le due citazioni più recenti di altri autori che citano quello stesso articolo e gli articoli correlati a quello visualizzato. In alto sulla pagina è possibile trovare informazioni bibliografiche relative all'articolo come il titolo della fonte (cliccabile per maggiori dettagli), volume, numero e anno di pubblicazione.

2. "Advanced Search": ricerca avanzata su tutti i campi del database; non essendo una ricerca guidata (tipo "Document Search") è necessario immettere nel riquadro di inserimento i codici di campo e gli operatori booleani o cliccare sugli stessi nell'elenco che riporta tutti quelli a disposizione. I risultati si possono ulteriormente perfezionare per rivista/fonte, autore, anno e area tematica.

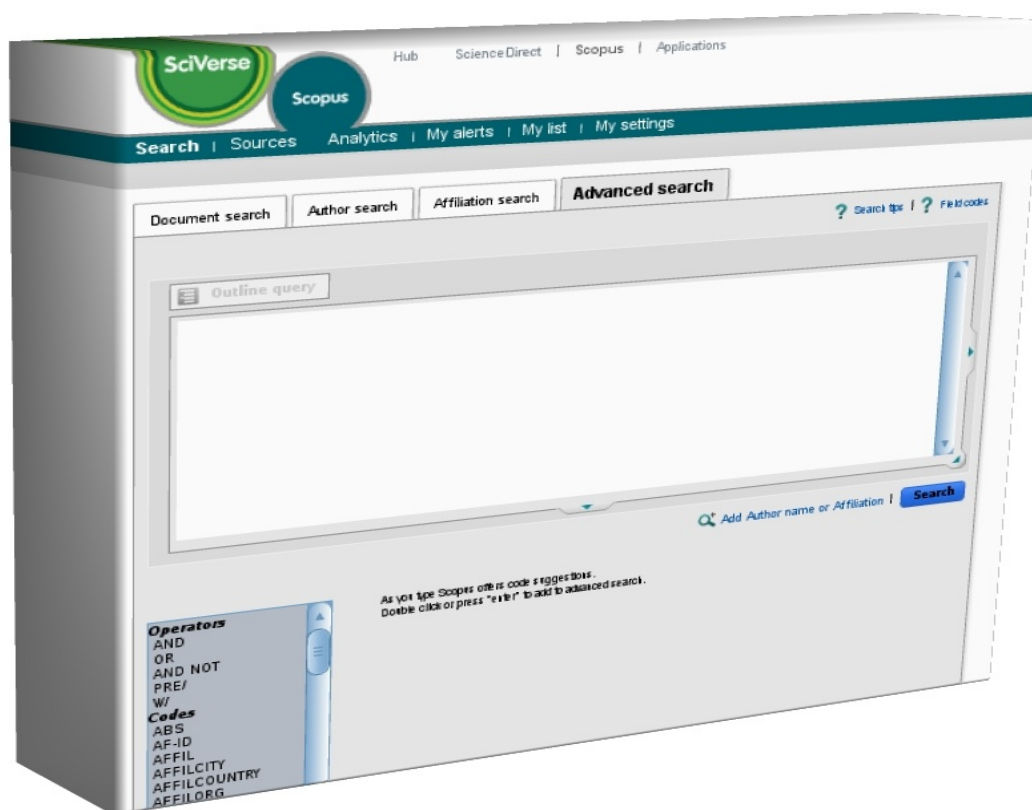


Figura 12: Schermata di ricerca avanzata (rielaborazione da fonte Scopus)

3. “Author Search”: ricerca che consente di recuperare agevolmente le citazioni collegate ai contributi di un determinato autore.

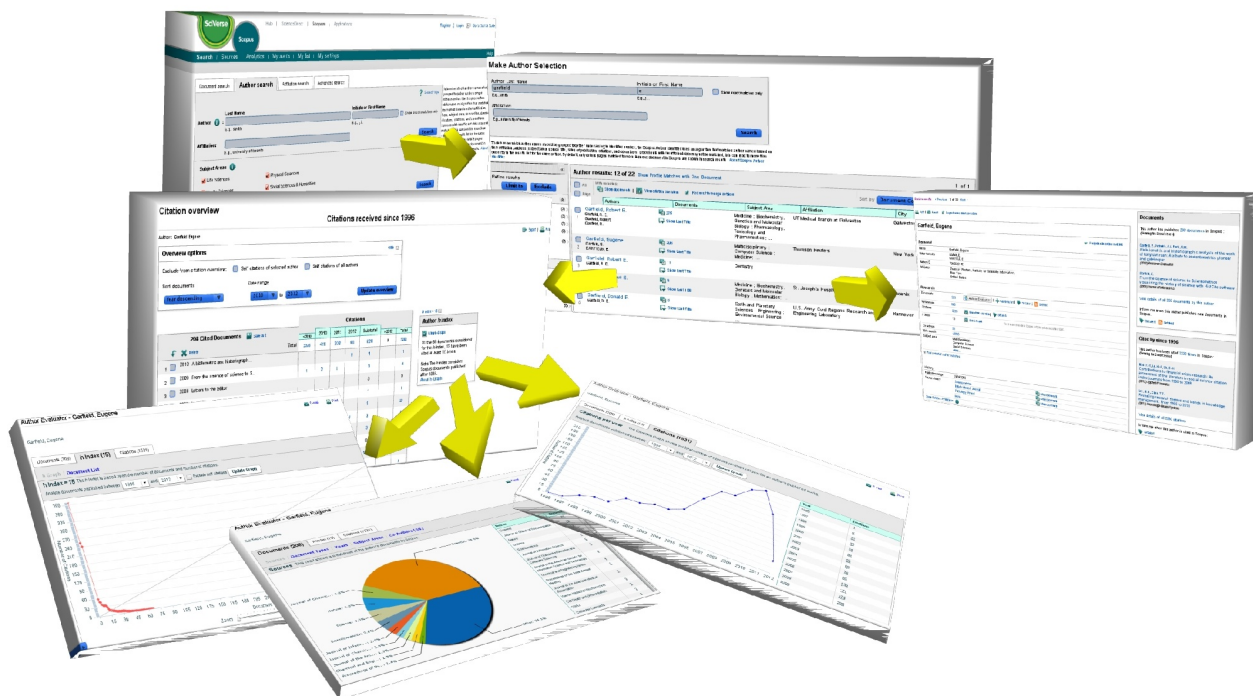


Figura 13: Schermate di ricerca per autore (rielaborazione da fonte Scopus)

La ricerca si fa con cognome e nome (completo o anche solo iniziali) e può essere completata con l'indicazione dell'ente a cui l'autore afferisce. Tutti i risultati includono il numero di documenti pubblicati da un autore e il collegamento a “Show Last Title” (mostra ultimo titolo pubblicato dall'autore). È possibile scegliere di visualizzare i risultati in ordine alfabetico, per numero di documenti, per città e per paese di origine. La pagina “Author Details” (dettagli dell'autore), a cui si accede cliccando sul nome dell'autore nella lista dei risultati, contiene informazioni contestuali relative a un autore utili per verificare che si tratti della persona cercata superando così il problema delle omonimie. Per ogni autore è anche possibile creare una tabella delle citazioni degli articoli e del numero di volte che sono stati citati dal 1996. È possibile ordinare la tabella delle citazioni in ordine crescente o decrescente in base all'anno e al numero di citazioni, modificare il limiti cronologici ed escludere le autocitazioni selezionando “Exclude from Cita-

tion Overview" ("escludi dalla tabella delle citazioni"). Portando il mouse su un determinato risultato viene visualizzato il riferimento completo, mentre cliccando sul titolo dell'articolo si accede al record completo. Con un clic sul numero delle citazioni è possibile visualizzare un'altra tabella di tutte le citazioni relative al record nel periodo di tempo selezionato. Sempre nella stessa schermata è visualizzato l'H-index, indicatore bibliometrico utile per valutare la produzione di un autore, in questo caso dal 1996 ad oggi. *Scopus* può fornire una rappresentazione grafica non solo di quest'ultimo indicatore bibliometrico (cliccando su "view h-graph"), ma anche degli articoli pubblicati (cliccando su "Documents") e delle citazioni (cliccando su "Citations") utili rispettivamente per avere una visione generale della cronologia delle pubblicazioni dell'autore e per analizzare il numero delle citazioni ricevute nel corso del tempo.

4. "Affiliation Search": simile alla ricerca per autore, con la ricerca per affiliazione è possibile individuare e recuperare tutte le citazioni collegate a un ente inserendo la sua denominazione.

Scopus come *WoS* offre alcune opzioni per la personalizzazione della ricerca. Registrandosi è possibile impostare messaggi di allerta per ricevere i nuovi risultati della ricerca via *e-mail* o tramite l'utilizzo del proprio *feed RSS*, e salvare la cronologia di una ricerca. La registrazione serve anche per configurare un proprio profilo personale attraverso cui, tra le molte opzioni, si possono gestire: le ricerche salvate; i messaggi di allerta delle ricerche e delle citazioni di documenti; gli elenchi salvati; gli autori raggruppati; il profilo personale; le impostazioni di *Scopus* e di *RefWorks* (dove disponibile).

Scopus offre inoltre la possibilità di avviare una ricerca partendo dalle fonti. Cliccando infatti "Sources" sulla barra di navigazione è possibile sfogliare l'elenco dei periodici per titolo, area tematica o tipo di fonte. La pagina visualizza non solo i titoli attivi e inattivi, ma anche la loro cronologia sottolineando alcuni dettagli come gli eventuali cambiamenti del titolo stesso e i numeri più recenti indicizzati dal sistema.

Sempre utilizzando la barra di navigazione ma cliccando su "Analytics" si accede al "Journal Analyser" ("analizzatore dei periodici"), uno strumento che permette l'analisi anche di tipo comparativo delle fonti presenti in *Scopus* su una serie di parametri bibliometri-

ci: SJR (SCImago Journal & Country Rank), SNIP (Source Normalised Impact per Paper), numero di citazioni, numero di documenti e percentuale di documenti non citati.



Figura 14: Schermate del "Journal Analyser" (rielaborazione da fonte Scopus)

4.3.3. Google Scholar

Google Scholar,¹⁹ la cui prima realizzazione in beta è del novembre 2004, è uno speciale motore di ricerca dedicato alla letteratura accademica e scientifica che, pur rimanendo distinto dall'archivio principale, è accessibile gratuitamente a partire proprio dalla *homepage* di Google. Esso si configura come uno strumento innovativo dalle grandi potenzialità sotto il profilo dell'integrazione bibliografica perché affianca al tradizionale metodo di ricerca per metadati, ai più recenti metodi di ricerca e *linking* citazionale e alle molte funzionalità di tipo bibliometrico, la possibilità di individuare e recuperare il *full-text* digitale di un'ampia gamma di tipologie documentarie ad accesso gratuito, ma anche ad accesso tariffato, i

19 <<http://scholar.google.it/>>

riferimenti bibliografici di articoli e libri cartacei e digitali non disponibili gratuitamente online ma localizzati attraverso opac come *WorldCat* o *ACNP* nel possesso di molte biblioteche e quindi facilmente recuperabili anche attraverso servizi di *document delivery*.

Come si legge nella guida, *Google Scholar* permette di “effettuare ricerche tra molte discipline e fonti: documenti approvati per la pubblicazione, tesi, libri, *abstract* e articoli di case editrici accademiche, ordini professionali, database di studi non ancora pubblicati, università e altre organizzazioni accademiche ... ti consente di identificare gli studi più rilevanti nel campo della ricerca accademica mondiale”.²⁰



Figura 15: Schermate di ricerca e visualizzazione dei risultati (rielaborazione da fonte *Google Scholar*)

Google Scholar fornisce quindi sì informazioni circa la tipologia di materiale che entra a far parte del suo archivio, ma non fa alcun accenno alle modalità di selezione dei materiali,

²⁰ Guida di *Google Scholar* disponibile online: <<http://scholar.google.it/intl/it/scholar/about.html>>

ai criteri di inclusione, ai tempi di aggiornamento, alle modalità di indicizzazione che erano e restano coperte da segreto e quindi, di conseguenza, non è neanche possibile controllare e valutare affidabilità e copertura della ricerca che si sta effettuando (Jacsò 2005 p. 209; Tarantino 2006; Ridi – Metitieri 2008; Baccini 2010 p. 75).

Rispetto a *WoS* e *Scopus*, *Google Scholar* presenta un sistema di ricerca molto sbrigativo che mal si accompagna al bisogno di accuratezza e approfondimento che si riscontrano nella maggior parte delle indagini accademiche.

L'*homepage* si presenta infatti pressoché identica a quella di *Google* che, allo stesso modo, lascia all'utente soltanto la possibilità di una interrogazione per parola nel *full-text*, perfezionabile con l'impiego di operatori avanzati come quelli booleani. Anche le opzioni avanzate, raggiungibili alla pagina dedicata, sono molto scarse in quanto limitate alla semplice ricerca per autore, per pubblicazione e per data. La pagina dei risultati inoltre prevede unicamente la possibilità di raffinare i documenti trovati per data e per lingua.

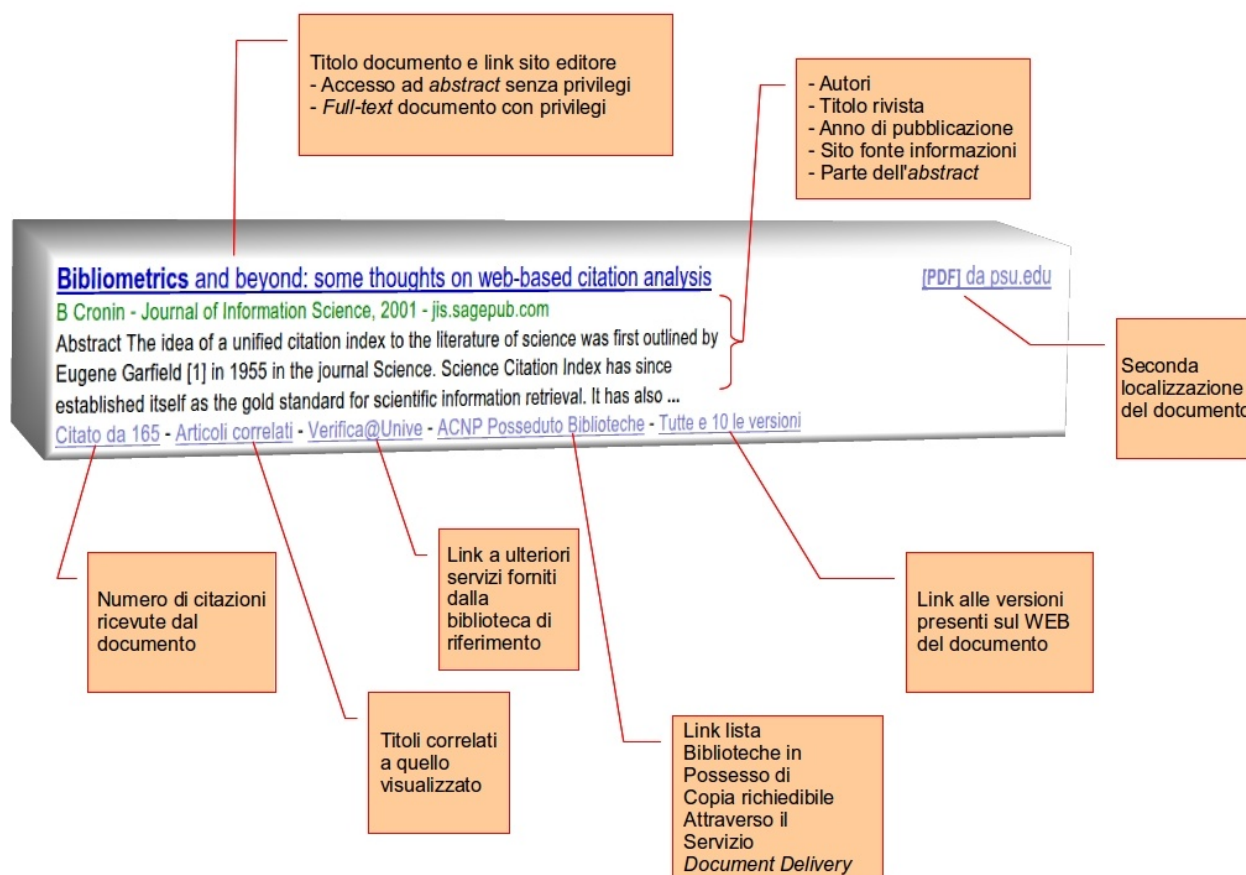


Figura 16: Dettagli di un singolo risultato di ricerca (rielaborazione da fonte *Google Scholar*)
 Per aumentare la visibilità e l'accessibilità dei suoi contenuti, *Google Scholar* collabora sia

con editori accademici sia con biblioteche. Per queste ultime in particolare i vantaggi, in termini di accessibilità e fruibilità, sono tangibili. In sostanza il motore di ricerca permette alle biblioteche che mettono a disposizione le proprie risorse tramite un *link resolver*²¹ di includere un collegamento per i propri iscritti a tali risorse come parte dei risultati della ricerca di *Google Scholar*. Gli utenti delle biblioteche che partecipano al programma visualizzano ulteriori link nella pagina dei risultati di ricerca di *Google Scholar* che puntano al *server* della biblioteca che, a sua volta, indirizza gli utenti al testo completo del documento se disponibile.

Google Scholar, a differenza di *WoS* e *Scopus*, pur avendo a disposizione tutti i dati necessari per la costruzione di indicatori bibliometrici basati sull'analisi delle citazioni, non presenta questo tipo di funzionalità, che invece è possibile soltanto grazie a un *tool* esterno. Si tratta del software realizzato da Anne-Wil Harzing, docente di management internazionale all'Università di Melbourne, distribuito gratuitamente in rete, chiamato *Publish or Perish (PoP)*,²² che calcola, utilizzando i dati di *Google Scholar*, indici citazionali per individui, riviste e istituzioni, e ha quindi l'obiettivo di entrare in concorrenza, per questa via, con i prodotti proprietari visti in precedenza (Harzing 2008; Baccini 2010 p. 75). Il *tool PoP* visualizza come schermata principale la pagina di ricerca per autore, la quale permette di effettuare una rapida indagine sull'impatto delle pubblicazioni di un autore.²³

Nel pannello di interrogazione i parametri di ricerca disponibili sono: il nome dell'autore che si intende ricercare ed eventualmente quelli da escludere dall'indagine; e i limiti cronologici nella produzione di un autore, esprimibili immettendo anno di inizio e fine, che si vuole prendere in considerazione per la ricerca. Il sistema non permette tuttavia una disambiguazione efficace delle omonimie, possibile invece solo a posteriori sui risultati di ricerca. È inoltre possibile restringere la ricerca per ambito disciplinare spuntando le caselle corrispondenti. Quest'ultima opzione deve però essere usata con cautela poiché la classificazione dei soggetti in *Google* non è molto accurata.

21 Per *link resolver* si intende un puntatore ad una risorsa web.

22 In realtà, *PoP* è soltanto il più diffuso di questa tipologia di strumenti, sono infatti numerosi i software disponibili in rete che utilizzano i dati di *Google Scholar* per la costruzione di indicatori bibliometrici. La prima versione di *PoP* è stata lanciata il 17 ottobre 2006. Il software è scaricabile liberamente all'indirizzo: <<http://www.harzing.com/pop.htm>>. Per maggiori informazioni sul prodotto utilissima è la sua guida <http://libris.nie.edu.sg/sites/default/files/UserGuide_Publish_or_Perish.pdf>, mentre per altri approfondimenti cfr. Noruzi (2005), Harzing – Van der Wal (2007), Meho – Yang (2007) e Harzing (2008 e 2010).

23 Cfr. traduzione italiana dell'*help* interno al programma: <<http://www.cab.unipd.it/system/files/publish-or-perish.doc>>

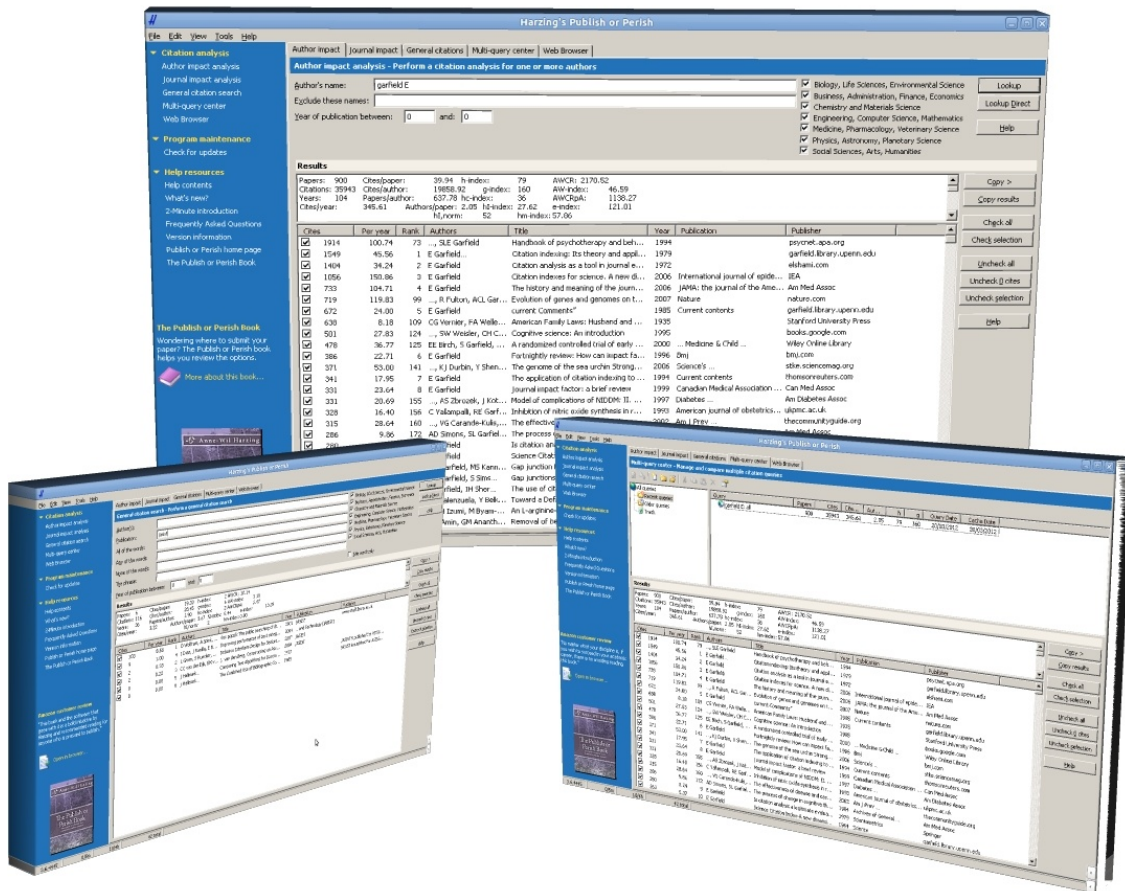


Figura 17: Schermate di ricerca e visualizzazione dei risultati (rielaborazione da fonte *Publish or Perish*)

Con un clic su "Lookup" in alto a destra, il programma andrà a recuperare le citazioni in *Google Scholar* e ricaverà i valori degli indicatori bibliometrici²⁴ che saranno visibili nel pannello "Results". Le citazioni complete invece saranno elencate nel riquadro sottostante i valori bibliometrici, da cui l'utente può ulteriormente raffinare la ricerca selezionando o deselegnando le caselle poste in corrispondenza di ogni risultato. I risultati possono poi essere ordinati in base al numero di citazioni, all'autore, al titolo, all'anno, al numero di pubblicazioni e all'editore, semplicemente cliccando sull'intestazione della colonna corri-

²⁴ Gli indicatori bibliometrici calcolati da *PoP* sono: numero totale di pubblicazioni; numero totale di citazioni per autore o per rivista; numero di anni dalla pubblicazione del primo contributo dell'autore o del primo numero della rivista; media numero di citazioni per documento, per anno, per autore; media numero di documenti per autore; media numero di autori per documento; *H-index*; *G-index*; *E-index*, *Contemporary H-index*; *individual H-index*; *Age-weighted Citation Rate*; e *Normalised Individual H-index*.

spondente. Cliccando poi sul singolo risultato visualizzato nell'elenco si aprirà una pagina di *Google Scholar* che mostra tutti i documenti che citano quello di partenza, cioè quello selezionato nella lista dei risultati.

I risultati possono inoltre essere copiati sulla *clipboard* di *Windows* (e utilizzati in altre applicazioni), salvati in un file di testo o in un foglio elettronico, esportati in vari formati (*EndNote format*, *BiBtex*, *CSV* e *RefMan/Ris*) e integrati nel programma per la gestione di bibliografie *EndNote*.

Oltre ad una ricerca per autore, *PoP* consente di effettuare anche ricerche per rivista e per singola citazione. Ogni ricerca effettuata sarà automaticamente aggiunta alla cartella "Recent queries" nella pagina "Multi-queries center", così da poter essere comodamente ripresa e confrontata in un secondo momento.

4.4. Altri indici citazionali specializzati in specifici settori disciplinari

Oltre agli indici di citazioni multidisciplinari visti nei paragrafi precedenti, ne esistono altri specializzati in determinati ambiti disciplinari. Di seguito viene proposta un'elencazione sommaria dei principali.

Chemical Abstract Service (CAS),²⁵ la banca dati online a pagamento consultabile attraverso l'interfaccia di ricerca *SciFinder Scholar*, contiene 30 milioni di record bibliografici relativi a periodici scientifici e tecnici, brevetti, rassegne, rapporti tecnici, monografie, atti di congressi, tesi di laurea nei settori della biochimica, chimica organica e inorganica, chimica fisica, chimica delle macromolecole, chimica analitica e industriale. La copertura cronologica è a partire dal 1907 e in alcuni casi anche precedente.

*Citebase Search*²⁶, per la fisica, la matematica, l'informatica ed alcune aree della biologia, è una base di dati citazionale ad accesso gratuito che utilizza un sistema di indicizzazione automatica di archivi *open access* tra cui *arXiv.org*, *PubMed Central*, *Biomedcentral* e *Cogprints*. *Citebase* permette a l'utente di ricercare le citazioni da e verso i documenti indicizzati, di vi-

25 <<http://www.cas.org>>

26 <<http://www.citebase.org/search>>

sualizzare alcuni indicatori bibliometrici basati sulle citazioni e sul *download* dei documenti, e di scaricare il *full-text* dei documenti trovati.

CiteSeerX,²⁷ disponibile gratuitamente online, è un sistema automatico di gestione di una biblioteca digitale che indicizza le pubblicazioni scientifiche di ambito accademico nei settori disciplinari dell'informatica e delle scienza dell'informazione disseminati nel Web. Il sistema permette di recuperare il *full-text* e le citazioni da e verso i documenti indicizzati, di visualizzare alcuni indicatori bibliometrici e di correlare documenti sulla base delle citazioni condivise. Complessivamente *CiteSeerX* indicizza 1,5 milioni di documenti e 30 milioni di citazioni.

*INSPIRE/SPIRES-HEP*²⁸ è un database citazionale online ad accesso gratuito, creato e mantenuto dallo *Stanford Linear Accelerator Center (SLAC)* con altri centri di ricerca, che indicizza oltre 500000 articoli nel campo della fisica delle alte energie (*High Energy Physics Literature*) tra cui articoli di riviste, *pre-print*, rapporti tecnici, atti di convegni e tesi di laurea, con copertura dal 1974 a oggi. Permette di recuperare le citazioni da e verso gli articoli indicizzati e di scaricare il *full-text* da *arXiv.org*.

MathSciNet,²⁹ banca dati citazionale online gratuita e a pagamento, creata e mantenuta dall'*American Mathematical Society (AMS)*, contiene le registrazioni bibliografiche e le recensioni apparse su *Mathematical Reviews* e *Current Mathematical Publications*. Vi è recensita la letteratura matematica mondiale a partire dal 1940 e contiene più di 2,5 milioni di record relativi a riviste internazionali, atti di convegni, monografie, tesi di dottorato, rapporti tecnici in settori disciplinari che comprendono la matematica pura e applicata, l'informatica teorica, la fisica matematica, la statistica, la ricerca operativa, la storia e la didattica della matematica. Alla ricerca tradizionale per autore e rivista affianca quella per citazione, consentendo il calcolo di alcuni indicatori bibliometrici, l'esportazione delle citazioni in vari formati e il link agli articoli *full-text* presenti nelle banche dati *ScienceDirect*, *JSTOR* e *Springer*.

*RePEc (Research Paper in Economics)*³⁰ è un database centralizzato che contiene informazioni su *working papers*, *e-prints*, articoli pubblicati su riviste, libri di economisti di 75 paesi

27 <<http://citeseerx.ist.psu.edu/index>>

28 <<http://inspirehep.net>>

29 <<http://www.ams.org/mathscinet/search.html>>

30 <<http://repec.org>>

appartenenti a istituzioni che partecipano al progetto per un totale di più di un milione di record. Quasi l'80% del materiale è accessibile gratuitamente in *full-text*. Il database è completato da un servizio che fornisce informazioni dettagliate sulle statistiche di accesso e citazionali riferite ai singoli prodotti di ricerca, agli autori e alle istituzioni di ricerca che partecipano a *RePEc*.

*SAO/NASA Astrophysics Data System (ADS)*³¹ è uno strumento bibliografico ad accesso gratuito composto da tre database contenenti complessivamente più di nove milioni di record nei settori dell'astronomia e dell'astrofisica. Permette la visualizzazione degli *abstract*, il *download* del *full-text* di articoli scientifici, link alle principali riviste del settore, ad atti di convegni, e a bollettini e monografie.

Scitation,³² online e ad accesso gratuito, è la piattaforma di *AIP (American Institute of Physics)* per la ricerca di contenuti tecnico-scientifici da riviste, atti di convegni, libri elettronici, database documentali, brevetti e standard. Include più di 2 milioni di record in settori che comprendono la fisica, la chimica, le scienze geologiche, l'acustica e diversi settori dell'ingegneria. *AIP* include importanti istituzioni tecnico-scientifiche sia come membri che come affiliati. Le ricerche su *Scitation* restituiscono liste unificate di risultati provenienti dai database delle varie collezioni bibliografiche delle società consorziate con *AIP*. Ogni record è anche un link alle interfacce native da cui l'utente può scaricare il *full-text* dei documenti e le citazioni ricevute dagli stessi in vari formati.

4.5. Considerazioni sulla copertura degli indici di citazioni

Se si considera che gran parte dell'affidabilità nell'utilizzo e della correttezza nella costruzione degli indicatori bibliometrici e citazionali si devono alle caratteristiche dei dati estrapolati dagli strumenti bibliografici analizzati nei paragrafi precedenti, diventa stringente una serie di considerazioni sul grado di copertura degli indici citazionali stessi, inteso come rapporto tra la quantità di pubblicazioni incluse in una base di dati e la quantità totale di quelle prodotte in un determinato settore disciplinare. È chiaro che diventa fondamentale sotto questo profilo la comprensione e l'analisi dei criteri di selezione o esclusione

31 <<http://adswww.harvard.edu>>

32 <http://scitation.aip.org/for_users/what_is_scitation>

dei materiali bibliografici utilizzati per la costruzione degli indici che direttamente influenzano sul loro grado di copertura. I parametri, sulla cui base avviare un'operazione di filtraggio del materiale, si ricavano considerando: la qualità della pubblicazione; la sua natura; la tipologia di citazione.

In relazione alla qualità delle pubblicazioni, il criterio adottato può essere quello di includere solo le pubblicazioni che rispettino qualche standard di qualità scientifica riconosciuto e accettato dalla comunità. L'esempio più noto e diffuso è quello che si basa sulla revisione dei pari: in questo caso, vengono incluse solo le pubblicazioni che abbiano superato un rigoroso processo di questo tipo. Gli strumenti citazionali però non sono sempre costruiti basandosi sul criterio della qualità. È questo il caso di *Google Scholar* che, in linea con quanto auspicato dai sostenitori più radicali del movimento *open access*, non prevede alcun tipo di filtraggio preventivo basato sulla qualità delle pubblicazioni inserite. Una soluzione di compromesso tra queste due punte estreme è il controllo di qualità sugli *e-prints* attuati per gli *open archives*, che viene effettuato da specifici comitati di redazione che si riservano di accettare o meno le proposte di *e-print* da parte degli autori.³³

Come già detto nei paragrafi precedenti, le procedure di filtraggio delle pubblicazioni per i database *WoS* e *Scopus* sono abbastanza vaghe. Per *WoS*, la selezione delle pubblicazioni è il risultato di una decisione presa da un gruppo di *editors* di *Thomson-Reuters* basata «su criteri di regolarità della pubblicazione, di standard editoriali adottati, tra cui la revisione dei pari, di grado di internazionalizzazione degli autori che pubblicano sulla rivista, e soprattutto dei dati citazionali» (Baccini 2010 p. 82). Anche *Scopus* si è dotato di un comitato consultivo composto da specialisti nei vari ambiti disciplinari che ha il compito di valutare le fonti da includere nel database tra quelle che presentano pubblicazioni regolari e qualche metodo di controllo della qualità che possa rendere conto anche della loro autorevolezza, popolarità e accessibilità.³⁴

33 È questo il caso di *arXiv* che però non seleziona ogni singolo contributo, bensì prevede una procedura ad un livello superiore, cioè l'accettazione in via definitiva dell'autore del contributo, a questo proposito cfr. <<http://arxiv.org/help/endorsement>>

34 Baccini (2010 p. 82), confrontando i dati relativi al numero di riviste indicizzate dai due database e calcolando il numero di riviste presenti in entrambi negli anni 2006-2007, ha scoperto che «poco meno della metà delle riviste sono indicizzate in entrambi i database; di conseguenza, l'utilizzazione di questi archivi per il calcolo di uno stesso indicatore può dare luogo a risultati molto diversi, a causa del diverso grado di copertura della letteratura. Uno studio di caso condotto confrontando le citazioni presenti nei due archivi per un gruppo di 15 studiosi affiliati all'Università dell'Indiana ha mostrato che delle circa 2.000 citazioni registrate solo il 58% è presente in entrambi [Yang e Meho 2007]».

Un criterio di selezione deve tener conto poi della natura della pubblicazioni incluse. Chiunque intenda utilizzare uno strumento citazionale di questo tipo deve considerare attentamente le tipologie di pubblicazioni incluse e, soprattutto, quanto esse riflettano il modo di comunicazione tipico o prevalente della comunità scientifica d'interesse. In generale gli archivi citazionali indicizzano prevalentemente articoli pubblicati su riviste *peer-reviewed* di lingua inglese, tipologia documentaria quella della rivista in cui frequentemente si concretizza la ricerca delle scienze della natura e dell'ingegneria, e le citazioni riguardano maggiormente articoli recenti apparsi in un nucleo molto ristretto di riviste. È ovvio quindi che il grado di copertura degli indici citazionali è molto elevato per queste discipline, mentre è decisamente inferiore per le scienze sociali e umane, per settori cioè che impiegano come forma di comunicazione tipologie documentarie diverse dall'articolo su rivista in lingua inglese.³⁵ Ma anche per la matematica all'articolo su rivista si affianca l'uso di *pre-prints* e *e-prints*, e nelle scienze applicate e nell'ingegneria gli atti di conferenze e *technical reference works*. Nelle scienze sociali e umane la situazione invece è più variegata. La monografia specialistica è la forma di comunicazione prevalente, la lingua in cui gli studiosi preferiscono esprimersi è la loro lingua madre. Per questi ambiti disciplinari, oltre a tenere conto della tipologia documentaria maggiormente impiegata, è anche necessario conoscere la rilevanza della letteratura nazionale in uno specifico settore e se questa letteratura è contenuta negli indici di citazioni che si intendono utilizzare. Un'altra forma di letteratura che caratterizza le scienze sociali e umane è quella di taglio generalista, cioè tutti quei contributi pubblicati su riviste, giornali o libri destinati a un pubblico non accademico. Rientrano in questa letteratura tutti i risultati degli studi condotti per committenti pubblici o privati condotti da accademici o esperti in un ambito disciplinare.

Il grado di copertura poi non riguarda soltanto le tipologie documentarie selezionate per l'inclusione ma naturalmente anche le citazioni bibliografiche che sono incluse al loro interno.

Se la preponderante tipologia documentaria per un archivio citazionale, come nel caso di *WoS*, è l'articolo di rivista in lingua inglese, diventa automaticamente scontato il fatto

³⁵ Moed (2005 p. 119-136), utilizzando come fonte dei dati il database ISI anno 2002, ha condotto uno studio sull'importanza delle riviste nei diversi ambiti disciplinari riportando le seguenti percentuali di riferimenti ad articoli su rivista: 96% nella biologia molecolare e biochimica, 95% nella biologia umana, 93% nella medicina, 90% nella chimica, 89% fisica e astronomia, 70% matematica, 59% economia, 41% scienze sociali e 34% scienze umane.

che mancando gran parte delle altre fonti citazionali come i libri, i capitoli di libri, gli atti di convegni e le pubblicazioni di taglio generalista, si abbassa anche il grado di copertura delle citazioni soprattutto a discapito di quelle discipline caratterizzate proprio da quelle forme di pubblicazione che generalmente rientrano in modo ridotto tra le fonti citazionali dell'indice. A questo si deve poi aggiungere il fatto che le scienze sociali e umane presentano una percentuale molto più bassa (circa il 50%), rispetto alle scienze naturali e all'ingegneria (circa 90%), di pubblicazioni considerate come fonti che citano prevalentemente articoli su riviste (Glänzel e Schoepflin 1999).

Dalle considerazioni fatte sin qui si ricava che, essendo il grado di copertura degli indici di citazioni del tutto inadeguato per le scienze sociali e umane, anche i dati estratti risultano completamente falsati rispetto a quello che è il reale quadro complessivo, dunque necessitano di essere maneggiati e interpretati con cura prima di essere utilizzati nella costruzione di indicatori bibliometrici.

5. Gli indicatori bibliometrici

Lo studio e la messa a punto di indicatori è uno dei filoni più fiorenti e promettenti della bibliometria, tanto che negli ultimi anni il loro numero è cresciuto vertiginosamente¹ e ha messo in competizione tra loro istituzioni, imprese – a cominciare dalla *Thomson-Reuters* – e molti centri di ricerca universitari, sempre al costante inseguimento dello strumento più efficace, più approfondito, più adatto a rappresentare la complessità dei risultati dell'attività di ricerca.

Una definizione semplice ma efficace di “indicatore bibliometrico” è stata proposta da Baccini che scrive (Baccini 2010 p.68): «un indicatore bibliometrico è un indicatore di qualità o impatto costruito, con opportune tecniche statistiche, a partire da informazioni elementari ricavate da riferimenti bibliografici contenuti in pubblicazioni scientifiche o in archivi creati appositamente». Un posto d'onore tra gli indicatori bibliometrici è riservato a quelli citazionali proprio per la significatività dell'unità di base di cui sono composti, ampiamente trattata nei capitoli precedenti. È bene tuttavia ricordare che esistono sia indicatori bibliometrici non citazionali, è questo il caso per esempio degli indicatori incentrati sul conteggio del numero di pubblicazioni oppure delle metriche del web (di cui si parlerà nel capitolo 6), sia indicatori non bibliometrici, cioè non basati su informazioni di tipo bibliometrico, ma, per esempio, su dati brevettuali o su dati relativi ai finanziamenti ottenuti dai progetti di ricerca o sul numero di partecipazioni di un determinato ricercatore a conferenze ecc.

Questo capitolo sarà dedicato alla trattazione dei più noti e più diffusi indicatori bibliometrici di tipo citazionale² suddivisi in primari (paragrafo 5.1) e secondari o derivati (paragrafo 5.2), di cui verranno discussi i punti di forza e le criticità.

Prima però di procedere a tale discussione sono necessarie alcune considerazioni che riguardano indistintamente qualsiasi tipo di indicatore bibliometrico e che permettono di in-

1 Sono stati contati ben 39 indicatori bibliometrici cfr. Cassella – Bozzarelli (2011 p.68) e Bollen – Chute – Hagberg – Van de Sompel (2009).

2 Per comodità di trattazione, viste le analogie con l'indicatore basato sul conteggio del numero di citazioni, si è deciso di inserire in questo capitolo anche un indicatore bibliometrico di tipo non citazionale, quello basato sul conteggio del numero di pubblicazioni prodotte da un ricercatore, da un gruppo di ricerca, da un'istituzione.

quadrare tali strumenti nell'ottica di un contesto più ampio di valutazione al fine di evitare inutili semplificazioni, o peggio, di cedere alla tentazione di assolutizzare i risultati.

Rispetto alla tradizionale revisione dei pari, l'analisi valutativa attraverso l'uso di indicatori bibliometrici è più veloce, più economica e per certi aspetti più obiettiva, ma procede inevitabilmente, e non può essere altrimenti, ad una semplificazione di una realtà che invece è molto più complessa. Se si pensa bene infatti gli indicatori non fanno altro che misurare i dati contenuti in uno o più archivi bibliografici che, come è noto, sono costruiti sulla base di una selezione di pubblicazioni e non certo sulla totalità della letteratura esistente. Al massimo si potrà parlare di maggiore o minore esaustività del database citazionale, e questo non solo per l'impossibilità materiale di indicizzare l'intero docuverso, «ovvero la totalità, di tutti i documenti disponibili alla fruizione, prodotti in qualsiasi codice o linguaggio e su qualsiasi tipo di supporto» (Ridi 1998), ma anche semplicemente perché non è detto che tutto ciò che viene prodotto venga poi anche citato.

Ne consegue quindi che l'*output* bibliometrico non può in nessun caso venire considerato come un dato assoluto ma deve sempre essere relativizzato alla fonte dei dati su cui viene calcolato e, ai fini dell'analisi valutativa, inserito in un contesto dove trovano spazio altre metriche bibliometriche e metodologie di valutazione.

5.1. Indicatori primari

Gli indicatori bibliometrici primari sono misure quantitative elementari sia perché consistono nel semplice conteggio delle pubblicazioni prodotte da un autore o gruppo di ricerca, o delle citazioni ricevute in un determinato intervallo di tempo da un autore o da una rivista (ma anche degli accessi a una pagina web o dei *download* di un documento web), sia perché rappresentano il punto di partenza di molte analisi bibliometriche, le loro combinazioni infatti danno origine a quelli che vengono definiti indicatori bibliometrici secondari o derivati.

Con il numero di pubblicazioni e numero di citazioni è possibile misurare la produttività di un autore (di una rivista, di una istituzione ecc.), ma, vista la loro natura, tali misure sono estremamente approssimative e sicuramente poco adatte a rappresentare la complessità nei vari contesti d'applicazione sia a livello di autore sia di rivista o di istituzione. Per

queste ragioni esse vengono scarsamente utilizzate ma la loro comprensione è tuttavia fondamentale non tanto nell'ottica del loro utilizzo come misure quantitative autonome, bensì perché costituiscono le basi su cui vengono costruiti proprio quegli indicatori bibliometrici da cui sono state ormai quasi completamente sostituite.

5.1.1. Il conteggio del numero di pubblicazioni

Il numero totale di pubblicazioni è un indicatore di produttività scientifica calcolabile sia a livello di singoli autori o di riviste sia a livello di istituzioni o di interi sistemi nazionali di ricerca.³

Vista la sua natura, fornisce informazioni di carattere puramente quantitativo, non esprime cioè, diversamente dall'indicatore costituito dal numero totale di citazioni ricevute, alcuna informazione circa l'impatto che una pubblicazione ha sulla comunità scientifica di riferimento.

La sostanziale semplicità dell'indicatore, il quale si presenta come la somma delle pubblicazioni prodotte in un determinato intervallo temporale, contrasta con tutta una serie di criticità che sottostanno alla sua stessa costruzione.

Il primo problema che si incontra è relativo alle caratteristiche del database bibliografico su cui viene effettuato il conteggio. Se infatti sempre più spesso sono le stesse istituzioni di ricerca a costruire dei database bibliografici in cui i ricercatori possano inserire i propri contributi, rendendo in tal modo molto facile il calcolo di questo specifico indicatore bibliometrico, è anche vero che la difficoltà aumenta nel caso di istituzioni molto complesse o di gruppi di ricerca appartenenti a paesi diversi. In quest'ultimo caso si ricorre generalmente ad archivi costruiti da terzi, vale a dire agli archivi bibliografici citazionali visti nel capitolo precedente, ben sapendo che alla fine si tratta non dell'elenco completo delle pubblicazioni di un ricercatore, ma solo di una selezione di esse, modellata in base ai criteri di selezione utilizzati per la costruzione del database stesso.

Il secondo problema riguarda le tipologie di pubblicazioni da includere nel conteggio, che è strettamente collegato al tema delle forme di comunicazione adottate nei vari settori disciplinari. Stabilire quali tipologie di pubblicazione sono più rilevanti per il calcolo di

³ Glänzel (2003) ha proposto la distinzione tra livello micro (individui), meso (istituzioni, discipline) e macro (regioni, nazioni).

questo indicatore è più facile per le discipline scientifiche, per le quali sappiamo che la ricerca si concretizza in una forma prevalente di comunicazione (l'articolo su rivista), mentre è molto più complicato per quelle discipline come le scienze sociali e umane dove ciò non accade.

Il terzo aspetto problematico riguarda le modalità di conteggio delle pubblicazioni con più di un autore. Come ben spiegato da Baccini (2010 p. 102-103) sono tre le possibili soluzioni: attribuire l'intera pubblicazione a ciascuno degli autori; attribuire solo una frazione del contributo a ciascun co-autore; o attribuire l'intera pubblicazione solo al primo autore.

Anche la definizione della base temporale su cui calcolare l'indicatore può creare qualche difficoltà soprattutto se si considera che molto spesso la produttività di un ricercatore, anche in relazione a un determinato intervallo di tempo, è direttamente proporzionale alla sua età anagrafica.

Altro aspetto problematico da tenere in considerazione è il diverso peso, in termini di impegno per stesura e qualità, delle tipologie di pubblicazione: sarà molto più elevato l'impegno profuso per scrivere una monografia rispetto a quello per scrivere un articolo su rivista o una breve presentazione per un convegno. In questo caso sarà utile procedere in modo tale da presentare, per ogni singolo ricercatore, più indicatori riferiti ognuno a un prodotto di ricerca diverso, o attribuendo pesi diversi alle varie tipologie di pubblicazione: alla monografia per esempio può essere dato un valore più alto rispetto all'articolo su rivista, costruendo di fatto in tal modo un indicatore di produttività ponderato sulla base della qualità intrinseca della tipologia documentaria.

Quest'ultimo punto introduce poi a tutti quei problemi connessi all'aggregazione dei risultati individuali. La semplice somma delle misure di produzione individuale non può essere considerata una buona tecnica di aggregazione in quanto non si può essere certi che la singola pubblicazione venga conteggiata come singola unità di aggregazione, esiste sempre il rischio di duplicazione. Se prendiamo a titolo di esempio il caso di un articolo con molti co-autori afferenti a dipartimenti diversi, il conteggio dell'intera pubblicazione per ciascun autore è improponibile poiché il numero delle pubblicazioni conteggiate per ciascun dipartimento sarebbe maggiore del totale delle pubblicazioni effettivamente prodotte. Per le misure aggregate quindi si preferisce costruire un indicatore che tenga conto del nu-

mero delle pubblicazioni pro capite prodotte in un intervallo temporale di riferimento. Negli altri casi, vale a dire come misura semplice di produzione, si presenta o un valore assoluto, come numero complessivo di pubblicazioni, o in forma relativa, come media delle pubblicazioni prodotte in un arco temporale prestabilito.

5.1.2. Il conteggio del numero di citazioni

Il secondo fondamentale indicatore bibliometrico primario è il *Citation Index*, costruito contando semplicemente il numero di citazioni ricevute da una pubblicazione, da un autore, gruppo di ricerca o istituzione in un intervallo temporale di riferimento.

Il *Citation Index*, pur essendo una misura di carattere quantitativo, fornisce importanti informazioni che sono valide non tanto allo scopo di stabilire la qualità intrinseca dell'oggetto dell'analisi bibliometrica ma per definire l'impatto sulla comunità scientifica di riferimento. Diversamente dall'indicatore basato sul conteggio del numero di pubblicazioni, il *Citation Index* non è di facile reperibilità a livelli elevati di aggregazione e nel suo calcolo non solo permangono, ma in un certo senso vengono amplificati, tutti i punti di criticità insiti nel comportamento citazionale che di fatto possono distorcere non solo questo, ma ogni indicatore citazionale (vedi paragrafo 3.1.2).

Il conteggio del numero di citazioni può essere calcolato a qualsiasi livello di aggregazione. Tuttavia è da tener presente che a livello micro, quello cioè che corrisponde al conteggio del numero di citazioni della singola pubblicazione o del singolo autore, pur essendo più facile la reperibilità del dato, «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» (De Bellis 2005 p. 129).

È infatti impensabile, sulla base di questo indicatore, procedere a un'analisi comparativa tra un autore giovane e un autore anziano, come anche tra due dipartimenti o gruppi di ricerca di dimensioni diverse, perché è quasi dato per scontato il fatto che il numero di citazioni dipende molto anche dalla quantità di pubblicazioni prodotte. In questi casi, una possibile soluzione, è l'impiego dell'indicatore noto come *Citations Per Publications (CPP)* che non è altro che il numero medio di citazioni per pubblicazione, ottenuto dal rapporto

tra il numero totale di citazioni ricevute da un *set* di pubblicazioni, e il numero totale di pubblicazioni. In quest'ultimo caso tuttavia entrano in gioco altre problematiche che sono sostanzialmente legate alla costruzione dei database citazionali, come la sostanziale distribuzione asimmetrica delle citazioni, sbilanciate, per quantità, verso gli anni più recenti, che chiaramente potrebbero distorcere la valutazione comparativa basata sui valori di questo indicatore specie se calcolati per archi temporali distanti fra loro, con la conseguenza che, il dato relativo agli anni più recenti può risultare arbitrariamente più elevato rispetto a quello relativo agli anni meno recenti. Per ovviare a questi inconvenienti il numero medio di citazioni per pubblicazione viene solitamente affiancato ad altri due indicatori citazionali (Van Raan 2000; Van Leeuwen et al. 2003; Baccini 2010 p. 159-160): l'impatto citazionale atteso delle riviste (*JCS, Journal Citation Score*), ossia il tasso medio di citazioni ricevute dagli articoli contenuti nelle riviste su cui il ricercatore o i membri di un gruppo di ricerca hanno pubblicato nell'intervallo temporale di riferimento; e l'impatto citazionale atteso nel campo di ricerca (*FCS, Field Citation Score*), ossia il tasso medio di citazioni ricevute dalle pubblicazioni appartenenti al settore disciplinare del ricercatore o gruppo di ricerca oggetto di valutazione.

L'accostamento del numero di citazioni in particolare con i valori medi del campo di ricerca rende il dato più chiaro, proprio perché contestualizzato, consentendo al valutatore di cogliere immediatamente l'impatto relativo delle pubblicazioni. Quest'ultimo punto appare particolarmente importante se si considera che più alto è il numero di pubblicazioni, più alto sarà anche il numero di riferimenti bibliografici contenuti in esse, dunque maggiore sarà la probabilità di essere citati. Ne consegue che, variando il tasso di produttività e il numero medio di citazioni per pubblicazione, varia anche la probabilità di essere citati. Sapendo poi che la produttività, o meglio, il tasso di crescita della letteratura, unitamente al numero medio di citazioni per pubblicazione, varia da settore disciplinare a settore disciplinare, diventa ancora più stringente contestualizzare il numero di citazioni rispetto all'ambito disciplinare di riferimento.

Le basi di dati citazionali esaminate nel capitolo precedente riportano tutte il numero totale di citazioni ricevute da una singola pubblicazione nell'arco temporale che va dalla data di pubblicazione a quella dell'interrogazione, altre guardano più lontano, e, oltre a in-

dicare per ogni singola pubblicazione indicizzata, il numero totale di citazioni ricevute nello stesso database, presentano anche il numero totale di quelle ricevute dal web. Se però si analizzano i valori riferiti alla stessa pubblicazione, ma recuperati nei diversi database, si scopre che il dato non è lo stesso, anzi, a volte differisce anche di molto. Le differenze che si riscontrano sono dovute al fatto che l'indicatore, non essendo un prodotto proprietario (come ad esempio l'*Impact Factor*), può essere recuperato in ogni banca dati citazionale e quindi il suo valore è strettamente dipendente dai criteri di selezione e implementazione dei documenti indicizzati, che si traducono nella effettiva copertura disciplinare e temporale dei singoli database. Il recupero del numero di citazioni dovrebbe quindi tener conto delle specificità di ciascuna banca dati e essere effettuato almeno sui tre principali database citazionali (*Web of Science*, *Scopus* e *Google Scholar*). Nel caso in cui i risultati siano molto diversi tra loro, è consigliabile riportare tutti i valori recuperati con l'indicazione dei database su cui sono stati effettuati i conteggi (Piazzini 2010 p. 76).

5.2. Indicatori secondari o derivati

5.2.1. L'*Impact Factor* (IF)

L'*Impact Factor* è il più noto ma anche il più discusso degli indicatori bibliometrici elaborati dalla *Thomson-Reuters*. Come già ricordato, venne ideato da Garfield in fase di realizzazione del *Science Citation Index* come strumento utile per l'individuazione del nucleo di periodici da inserire nello stesso indice citazionale.⁴

La sua diffusione al di fuori dell'*ISI* a supporto delle politiche documentarie delle biblioteche è però stato un processo molto lento soprattutto se si pensa che il *Journal Citation Reports* (*JCR*), la pubblicazione in cui è contenuto l'*Impact Factor*, venne distribuito per la prima volta nel 1975. Il cambio di passo avviene nei primi anni '90 ed è riconducibile alla pubblicazione dello strumento prima su microfiche (1990) e poi su CD-Rom (1995), fino ad approdare sul Web e a venire integrato nella piattaforma *Web of Knowledge*.

4 Garfield utilizza per la prima volta il termine *Impact Factor* in un articolo del 1955 (Garfield 1955), nel quale fa specifico riferimento alla tecnica citazionale impiegata dai coniugi Gross nel loro lavoro del 1927: «such an impact factor may be much more indicative than an absolute count of the number of a scientist's publications. The impact factor is similar to the quantitative measure obtained by Gross & Gross in evaluating the relative importance of scientist journals, a method later criticized by Brodman but used again by Fussler».

Il JCR, distribuito in due edizioni, la *Science Edition* e la *Social Science Edition*, partendo dai dati delle riviste indicizzate nello *Science Citation Index* e nel *Social Sciences Index* (ma non nell'*Arts & Humanities Citation Index*) contenuti in *Web of Science*, stila una classifica annuale della performance citazionale delle riviste ivi contenute sulla base non solo dell'*Impact Factor* ma anche di diversi altri parametri (*Total Cites*, *Immediacy Index*, *5-Year Impact Factor*, *Citable Items*, *Cited Half-Life*).

L'algoritmo per il calcolo dell'*Impact Factor* è molto semplice. Esso si ottiene dividendo il numero delle citazioni ricevute in un determinato anno dagli articoli pubblicati dalla rivista nei due anni precedenti, per il numero totale degli articoli pubblicati nello stesso periodo di tempo e sulla stessa rivista.

Per esempio, se nel 2011 gli articoli pubblicati nel 2009 e 2010 su una rivista (X) hanno ricevuto rispettivamente 77 e 110 citazioni, e se il numero complessivo degli articoli usciti su quella stessa rivista nel 2009 e 2010 è 52, l'*IF* calcolato per il 2011 per la rivista X è pari a 3,59:

$$IF(X) = \frac{77 + 110}{52} = 3,59$$

Questo significa che i 52 articoli pubblicati nel 2009 e nel 2010 nella rivista X sono stati citati nel 2011 in media oltre 3 volte.

Come si evince dalla sua costruzione l'*IF* è sostanzialmente un indicatore della performance citazionale delle riviste scientifiche presenti nel database *ISI*, molto utile quindi in campo biblioteconomico come supporto alle decisioni di politica documentaria (Altmann – Gorman 1996), ma utilizzato in modo del tutto improprio come strumento di valutazione dei singoli ricercatori e dei singoli dipartimenti e gruppi di ricerca. E questa è soltanto una delle tante critiche che, dal lancio dell'indicatore, si è levata contro l'*Impact Factor*. Riassumerle tutte sarebbe impossibile e, vista la ridondanza spesso riscontrata, forse anche inutile. Schematicamente possono tuttavia essere indicati i punti di maggior dissenso che emergono più frequentemente nella letteratura del settore (Figà-Talamanca 2000; Tammaro 2001;

Di Cesare 2002 p. 145; De Bellis 2005 p. 118; Morriello 2007; Cassella – Bozzarelli 2011 p. 71).

1. la copertura dei database *Web of Science* non è adeguata e omogenea per tutti gli ambiti disciplinari (vedi paragrafo 4.3.1. e 4.5), in particolare per le scienze umane e sociali. A ciò si deve poi aggiungere il fatto che, pur contenendo *Web of Science* un indice specificatamente dedicato alle riviste di arte e discipline umanistiche (*Arts & Humanities Citation Index*), non esiste un'edizione specifica di *JCR* dedicata a tali discipline. *JCR* riporta l'*IF* delle riviste indicizzate in *Science Citation Index* e in *Social Science Citation Index* ma non di quelle contenute in *Arts & Humanities Citation Index*, lasciando di fatto priva del dato bibliometrico un'importante fetta di letteratura pur indicizzata nel database;
2. non considera il comportamento citazionale diffuso nelle diverse aree disciplinari, impedendo di fatto un affidabile confronto tra l'*Impact Factor* non solo di riviste appartenenti ad ambiti differenti,⁵ ma anche di quelle considerate trasversali, che coprono cioè ambiti disciplinari diversi (Tammaro 2001 p. 105), e di quelle che, pur riconducibili allo stesso ambito disciplinare, sono incentrate su specifiche aree tematiche (per esempio la biologia teorica rispetto a quella sperimentale e applicata) (Morriello 2007 p. 25);
3. non c'è nessuna segnalazione circa eventuali scostamenti rispetto ai valori medi e, di conseguenza, mancando un termine di confronto, diventa difficile per il valutatore saper attribuire il giusto significato al semplice dato che, in tal modo, non risulta contestualizzato nella sequenza statistica;
4. la base temporale di due anni su cui è calcolato l'*IF* è penalizzante per alcune discipline la cui performance citazionale, se si considera il tempo medio che passa tra la pubblicazione di un articolo e il picco del numero di citazioni ricevute, si presenta più precoce o più tardiva; alcuni studi (Adler – Ewing – Taylor 2008; Figà-Talamanca 2000) hanno registrato nelle riviste di matematica un picco altissimo (90%) di citazioni anteriore all'intervallo di due anni utilizzato per il calcolo dell'indicatore;

5 «Le riviste che si occupano di scienze della vita, di neuroscienze e di medicina hanno in media un *IF* da tre a sei volte superiore a quello di riviste di matematica, informatica e scienze sociali e fattore di impatto più elevato di un'area può essere inferiore all'*IF* più basso di un'altra area» (Di Cesare 2002 p. 146).

5. l'impossibilità di controllare i dati utilizzati per il calcolo dell'indicatore soprattutto alla luce di evidenti e documentati casi di errore;⁶
6. rischio di distorsione dei risultati dovuto all'incertezza nella definizione delle grandezze, cioè delle tipologie documentarie citabili, al numeratore e al denominatore (Seglen 1997 p. 500; Amin 2000 p. 5). Nel numeratore, che consiste nella somma delle citazioni ricevute da una rivista, si prendono infatti in considerazione, senza esercitare alcun tipo di selezione, tutte le tipologie di articolo da cui proviene la citazione. Nel denominatore invece si tiene conto soltanto degli articoli scientifici, degli atti di convegni e delle rassegne. Non vengono quindi conteggiati al denominatore alcune tipologie come le note, gli editoriali e le lettere agli editori, che tuttavia in molti casi contribuiscono massicciamente all'aumento del numero delle citazioni del numeratore. Ne consegue che le riviste con molte note, editoriali e lettere, magari anche molto citati, vedranno aumentato il valore del numeratore, senza però un corrispondente aumento del denominatore, situazione questa che chiaramente comporterà un conseguente e ingiustificato aumento dell'*Impact Factor* (Van Leeuwen – Moed 2002);
7. non si tiene conto né della grandezza del periodico, ovvero del numero medio di articoli per rivista, né delle tipologie di scritti che vi appaiono. Sembra infatti quasi scontato che più alto è il numero di articoli pubblicati da una rivista, più probabilità avrà quest'ultima di venire citata. Lo stesso vale per le tipologie di scritti: le rassegne sono più citate degli articoli e quest'ultimi, se sono più lunghi, sono citati di più di quelli più brevi (Seglen 1997). Le riviste quindi con un numero medio di articoli più alto, che pubblicano rassegne o articoli mediamente più lunghi riportano un valore d'*IF* più elevato;
8. l'uso pilotato delle autocitazioni, cioè delle citazioni ad articoli della rivista provenienti da altri articoli della rivista stessa, da parte dei comitati editoriali delle riviste allo scopo di gonfiare i dati citazionali e ottenere così un valore più alto dell'*IF*, è un problema ancora presente ma che è stato in parte risolto grazie alla possibilità offer-

⁶ A tal proposito Baccini (2010 p. 171) riporta l'esempio della rivista *Educational Research*, inserita nel database *ISI*, che per 20 anni ha beneficiato delle citazioni ricevute da un'altra rivista, *Educational Researcher*, simile nel titolo ma non inclusa nel database!!

ta dalla *Thomson-Reuters* di calcolare il valore dell'indicatore escludendo le autocitazioni;⁷

9. la mancata pulizia di tutti quegli articoli considerati errati dalla comunità scientifica o dagli stessi autori, è un elemento di grande distorsione dell'indicatore poiché tali articoli continuano ad essere citati e, di conseguenza, contribuiscono anch'essi, pur essendo destituiti di ogni credenziale, a formare il valore dell'*IF*;

Nonostante i punti di criticità appena descritti, l'*IF* presenta anche molte caratteristiche positive che hanno contribuito a farlo diventare uno degli indicatori citazionali più utilizzati. Baccini (2010 p. 167) ha individuato tra i fattori di maggior successo:

- a) la semplicità del suo algoritmo rende chiari e quindi facilmente comprensibili i meccanismi del suo funzionamento;
- b) la stabilità della misura nel tempo. I valori dell'indicatore non registrano mai grandi variazioni da un anno all'altro e ciò permette una minore obsolescenza del dato, che quindi può essere utilizzato anche se vecchio di uno o due anni senza particolari rischi d'errore;
- c) il costante aggiornamento del dato;
- d) l'idea della facile controllabilità del dato, ovvero la sicurezza purtroppo solo apparente, che sia possibile controllare in ogni momento i dati su cui si basa il calcolo dell'indicatore;
- e) la possibilità di confrontare riviste con periodicità di pubblicazione e dimensioni diverse.

5.2.2. Altri indicatori contenuti nel *Journal Citation Reports*

Oltre all'*Impact Factor* nel *Journal Citation Reports* sono pubblicati altri indicatori, calcolati sempre a partire dagli archivi *ISI*, che in parte rispondono alla numerose critiche mosse al più noto degli indicatori bibliometrici.

Nel 2007 *Thomson-Reuters* ha risposto con il *5-Year Impact Factor* alle critiche sulla ristrettezza temporale su cui viene calcolato l'*Impact Factor* che, a detta di molti, sarebbe incapace

⁷ Un caso clamoroso riportato sia da Baccini (2010 p. 172) sia da Cassella – Bozzarelli (2011 p. 70) è quello della rivista internazionale *Leukemia*, la quale nel 1997 venne accusata dal direttore di una rivista concorrente, la *Leukemia Research*, di spingere, chiedendo esplicitamente, ai propri potenziali autori di citare soprattutto altri contributi usciti sulla stessa rivista, al fine di manipolare artificiosamente, aumentandolo, il valore dell'*IF*.

di cogliere le specificità disciplinari del comportamento citazionale.

L'*IF* a 5 anni presenta lo stesso algoritmo dell'*IF* di base e si differenzia da quest'ultimo soltanto per la cornice temporale di riferimento, che da due passa a cinque anni.

In direzione opposta si muove l'*Immediacy Index*, che calcola invece il numero medio di citazioni ricevute in un determinato anno dagli articoli pubblicati in quello stesso anno nella rivista oggetto di analisi. L'indicatore ha l'obiettivo di misurare la velocità, e quindi l'impatto sulla letteratura scientifica, con cui l'articolo viene mediamente citato.

Il *Cited Half-Life* (semivita delle citazioni), già ricordato nel paragrafo 2.1.2, misura sostanzialmente l'obsolescenza degli articoli citati o la stabilità nel tempo delle citazioni ricevute da una rivista. Esso calcola il numero di anni, partendo dall'anno in corso e andando a ritroso nel tempo, in cui si raggiunge la metà del numero totale di citazioni ricevute dalla rivista nell'anno corrente.

5.2.3. *H-Index* e indici derivati

H-Index o Indice di Hirsch viene ufficialmente presentato dal suo ideatore il fisico Jorge E. Hirsch, docente presso l'Università della California di San Diego, nel *preprint* di un articolo pubblicato il 3 agosto 2005 sul *repository ArXiv* (Hirsch 2005).

L'innovazione apportata dallo strumento alle analisi bibliometriche è chiara fin da subito agli occhi sia dei ricercatori⁸ sia dei bibliometri⁹ che lo accolgono immediatamente con grande favore, tanto che oggi, a distanza di pochi anni dal suo lancio, è diventato uno degli indicatori più usati «in particolare nell'ambito delle scienze fisiche e matematiche e delle scienze mediche e soprattutto in Nord America e nei paesi scandinavi» (Morriello 2007 p. 29).

H-Index è un indicatore capace di esprimere in un unico valore numerico sia la produttività (numero di pubblicazioni) sia l'impatto sulla comunità scientifica (numero di citazioni) riferiti al singolo ricercatore.¹⁰

8 Cfr. Batista – Campiteli – Kinouchi – Martinez (2005) e Popov (2005).

9 Cfr. Braun – Glänzel – Schubert (2006); Bornmann – Daniel (2005); Alonso – Cabrerizo – Herrera-Viedma – Herrera (2009). Interessante inoltre la ricca (822 contributi) e aggiornata bibliografia curata da F. J. Cabrerizo sull'*H-Index* e varianti disponibile al sito: <<http://sci2s.ugr.es/hindex/biblio.php>>

10 Hirsch afferma che questo indicatore non solo «dà una stima dell'importanza, della significatività e del generale impatto dei contributi di ricerca cumulati di uno scienziato», ma «rappresenta un metro di paragone non distorto per comparare individui diversi che competono per la stessa risorsa quando un criterio importante di valutazione sono le conquiste scientifiche» (Hirsch 2005 p. 5).

Hirsch definisce l'indicatore riferendolo direttamente all'unità oggetto di misurazione, vale a dire al ricercatore che «ha un indice h se h dei suoi articoli pubblicati in n anni (Np) hanno ricevuto ciascuno almeno h citazioni, e gli altri ($Np - h$) articoli hanno ricevuto non più di h citazioni ciascuno» (Hirsch 2005 p. 1).

Per calcolare l'indice di Hirsch per un determinato autore (ma è possibile anche per un gruppo di ricerca o una rivista) è necessario: definire l'intervallo temporale di riferimento e attribuire alle pubblicazioni un ordine decrescente basato sul numero di citazioni ricevute. A questo punto, l'*H-Index* di quell'autore sarà determinato dall'ultima pubblicazione in tale classifica che riuscirà a far corrispondere il numero di posizione occupato in tale elenco con una quantità almeno pari di citazioni ricevute.

Un autore quindi con un *H-Index* pari a 3 ha almeno 3 articoli pubblicati che hanno ricevuto almeno 3 citazioni ciascuno.

Per capire meglio come venga individuato *H-Index* di un autore è necessario fare un esempio più concreto che faccia risaltare le fasi operative di calcolo. Supponiamo di avere un autore che ha pubblicato 8 articoli, i quali hanno riportato complessivamente un totale di 45 citazioni.

Set dati di partenza

P	NC
A	4
B	7
C	6
D	7
E	9
F	1
G	11
H	0

P= pubblicazioni
NC= numero di citazioni



formiamo un elenco disponendo le pubblicazioni in ordine decrescente in base alle citazioni ricevute

Set dati ordinati

P	POS.	NC
G	1	11
E	2	9
B	3	7
D	4	7
C	5	6
A	6	4
F	7	1
H	8	0

H-Index = 5

POS.= posizione
P= pubblicazioni
NC= numero di citazioni

H-Index è pari a 5 perché ci sono 5 pubblicazioni che hanno ricevuto almeno 5 citazioni ciascuna

Tabella 2: Esempio di calcolo di H-Index

L'*H-Index* non è un prodotto proprietario ed è ormai inserito tra gli strumenti bibliometrici sia di molti database citazionali come *Web of Science* e *Scopus*, sia di motori specialistici come *Google Scholar*, in quest'ultimo caso l'indicatore è calcolato grazie all'impiego di specifici *tools* come il già ricordato *Publish or Perish* di Anne-Wil Harzing (vedi paragrafo 4.3.3).¹¹ Naturalmente è da tenere sempre in mente che l'indice *h* riferito a un determinato autore potrebbe assumere valori anche molto diversi se calcolato su database citazionali differenti. È auspicabile quindi riportare il valore bibliometrico indicando sempre la fonte su cui è avvenuto il calcolo.

Come altri indicatori, anche *H-Index* presenta contemporaneamente punti di forza e punti di criticità. Tra i vantaggi si annoverano (Baccini 2010 p. 176):

- a) la semplicità dell'algoritmo;
- b) la capacità di cogliere entrambi gli aspetti fondamentali dell'attività scientifica: produttività e impatto della ricerca;
- c) può essere applicato facilmente dal micro al macro livello d'indagine: dal singolo ricercatore alla nazione;
- d) gli eventuali piccoli errori presenti nel database citazionale di riferimento non influenzano, o comunque, non modificano in modo determinante, i risultati espressi dall'indicatore;
- e) i picchi citazionali di una o poche pubblicazioni non modificano in nessun caso il valore dell'indicatore;
- f) le pubblicazioni che hanno ricevuto poche o nessuna citazione non condizionano il calcolo dell'indicatore.

Uno degli svantaggi invece più evidenti sta proprio nel suo principale punto di forza. La semplicità di rilevazione del valore bibliometrico infatti, se da una parte facilita la comprensione e l'applicazione dell'indicatore, dall'altra non permette di evidenziare completamente tutte le sfumature informative presenti nel dato citazionale del singolo ricercatore, con la conseguente perdita di importanti informazioni su cui basare proprio la sua valuta-

¹¹ Altri strumenti di rilevazione dell'indice di Hirsch e dei suoi derivati sono: *Scholarometer* <<http://scholarometer.indiana.edu>>, creato dall'Università dell'Indiana; *HView Visualizer* <<http://hview.limsi.fr/>> del *Centre national de la recherche scientifique (CNRS)*; *QuadSeach* <<http://delab.csd.auth.gr/~lakritid/index.php?lan=1&s=2>>, sviluppato dall'Università Aristotele di Tessalonica a Salonicco; *scholar Index* <<http://www.cs.odu.edu/~mln/pubs/2007-10-09-h-index.html>>, messo a punto da l'*Institut National de Recherche en Informatique et en Automatique (INRIA)*, Istituto nazionale per la ricerca nell'informatica e nell'automazione).

zione. Se prendiamo per esempio due ricercatori con un valore dell'indice h pari a 15, quindi entrambi con 15 pubblicazioni che hanno ricevuto almeno 15 citazioni ciascuna, e supponiamo poi che uno dei due autori abbia anche altre 100 pubblicazioni con 14 citazioni ciascuna, o che abbia sì 15 pubblicazioni con almeno 15 citazioni ciascuna ma che effettivamente le citazioni siano 100 per ciascuna delle 15 pubblicazioni, capiamo che l'indice h non riesce a estrapolare alcuni elementi che permetterebbero invece una più efficace valutazione, specie se di tipo comparativo, tra due o più autori.

Altri punti di criticità sono: la sostanziale penalizzazione dei ricercatori più giovani o in formazione e di tutti quelli che pubblicano pochi articoli di grande impatto scientifico, dunque molto citati; l'impossibilità di calcolare l'indice al netto delle autocitazioni; non tiene conto degli eventuali cambiamenti, in termini di produttività e impatto, di un'unità di ricerca poiché non può decrescere; nel caso di pubblicazioni con più di un autore, attribuisce l'intera pubblicazione a ciascuno degli autori, trascurando quindi totalmente il contributo individuale al lavoro; influisce negativamente sul comportamento del singolo ricercatore perché se da una parte è incoraggiato a pubblicare lavori di qualità che facciano guadagnare molte citazioni per contributo, dall'altra però lo spinge, una volta raggiunto un valore dell'indice elevato, a «cullarsi sugli allori» visto che il suo indice può crescere anche senza pubblicare (Baccini 2010 p. 180), o peggio, può essere portato a frammentare la produzione scientifica in unità informative minime che avranno come conseguenza tante brevi pubblicazioni ingiustificatamente slegate le une alle altre con il rischio di far perdere organicità al disegno complessivo del contributo scientifico.

Tutti questi punti deboli insiti nell'indicatore hanno contribuito allo sviluppo di altri strumenti bibliometrici sempre derivati da esso ma capaci di correggere, almeno in parte, i problemi appena affrontati.¹² Si tratta di un filone degli studi bibliometrici in forte espansione, con risultati apprezzabili sul piano teorico, cioè come numero di proposte di nuovi indicatori, ma deludente sul piano applicativo in quanto tali proposte, con l'eccezione dell'indice g , sono guardate da più parti con molta prudenza (Bornmann – Mutz – Daniel 2008; Bornmann – Daniel 2009; Baccini 2010 p. 182-183; Cassella – Bozzarelli 2011 p. 73).

¹² Molte sono le proposte che ruotano attorno al perfezionamento dell'indice di Hirsch, tra cui cfr: Vaidya (2005); Sidiropoulos – Katsaros – Manolopoulos (2006, 2007); Anderson – Hankin – Killworth (2008); Baldock – Ma – Orton (2009); Franceschini – Maisano (2010).

Come già ricordato, uno dei punti deboli dell'indicatore è quello di favorire i ricercatori con più anni di carriera scientifica alla spalle in ragione del fatto che esso cresce proporzionalmente con il passare del tempo, penalizzando in tal modo i ricercatori più giovani. È lo stesso Hirsch (2005) a proporre un indice normalizzato che prenda in considerazione l'età scientifica degli studiosi.

Tale variante è nota con il nome di quoziente m e, nella sostanza, è il rapporto tra *H-Index* calcolato per un determinato ricercatore e la somma degli anni trascorsi dalla pubblicazione del primo contributo scientifico valido per il calcolo dell'indice stesso.¹³

Un'altra variante dell'indice h è quella che è stata proposta da Leo Egghe (2006a, 2006b, 2006c) nota come *G-Index*, che nasce allo scopo di correggere l'incapacità dell'indice di partenza di tener conto degli eventuali cambiamenti sia in positivo (pubblicazioni *h-core* con molte citazioni) sia in negativo (pubblicazioni con poche o nessuna citazione) rispetto al numero medio di citazioni ricevute dalle pubblicazioni di un ricercatore. In particolare, *H-Index* non è in grado di attribuire un peso maggiore a tutti quei contributi che invece meriterebbero di essere rivalutati sulla base dell'alto numero di citazioni ricevute, con il rischio reale che, appiattendolo il valore di tali contributi, vadano perse informazioni preziose ai fini dell'analisi valutativa. L'idea di Egghe è quella di far emergere tutte quelle pubblicazioni che, pur appartenendo al gruppo isolato dall'indice h , hanno un numero di citazioni molto più alto rispetto al valore calcolato con l'indice di Hirsch. Secondo la definizione data dallo stesso Egghe (2006a) «l'indice g è il numero d'ordine più grande (quando gli articoli sono classificati in ordine decrescente per numero di citazioni ricevute) tale che i primi g articoli abbiano ricevuto [cumulativamente] almeno g^2 citazioni». Per comprendere meglio la differenza tra *H-Index* e *G-Index* è necessario un esempio pratico. Supponiamo di formare una tabella in cui si dispongono le singole pubblicazioni in ordine decrescente sulla base delle citazioni ricevute.

13 Harzing (2008) ha sollevato dei dubbi sull'impiego del quoziente m : la data del primo articolo cui far riferimento per stabilire l'anzianità della carriera può risultare inappropriata se la scelta dovesse cadere su un articolo di scarsa importanza pubblicato prima della "maturità" scientifica dello studioso; non tiene conto dei ricercatori che lavorano *part-time* o di quelli che, come spesso succede per le donne, registrano periodi di discontinuità lavorativa.

	NC	P (x)	SNC	x ²	
	52	1	52	1	
	31	2	83	4	
	28	3	111	9	
	21	4	132	16	
	20	5	152	25	
	18	6	170	36	
	16	7	186	49	
	15	8	201	64	
	14	9	215	81	
	12	10	227	100	
<i>H-Index</i>	11	11	238	121	
	10	12	248	144	
	10	13	258	169	
	9	14	267	196	
	9	15	276	225	
	8	16	284	256	
	7	17	291	289	<i>G-Index</i>
	6	18	297	324	
	5	19	302	361	
	4	20	306	400	



NC= numero di citazioni ricevute per pubblicazione
P (x)= posizione
SNC= somma il numero di citazioni
X²= posizione al quadrato
 *H-Index*
 *G-Index*

Tabella 3: Esempio di calcolo di H-Index e G-Index

La prima colonna (NC) contiene il numero di citazioni ricevute da ciascuna pubblicazione; La seconda colonna (P(x)) indica il numero di posizione (x) della pubblicazione; La terza colonna (SNC) contiene la somma del numero delle citazioni; La quarta colonna (x²) il quadrato del numero di posizione (x).

H-Index, come già spiegato, si ricava confrontando prima e seconda colonna, e si identifica con il numero di posizione più elevato dell'articolo che ha ottenuto un numero di citazioni pari o superiore al numero di posizione, in quest'esempio *H-Index* è pari a 11.

G-Index si ricava invece confrontando le altre due colonne; esso è individuato dal numero di posizione più elevato tale per cui il numero che esprime la somma delle citazioni delle pubblicazioni è almeno pari al quadrato del numero di posizione, nell'esempio *g*= 17.

In altre parole, un *G-Index* pari a 17 sta ad indicare che il numero di citazioni cumulate è pari almeno a 289 (17*17).

Come si può capire dall'esempio, dilatando la base, cioè il numero di articoli, su cui poggia il suo calcolo, il *G-Index* riesce, diversamente da quanto accade usando *H-Index*, a far pesare maggiormente nel suo calcolo le pubblicazioni che riportano un numero elevato di citazioni.

Anche il *G-Index* può essere comodamente calcolato utilizzando il software *Publish or Perish (PoP)*. L'unico limite (ma può essere visto anche come punto di forza) attribuito a questo indicatore è l'impiego, rispetto all'indice *h*, di un maggior numero di dati, che quindi lo espongono maggiormente agli errori e alle distorsioni che spesso sono riconducibili alle banche dati da cui provengono i dati necessari al suo calcolo (vedi paragrafi 3.1.2 e 4.5).

H-B-Index (Hirsch-Banks Index) è definito dallo stesso ideatore, Michael Banks del *Max Planck Institute for Solid State Research*, un'estensione dell'indice *h* (Banks 2006). Banks non ha fatto altro infatti che riprendere *H-Index*, così come sviluppato da Hirsch, e applicarlo, invece che al singolo ricercatore, ai composti usati in fisica dello stato solido o, più in generale, a un determinato argomento scientifico.

L'*A-Index*, inventato dalla studiosa cinese Jin Bihui, utilizza le stesse informazioni necessarie al calcolo di *H-Index* e consiste nel numero medio di citazioni ricevute dagli articoli inclusi nell'*h-core* (Bihui 2006 p. 8).

Esistono anche varianti dell'*A-Index*: l'indice *r* e l'indice *ar*.

Il primo, proposto da Bihui e alcuni suoi collaboratori, nasce con lo scopo di misurare la portata delle citazioni ricevute dalle pubblicazioni di un'unità di ricerca che hanno ottenuto più citazioni. Esso consiste nel quadrato del numero complessivo delle citazioni ricevute dalla pubblicazioni che formano *h-core* (Bihui – Liming – Rousseau – Egghe 2007).

Il secondo, proposto dalla stessa Bihui in una breve nota (Bihui 2007 p. 6), ha l'obiettivo di correggere l'incapacità dell'indice *h* di tener conto delle eventuali variazioni della performance citazionale di un ricercatore o unità di ricerca. Come già visto infatti *H-Index*, oltre a non prendere in considerazione il numero effettivo di citazioni, non può mai decrescere. L'*AR-Index* è definito come la radice quadrata della somma del numero medio di citazioni per anno delle pubblicazioni incluse nell'*h-core*.

Altri indicatori derivati dall'indice *h* sono: *HC-Index*, *HL-Index* e *HL norm*.

HC-Index (*contemporary H-Index*) sviluppato da Sidiropoulos, Katsaros e Manolopoulos (Sidiropoulos – Katsaros – Manolopoulos 2006) allo scopo di dare più rilevanza, rispetto all'*H-Index*, alle pubblicazioni recenti in modo tale da far emergere l'impegno di tutti quei ricercatori che lavorano e pubblicano con continuità.

HL-Index (*individual H-Index*), sviluppato da Batista, Campitelli, Kinouchi e Martinez (Batista – Campitelli – Kinouchi – Martinez 2006), corregge il difetto dell'indice *h* di attribuire, nel caso di pubblicazioni con più di un autore, l'intero contributo a ciascuno degli autori. Esso consiste nel rapporto tra il valore dell'*H-Index* e la media del numero di autori delle pubblicazioni che formano la base per il calcolo dell'indice *h*.

HL norm (*Normalized individual H-Index*), costruito e implementato dal tool *Publish or Perish* (*PoP*), è una variante del precedente. Esso prima normalizza il numero di citazioni per ogni pubblicazione mediante una divisione tra il numero di citazioni e quello degli autori che hanno contribuito alla pubblicazione, poi individua l'indice *h* utilizzando però come base per il calcolo il numero di citazioni normalizzate.

5.2.4. *Eigenfactor* (EF) e *Article Influence Score* (AIS)

L'*Eigenfactor* (EF)¹⁴ è un indicatore bibliometrico non proprietario sviluppato nel 2007 all'interno di un progetto di ricerca promosso dal laboratorio *Bergstrom* presso l'Università di Washington.

L'indicatore nasce con l'obiettivo di contestualizzare il valore di una rivista sia sulla base del comportamento citazionale specifico per ogni disciplina sia su quella del diverso peso in termini di prestigio delle riviste citanti nei vari ambiti disciplinari.

Eigenfactor, che si pone come strumento di valutazione delle riviste alternativo all'*Impact Factor*, sfrutta, adattandolo all'analisi delle citazioni ricevute dalle riviste, *Pagerank*, l'algoritmo di *ranking* dei siti web di *Google*. Il *ranking* dell'*Eigenfactor* avviene sulla base del tempo, espresso in percentuale, che il lettore impiega nel consultare e selezionare i riferimenti bibliografici che ritiene più utili per la propria ricerca.

14 <<http://www.eigenfactor.org>>. Una ricca bibliografia sull'indicatore è disponibile alla pagina web: <<http://www.eigenfactor.org/papers.php>>

Il valore percentuale che permette il *ranking* è anche il valore numerico dell'*Eigenfactor*. Se per esempio l'*Eigenfactor* di una rivista X nel 2011 è pari a 2,45 significa che il lettore ha impiegato il 2,45% del tempo a consultare e selezionare i riferimenti bibliografici sulla rivista X.

La fonte citazionale per il calcolo dell'*Eigenfactor* è il *Journal Citation Reports (JCR)* dell'*ISI* integrato nella piattaforma *Web of Knowledge*. La dipendenza dei dati dal *JCR* rendono l'*Eigenfactor*, come l'*Impact Factor*, soggetto a tutte quelle possibili distorsioni causate dal grado di copertura della fonte citazionale. Questo problema e l'impossibilità di calcolare l'indicatore al netto delle autocitazioni sono i principali punti deboli di questo indicatore.

Oltre a quelli già menzionati, tra i punti a favore si ricordano: la gratuità dello strumento; l'ampio *frame* temporale su cui avviene il calcolo che va dai due ai cinque anni; e, grazie alla collaborazione con il *Journal Cost-Effectiveness*,¹⁵ le informazioni che esso fornisce sui prezzi di moltissime riviste a carattere scientifico.

L'*Article Influence Score (AIS)* è l'altro indicatore disponibile nel sito <<http://www.eigenfactor.org>>¹⁶ che nasce allo scopo di misurare il livello medio d'influenza di un articolo contenuto in una rivista. Esso può essere definito una variante, o meglio un'estensione, dell'*Eigenfactor* poiché consiste nel valore dell'*Eigenfactor* diviso per il numero di articoli di una rivista.

5.2.5. *SCImago Journal Rank (SJR)* e *Source Normalised Impact per Paper (SNIP)*

*SCImago Journal Rank (SJR)*¹⁷ è l'indicatore citazionale da cui prende il nome la piattaforma *Scimago Journal & Country Rank*,¹⁸ un portale che include indicatori relativi sia alle riviste sia alla produzione scientifica di intere nazioni e che vuole porsi come valida alternativa ai *Journal Citation Reports* della *Thomson-Reuters*, sviluppato dal gruppo di ricerca *SCImago*¹⁹ del *Consejo Superior de Investigaciones Cientificas (CSIC)* in collaborazione con le uni-

15 <<http://www.journalprices.com/>>

16 Per un approfondimento sui metodi di *ranking* utilizzati nel sito <<http://www.eigenfactor.org>> si rimanda alla pagina web: <<http://www.eigenfactor.org/methods.pdf>>.

17 <<http://www.scimagojr.com/journalrank.php>>

18 <<http://www.scimagojr.com/index.php>>

19 <<http://www.scimago.es/>>

versità spagnole di Granada, dell'Estremadura, di "Carlo III" di Madrid e dell'Alcalà de Henares.

SJR^{20} è calcolato sulla base dei dati, relativi a riviste *peer-reviewed* e ad atti di convegni, indicizzati dal 1996 in *Scopus* (vedi paragrafo 4.3.2) e filtrati utilizzando un algoritmo del tutto simile al modello *PageRank* di *Google*. Scopo dell'indicatore, tra l'altro molto simile a quello di *Eigenfactor*, a cui viene accomunato anche per metodologia e per gratuità, è misurare, all'interno di una cornice temporale di tre anni, il prestigio di una rivista sulla base della provenienza delle citazioni ricevute dalla stessa rivista: le citazioni provenienti da riviste prestigiose, trasferendo parte del prestigio della fonte alla rivista citata, hanno maggiore rilevanza rispetto a quelle provenienti da riviste meno prestigiose, dunque anche maggior peso ai fini del calcolo dell'indicatore.

Il valore di *SJR* è oggi ricercabile anche nel database *Scopus*, così come lo è un altro indicatore: il *Source Normalised Impact per Paper (SNIP)*.

Creato da *Henk Moed al Centre for Science and Technology Studies (CWTS)*²¹ dell'Università di Leiden, *SNIP* misura l'impatto citazionale di una rivista mediante una ponderazione basata sul numero totale di citazioni appartenenti al contesto disciplinare di riferimento. Attribuisce quindi un valore più alto a una citazione che rientra in un settore disciplinare in cui le citazioni sono generalmente meno probabili, e viceversa.

È definito come il rapporto tra il numero di citazioni per articolo di una rivista e la capacità di attrarre citazioni del campo disciplinare d'appartenenza (Moed 2009).

20 *SJR* è soltanto uno dei numerosi indicatori calcolati da *SCImago* che calcola anche: l'indice di Hirsch, il numero di autocitazioni, il numero di documenti pubblicati nei quattro anni precedenti a quello in corso, il numero dei riferimenti bibliografici degli articoli pubblicati nelle riviste l'anno precedente, *Citation Index* degli articoli pubblicati nelle riviste nei tre anni precedenti, i documenti citabili (riferiti sempre ai tre anni precedenti) e il numero medio di citazioni per documento. Per una trattazione più approfondita cfr. *SCImago* (2007).

21 <<http://www.journalindicators.com/>>

6. Le metriche del Web

6.1. Dall'analisi citazionale all'analisi del *link* (*PageRank*)

Quando un sito o una pagina web punta un *link* ad un'altra risorsa web non fa altro che attivare lo stesso meccanismo di rinvio del documento che nei riferimenti bibliografici ne cita un altro: «un *link* rappresenta, con buona approssimazione, la versione *high-tech* della citazione bibliografica» (De Bellis 2005 p. 154; Jacsò 2004 p. 454). Le affinità di contenuto che nell'analisi citazionale sta alla base delle interrelazioni tra documenti citanti e citati è, a ben guardare, la stessa che si riscontra nell'analisi dei *link* tra pagine web.

Tale analogia ha permesso l'applicazione di gran parte del bagaglio teorico e pratico accumulato in molti decenni dalla bibliometria allo studio della struttura del Web e, come è accaduto per l'analisi citazionale, l'ideazione e la messa a punto di modelli matematici e metriche appositamente studiate per l'analisi del *link*.

Se infatti è vero che un documento citato deve pur avere qualche punto di contatto con il documento citante, è altrettanto vero che la pagina web verso cui si punta un *link* deve pur avere qualche affinità sotto il profilo del contenuto con la pagina che crea tale *link*. Lo studio delle interrelazioni tra pagine web fornisce informazioni riguardo sia al significato sia alla rilevanza della pagina web. In altre parole, con l'analisi dei *link*, allo stesso modo, si può desumere la rilevanza della pagina web guardando il numero di *link* che la separano da un'altra considerata rilevante: minore sarà il numero dei *link*, maggiore sarà la probabilità che la pagina web sia considerata rilevante (De Bellis 2005 p. 155).

È proprio su questi principi, derivati dall'analisi citazionale, che si basa *PageRank*,¹ l'algoritmo per l'ordinamento (*ranking*) delle pagine web nel motore di ricerca *Google* sviluppato da Larry Page alla fine degli anni '90. *PageRank* attribuisce maggior peso alle pagine

¹ *PageRank* è un marchio di proprietà di *Google* e il suo algoritmo un brevetto (U.S. Patent 6.285.999) registrato negli Stati Uniti. Una descrizione del brevetto è disponibile all'indirizzo <<http://www.google.com/patents/US6285999?dq=6285999&hl=it>>.

web che ricevono più *link* e, tra queste, favorisce quelle che vengono linkate da siti web che a loro volta sono molto linkati (Brin – Page 1998).

Quando si effettua una qualsiasi ricerca su *Google* l'ordinamento dei risultati è basato sul *PageRank* associato alle pagine; questo però non dipende dal numero assoluto dei *link* che rimandano a una pagina, ma dal *PageRank* dei siti che rimandano alla pagina in questione. In sostanza, l'idea è che il *PageRank* di un sito web si trasmette, almeno parzialmente, ai siti cui rimanda. L'efficienza dell'algoritmo di *PageRank* è testimoniata dall'enorme e crescente successo del motore di ricerca *Google*.

6.2. Web Impact Factor (WIF)

Il *Web Impact Factor*, sviluppato nel 1998 da Peter Ingwersen (Ingwersen 1998)² sull'analogia esistente tra *link* e citazioni bibliografiche, è considerato l'adattamento al web dell'*Impact Factor* sviluppato da Garfield.

WIF permette il *ranking*, la valutazione e il confronto tra siti web, domini di primo livello e sottodomini. Esso si calcola dividendo il numero totale di pagine web esterne che contengono un collegamento ipertestuale al sito oggetto di analisi (al netto dei link interni tra le pagine) per il numero di pagine web pubblicate nel sito raggiungibili dai motori di ricerca; quindi sono escluse dal calcolo tutte quelle pagine che, al contrario, non sono direttamente accessibili ai motori di ricerca.³ *WIF* misura quindi la frequenza con cui la “pagina web media” di un sito web riceve link da altri siti web in un dato intervallo di tempo.

WIF è in sostanza una misura dell'impatto di un sito web, utile soprattutto per determinare il posizionamento relativo di un sito (*ranking* del sito) all'interno di una classifica valida per un determinato settore o paese.⁴ Più alto è il fatto di impatto, maggiore è la reputazione percepita del sito web; un sito web con un *impact factor* superiore può essere considerato più prestigioso o di qualità superiore rispetto a quei siti web con un fattore d'impatto minore.

2 In realtà, prima di Ingwersen, già Rodriguez i Garin aveva parlato di impatto dell'informazione in Internet introducendo il concetto in un articolo apparso nel 1997 in una pubblicazione in lingua spagnola (Rodriguez i Garin 1997), contributo però che ha riscosso meno successo sulla comunità scientifica rispetto a quello dell'autore danese.

3 Questa è la definizione data da Ingwersen, ma esistono anche altre formulazioni dell'indicatore cfr. Li (2003).

4 Un esempio dell'applicazione di *WIF* è il *ranking* dei depositi istituzionali a livello mondiale : <<http://repositories.webometrics.info/en/world>>.

Come per gli indicatori citazionali visti nel capitolo precedente, anche l'impiego del *WIF* presenta punti a favore e punti a sfavore. Tra i principali vantaggi e utilità del *WIF* si contano:

1. possibilità di analizzare, sfruttando una metodologia di tipo comparativo, non solo la "visibilità internazionale" dei siti web istituzionali e accademici, ma anche la "visibilità relativa" di imprese, organizzazioni e paesi sul Web;
2. possibilità di valutare l'importanza e l'impatto di un sito web nel contesto di uno specifico ambito disciplinare;
3. applicabilità al macro livello di indagine; con *WIF* è possibile valutare interi segmenti del Web e più domini di primo livello raggruppati sulla base di un criterio geografico o disciplinare;
4. fornisce nuove intuizioni e spunti di miglioramento nell'ambito del processo di recupero dell'informazione sul Web; *WIF* può essere utile come strumento per misurare l'accuratezza sia delle prestazioni del motore di ricerca sia dell'organizzazione di un sito web, soprattutto per quanto riguarda la strutturazione delle pagine web e dei collegamenti ipertestuali nei contenuti.

Quando si usa *WIF* poi è importante tenere presente che il risultato è spesso influenzato, o peggio distorto, da una pluralità di fattori e variabili, come quelli indicati schematicamente nell'elenco seguente:

1. come avviene per gli indicatori citazionali i cui risultati variano a seconda del database utilizzato per il calcolo, anche il fattore d'impatto di qualsiasi sito web dipende dalla copertura dei motori di ricerca, che varia sia sotto il profilo del contenuto sia sotto quello temporale in quanto i dati relativi ai siti web contenuti nei database dei motori di ricerca non sono costanti nel tempo, ma possono cambiare considerevolmente per composizione anche di mese in mese. Se poi a ciò si aggiunge il fatto che i dati forniti dai motori di ricerca sono intrinsecamente incompleti – e qui è chiaro il parallelo con gli indici di citazioni sulla cui base gli indicatori bibliometrici vengono calcolati – perché è il Web stesso che non può garantire un'assoluta esaustività informativa (si dovrebbe sfatare una volta per tutte l'idea, purtroppo abbastanza radicata, secondo la quale tutto quello che non si trova nel Web non esiste nemmeno nel

docuverso ...), appare quindi chiaro che i risultati delle analisi dei motori di ricerca non possono essere considerati come dei valori assoluti sulla cui base trarre conclusioni definitive, bensì soltanto come indicazioni approssimative che dovrebbero essere abbinate ad altre metodologie di valutazione;⁵

2. i siti web di grandi dimensioni che pubblicano molte risorse, pur apportando un contributo effettivo più consistente alla letteratura del settore, possono anche non far registrare un *impact factor* più elevato rispetto a più piccoli siti web scientifici o accademici: l'elevato numero di risorse web è controbilanciato dall'elevato numero di *link* ricevuti da siti web esterni. Per chiarire questo punto di criticità è necessario fare un esempio. Supponiamo di avere soltanto due siti web in una ambito di interesse: A e B. A pubblica 10 risorse web e riceve 10 *link* da siti esterni in un dato intervallo temporale; B invece pubblica 100 risorse web e riceve 100 *link* da siti esterni nello stesso arco temporale di riferimento. Da questo esempio è ovvio che entrambi i siti web riportano lo stesso valore di *impact factor*, ma vista la sostanziale differenza in termini di influenza sulla letteratura del settore, è altrettanto evidente che l'indicatore non riesce a estrapolare alcuni elementi che permetterebbero invece una più efficace valutazione comparativa tra due o più siti web;
3. l'*impact factor* di un sito non è statisticamente rappresentativo delle pagine web che lo compongono. *WIF* varia in funzione del numero di siti web esterni che puntano alla risorsa da analizzare nel suo complesso, ma non in funzione del numero di volte la singola pagina web viene linkata da risorse esterne; e quindi, per esempio, se una risorsa web esterna contiene due *link* a pagine diverse del sito da analizzare, il fattore d'impatto non verrà aumentato di due unità, bensì di una sola;
4. i siti web pubblicati in lingue diverse dall'inglese registrano indipendentemente dal valore del loro contenuto, un *impact factor* molto più basso rispetto ai siti pubblicati in quella lingua. Ciò è dovuto parte agli autori e *webmaster* anglofoni che raramente creano collegamenti a siti pubblicati in lingue diverse dall'inglese, anche a prescindere dall'appropriatezza del *link* stesso, e parte agli autori e *webmaster* non di lingua inglese, o comunque appartenenti a contesti dove si utilizzano altre lingue, che cer-

5 È lo stesso Ingwersen (1998 p. 238) ad affermare: «in principle almost all searches will thus be incomplete».

cano di pubblicare i loro contributi nel Web in lingua inglese. Ad aggravare questo squilibrio sono i motori di ricerca stessi che, preferendo garantire copertura soprattutto ai siti web in lingua inglese, contribuiscono ad abbassare il *WIF* dei pochi siti non in lingua inglese che comunque vi sono inclusi poiché la maggior parte dei *link* che puntano a siti pubblicati in lingue diverse dall'inglese provengono da altri siti pubblicati in quelle lingue;⁶

5. così come gli indicatori di tipo citazionale possono soltanto misurare l'impatto della ricerca e non la sua qualità (Seglen 1997; Moed 2002) anche *WIF* non può che essere considerato un indicatore del prestigio o della popolarità che un sito web ottiene in rete, e non certo come una misura della qualità intrinseca della risorsa. Non sempre infatti un numero maggiore di *link* ottenuti da un sito corrisponde ad una maggiore qualità del suo contenuto;
6. *WIF* viene spesso utilizzato senza tener conto delle specificità disciplinari dei siti web. È sconsigliato infatti impiegare l'indicatore per valutare siti web dal contenuto trasversale a più discipline perché se in generale è vero che il numero di *link* ricevuti varia molto in funzione di lingua, stile comunicativo, numero di lettori e diffusione del sito web (più sono i lettori e più ampia è la diffusione e la probabilità per il sito di ricevere *link* da risorse web esterne), ma anche in base al conformismo di autori e *webmaster* (è diffusa infatti la tendenza a creare collegamenti a siti generalmente già molto linkati) e alle tendenze conoscitive, cioè ai temi più alla moda, in un determinato momento storico, è altrettanto vero che i siti web caratterizzati da un elevato grado di interdisciplinarietà e quelli appartenenti a settori dove la ricerca è più attiva o lo sviluppo tecnologico è più marcato, ricevono mediamente più *link* esterni rispetto agli altri siti web. È quindi auspicabile applicare sempre l'indicatore *WIF* a quei siti che possono essere ricondotti allo stesso ambito disciplinare;
7. *WIF* dipende dalle dinamiche di espansione e contrazione del sito web tanto che, diversamente dall'*Impact Factor*, basato sulla citazione bibliografica, il calcolo retrospettivo dell'indicatore non è ripetibile (Ingwersen 1998). Infatti, a differenza del nu-

6 Alcuni studi (Bharat et al. 2001; Thelwall 2002; Vaughan – Thelwall 2004) hanno dimostrato che i *webmaster* e *web bloggers* preferiscono creare collegamenti a risorse pubblicate nella lingua del sito web, e che, in genere, i siti linkano molto di più risorse del proprio paese che non quelle estere.

mero di citazioni bibliografiche a riviste scientifiche, istituzioni o singoli autori, che può essere stabile o aumentare costantemente nel corso del tempo, il numero di pagine web che si collegano a una determinata risorsa web può invece diminuire fino anche ad azzerarsi nel corso del tempo a causa, ad esempio, di chiusure o ristrutturazioni dei siti;

8. l'uso improprio delle autocitazioni nel calcolo dell'*Impact Factor* classico (*JIF*) da parte dei comitati editoriali delle riviste allo scopo di gonfiare i dati citazionali e ottenere così un valore più alto dell'indicatore (Seglen 1997) è un rischio, ma sarebbe meglio parlare di realtà, che riguarda anche il calcolo del *WIF*, visto che in questo caso i sono i *webmaster* che spesso fanno esplicita richiesta ad altri *webmaster* e autori di collegarsi ai loro siti web.

6.3. Indicatori d'uso non basati sulle citazioni

6.3.1. Progetti internazionali per la definizione di standard comuni e di metodi condivisi di raccolta e analisi dei dati sull'utilizzo

Negli ultimi anni la crescita esponenziale della letteratura scientifica online oltre a rendere più facile l'analisi valutativa e comparativa dei dati d'uso dei documenti elettronici (visualizzazione e *download*) raccolti esaminando i *logfiles* dei *server web*, ha reso sempre più stringente la necessità di definire standard comuni e metodologie condivise di raccolta e analisi dei dati di *log*. L'esigenza di normalizzazione è sfociata nell'avvio e sviluppo dei progetti internazionali *COUNTER*, *MESUR*, *SUSHI*.

6.3.1.1. COUNTER

Lanciato nel marzo del 2002, *COUNTER* (*Counting Online Usage of Networked Electronic Resources*)⁷ è un progetto internazionale nato allo scopo di facilitare la raccolta e la presentazione dei dati d'uso relativi alle risorse digitali online mediante la definizione di standard comuni a beneficio soprattutto di bibliotecari, editori e aggregatori, cioè dei principali *stakeholders* dell'iniziativa. Il progetto si è rilevato particolarmente utile in ambito biblioteco-

⁷ <<http://www.projectcounter.org/about.html>>

nomico, dove il problema della standardizzazione dei dati è emerso abbastanza precocemente visto che per lungo tempo i bibliotecari hanno dovuto procedere autonomamente e con grande dispendio di tempo ed energia, pena l'impossibilità di un loro utilizzo, alla normalizzazione dei dati d'uso forniti dagli editori nei formati più disparati.

Avere a disposizione delle statistiche d'uso standardizzate e quindi facilmente interpretabili e interoperabili ha giovato sia ai bibliotecari sia agli editori e aggregatori.

Per i primi ha voluto dire avere informazioni utili per uno sviluppo consapevole delle collezioni sia cartacee che digitali a supporto di decisioni quali la selezione del materiale informativo, la cancellazione e il rinnovo degli abbonamenti alle risorse digitali e non, e il passaggio all'*e-only*; la creazione di un modello di ripartizioni dei costi (costo – articolo) basato sul monitoraggio delle risorse realmente utilizzate e utile per l'ottimizzazione delle risorse a disposizione delle biblioteche.

Per i secondi ha significato poter: confrontare i dati d'uso relativi a canali di distribuzione diversi; aggregare i dati d'uso relativi a clienti che utilizzano più canali di distribuzione; migliorare i servizi offerti alla propria clientela sviluppando modelli di valutazione basati sulle statistiche d'uso.

Nel 2003 il progetto *COUNTER* ha pubblicato la prima versione del *Code of Practice*,⁸ una guida di livello internazionale con indicazioni pratiche per la produzione di statistiche d'uso semplici e attendibili relative a rilevazioni condotte su riviste e banche dati bibliografiche digitali. In particolare in questa prima versione il codice contiene:

- una lista controllata dei termini utilizzati e degli elementi che identificano i dati da misurare allo scopo di ridurre errori e inutili duplicazioni;
- i conteggi delle operazioni effettuate intenzionalmente dagli utenti, quelle accidentali infatti vengono eliminate dal calcolo per non distorcere il risultato finale;
- numero di report previsti (due per le riviste elettroniche e tre per le banche dati) con indicazione del loro formato (che deve essere compatibile con Microsoft Excel) e cadenza temporale con cui è consigliato vengano prodotti.

Dal 2003 sono uscite altre versioni di *COUNTER* dedicate ad altre risorse digitali; la più recente, *Code of Practice for e-Resources*, è stata rilasciata nell'aprile del 2012 (che assorbe e

8 <http://www.projectcounter.org/code_practice.html>

rinnova le versioni precedenti). Attualmente è possibile avere statistiche *COUNTER* per i principali editori e aggregatori internazionali tra i quali si ricorda Blackwell, Elsevier, Springer e Wiley.

6.3.1.2. MESUR

Il progetto *MESUR* (*Mesur from Scholarly Usage of Resources*),⁹ avviato nell'ottobre 2006 da Johan Bollen, Marko Rodriguez e Herbert Van de Sompel (Bollen – Rodriguez – Van de Sompel 2007 e 2008) del *Los Alamos National Laboratory* e finanziato dalla *Andrew W. Mellon Foundation*, nasce allo scopo di mettere a punto e validare delle metriche per la valutazione della ricerca scientifica basate sull'impatto dei dati d'uso delle risorse digitali e di sviluppare una cornice di procedure di riferimento sensibile al contesto d'applicazione di tali indicatori che concretamente si costruiscono conteggiando il numero di download dei documenti elettronici dalla rete.

Con l'avvento del Web, infatti, l'impatto di un documento può essere misurato sia sfruttando i tradizionali indicatori citazionali, retaggio del mondo cartaceo, costruiti su informazioni prodotte dagli autori, sia sfruttando gli indicatori d'uso costruiti invece su informazioni generate dai lettori. Come sottolineato da Cassella – Bozzarelli (2011 p. 75) sono molti i vantaggi legati all'impiego dei dati d'uso per la costruzione di indicatori bibliometrici:

- il dato sull'utilizzo di un documento in rete può essere rilevato molto prima del dato relativo al numero di citazioni ricevute dallo stesso; l'utilizzo può infatti essere tracciato non appena il documento viene messo in rete, mentre il numero di citazioni, è calcolabile soltanto quando è trascorso un certo lasso di tempo, che può variare dai tre ai sei mesi dopo la pubblicazione;
- diversamente dal dato citazionale che spesso è limitato al solo articolo scientifico in lingua inglese, il dato sull'utilizzo può essere rilevato per tutte le tipologie documentarie presenti nella rete;
- rispetto al dato citazionale inoltre, il dato sull'utilizzo è più efficace nel rendere conto dell'effettivo interesse con cui un documento è stato accolto dalla comuni-

9 <<http://mesur.informatics.indiana.edu/>>

tà scientifica: non sempre quello che viene letto e ritenuto interessante viene anche citato.

Il progetto è sfociato nella costruzione di un software e di una base di dati semantica e relazionale¹⁰ che integra informazioni di tipo bibliografico, citazionale e d'utilizzo dei documenti,¹¹ e che «consente di mappare, misurare e validare la massa critica di download degli articoli che creano una struttura sindetica nella rete» (Cassella – Bozzarelli 2011 p. 75).

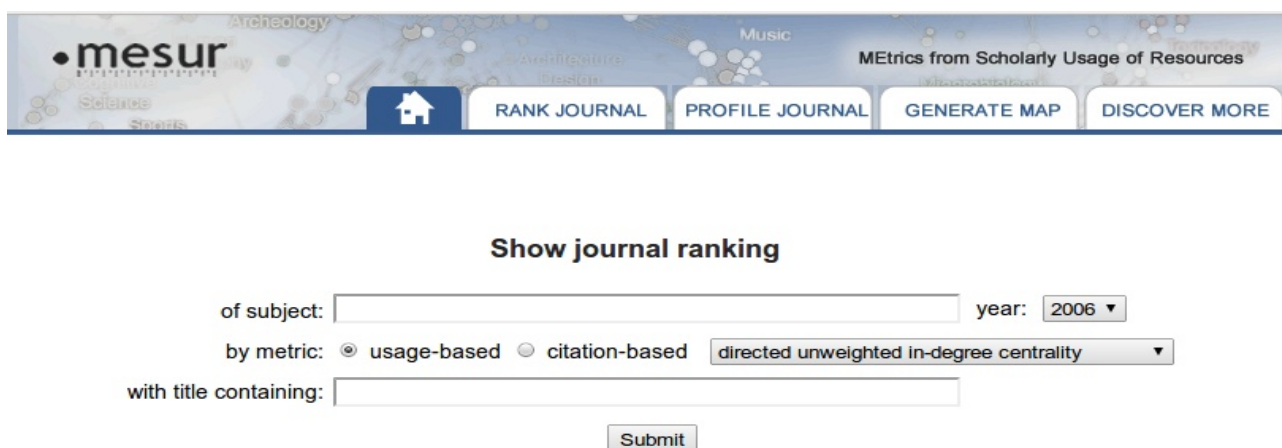


Figura 18: Schermata di ricerca di MESUR (fonte MESUR)

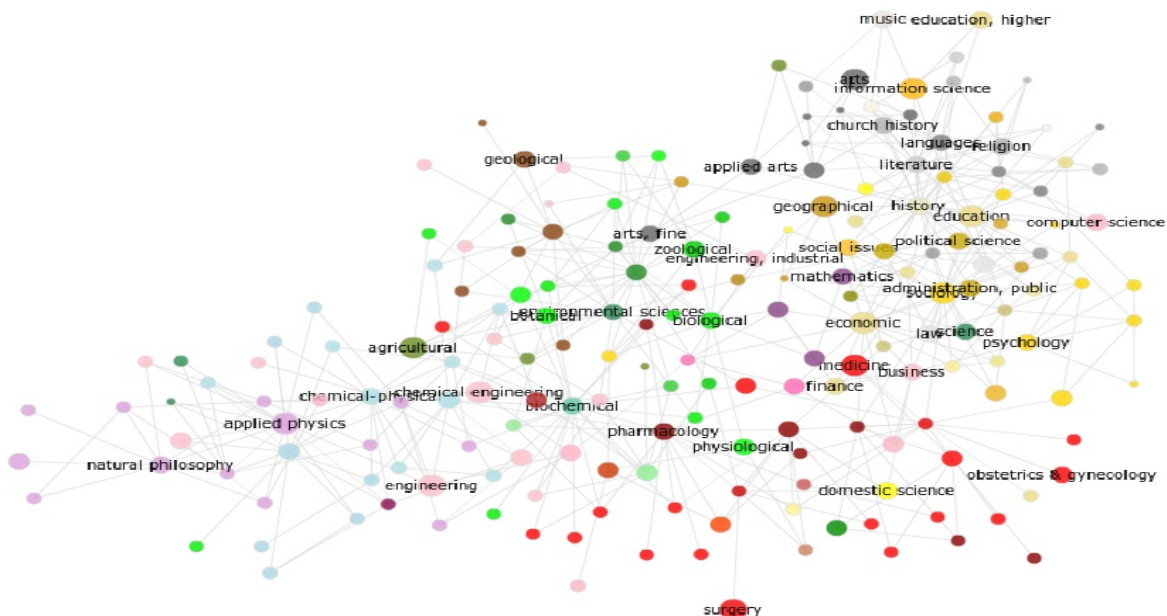


Figura 19: Esempio di mappa delle discipline generata da MESUR (fonte MESUR)

¹⁰ <<http://mesur.informatics.indiana.edu/>>

¹¹ I dati sono stati ricavati analizzando i file di log dei download dai siti web di importanti editori, aggregatori e consorzi di biblioteche che si riferiscono a un corpus selezionato di 100.000 periodici, per 10.000 dei quali è riportato e mappato anche il dato citazionale. MESUR può inoltre contare su milioni di report COUNTER prodotti da migliaia di istituzioni internazionali.

6.3.1.3. SUSHI

Il progetto *SUSHI* (*Standardized Usage Statistics Harvesting Initiative*),¹² sviluppato a partire dal 2004 prima come iniziativa congiunta di alcune università americane e importanti editori e aggregatori, poi come standard *NISO* (*National Information Standards Organization*),¹³ può essere definito «un modello basato su un servizio web per la richiesta di dati statistici» (Dellisanti 2007 p. 18).

L'iniziativa nasce per ovviare ad alcune problematiche legate all'utilizzo dai dati d'uso elaborati in base allo standard *COUNTER* da parte di programmi chiamati *ERM* (*Electronic Resource Management*) creati per la gestione di tutte le fasi del ciclo di vita delle risorse elettroniche in biblioteca.

I report *COUNTER* infatti oltre a essere disponibili soltanto nel formato Microsoft Excel, presentano dati difficilmente aggregabili poiché sono elaborati su tanti formati file quanti sono gli editori che li forniscono.

SUSHI si presenta invece un modello più efficace, capace di facilitare il processo di acquisizione e aggregazione dei dati; i bibliotecari infatti non devono più scaricare i dati d'uso dai siti web dei diversi editori e aggregatori poiché questa operazione viene fatta direttamente da *SUSHI* (*SUSHI Client*) che invia una richiesta al provider (*SUSHI Server*) che, a sua volta, reperisce e fornisce la risposta. Questa operazione inoltre può essere ulteriormente facilitata se il *SUSHI Client*, normalmente integrato nel server della biblioteca, è associato a un software *ERM* che provvede a immettere i dati direttamente nel ciclo di gestione delle risorse elettroniche. Altri vantaggi offerti da *SUSHI* sono il numero illimitato di richieste che la biblioteca può effettuare; l'elaborazione di report a livello consortile; e la possibilità di richiedere report preimpostando gli intervalli temporali a cui si vuole vengano riferiti i dati d'uso.

I dati statistici che *SUSHI Server* elabora sono in formato *COUNTER* di cui è stato mantenuto lo standard *XML*, noto per le sue caratteristiche di flessibilità, compatibilità e portabilità.

¹² <<http://www.niso.org/workrooms/sushi/>>

¹³ *NISO* è un'organizzazione *non-profit* statunitense che si occupa di sviluppare, implementare e pubblicare norme tecniche relative ad applicazioni editoriali, bibliografiche e biblioteconomiche.

6.3.2. Usage Factor (UF)

Con la sempre più massiccia presenza online di rilevanti riviste scientifiche e accademiche unitamente all'ampia disponibilità di affidabili statistiche d'uso online compatibili con *COUNTER* e alla spinta propositiva del progetto *MESUR*, che ha dimostrato l'alto valore potenziale di un'ampia gamma di metriche basate su dati d'uso per la valutazione a livello globale dell'impatto dei periodici elettronici, si è fatta avanti, sempre più prepotentemente, l'idea di affiancare al tradizionale *Impact Factor* un indicatore capace di misurare la performance di una rivista utilizzando come unità costitutiva non più la citazione bibliografica, bensì il dato sull'uso della risorsa.

Tale indicatore, chiamato *Journal Usage Factor (JUF)* è stato messo a punto all'interno del progetto *Usage Factor* avviato in collaborazione con *COUNTER* dall'*UK Serials Group (UKSG)*,¹⁴ che dal 2006 studia le potenzialità offerte alla valutazione scientifica da una metrica basata sulle statistiche d'uso delle riviste e pubblicazioni online.

Nello specifico il progetto mira a valutare: quali sono i metodi generalmente impiegati per valutare la qualità di una rivista; la significatività statistica dello *Usage Factor*; il grado di interesse con cui tale metrica può essere accolta da ricercatori, editori, bibliotecari e istituti di ricerca; il suo grado di credibilità, affidabilità e robustezza; e, infine, il modello organizzativo ed economico più adatto e accettato dalle principali parti interessate entro cui l'indicatore può essere messo a punto e implementato.

Il progetto si è sviluppato in più fasi di cui attualmente è stata completata la terza.

La prima fase, conclusasi nel giugno 2007, è stata dedicata a una ricerca di mercato rivolta ai principali *stakeholders* ovvero ai bibliotecari, editori e autori, che ha avuto lo scopo di sondare l'accettabilità e fattibilità complessiva del principio su cui è costruito l'indicatore.

Lo *Usage Factor*, oggetto principale dell'indagine, si ottiene dividendo il numero totale di download avvenuti in un determinato arco temporale degli articoli pubblicati da una rivista in determinato anno e il numero totale di articoli pubblicati nella rivista nello stesso anno.

Dall'indagine è emerso che sebbene le principali parti interessate al progetto sono favorevoli all'adozione di un indicatore complementare o alternativo all'*Impact Factor*, tuttavia

14 <<http://www.uksg.org/>>

si sono palesati alcuni punti di criticità che possono essere riassunti schematicamente nel modo seguente:

- senza miglioramenti, le statistiche *COUNTER* attuali sono inadeguate a fungere da unità di base per tale indicatore;
- le modalità di calcolo dell'indicatore non sono condivise universalmente da tutti gli interessati ad un suo impiego;
- da alcune parti viene sottolineato che il dato d'uso non sarebbe rappresentativo del reale impatto esercitato da una risorsa elettronica sulla comunità in quanto non è detto che tutto ciò che viene “scaricato” venga anche letto;
- come per il dato citazionale, anche per il dato relativo ai download esiste il rischio di manipolazione da parte dell'autore;
- e, infine, l'assenza di segnalazione circa eventuali scostamenti rispetto ai valori medi rende difficile attribuire il giusto significato al semplice dato che oltre a non essere contestualizzato nella sequenza statistica non lo è neppure rispetto all'ambito disciplinare di riferimento.

La fase due del progetto, che si è conclusa nel 2010, ha riguardato la costruzione di un modello d'analisi concreto di applicazione dell'indicatore. In particolare, è stato concretamente definito quali debbano essere i dati forniti dagli editori nel formato *COUNTER* e quale debba essere la cadenza temporale con cui deve avvenire la raccolta dei dati. I dati d'uso reali sono stati poi utilizzati per testare la formula per il calcolo dell'*UF*.

La terza fase iniziata nell'ottobre del 2011 non è ancora terminata, ma appare chiaro già oggi che nonostante il grande entusiasmo con cui l'indicatore è stato accolto dai bibliotecari, editori e autori, esso non sia ancora “maturo” per essere considerato una stabile alternativa ai tradizionali indicatori citazionali, bensì una misura a essi complementare nella valutazione dell'impatto – e di conseguenza, del valore – delle pubblicazioni scientifiche e accademiche.

6.4. Aggregazione di metriche a livello dell'articolo (*Article Level Metrics*): il caso *PLoS*

Un punto di convergenza tra le diverse metodologie di valutazione della ricerca, capace di rispondere, almeno in parte, alla necessità di un uso combinato di indicatori differenti, unica vera garanzia per un'analisi efficiente e affidabile, sembra essere stato raggiunto nel marzo 2009 con l'introduzione del servizio *Article Level Metrics (ALM)*¹⁵ da parte della *Public Library Science (PLoS)* per tutte le sette riviste della sua piattaforma.

L'idea innovativa di *PLoS* non è stata inventare un nuovo indicatore alternativo o complementare a quelli già esistenti, ma quella di aver saputo aggregare a livello dell'articolo ciò che già c'era, contribuendo in questo modo sia a colmare quell'esigenza di integrazione, sentita da sempre ma mai concretamente affrontata, di integrazione tra valutazione quantitativa e valutazione qualitativa (Cassella – Bozzarelli 2011 p. 77), sia a sviluppare un sistema di metriche capaci di misurare l'impatto a livello dell'articolo anche in contrapposizione allo schema dell'indicatore bibliometrico più noto, l'*Impact Factor*, che nel tentativo di misurare l'impatto dell'articolo misura di fatto l'impatto della rivista in cui l'articolo è pubblicato.

Come si può leggere nel sito web di *PLoS*,¹⁶ *ALM* è disponibile per ogni singolo articolo contenuto nella piattaforma e comprende:

- statistiche d'uso dei documenti: numero di visite alle pagine *HTML*, numero di download dei documenti in formato *PDF* e *XML*;
- dati citazionali: numero di citazioni bibliografiche calcolato dagli indici *PubMed Central*, *Scopus*, *Web of Science* e *CrossRef*;
- informazioni provenienti dai *social networks* *CiteULike*, *Connotea*, *Twitter*, *Facebook* e *Mendeley*;
- *Blog Posts*: numero di post contenuti in *Post genomic*, *Nature Blogs* e *Bloglines*;

15 Per maggiori informazioni su *Article Level Metrics (ALM)* cfr. <<http://article-level-metrics.plos.org/>> e <<http://article-level-metrics.plos.org/alm-info/>>.

16 <<http://www.plosone.org/home.action>>

- note e commenti: la piattaforma *PLoS* permette al lettore di partecipare alla valutazione dell'articolo attraverso l'aggiunta di note e commenti, e l'assegnazione di un punteggio da 1 a 5 a ogni singolo articolo.



Figura 20: Schermata servizio *Article Level Metrics (ALM)* di *PloS* (fonte *PLoS*)

7. Le nuove sfide per le biblioteche e i bibliotecari

La crescente importanza della bibliometria nella società attuale dovuta sia allo sviluppo delle moderne tecnologie dell'informazione sia ai significativi cambiamenti che stanno interessando la politica e l'organizzazione della ricerca scientifica, oltre a rappresentare certamente una grande opportunità, pone nuove sfide per le biblioteche accademiche e scientifiche (Ball – Tunger 2004; Gorraiz – Wieland 2009a, b, c; MacColl 2010; Haest 2010; Gerrietsma et al. 2010) specie sotto il profilo del ruolo da loro svolto all'interno del processo di comunicazione scientifica che, nel complesso, risulta notevolmente ridefinito e ampliato (Gumpenberger et al. 2012 p. 175).

In una società in cui le risorse destinate alla ricerca scientifica stanno diventando sempre più scarse per effetto della crisi economica mondiale ma anche della crescita della popolazione scientifica che contribuisce a decrementarne proporzionalmente la quantità, sta determinando una sempre più forte competizione tra scienziati, tra e all'interno di università e centri di ricerca per l'accaparramento dei fondi di ricerca. Tutto ciò ha fatto emergere una doppia necessità: da una parte quella sentita dai governanti di ottimizzare le poche risorse a disposizione attribuendo finanziamenti pubblici ai progetti davvero meritevoli e di qualità, quelli cioè capaci di favorire lo sviluppo del paese; dall'altra, quella di scienziati, università e istituti di ricerca di poter contare su metodologie di valutazione e indicatori quantitativi davvero imparziali e oggettivi, come lo sono, fatte le dovute eccezioni, l'analisi citazionale e gli indicatori bibliometrici.

È proprio in questo contesto che emerge il ruolo svolto dalle biblioteche nelle università e nei centri di ricerca. Per tali biblioteche infatti si profila un impiego della bibliometria non solo a supporto delle politiche documentarie finalizzate allo sviluppo e alla gestione delle raccolte, ambito d'applicazione questo già ampiamente collaudato e consolidato (vedi capitolo 2), ma anche per l'elaborazione di dati d'alta qualità a sostegno della valutazione della ricerca svolta all'interno delle stesse università e istituti di ricerca (Ball – Tunger 2004; Gorraiz et al. 2010; Piazzini 2010 p. 82; Cassella – Bozzarelli 2011 p. 77-78).

Per le biblioteche accademiche e scientifiche si tratta, in sostanza, di un servizio nuovo, ben calato però nella realtà di quelli tradizionalmente offerti; da sempre infatti le biblioteche hanno saputo offrire alla ricerca risorse informative adeguate e integrate da servizi quali il prestito interbibliotecario e di fornitura di documenti, opac ordinati e sempre più arricchiti con nuove risorse e servizi, accessi a risorse elettroniche locali e remote robusti e costantemente aggiornati; e infine, ma non perché meno importanti, le competenze e la professionalità dei bibliotecari.

In stretta collaborazione con i nuclei di valutazione interni alle università o, più in generale, con il settore o l'ufficio preposto alla valutazione della ricerca, le biblioteche si occupano di raccolta e conservazione di *set* di dati primari, di rilevazioni bibliometriche, di analisi quantitative effettuate mediante l'impiego degli indicatori bibliometrici (citazionali e non) più consoni per la specificità dei dati di riferimento, fino ad arrivare – ma oggi ciò rappresenta una rarità – ad essere coinvolte in un'attività vera propria di consulenza rivolta ai ricercatori e finalizzata per esempio alla scelta del canale di pubblicazione più adatto o alla valutazione della performance scientifica di uno o più contributi pubblicati sotto il profilo sia dell'impatto prodotto nella comunità disciplinare di appartenenza sia dell'incisività ponderata in riferimento all'assegnazione dei fondi di ricerca all'interno dell'università o dell'istituto di ricerca di riferimento.

Da quest'ultima considerazione, emerge quindi che la bibliometria in biblioteca ha un triplice impiego:

- a sostegno delle politiche documentarie finalizzate allo sviluppo e alla gestione delle collezioni bibliotecarie;
- a sostegno degli obiettivi generali in tema di valutazione della ricerca dell'ente di appartenenza;
- a sostegno del bisogno informativo di tipo bibliometrico dell'utenza composta, in questo caso, soprattutto da ricercatori ma estendibile anche a studenti e dottorandi.

All'ultima applicazione elencata, che come detto è poco sviluppata in Italia, dovrebbe essere dedicato uno spazio specifico all'interno della biblioteca, contiguo ma separato da quello occupato dal *reference librarian*. Contiguo perché i metodi bibliometrici potrebbero

apportare un valore aggiunto al servizio di *reference*; separato perché le competenze, gli strumenti e la metodologia operativa del bibliotecario esperto di bibliometria sono fondamentalmente diversi da quelli del *reference librarian*. Si potrebbe parlare quindi di una sorta di “*reference service* di tipo bibliometrico” dove il personale addetto avrebbe un contatto diretto con l'utenza, di cui cercherebbe di soddisfare uno specifico bisogno informativo che, a sua volta, potrebbe anche variare molto in termini di complessità; il bibliotecario cioè potrebbe trovarsi a rispondere a richieste che vanno dalla semplice assistenza nell'uso delle banche dati bibliometriche fino ad arrivare ad includere argomenti più sofisticati e complessi come la strategia di pubblicazione, vale a dire la scelta del canale di pubblicazione più adatto dal punto di vista bibliometrico.

Altro compito fondamentale del bibliotecario esperto di bibliometria all'interno di università e di istituti di ricerca è quello di organizzare attivamente momenti formativi rivolti non solo ai ricercatori e agli studenti ma anche ai membri dello staff che si occupa di ricerca nell'ente e ad altri bibliotecari. L'obiettivo generale è quello di far aumentare la consapevolezza dell'impiego degli strumenti bibliometrici in ambito biblioteconomico e della valutazione della ricerca. Nello specifico, il bibliotecario dovrà illustrare in modo semplice e chiaro:

- le metodologie d'analisi qualitativa e quantitativa utili alla valutazione della ricerca;
- i punti di forza e di criticità dei principali strumenti bibliometrici;
- i vantaggi, in termini di affidabilità e correttezza dei risultati, di un uso combinato di metriche diverse, di converso, i limiti e le possibili distorsioni dei risultati derivanti dall'impiego di una singola metrica di valutazione;
- le caratteristiche, specificità e potenzialità dei più importanti strumenti di rilevazione dei dati bibliometrici;
- le tendenze di sviluppo della bibliometria e delle metriche bibliometriche anche in relazione all'innovazione tecnologica e alle trasformazioni riguardanti il processo di produzione e pubblicazione della ricerca scientifica.

Le occasioni d'apprendimento proposte dai bibliotecari possono consistere in seminari, *workshop*, conferenze fino anche ad arrivare a veri e propri corsi strutturati, o in soluzioni

di tipo auto-formativo come guide e *slides* online, *tutorial*, opuscoli orientativi e prontuari sintetici etc.

A tal proposito, all'interno del sito web della biblioteca sarebbe utile organizzare una sezione appositamente dedicata alla bibliometria che riporti finalità, modalità di espletamento del servizio, orari ed eventuale materiale didattico-informativo messo a disposizione dell'utenza.

Sembra interessante a questo punto della trattazione, capire quali siano le ragioni che spingono le università e gli istituti di ricerca a considerare la biblioteca come la struttura organizzativa locale più adatta per le attività bibliometriche. Un motivo particolarmente rilevante è dato dall'importanza che riveste la bibliometria all'interno della ricerca biblioteconomica che rende quindi del tutto naturale una sua applicazione nella pratica professionale. Altre motivazioni sono legate alle competenze dei bibliotecari: l'esperienza nello sviluppo e nella gestione di grandi quantità di dati bibliografici; la capacità di organizzare e gestire, attraverso classificazione e soggettazione, grandi quantità di documenti; la possibilità di aver accesso, e di sapere naturalmente utilizzare in modo efficiente, i principali database bibliografici e molti dei loro strumenti analitici integrati; l'esperienza nella rappresentazione del contenuto concettuale di un documento mediante un linguaggio normalizzato, e più in generale nella produzione di metadati allo scopo di facilitare la ricerca e il recupero dell'informazione.

Altri punti a favore delle biblioteche sono la loro grande flessibilità e versatilità; i bibliotecari infatti non solo hanno avuto e hanno la capacità di creare e implementare una vasta gamma di servizi anche in funzione dell'innovazione tecnologica, ma contribuiscono anche al progresso scientifico generale di una qualsiasi disciplina partecipando a progetti e collaborazioni, all'organizzazione di conferenze e seminari, e pubblicando attivamente i risultati più significativi raggiunti.

Da non trascurare poi tra le motivazioni appena ricordate la posizione della biblioteca all'interno della più ampia organizzazione universitaria o di un istituto di ricerca. La biblioteca è considerata una struttura stabile e indipendente all'interno di un'università o istituto di ricerca e, cosa ancora più importante, non influenzata dai risultati delle analisi bibliometriche, che le permette così di assumere una posizione neutrale e oggettiva rispet-

to a ricercatori, a facoltà e dipartimenti soggetti a valutazione della performance scientifica.

Non bisogna infine dimenticare che sempre più spesso la biblioteca è parte fondamentale nei processi di comunicazione scientifica. La biblioteca infatti, oltre ad assumere un ruolo attivo nelle tematiche inerenti la pubblicazione *open access*, è frequentemente responsabile dello sviluppo e della gestione di depositi istituzionali (De Robbio 2007 p. 28; Ridi 2007 p. 102-103; Deana 2008; Galimberti 2010a), cioè di archivi digitali ad accesso aperto che permettono la tempestiva auto-archiviazione, mediante il deposito dei *pre-prints* o dei *post-prints*, dei contributi di ricerca (ma anche della letteratura grigia prodotta ai fini della ricerca e della didattica) prodotti in un'università o, più in generale, in un ente. Nel contesto della valutazione della ricerca i depositi istituzionali costituiscono una preziosa fonte di dati sulla cui base misurare la performance scientifica non solo di un'università nel suo complesso ma anche dei singoli dipartimenti che la compongono. Possono poi diventare delle piattaforme su cui oltre a poter sperimentare l'applicazione delle metriche rientranti nel filone più innovativo della bibliometria, o meglio della webmetrica, cioè quelle basate sui dati d'uso, poter anche applicare degli indicatori specificatamente calibrati sulla realtà effettiva della ricerca della singola università o del singolo istituto. Spesso si è parlato dell'inadeguatezza dei normali strumenti di rilevazione del dato bibliometrico per settori disciplinari come le scienze umane e sociali e della conseguente impossibilità di applicare senza modifiche le metriche utilizzate per altri settori scientifici; il dato sui depositi istituzionali sarebbe invece in grado di rappresentare più efficacemente la realtà della ricerca in tali ambiti proprio perché elaborato su una base contenente l'elenco completo dei contributi di docenti e ricercatori afferenti a un'università o istituto di ricerca.

Per i bibliotecari quindi l'impiego della bibliometria ai fini della valutazione della ricerca rappresenta una sfida che assolutamente deve essere raccolta. Sviluppando nuove competenze, essa contribuisce a rafforzare il ruolo delle biblioteche in relazione all'utenza ma anche nel più ampio contesto dell'università o dell'istituto di ricerca a cui afferiscono.

Dal punto di vista della professione, l'allargamento delle competenze può portare un potenziale miglioramento dello status della biblioteconomia sia perché gli indicatori quantitativi sono sempre stati tenuti in grande considerazione nel mondo accademico sia per-

ché con l'utilizzo della bibliometria i bibliotecari arriverebbero ad esercitare maggiore influenza sui processi politici interni all'istituto di appartenenza.

La bibliometria però per rappresentare davvero una sfida per le biblioteche, e di conseguenza un'opportunità di inserimento nei processi della valutazione e della comunicazione scientifica, deve prima diventare una parte stabile del bagaglio professionale dei bibliotecari, dunque non relegata all'interesse del singolo. In questo contesto quindi sarebbe auspicabile integrare le abilità e le competenze dei bibliotecari con dei corsi di formazione mirati all'acquisizione di questa specifica metodologia e, a livello universitario, costituire una parte fondamentale della formazione degli studenti che scelgono un percorso bibliografico-biblioteconomico.

8. Caso di studio: antropologia a Ca' Foscari

8.1. Introduzione

Le analisi bibliometriche che seguono sono state costruite su un *set* di dati che pur essendo molto specifico e delimitato permette un buon grado di generalizzazione dei risultati; i dati qui analizzati presentano infatti tutte le caratteristiche, e se vogliamo, tutte le problematiche se considerati sotto il profilo dell'applicazione degli indicatori bibliometrici, tipiche del campo disciplinare d'appartenenza e, più in generale, delle scienze umane e sociali. I dati raccolti si riferiscono ad un'ampia area disciplinare riconducibile alle scienze etnoantropologiche. Nello specifico, si è scelto di analizzare tutti i riferimenti bibliografici che costituiscono le bibliografie d'esame per gli anni accademici 2010/11 e 2011/12 di alcuni insegnamenti previsti all'interno del Corso di Laurea Magistrale in "Antropologia culturale, etnologia, etnolinguistica"¹ dell'Università Ca' Foscari di Venezia.

Il *set* di dati considerato comprende 298 pubblicazioni per l'anno accademico 2010/11 e 285 pubblicazioni per l'anno accademico 2011/2012 suddivise per i seguenti insegnamenti:

	INSEGNAMENTO	A.A. 2010/2011	A.A. 2011/2012
1	Antropologia culturale	10	23
2	Antropologia culturale, introduzione	7	6
3	Antropologia culturale, storia	5	5
4	Antropologia del Medio Oriente	22	14
5	Antropologia dell'arte	26	26
6	Antropologia della salute	10	11
7	Antropologia filosofica	30	42
8	Antropologia sociale	21	27
9	Antropologia teatrale	6	10
10	Antropologia visiva	4	10
11	Etnografia	20	22
12	Etnolinguistica	59	50
13	Etnologia	78	37
14	Etnologia della musica	-	2
	TOTALE	298	285

Tabella 4: Numero pubblicazioni insegnamento per anno accademico

1 <http://www.unive.it/nqcontent.cfm?a_id=93122>

8.2. Gli strumenti utilizzati per la rilevazione dei dati

bibliometrici

Per la rilevazione dei dati bibliometrici sono stati interrogati i seguenti indici di citazioni multidisciplinari online: *Web of Science*, *Scopus* e *Google Scholar*.

Pur venendo universalmente considerati i principali strumenti di recupero di informazioni bibliometriche, ad una veloce ricerca finalizzata a stabilire la presenza o meno tra i documenti indicizzati dei riferimenti bibliografici oggetto di analisi, emerge tuttavia palesemente la scarsa copertura dei database citazionali *WoS* e *Scopus* nelle scienze sociali.

Come già ricordato nel capitolo 4, tali limitazioni sono dovute al fatto che questi indici di citazioni, in particolare *WoS* e *Scopus*, indicizzano nella stragrande maggioranza dei casi articoli che appartengono ai settori delle cosiddette “scienze dure”, mentre come è noto nelle discipline sociali e umane le pratiche di pubblicazione propendono più verso la monografia specialistica che non verso la rivista che, in questi ambiti disciplinari, come mezzo di comunicazione della ricerca, gioca per forza un ruolo minore.

Anche l'aspetto linguistico gioca un ruolo importante se si considera che *WoS* e *Scopus* indicizzano soprattutto articoli in lingua inglese, mentre la produzione scientifica nelle scienze umane e sociali ha prevalentemente carattere nazionale e quindi utilizza la lingua nazionale che nel caso dei documenti oggetto di questo studio è l'italiano.

Nello specifico, rispetto al *set* di dati di riferimento la copertura è stata:

- in *Scopus*: scarsa; non trova conferma in questo caso l'opinione comune su *Scopus*, ribadita da Biolcati-Rinaldi (2010 p. 9) «di essere, rispetto a *Web of Science*, maggiormente inclusivo delle scienze sociali e delle scienze nazionali», in quanto, se si guarda al numero di documenti indicizzati, l'interrogazione ha dato risultati pressoché sovrapponibili; se si guarda invece al recupero del dato bibliometrico – e questo spiega le diverse percentuali di recupero nei due database – *Scopus* presenta un numero di documenti indicizzati molto inferiore rispetto a *WoS*;
- in *Web of Science*: scarsa, ma maggiore rispetto a *Scopus*; pur avendo *WoS* un numero di documenti indicizzati inferiore a quello di *Scopus*, ha permesso tuttavia

il recupero di un maggior numero di dati bibliometrici. Ciò è dovuto parte alla recente introduzione all'interno della piattaforma *WoS* del *Book Citation Index*, uno specifico database citazionale dedicato a libri e capitoli di libri, parte al fatto che è possibile recuperare il numero di citazioni ricevute anche dai documenti non indicizzati in *WoS* ma comunque presenti tra i riferimenti bibliografici di altri documenti indicizzati;

- il *Google Scholar*: molto estesa; rispetto ai precedenti database commerciali, *GS* indicizza un numero molto più ampio di tipologie documentarie e quindi ha il pregio di essere maggiormente inclusivo a discapito però della qualità dei dati che, viceversa, è molto più elevata in *WoS* e *Scopus*. Per ovviare a tale inconveniente e sfruttare al meglio *GS* è stata utilizzata per il recupero dei dati l'interfaccia di interrogazione *Publish or Perish* (vedi paragrafo 4.3.3.), grazie alla quale è stata possibile un'accurata pulizia dei dati di partenza.

8.3. Metodologia d'analisi

Ogni singolo riferimento bibliografico inserito nelle bibliografie d'esame degli insegnamenti oggetto d'analisi per il periodo considerato è stato analizzato. Il lavoro d'analisi, avvenuto durante l'estate del 2012, è stato sviluppato in due fasi: la prima di raccolta e individuazione dei dati primari; la seconda di rilevazione dei dati bibliometrici, preceduta da una approfondita fase di scelta degli indici di citazioni e indicatori bibliometrici più adatti al *set* di dati di riferimento.

I dati primari sono stati raccolti consultando i programmi di ogni singolo insegnamento disponibile online sul sito dell'Università di Venezia. I dati così raccolti sono stati numerati, mantenendo la suddivisione per insegnamento, e trasferiti in un foglio di calcolo, riportando nome dell'autore, titolo del documento e anno di pubblicazione.

La fase di rilevazione dei dati bibliometrici si è concretizzata, a sua volta, in due fasi distinte ma complementari. Nella prima fase, si è considerato il *set* di dati nella sua interezza senza procedere a suddivisioni interne; l'obiettivo è stato quello di descrivere in generale, o meglio di tratteggiare a grandi linee, la fisionomia del *set* dati di riferimento utilizzando l'analisi bibliometrica. In questa fase sono stati applicati gli indicatori più elementari, quel-

li noti come primari. Il conteggio del numero di pubblicazioni è servito a stabilire per i documenti oggetto di analisi: la proporzione tra libri e riviste; la suddivisione in lingue; l'età; e la percentuale di presenza nei vari indici di citazioni analizzati. Il conteggio del numero di citazioni, rilevato per ogni singolo documento e su tutti i tre gli indici di citazioni, è stato però rappresentato graficamente soltanto per i dati provenienti da *GS* poiché, vista la scarsa copertura di *WoS* e *Scopus* rispetto al *set* di dati a disposizione, si è scelto di privilegiare il database più completo in modo tale da fornire un quadro citazionale più rappresentativo possibile dei documenti oggetto d'analisi.

È stato calcolato anche la media di citazioni per anno di ogni singolo documento.

Nella seconda fase si è proceduto ad un *ranking*, sulla base delle citazioni ricevute, degli autori e delle riviste più rappresentative all'interno del *set* di documenti a disposizione e per entrambi gli anni accademici. In sostanza analizzando le citazioni provenienti dai tre database si sono estratti i dieci autori più citati. Dopo una ricerca preliminare del numero di articoli di ogni singolo autore presenti nei tre database, a questo *set* di dati più ristretto sono stati applicati i seguenti indicatori sempre riferiti al singolo autore: *Citation Index*; *H-Index*; media citazioni per documento; media citazioni per anno; *G-Index*; *hc-Index*; *hl, norm.*

La stessa procedura è stata seguita per le dieci riviste più citate alle quali sono stati applicati i seguenti indicatori: numero totale di citazioni, *Impact Factor*; *5-Year Impact Factor*; *Immediacy Index*; *Cited Half-life*; *Eigenfactor*; *Article Influence Score*; *SCImago Journal Rank (SJR)*; e *Source Normalised Impact per Paper (SNIP)*.

Per il calcolo dell'*IF* e degli indicatori ad esso collegati e da esso derivanti è stata utilizzata l'edizione del *Social Sciences Citation Index* del 2011, così come sono basati sui dati del 2011 gli indicatori *SJR* e *SNIP* calcolati in *Scopus*.

In questa seconda fase i dati bibliometrici riferiti sia agli autori sia alle riviste sono stati limitati al periodo compreso nei 16 anni che vanno dal 1996 al 2012 allo scopo di uniformare i dati di *WoS* e *GS* a quelli di *Scopus* che calcola le citazioni ricevute dagli articoli soltanto a partire dal 1996.

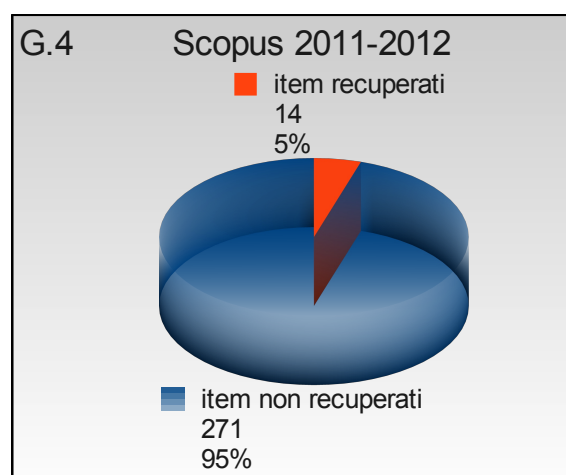
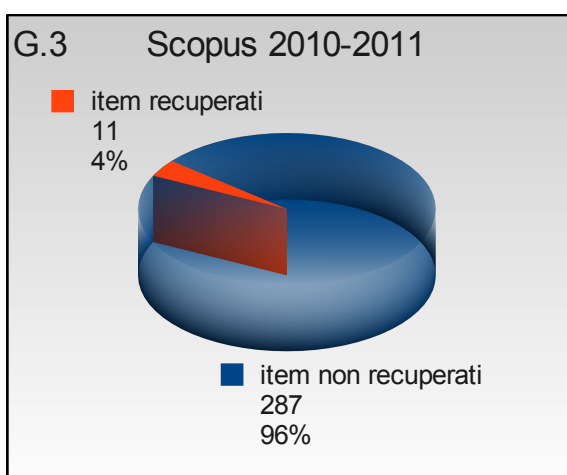
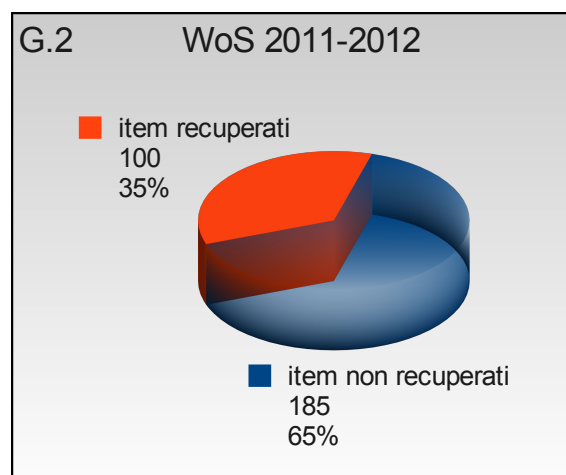
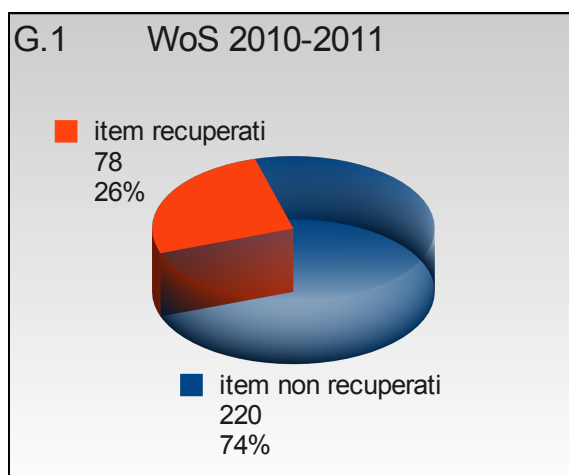
Come esercitazione finale è stato interrogato il catalogo di Ca' Foscari (<<http://polovea.sebina.it/SebinaOpac/Opac?sysb=cafoscari>>) per entrambi gli anni accademici e rilevato il numero e la percentuale di documenti recuperati.

8.4. I risultati

Numero e proporzione item recuperati e non recuperati

DB per anno	Item recuperati	Percentuale	Item non recuperati	Percentuale
WOS 10/11	78	26	220	74
SCOPUS 10/11	11	4	287	96
GS 10/11	271	91	27	9
WOS 11/12	100	35	185	65
SCOPUS 11/12	14	5	271	95
GS 11/12	220	77	65	23

Tabella 5: Copertura dei database rispetto set dati di riferimento



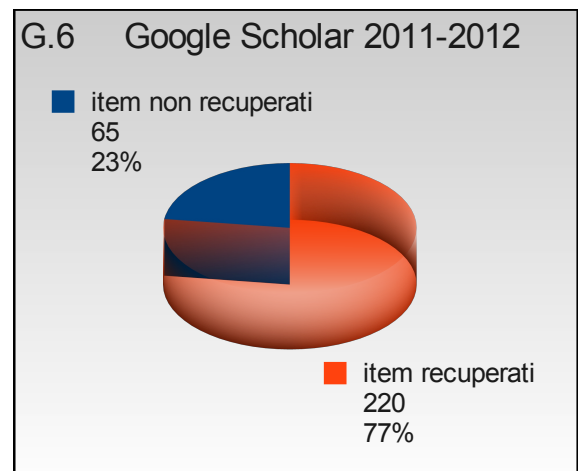
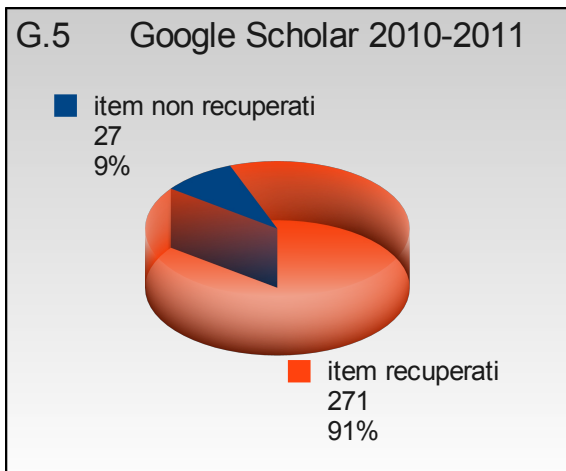


Figura 21 (grafici 1-6): Copertura dei database rispetto set dati di riferimento

La tabella e i grafici mostrano il numero e la proporzione degli *item* recuperati e non recuperati, distinti per anno accademico, interrogando i tre database.

Emerge chiaramente l'ampia copertura di GS che, anche se non completa, si attesta al 91% del totale dei documenti riferiti all'anno accademico 2010/11 e al 77% per quello del 2011/12. Scarse o scarsissime le coperture di *WoS* e *Scopus* che sono pari rispettivamente al 26% e al 4% del totale dei documenti riferiti all'anno accademico 2010/11, al 35% e al 5% per quello 2011/12.

Proporzione tra libri e riviste

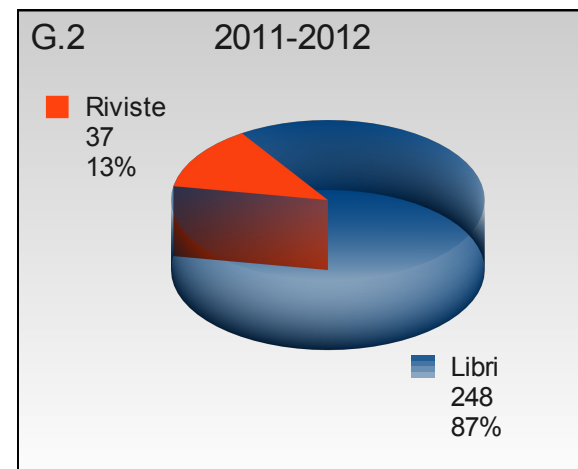
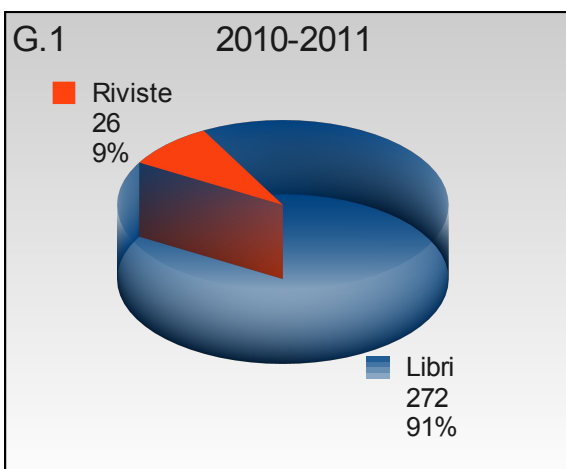
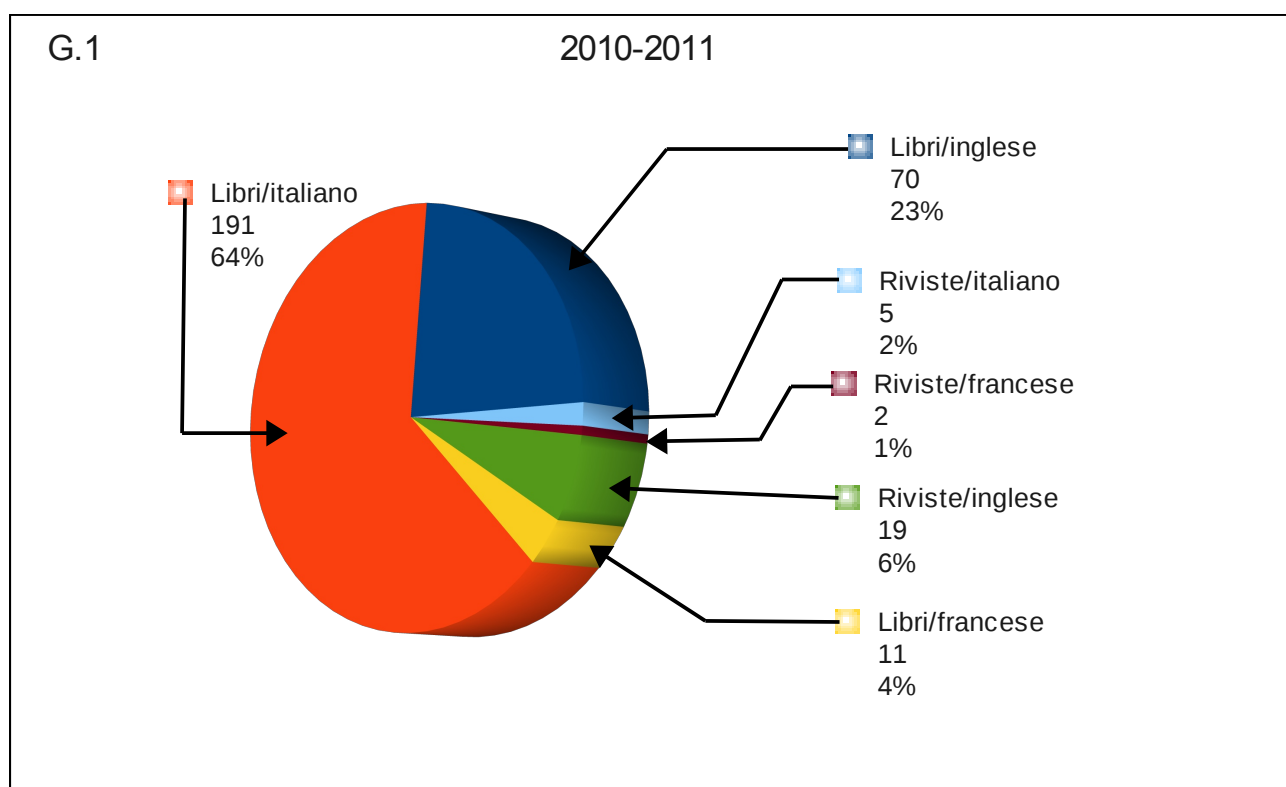


Figura 22 (grafici 1 e 2): Proporzione tra libri e riviste

I due grafici mostrano la predominanza schiacciante dei libri rispetto alle riviste che è pari al 91% del totale per l'anno accademico 2010/11 e al 87% del totale per l'anno accademico 2011/12. La proporzione conferma la pratica di pubblicazione tipica delle scienze sociali, comune anche a quelle umane, orientata più verso il libro e alla monografia specialistica che non verso la rivista.²

Suddivisione in lingue



² Numerosi studi (Cullars 1989 e 1990; Knievel – Kellsey 2005) riportano percentuali molto vicine a quelle rilevate in questo lavoro che si aggirano intorno al 75-80% per i libri e al 15-20% per le riviste.

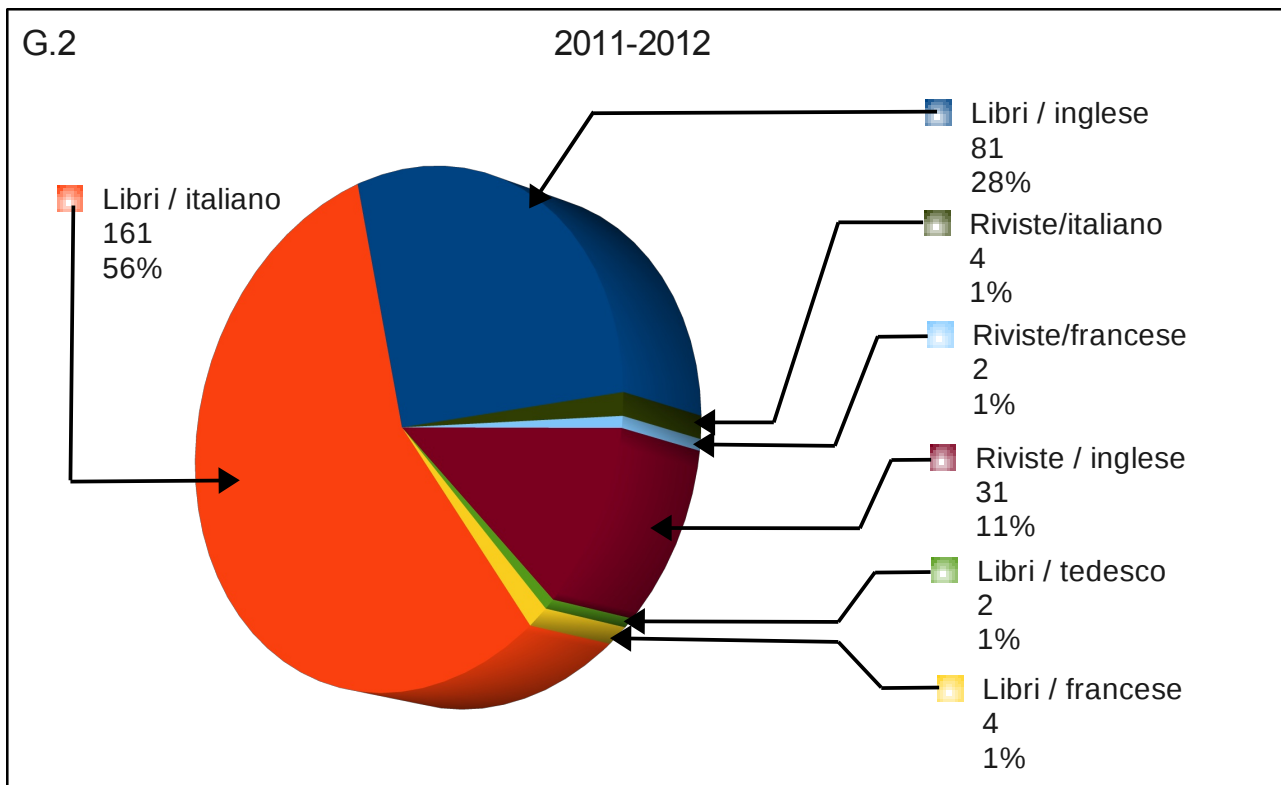


Figura 23 (grafici 1 e 2): Suddivisione in lingue

Anche se è necessario prendere in considerazione il fatto che la bibliografia d'esame proviene da un insegnamento impartito in un'università italiana, almeno in parte è confermato il carattere nazionale della produzione scientifica nelle scienze sociali: per l'anno accademico 2010/11, il 66% del totale dei documenti è composto da libri (64%) e riviste (2%) in italiano; per l'anno accademico 2011/12, il 57% del totale dei documenti è composto da libri (56%) e riviste (1%) in italiano. Guardando i grafici emergono anche altre informazioni interessanti. Se si isola il dato riferito alle riviste è evidente che sono, per entrambi gli anni accademici, le riviste in inglese, e non quelle in italiano, ad essere la maggioranza. Se infatti il totale delle riviste per l'anno accademico 2010/11 è pari al 9% del totale dei documenti, ben il 6% è composto da riviste in lingua inglese e solo il 2% in italiano; dato che non solo è confermato ma tende anche ad aumentare per l'anno accademico successivo, dove su un totale di riviste pari al 13% del totale dei documenti, l'11% è fatto di riviste in inglese e soltanto l'1% è composto da riviste in italiano.

È evidente quindi che l'autore appartenente all'ambito disciplinare delle scienze sociali e nello specifico delle scienze etnoantropologiche, pubblica generalmente i propri risultati di

ricerca in italiano utilizzando la monografia specialistica, ma è anche ipotizzabile che in alcuni casi, legati magari alla necessità di inserirsi in una comunità più ampia, quale lo è quella internazionale o, semplicemente, rendere noti più velocemente i propri risultati, scelga di pubblicare in inglese utilizzando la rivista accademica, avvicinandosi così alle pratiche di comunicazione tipiche degli autori nelle scienze naturali e ingegneristiche.

Età dei documenti

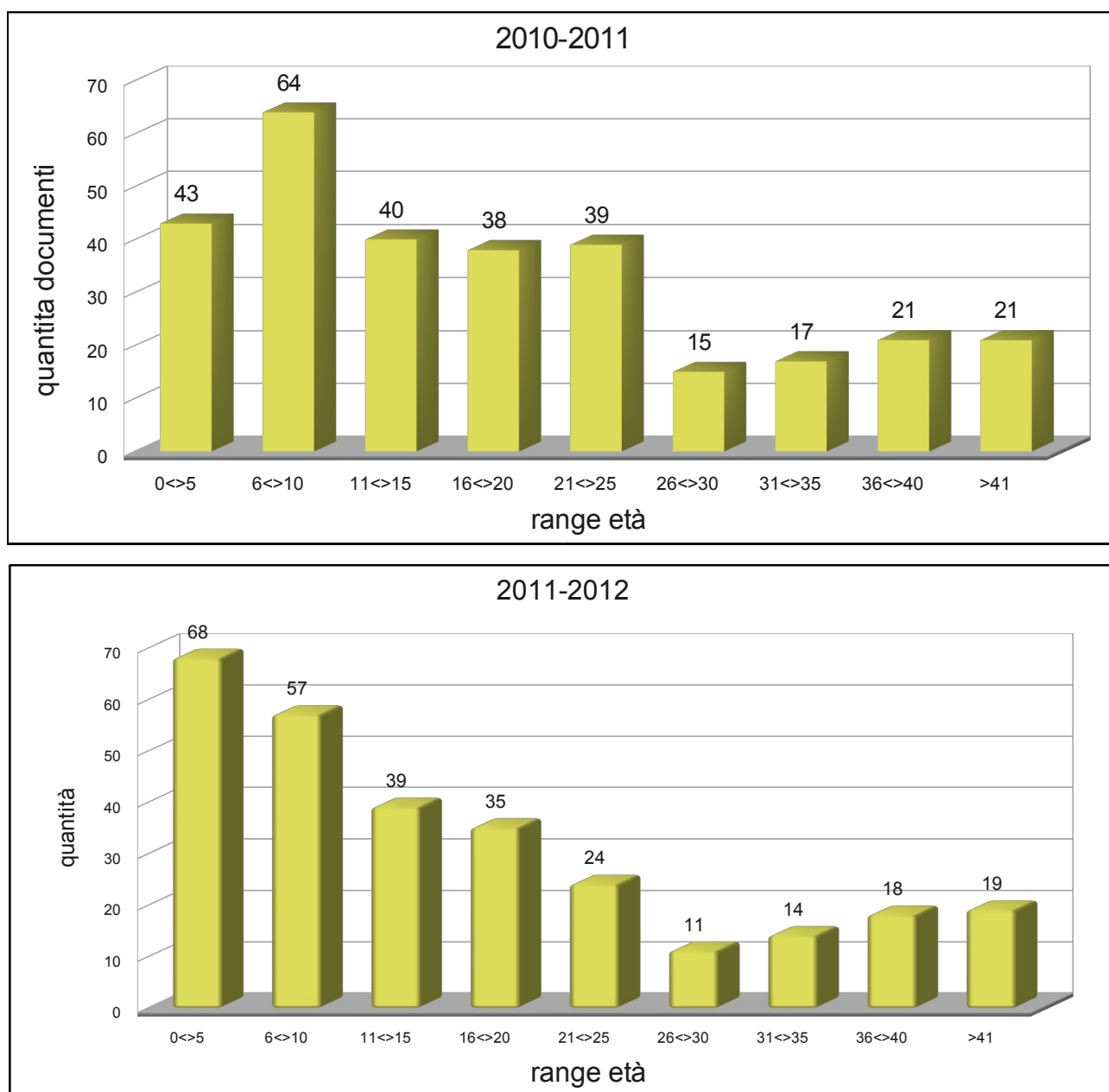


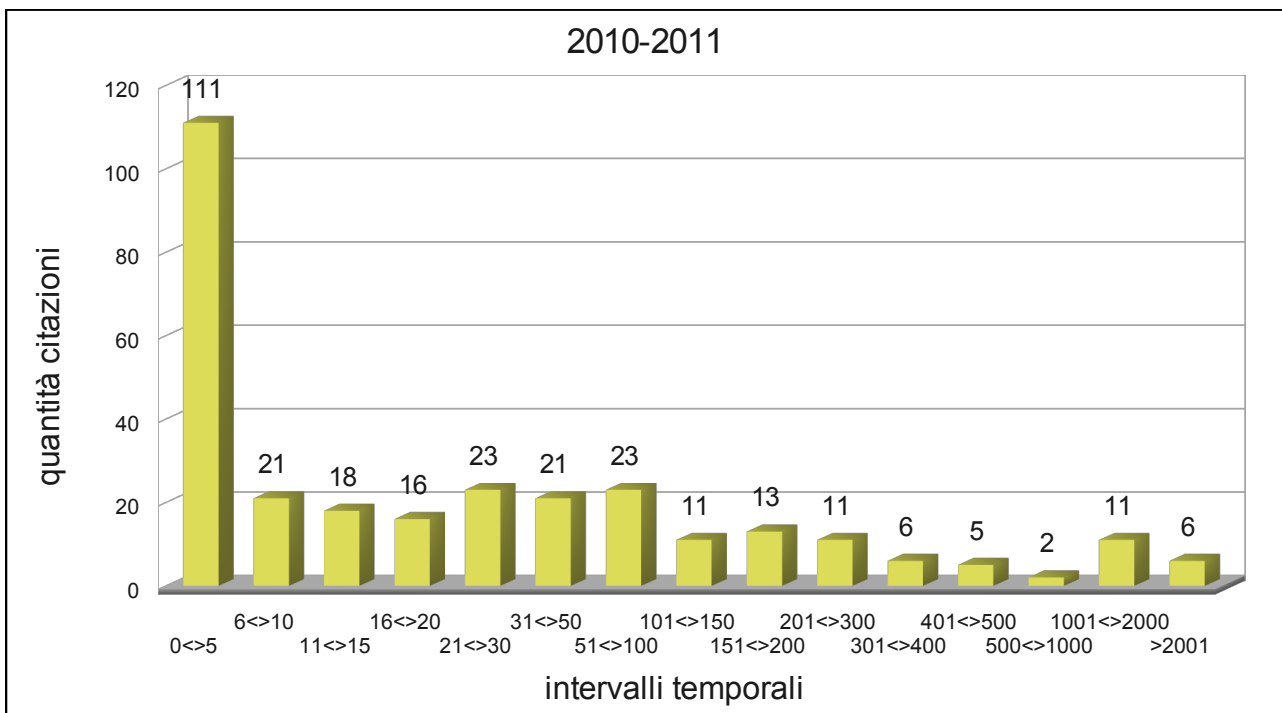
Figura 24 (2010-2011, 2011-2012): Distribuzione dei documenti per età

Su un totale di 298 documenti riferiti all'anno accademico 2010/11 solo 107 (di cui 43 con un'età da 0 a 5 anni e 64 con un'età da 6 a 10 anni), pari al 35,9% del totale, hanno un'età compresa entro i 10 anni, i restanti 191 documenti, pari al 64,1% del totale, hanno un'età che supera i 10 anni, di cui 21 documenti superano addirittura i 40 anni.

Per i documenti appartenenti all'anno accademico 2011/12 la situazione cambia di poco: su un totale di 285 documenti, 125 (di cui 68 con un'età compresa entro i 5 anni e 57 entro i 10 anni), pari al 43,8% del totale, hanno un'età inferiore o pari a 10 anni, i restanti 160 documenti, pari al 56,2% del totale, superano i 10 anni e 19 documenti (6,6%) superano addirittura i 40 anni.

I dati confermano la bassa obsolescenza delle pubblicazioni nelle scienze sociali, tipica anche delle scienze umane.

Impatto citazionale



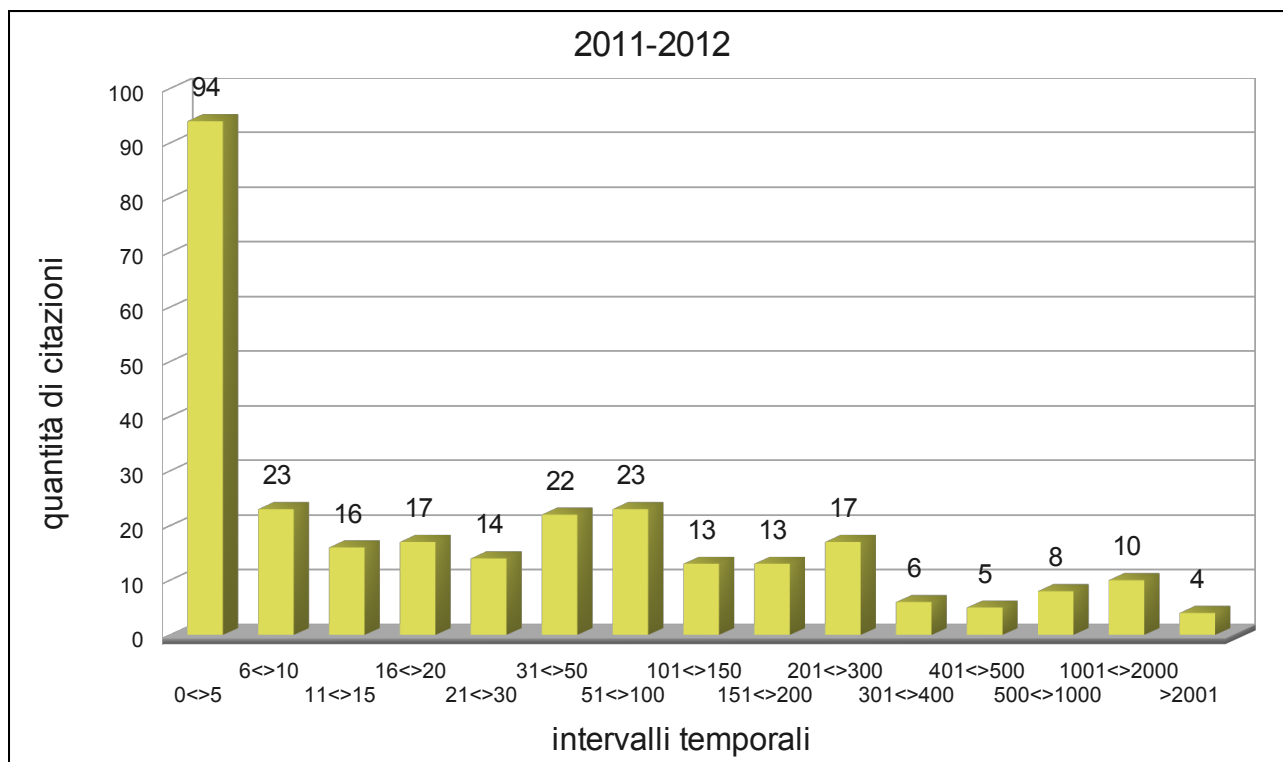


Figura 25 (2010-2011, 2011-2012): Distribuzione citazioni a intervalli temporali irregolari

I documenti oggetto d'analisi mostrano, come sintetizzato nei due grafici, un buon impatto citazionale.

Per l'anno accademico 2010/11 su 298 documenti, 111 hanno ottenuto da 0 a 5 citazioni, 99 da 6 a 50 citazioni, 47 da 51 a 200 citazioni, 22 da 201 a 500 citazioni, 13 da 501 a 2000 citazioni e 6 superano le 2000 citazioni.

Per l'anno accademico 2011/12 su 285 documenti, 94 hanno da 0 a 5 citazioni, 92 da 6 a 50 citazioni, 49 da 51 a 200 citazioni, 28 da 201 a 500 citazioni, 18 da 501 a 2000 citazioni e 4 superano le 2000 citazioni.

Citazioni per anno

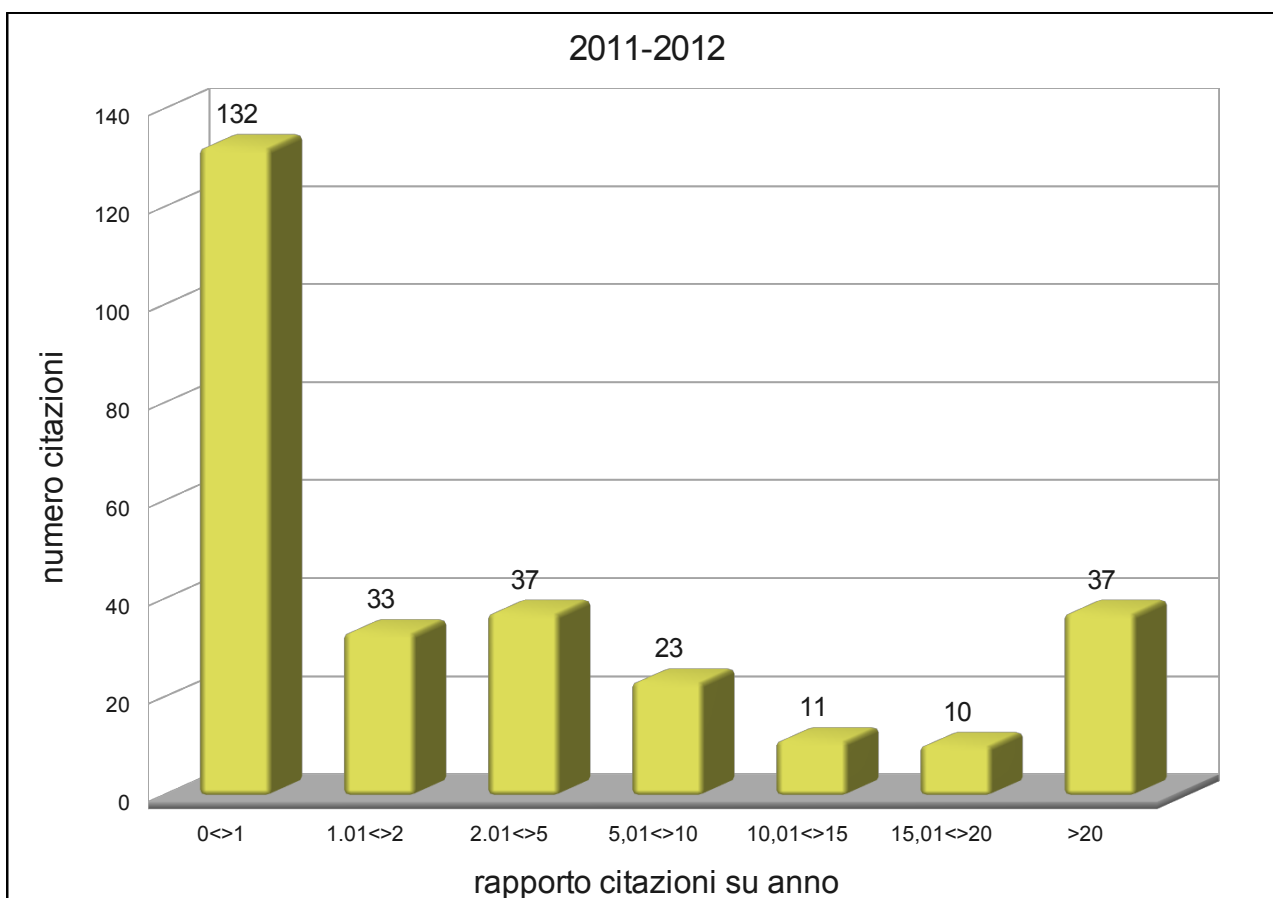
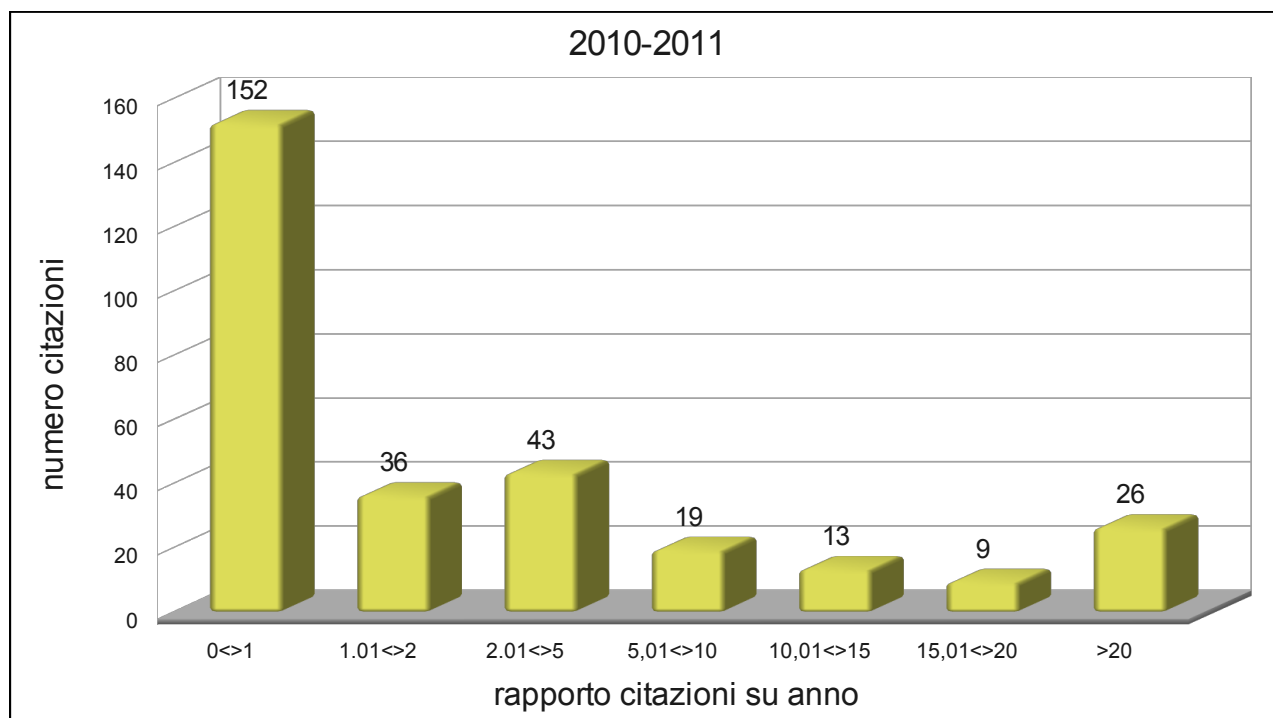


Figura 26 (2010-2011, 2011-2012): Distribuzione documenti per media citazioni su anno

Il rapporto tra citazioni e anni di pubblicazione di ogni singolo documento restituisce un quadro uniforme per entrambi gli anni accademici considerati. È alto per entrambi infatti il numero di documenti (152 per l'anno 2010/11 e 132 per l'anno 2011/12) che hanno ottenuto meno di una citazione all'anno. Il dato, se rapportato al numero molto elevato di citazioni ricevute dalla totalità dei documenti considerati, può sembrare a prima vista sorprendente, ma comprensibilissimo se poi si considera l'età dei documenti che è mediamente molto alta.

Indicatori bibliometrici per autore

Come esercizio a titolo esemplificativo dell'applicazione dei più importanti indicatori bibliometrici per autore, e di seguito per rivista, sono stati estratti dal *set* di dati di partenza i 10 autori dei documenti che hanno ottenuto il maggior numero di citazioni.

Per ogni singolo autore è stato estratto il *Citation Index* calcolato nei tre database.

	AUTORE	WOS *	WOS **	SCOPUS	GS
1	Asad, T.	475	162	28	2987
2	Carsten, J.	141	118	33	1129
3	Clifford, J.	437	174	76	2837
4	Cohen, E.	59	12	1183	1004
5	Hammersley, M.	443	312	546	10274
6	Labov, W.	454	264	84	6220
7	MacCannell, D.	53	17	54	478
8	Mahmood, S.	418	262	332	1894
9	Marcus George, E.	169	56	200	6756
10	Myers-Scotton, C.	276	70	85	2396

* WOS: somma nr. citazioni ricevute da tutte le pubblicazioni di un autore comprese quelle non indicizzate in WOS ma presenti tra i riferimenti bibliografici di altre pubblicazioni indicizzate

** WOS: somma nr. citazioni ricevute esclusivamente dalle pubblicazioni di un autore indicizzate in WOS

Tabella 6: Citation Index

Quello che sicuramente stupisce del calcolo di questo indicatore primario è la discrepanza tra i risultati provenienti dai diversi database. Per *WoS*, l'autore più citato, se si sommano le citazioni ricevute da tutte le sue pubblicazioni comprese quelle non indicizzate in *WoS* ma presenti tra i riferimenti bibliografici di altre pubblicazioni indicizzate, è Asad

(475 citazioni), ma se invece si considerano le citazioni ricevute esclusivamente dalle pubblicazioni di un autore indicizzate in *WoS*, è Hammersley (312 citazioni).

Per *Scopus* è Cohen con 1183 citazioni.

L'unico risultato che parzialmente si uniforma ai precedenti è quello di *GS* che, come *WoS*, indica Hammersley (10274 citazioni).

Per fornire una cornice valutativa più ampia è stato anche calcolato il numero di articoli di ogni singolo autore presenti nei tre database che proporzionalmente conferma i risultati per *Citation Index*:

	AUTORE	WOS	SCOPUS	GS
1	Asad, T.	24	6	60
2	Carsten, J.	11	3	26
3	Clifford, J.	18	7	127
4	Cohen, E.	16	62	85
5	Hammersley, M.	74	61	132
6	Labov, W.	14	9	87
7	MacCannell, D.	14	10	27
8	Mahmood, S.	14	11	33
9	Marcus George, E.	10	31	113
10	Myers-Scotton, C.	18	8	60

Tabella 7: Numero di articoli di ogni singolo autore presenti nei tre database

Sono stati inoltre calcolati la media delle citazioni per documento (in *WoS*, *Scopus* e *GS*) e la media delle citazioni per anni di pubblicazione (in *WoS* e *GS*):

	AUTORE	WOS	SCOPUS	GS
1	Asad, T.	6,75	4,66	49,78
2	Carsten, J.	10,73	11	43,42
3	Clifford, J.	9,83	10,85	22,34
4	Cohen, E.	0,94	19,08	11,81
5	Hammersley, M.	4,64	8,95	77,83
6	Labov, W.	19	9,33	71,49
7	MacCannell, D.	1,21	5,4	17,7
8	Mahmood, S.	18,86	30,18	57,39
9	Marcus George, E.	5,6	6,45	59,76
10	Myers-Scotton, C.	4,11	10,62	39,93

Tabella 8: Media delle citazioni per documento

	AUTORE	WOS	GS
1	Asad, T.	6,23	175,71
2	Carsten, J.	6,56	70,56
3	Clifford, J.	13,62	166,88
4	Cohen, E.	3,75	59,06
5	Hammersley, M.	22,87	604,35
6	Labov, W.	19	365,88
7	MacCannell, D.	1,7	28,12
8	Mahmood, S.	22	111,41
9	Marcus George, E.	9,33	397,41
10	Myers-Scotton, C.	5,69	140,94

Tabella 9: Media delle citazioni per anni di pubblicazione

H-Index e derivati

È stato calcolato in tutti e tre i database di riferimento l'indice h , l'indicatore capace di esprimere in un unico valore numerico sia la produttività (numero di pubblicazioni) sia l'impatto sulla comunità scientifica (numero di citazioni) riferiti al singolo ricercatore (vedi capitolo 5 p. 132 e segg.). Il risultato è stato identico in tutti i database (Hammersley: WoS 10; Scopus 14; GS 26).

	AUTORE	WOS	SCOPUS	GS
1	Asad, T.	8	3	21
2	Carsten, J.	5	2	10
3	Clifford, J.	9	2	22
4	Cohen, E.	3	8	14
5	Hammersley, M.	10	14	26
6	Labov, W.	6	5	25
7	MacCannell, D.	1	2	9
8	Mahmood, S.	6	6	12
9	Marcus George, E.	4	5	26
10	Myers-Scotton, C.	5	4	21

Tabella 10: H-Index

Seguono altri indicatori definiti varianti di *H-Index*, tutti estratti in GS (perché non sono calcolati in WoS e Scopus), che attribuiscono, se si dovesse procedere ad un *ranking*, il primo posto a Hammersley confermando in questo modo il valore già espresso dall'*H-Index* calcolato nei tre database.

	AUTORE	G-INDEX	HC-INDEX	HL,NORM
1	Asad, T.	54	14	20
2	Carsten, J.	26	8	10
3	Clifford, J.	52	16	21
4	Cohen, E.	30	10	13
5	Hammersley, M.	101	18	26
6	Labov, W.	78	15	22
7	MacCannell, D.	21	8	9
8	Mahmood, S.	33	9	12
9	Marcus George, E.	81	17	20
10	Myers-Scotton, C.	48	16	19

Tabella 11: Varianti di *H-index*

Indicatori bibliometrici per rivista

Sono state estratte dal *set* di dati di partenza, allo stesso modo degli autori, anche le 10 riviste che hanno ottenuto il maggior numero di citazioni.

Per ogni singola rivista sono stati calcolati i seguenti indicatori bibliometrici: in *WoS*, *Total Cites*, *Impact Factor*, *5-Year-Impact Factor*, *Immediacy Index*, *Cited Half-life*, *Eigenfactor Score*, *Article Influence Score*; in *Scopus*, *SJR* e *SNIP*.

	RIVISTA	Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Cited Half-life
1	<i>American Ethnologist</i>	1330	1,409	1,929	0,087	>10,0
2	<i>American Journal of Sociology</i>	10394	3,169	4,561	0,429	>10,0
3	<i>Annals of Tourism Research</i>	3646	3,259	3,356	0,183	>10,0
4	<i>Annual Review of Anthropology</i>	2112	2,553	3,342	0,1	>10,0
5	<i>Anthropological Quarterly</i>	341	0,333	0,651	0,176	9,1
6	<i>Anthropological Theory</i>	216	0,521	1,426	0,048	6,4
7	<i>Cultural Anthropology</i>	957	2,95	2933	0,281	9,3
8	<i>Journal of Material Culture</i>	231	0,793	1,267	-	7,1
9	<i>Journal of the Royal Anthropological Institute</i>	709	1,078	1,049	0,125	8,2
10	<i>Public Culture</i>	751	0,646	1,022	0,059	9,6

Tabella 12: Indicatori bibliometrici estratti da *WoS*

	RIVISTA	Eigenfactor Score	Article Influence Score
1	<i>American Ethnologist</i>	0,00406	1,052
2	<i>American Journal of Sociology</i>	0,01064	3,064
3	<i>Annals of Tourism Research</i>	0,00277	0,638
4	<i>Annual Review of Anthropology</i>	0,00411	1,964
5	<i>Anthropological Quarterly</i>	0,00095	0,348
6	<i>Anthropological Theory</i>	0,00108	-
7	<i>Cultural Anthropology</i>	0,00322	1,684
8	<i>Journal of Material Culture</i>	0,00095	0,689
9	<i>Journal of the Royal Anthropological Institute</i>	0,00263	0,582
10	<i>Public Culture</i>	0,00211	0,839

Tabella 13: Eigenfactor e Article Influence Score

	RIVISTA	SJR	SNIP
1	<i>American Ethnologist</i>	0,037	1,935
2	<i>American Journal of Sociology</i>	0,069	4,217
3	<i>Annals of Tourism Research</i>	0,04	3,218
4	<i>Annual Review of Anthropology</i>	0,161	2,076
5	<i>Anthropological Quarterly</i>	0,03	0,676
6	<i>Anthropological Theory</i>	0,029	0,866
7	<i>Cultural Anthropology</i>	0,044	3,779
8	<i>Journal of Material Culture</i>	0,03	1,116
9	<i>Journal of the Royal Anthropological Institute</i>	0,031	1,661
10	<i>Public Culture</i>	0,028	2,064

Tabella 14: Indicatori bibliometrici estratti da Scopus

Tutti gli indicatori, ad esclusione dell'*Impact Factor* e di *SJR*, attribuiscono il valore più alto alla rivista "*American Journal of Sociology*".

Numero e percentuali di recupero nel catalogo di Ca' Foscari

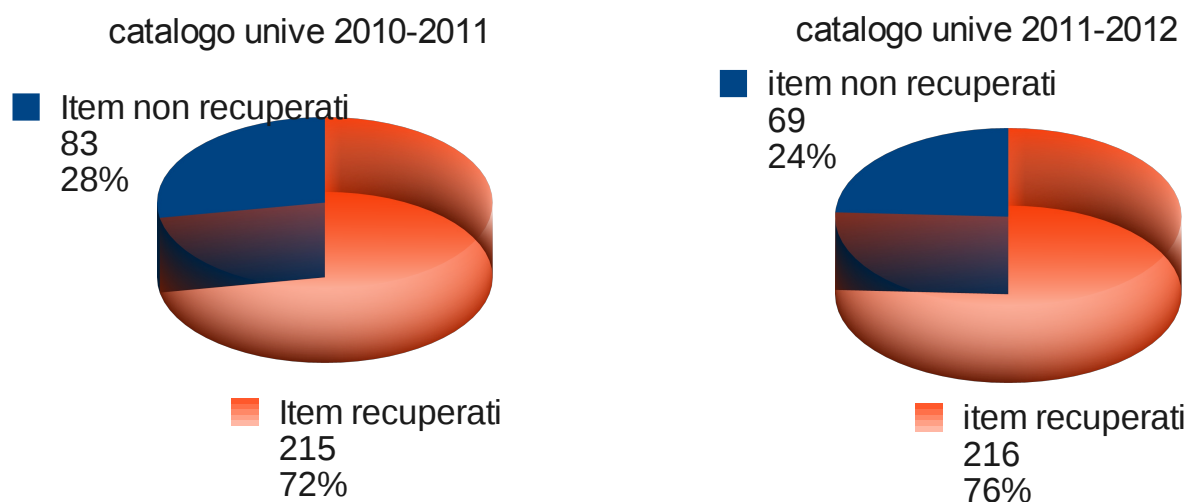


Figura 27 (2010-2011; 2011-2012): Numero e percentuali di recupero catalogo Ca' Foscari

Per l'anno accademico 2010-2011 i documenti recuperati sono 215 pari al 72% del totale; per l'anno accademico 2011-2012 i documenti recuperati sono 216 pari al 76% del totale.

8.5. Conclusioni

Riassumendo in modo molto schematico, dal lavoro di analisi precedente emerge:

- la copertura parziale degli indici citazionali rispetto al *set* di documenti oggetto di analisi;
- la copertura molto diversificata delle fonti di dati rispetto alla produzione scientifica dell'area disciplinare analizzata (scienze etnoantropologiche): molto estesa, anche se non totale in *GS*; scarsa o scarsissima in *WoS* e *Scopus*, ristretta agli articoli in lingua inglese pubblicati su riviste internazionali;
- la qualità dei dati è inversamente proporzionale al grado di copertura del database citazionale: *WoS* e *Scopus* presentano una copertura ristretta ma una elevata qualità dei dati; viceversa, *GS* presenta una copertura più ampia ma una qualità

più scarsa dei dati che infatti necessitano di una preliminare pulizia effettuata, in questo caso, attraverso l'interfaccia di interrogazione *Publish or Perish*;

- la diversa accessibilità degli indici di citazioni (gratuita per *GS*, a pagamento per *WoS* e *Scopus*) che influisce sulla piena possibilità di effettuare un confronto tra diverse rilevazioni bibliometriche.

Considerando i punti precedenti è possibile trarre le seguenti conclusioni riguardanti:

- la necessità di utilizzare più fonti di dati per la rilevazione del dato bibliometrico allo scopo di restituire il più completo quadro di valutazione e far fronte in qualche modo alle lacune riscontrate, singolarmente e nelle globalità, negli indici di citazioni;
- la scelta degli indicatori bibliometrici da utilizzare dipende tantissimo dalle finalità di valutazione e dal *set* di dati primari di riferimento che quindi devono essere considerati attentamente prima di procedere a qualsiasi tipo di analisi bibliometrica;
- la necessità sempre più stringente di avere a disposizione indici citazionali e, più in generale, strumenti bibliometrici specificatamente dedicati alle scienze sociali (e umane), capaci cioè di rendere conto delle diverse pratiche di pubblicazione e dei diversi caratteri nazionali diffusi in aree disciplinari diverse rispetto alle scienze naturali e ingegneristiche.

Conclusioni

L'esigenza di un approccio oggettivo alla valutazione della ricerca tanto sentita dai governanti, intenzionati ad ottimizzare le poche risorse a disposizione, e dagli scienziati, desiderosi, dal canto loro, di poter contare su strumenti di valutazione imparziali e davvero meritocratici, è la stessa sentita dai bibliotecari che qualche decennio fa hanno introdotto la bibliometria tra le metodologie utilizzate nella valutazione delle raccolte allo scopo di ridurre al minimo i problemi di soggettività che andavano ad influire direttamente, e spesso negativamente, sulle politiche documentarie.

La bibliometria nasce quindi sì in ambito biblioteconomico come strumento di analisi dei meccanismi di crescita e obsolescenza della letteratura, ma sarà però la sua applicazione alla valutazione della ricerca unitamente all'innovazione tecnologica ad espanderne i confini e le potenzialità anche in biblioteca, il cui ruolo all'interno del processo di comunicazione scientifica risulta notevolmente ridefinito e ampliato.

Tuttavia, le enormi potenzialità di questa metodologia sia nella valutazione delle collezioni sia in quella della ricerca possono definirsi tali solo se si conoscono i limiti e le possibili distorsioni dei metodi e degli strumenti di rilevazione del dato bibliometrico. In sostanza, per utilizzare efficacemente la bibliometria, e quindi superarne i limiti, è necessario prima averne conosciuto e capito i meccanismi e, cosa ancora più importante, avere sempre in mente che l'*output* bibliometrico non può in nessun caso venire considerato come un dato assoluto ma deve sempre essere relativizzato alla fonte dei dati su cui viene calcolato e, ai fini dell'analisi valutativa, inserito in un contesto dove trovano spazio altre metriche bibliometriche e metodologie di valutazione.

Il futuro della bibliometria corre su di un doppio binario ed interseca, e non può essere altrimenti, quello delle biblioteche.

Da una parte si assiste alla forte espansione di un filone degli studi bibliometrici che punta a correggere i problemi riscontrati negli indicatori tradizionali attraverso la creazione di metriche nuove e più raffinate applicate anche allo studio della comunicazione mediata dalla rete di *hyperlink* che popolano il World Wide Web e che molto hanno a che fare

col contesto delle biblioteche digitali poiché vengono spesso utilizzate per determinare impatto e *ranking* di documenti e riviste ad uso dell'utenza della biblioteca.

Dall'altra parte si assiste al proliferare di iniziative, di cui spesso sono responsabili le biblioteche che tra l'altro assumono anche un ruolo attivo nelle tematiche inerenti la pubblicazione *open access*, finalizzate allo sviluppo e alla gestione di depositi istituzionali, di repertori bibliografici cioè capaci di rendere conto, diversamente da quanto accade con i tradizionali strumenti di rilevazione del dato bibliometrico, delle caratteristiche peculiari di discipline come le scienze umane e sociali, e della realtà effettiva della ricerca nella singola università o nel singolo istituto.

Bibliografia

Adam, D. (2002). *The counting house*, «Nature», 415, 725–729.

Adams, J. (2005). *Early citation counts correlate with accumulated impact*, «Scientometrics», 63(3), 567–581.

Adler, R. - Ewing, J. - Taylor, P. (2008). *Citation Statistics: a report from the International Mathematical Union in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)*. Berlin, International Mathematical Union.

<<http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>>

Aguillo, I. F. - Granadino, B. - Ortega, J. L. - Prieto, J. A. (2006). *Scientific research activity and communication measured with cybermetrics indicators*, «Journal of the American Society for Information Science and Technology», 57(10), 1296–1302.

Aksnes, D. W. (2003a). *Characteristics of highly cited papers*, «Research Evaluation», 12(3), 159–170.

Aksnes, D. W. (2003b). *A macro study of self-citation*, «Scientometrics», 56(2), 235–246.

Aksnes, D. W. (2006). *Citation rates and perceptions of scientific contribution*, «Journal of the American Society for Information Science and Technology», 57(2), 169–185.

Aksnes, D. W. - Randi, T. - Randi, E. (2004). *Peer reviews and bibliometric indicators: a comparative study at a Norwegian university*, «Research Evaluation», 13(1), 33–41.

<<http://eres.library.manoa.hawaii.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=13088602&site=ehost-live>>

Almind, T. - Ingwersen, P. (1997). *Informetric analysis of the World Wide Web: Methodological approaches to "Webometrics"*, «Journal of Documentation», 53(4), 404–426.

Alonso, S. - Cabrerizo, F. J. - Herrera-Viedma, E. - Herrera, F. (2009). *H-index: A Review Focused in its Variants, Computation and Standardization for Different Scientific Fields*. «Journal of Informetrics», 3(4), 273–289.

<<http://sci2s.ugr.es/hindex/>> oppure

<<http://sci2s.ugr.es/hindex/pdf/JOI-3-4-273-289.pdf>>

Althouse, B. M. - West, J. D. - Bergstrom, T. C. - Bergstrom, C. T. (2009). *Differences in impact factor across fields and over time*. «Journal of the American Society for Information Science and Technology», 60(1), 27–34.

<<http://octavia.zoology.washington.edu/people/jevin/Documents/Althouse2008JASIST.pdf>>

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), 379–392.

Altmann, K. G. - Gorman, G. E. (1998). *The usefulness of impact factors in serial selection: A rank and mean analysis using ecology journals*, «Library Acquisitions: Practices & Theory», 22(2), 147-159.

<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1016/S0364-6408%2898%2900004-0>>

Amin, M. - Mabe, M. (2000). *Impact factors: Use and Abuse*. «Perspectives in Publishing», 1(1), 1-6, ristampato in (2003) *Medicina*, 63, 347-354, Buenos Aires.

<http://www.elsevier.com/framework_editors/pdfs/Perspectives1.pdf>

Anderson, T. R. - Hankin, R. K. - Killworth, P. D. (2008). *Beyond the Durfee square: Enhancing the h-index to score total publication output*, «Scientometrics», 76(3), 577-588.

Anon. (2002). *Errors in citation statistics*, «Nature», 415(6868), 101.

Åström, F. - Hansson, J. (2012). *How implementation of bibliometric practice affects the role of academic libraries*, «Journal of Librarianship and Information Science», 44(1), 1-22.

<<http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=2157285&fileId=2342851>>

Atkinson R. (1984). *The citation as intertext: Toward a theory of the selection process*. «Library Resources & Technical Services», 28(2), 109-119.

Baccini, A. (2010). *Valutare la ricerca scientifica: uso ed abuso degli indicatori bibliometrici*, Bologna, Il Mulino.

Bakkalbasi, N., Bauer, K., Glover, J., & Wang, L. (2006). *Three options for citation tracking: Google Scholar, Scopus and Web of Science*, «Biomedical Digital Libraries», 3, 7-14. Retrieved July 31, 2007.

<<http://www.bio-diglib.com/content/3/1/7>>

Baldazzi, A. (2002). *La biblioteca ideale nel 2045: Garfield e l'impresa della comunicazione scientifica in trasmissione d'élite*, in *Trasmissione d'élite o accesso alle conoscenze?: percorsi e contesti della documentazione e comunicazione scientifica*, a cura di Adriana Valente, Milano, Angeli, 98-129.

Baldi, S. (1998). *Normative versus social constructivist processes in the allocation of citations: A network-analytic model*, «American Sociological Review», 63(6), 829-46.

Baldock, C. - Ma, R. - Orton, C. G. (2009). *The h index is the best measure of a scientist's research productivity*, «Medical Physics», 36(4), 1043-1045.

Ball, R. - Tunger, D. (2004). *Bibliometrische Analysen – ein neues Geschäftsfeld für Bibliotheken?*, «BIT Online», 7(4), 271-278.

Ball, R. - Tunger, D. (2006). *Bibliometric analysis: A new business area for information professionals in libraries? Support for scientific research by perception and trend analysis*, «Scientometrics», 66(3), 561-577.

<<http://epub.uni-regensburg.de/5245/1/ball2.pdf>>

Banks, M. G. (2006). *An extension of the Hirsch Index: Indexing scientific topics and compounds*, «Scientometrics», 69(1), 161-168.

- Bar-Ilan, J. (2002a). *How much information do search engines disclose on the links to a web page? A longitudinal case study of the 'cybermetrics' home page*, «Journal of Information Science», 28(6), 455–466.
- Bar-Ilan, J. (2002b). *Methods for measuring search engine performance over time*, «Journal of the American Society for Information Science and Technology», 53(4), 308–319.
- Bar-Ilan, J. (2008). *Informetrics at the Beginning of the 21st Century - a Review*, «Journal of Informetrics», 2(1), 1-52.
- Bar-Ilan, J. - Peritz, B. (2002). *Informetric theories and methods for exploring the internet: an analytical survey of recent research literature*, «Library Trends», 50(3), 371–392.
- Barbieri, E. (a cura di) (2011). *La ricerca universitaria e la sua valutazione*, Rimini, Guaraldi.
- Batista, D. P. - Campitelli, M. - Kinouchi, O. - Martinez, A. S. (2006). *Is it possible to compare researchers with different scientific interests?*, «Scientometrics», 68(1), 179-189.
<<http://arxiv.org/abs/physics/0509048>>.
- Beile, P. - Boote, D. - Killingsworth, E. (2004). *A Microscope or a mirror? A question of study validity regarding the use of dissertation citation analysis for evaluating research collections*, «Journal of Academic Librarianship», 30(5), 347–353.
- Bensman, S. J. (1985). *Journal Collection Management as a Cumulative Advantage Process*, «College & Research Libraries», 46(1), 13-29.
<<http://garfield.library.upenn.edu/bensman/crl85cum.pdf>>
- Bensman, S. J. (2000). *Probability distributions in library and information science: A historical and practitioner viewpoint*, «Journal of the American Society for Information Science and Technology», 51(9), 816-833.
<<http://garfield.library.upenn.edu/bensman/jasistprobart.pdf>>
- Bensman, S. J. (2001a). *Bradford's Law and Fuzzy Sets: Statistical Implications for Library Analyses*, «International Federation of Library Association (IFLA) Journal», 27(4), 238-246.
<<http://garfield.library.upenn.edu/bensman/iflabradfordslawfuzzysets.pdf>>
- Bensman, S. J. (2001b). *Urquhart's and Garfield's Laws: The British controversy over their validity*, «Journal of the American Society for Information Science and Technology», 52(9), 714-724.
<<http://garfield.library.upenn.edu/bensman/jasisturqgar.pdf>>
- Bensman, S. J. (2005a). *Urquhart and probability: The transition from librarianship to library and information science*, «Journal of the American Society for Information Science and Technology», 56(2), 189-214.
<<http://garfield.library.upenn.edu/bensman/jasisturqprob.pdf>>
- Bensman, S. J. (2005b). *Urquhart's Law: Probability and the Management of Scientific and Journal Collections - Part 1. The Law's Initial Formulation and Statistical Bases*, «Science & Technology Libraries», 26(1), 31-68.
<<http://garfield.library.upenn.edu/bensman/urquhartlaw1.pdf>>

Bensman, S. J. (2005c). *Urquhart's Law: Probability and the Management of Scientific and Technical Journal Collections: Part 2. Probability in the Development and Management of a Central Document Delivery Collection*, «Science & Technology Libraries», 26(2), 5-32.

<<http://garfield.library.upenn.edu/bensman/urquhartlaw2.pdf>>

Bensman, S. J. (2005d). *Urquhart's Law: Probability and the Management of Scientific and Technical Journal Collections: Part 3. The Law's Final Formulation and Implications for Library Systems*, «Science & Technology Libraries», 26(2), 33-70.

<<http://garfield.library.upenn.edu/bensman/urquhartlaw3.pdf>>

Bensman, S. J. (2007). *Garfield and the Impact Factor*, «Annual Review of Information Science and Technology», 41, 93-155.

<<http://garfield.library.upenn.edu/bensman/bensmanegif2007.pdf>>

Bensman, S. J. (2008). *Distributional Differences of the Impact Factor in the Sciences Versus the Social Sciences: An Analysis of the Probabilistic Structure of the 2005 Journal Citation Reports*, «Journal of the American Society for Information Science», 59(9), 1366-1382.

<<http://garfield.library.upenn.edu/bensman/bensman072008.pdf>>

Bensman, S. J. - Wilder, S. J. (1998). *Scientific and technical serials holdings optimization in an inefficient market: A LSU Serials Redesign Project exercise*, «Library Resources & Technical Services», 42(3), 147-242.

<<http://garfield.library.upenn.edu/bensman/lrts1998oi.pdf>>

Bergstrom, C. T. (2007). *Eigenfactor: Measuring the value and prestige of scholarly journals*, «C&RL News», 68(52), 314-316.

<<http://crln.acrl.org/content/68/5/314.full.pdf>>

Bergstrom, C. T. (2010). *Use ranking to help search*, «Nature», 465, 871-872

<<http://octavia.zoology.washington.edu/publications/Bergstrom10.pdf>>

Bergstrom, C. T. - West, J. D. (2008). *Assessing Citations with the Eigenfactor™ Metrics*, «Neurology», 71, 1850-1851

<<http://octavia.zoology.washington.edu/people/jevin/Documents/Bergstrom2008Neurology.pdf>>

Bergstrom, C. T. - West, J. D. - Wiseman, M. A. (2008). *The Eigenfactor Metrics*, «Journal of Neuroscience», 28(45), 11433-11434.

<http://octavia.zoology.washington.edu/people/jevin/Documents/Bergstrom_J_neurosci_2008.pdf>

Bharat, K. - Chang, B.W. - Henzinger, M. - Ruhl, M. (2001), *Who links to whom: mining linkage between web sites*, in *Proceedings of the IEEE international conference on data mining (ICDM)*, San Jose, 51-58.

Bihui, J. (2007). *The AR-index: complementing the h-index*, «ISSI Newsletter», 3(1), 6.

<<http://sci2s.ugr.es/hindex/pdf/Jin2007.pdf>>

Bihui, J. - Liming, L. - Rousseau, R. - Egghe, L. (2007). *The R- and AR-indices: complementing the h-index*, «Chinese Science Bulletin», 52(6), 855-863.

Binfield, P. (2009). *PLoS One: background, future development, and articlelevel metrics*, North

America, ELPUB.

<<http://conferences.aepic.it/index.php/elpub/elpub2009/paper/view/114/51>>

Biolcati-Rinaldi, F. (2010). *Quali indicatori bibliometrici per le scienze sociali?*, «Working Paper 2», Dipartimento di Studi Sociali e Politici dell'Università degli Studi di Milano.

<http://www.socpol.unimi.it/papers/2010-02-09_Ferruccio%20biolcati-Rinaldi.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.

Black, S. (1999). *An assessment of social science coverage by four prominent full-text online aggregated journal packages*, «Library Collections, Acquisitions, & Technical Services», 23(4), 411–419.

<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1016/S1464-9055%2899%2900086-X>>

Bollen, J. - Luce, R. (2002). *Evaluation of digital library impact and user communities by analysis of usage patterns*, «D-Lib Magazine», 8(6).

<<http://www.dlib.org/dlib/june02/bollen/06bollen.html>>

Bollen, J. - Luce, R. - Vemulapalli, S. S. - Xu, W. (2003). *Usage analysis for the identification of research trends in digital libraries*, «D-Lib Magazine», 9(5).

<<http://www.dlib.org/dlib/may03/bollen/05bollen.html>>

Bollen, J. - Rodriguez, M.A. - Van de Sompel, H. (2007). *MESUR: usage-based metrics of scholarly impact*, *Proceedings of the Joint Conference on Digital Libraries*, Vancouver.

<http://mesur.lanl.gov/Documentation_files/JCDL07_bollen.pdf>

Bollen, J. - Van de Sompel, H. - Hagberg, A. - Chute, R. (2009). *A principal component analysis of 39 scientific impact measures*, «PLoS ONE», 4(6), e6022.

<<http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0006022>>

Bollen, J. - Van de Sompel, H. - Rodriguez, M. A. (2008). *Towards usage-based impact metrics: first results from the MESUR project*, «JCDL», *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, Pittsburgh, PA.

<<http://arxiv.org/abs/0804.3791v1>>

Bonitz, M. - Scharnhorst, A. (2001). *Competition in science and the Matthew core journals*, «Scientometrics», 51(1), 37–54.

Bookstein, A. (1976). *The Bibliometric Distributions*, «Library Quarterly», 46(4), 416-423.

Borgman, C. L. (1990). *Editor's Introduction*, in Borgman, C. L. (1990), *Scholarly Communication and Bibliometrics*, Newbury Park, Sage, 10-27.

Borgman, L. - Furner, J. (2002). *Scholarly communication and bibliometrics*, «Annual Review of Information Science and Technology», 36, 3-72.

<<http://polaris.gseis.ucla.edu/jfurner/arist02.pdf>>

<http://works.bepress.com/borgman/subject_areas.html>

Börner, K. - Chen, C. - Boyack, K. W. (2003). *Visualizing knowledge domains*, «Annual Review of

Information Science and Technology», 37, 179-255.

Bornmann, L. - Daniel, H. D. (2005). *Does the h-index for ranking of scientists really work?*, «Scientometrics», 65(3), 391-392.

Bornmann, L. - Daniel, H. D. (2009). The state of h index research. Is the *h index the ideal way to measure research performance*, «EMBO Reports», 10(1), 2-6.
<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2613214/>>

Bornmann, L. - Daniel, H. D. (2008). *What do citation counts measure? A review of studies on citing behavior*, «Journal of Documentation», 64(1), 45-80.
<<http://lutz-bornmann.de/icons/BornmannLutzCitingBehavior.pdf>>

Bornmann, L. - Mutz, R. - Daniel, H.-D. (2008). Are there better indices for *evaluation purposes than the h-index? A comparison of nine different variants of the h-index using data from biomedicine*, «Journal of the American Society for Information Science and Technology», 59(5), 830-837.
<<http://www.lutz-bornmann.de/icons/BornmannMutzDanielFinal.pdf>>

Bosman, J. - Mourik, I. - van, Rasch, M. - Sieverts, E. - Verhoeff, H. (2006). *Scopus reviewed and compared. The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar*, Utrecht, Utrecht University Library.

Bouyssou, D. - Marchant, T. (2010). *Consistent bibliometric rankings of authors and of journals*, «Journal of Informetrics», 4(3), 365-378.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1007/s11192-006-0147-4>>

Boyack, K. W. - Klavans, R. (2010). *Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?*, «Journal of the American Society for Information Science & Technology», 61(12), 2389-2404.

Bradford, S. C. (1934). *Sources of information on specific subjects*, «Engineering: An Illustrated Weekly Journal», Londra, Arthur Ernest Maw, 137(3550), ristampato (1985) *Journal Information Science*, 10, 176-180.

Bradford, S. C. (1950). *Documentation*, Washington, D.C., Public Affairs Press.

Braun, T. - Glänzel, W. - Schubert, A. (2006). *A Hirsch-type index for journals*, «Scientometrics», 69(1), 169-173.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1007/s11192-006-0147-4>>

Brennan, P. (2008). *Library assessment: Changing roles for the academic library in support of academic research evaluation*, unpublished presentation at the 2008 Library Assessment Conference. Seattle, WA.
<<http://libraryassessment.org/bm~doc/brennan.pdf>>

Brin, S. - Page, L. (1998). *The anatomy of a large-scale hypertextual Web search engine*, «Computer Networks and ISDN Systems», 30(1-7), 107-117.

Broadus, R. (1977). *The Applications of Citation Analyses to Library Collection Building*, in Voigt,

- M. J. - Harris, M. H., «Advances in Librarianship 7», New York, Academic Press, 299–335.
- Broadus, R. N. (1987a). *Early approaches to bibliometrics*, «Journal of the American Society for Information Science», 38(2), 127-129.
- Broadus, R. N. (1987b). *Toward a definition of “bibliometrics”*, «Scientometrics», 12(5-6), 373-379.
- Brodman, E. (1944). *Choosing physiology journals*, «Bulletin of the Medical Library Association», 32(4), 479–83
<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC194405/>>
- Brookes, B. C. (1970). *Obsolescence of Special Library Periodicals: Sampling Errors and Utility Contours*, «American Society for Information Science Journal», 21(5), 320-329.
- Brookes, B. C. (1977). *Theory of the Bradford Law*, «Journal of Documentation», 33(3), 180–209.
- Brookes, B. C. (1979). *The Bradford Law: A new calculus for the social sciences?*, «Journal of the American Society for Information Science», 30(4), 233–234.
- Brookes, B. C. (1984). *Ranking techniques and the empirical log law*, «Information Processing & Management», 20(1-2), 37–46.
- Brookes, B. C. (1990). *Biblio-, Sciento-, Infor-metrics? What are we talking about?* In Egghe, L. - Rousseau, R., *Informetrics 89/90. Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics*, Amsterdam, Elsevier, pp. 31-43.
- Bruno, F. (2009). *Economists in the PITS?*, «International Review of Economics», 56(4), 335-346.
<http://www.bsfrey.ch/articles/C_495_09.pdf>
- Bucchi, M. (2002). *Scienza e Società*, Bologna, Il Mulino.
- Bucchi, M. (2004). *Sociologia della scienza*, «Nuova informazione bibliografica», 3, 1-17.
<<http://www4.unitn.it/Ugcvp/it/Web/ProdottiAutore/PER0003016>>
- Buckland, M. K. - Liu, Z. (1995). *History of information science*. In Williams, M. E., «Annual review of information science and technology», 30, 385-416.
- Buckley, C. (1997). *Core conservation biology monographs: a citation analysis*, «Reference Services Review», 25(3-4), 147-158.
- Burton, R. E. - Kebler, R. W. (1960). *The half-life of some scientific and technical literature*, «American Documentation», 11(1), 18–22.
- Bush, V. (1945). *The Endless Frontier: A Report to the President*, ristampato New York, Arno Press, 1980.
- Butter, R. K. - Noyons, E. C. M. (2001). *Improving the functionality of interactive science maps*. «Scientometrics», 51, 55–68.

- Cahlik, T. (2000). *Comparison of the maps of science*, «Scientometrics», 49(3), 373–387.
- Callon, M. - Law, J. - Rip, A. (1986). *Mapping the Dynamics of Science and Technology*, *Sociology of Science in the Real World*, London, The Macmillan Press Ltd.
- Cameron, B. D. (2005). *Trends in the Usage of ISI Bibliometric Data: Uses, Abuses, and Implications*, «Portal-Libraries and the Academy», 5(1), 105-125.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1353/pla.2005.0003>>
- 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.
- Cason, H. - Lubotsky, M. (1936). *The influence and dependence of psychological journals on each other*, «Psychological Bulletin», 33, 95-103.
- Cassella, M. - Bozzarelli, O. (2011). *Nuovi scenari per la valutazione della ricerca tra indicatori bibliometrici citazionali e metriche alternative nel contesto digitale*, «Biblioteche Oggi», 29(2), 66-78.
<<http://www.bibliotecheoggi.it/content/n20110206601.pdf>>
- Cassella, M. (2006). *Le statistiche d'uso delle risorse elettroniche remote. Alcuni report a confronto*, «Biblioteche Oggi», 24(1), 48-61.
<<http://www.bibliotecheoggi.it/2006/20060104801.pdf>>
- Cassella, M. (2009). *La valutazione della ricerca nelle scienze umane*, «I Quaderni del CNBA».
<<http://hdl.handle.net/10760/14226>>
- Cassella, M. (2010). *Measure for measure: la valutazione dei depositi istituzionali nel modello della Balanced Scorecard*, «Biblioteche oggi», 28(10), 27-38.
<http://eprints.rclis.org/bitstream/10760/16205/1/IR_italiano_3ver.pdf>
- Chen, C. (1999). *Visualising semantic spaces and author co-citation networks in digital libraries*, «Information Processing and Management», 35, 401–420.
- Chen, C. M. (2003). *Mapping scientific frontiers: The quest for knowledge visualization*, London, Springer-Verlag.
- Chen, C. M. (2004). *Information visualization: Beyond the horizon*, 2nd ed., London, Springer-Verlag.
- Chongde, W. (1996). (In cinese). *On dispute in bibliometrics*, «Journal of Information, Communication, and Library Science», 2, 15-21.
- Christensen, J. O. (1992). *Cost of Chemistry Journals to One Academic Library, 1980–1990*, «Serials Review», 18, 19-34.
- Christie, A. - Laurel, K. (2001), *Developing an Online Science Journal Collection*, «Science and Technology Librarianship», 30.
<<http://www.istl.org/01-spring/article2.html>>

Civardi, M. - Maffenini, W. - Zavarrone, E. (2002). *Web based methodologies and techniques to monitor electronic resources use in university libraries*, «Proceedings ICIS 2002», The International Conference on improving surveys, Copenhagen, Denmark.

<<http://eprints.rclis.org/archive/00000265/>>

Cole, J. R. (2000). *A short history of the use of citations as a measure of the impact of scientific and scholarly work*, «The Web of Knowledge: a Festschrift in Honor of Eugene Garfield», 281-300.

Cole, J. R. - Cole, S. (1967). *Scientific output and recognition: a study in the operation of the reward system in science*, «American Sociological Review», 32(3), 377-90.

Cole, J. R. - Cole, S. (1971). *Measuring quality of sociological research – problems in use of science citation index*, American Sociologist, 6(1), 23-29.

<http://www.columbia.edu/cu/univprof/jcole/_pdf/1971ScienceCitationIndex.pdf>

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

Collins, H.M. (1999), Tantalus and the aliens: publications, audiences and the search for *gravitational waves*, «Social Studies of Science», 29(2), 163-197.

Comba, Valentina (2003). *La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives*, «Bollettino AIB», 43(1), 65-76.

<<http://eprints.rclis.org/archive/00000095/>>

Corsi, A. - Manzi, S. (2006). *Citare Internet: un repertorio di risorse in rete*, «ESB Forum», Vers. 3.0.

<<http://www.burioni.it/forum/citare.htm>>

Courtial, J. P. - Callon, M. (1984). *Is indexing trustworthy? Classification of articles through co-word analysis*, «Journal of Information Science», 9(2), 47-56.

Cronin, B. - Snyder, H. - Rosenbaum, H. et al., *Invoked on the Web*, «Journal of the American Society for Information Science», 49(14), 1319-1328.

CRUI – Commissione Biblioteche, Gruppo Open Access (2009). *L'Open Access e la valutazione dei prodotti della ricerca scientifica, Raccomandazioni*, Roma.

<http://www.sssup.it/UploadDocs/5356_OA_valutazione_1_.pdf>

Cui, L. (1999). *Rating health Web sites using the principles of citation analysis: A bibliometric approach*, «Journal of Medical Internet Research», 1(1), e4.

Cullars, J.M. (1989). *Citation characteristics of French and German literary monographs*, «The Library Quarterly», 59(4), 305-325.

Cullars, J.M. (1990). *Citation characteristics of Italian and Spanish literary monographs*, «The Library Quarterly», 60(4), 337-356.

Curti, M. - Pistotti, V. - Gabutti, G. - Klersy, C. (2001). *Impact factor and electronic versions of biomedical scientific journals*, «Haematologica», 86(10), 1015–1020.

De Bellis, N. (2005). La citazione bibliografica nell'epoca della sua riproducibilità tecnica: bibliometria e analisi delle citazioni dallo Science Citation Index alla Cybermetrica, ultima revisione 31/05/2005.

<<http://www.bibliotecheoggi.it/content/CITAZIONE.pdf>>

De Bellis, N. (2009). Materiale del seminario *Bibliometria e analisi delle citazioni*, Università Ca' Foscari, Venezia, 1/12/2009.

<lettere2.unive.it/ridi/sem091201.ppt>

De Bellis, N. (2009). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*, Lanham, MD, The Scarecrow Press.

De Robbio, A. (2007). *Analisi citazionale e indicatori bibliometrici nel modello Open Access*, «Bollettino AIB», 47(3), 257-287.

<<http://eprints.rclis.org/11999/>>

De Robbio, A. (2010). *L'Open Access come strategia per la valutazione delle produzioni intellettuali*, «CIBER 1999-2009», a cura di Paola Gargiulo e Domenico Bogliolo, Ledizioni, 104-124.

<<http://eprints.rclis.org/17742>>

Deana, D. (2008). *Gli archivi ad accesso aperto e l'impatto delle pubblicazioni: indicatori e tecniche di misurazione*, «Biblioteche Oggi», 26(1-2), 9-16.

<<http://www.bibliotecheoggi.it/content/20080100901.pdf>>

Dell'Orso, F. (2000). *Citazioni bibliografiche secondo il Chicago manual of style*, «AIB-WEB».

<<http://www.aib.it/aib/contr/dellorso1.htm>>

Dellisanti, D. (2007). *ICOLC 2006: SUSHI un nuovo strumento per la misurazione delle risorse elettroniche*, «Bollettino CILEA», 105.

<<http://bollettino.cilea.it/article/view/441/434>>

Deogan, M. S. (1987). *On-line bibliometrics*, «Lucknow Librarian», 19, 43-48.

Devin, R. B. - Kellogg, M. (1990). *The serials/monograph ratio in research libraries: Budgeting in light of citation studies*, «College & Research Libraries», 51(1), 46-54.

Di Cesare, R. (2002). *Alcune riflessioni su bibliometria e analisi delle citazioni*, in *Trasmissione d'élite o accesso alle conoscenze?: percorsi e contesti della documentazione e comunicazione scientifica*, a cura di Adriana Valente, Milano, Angeli, 131-150.

Ding, Y. - Chowdhury, G. G. - Foo, S. (2001). *Bibliometric cartography of information retrieval research by using co-word analysis*, «Information Processing and Management», 37(6), 817-842.

Di Girolamo, M. (2002). *Quelli che le risorse elettroniche... I bibliotecari delle università alle prese con la gestione e la valutazione delle collezioni elettroniche*, «AIB-WEB», Contributi.

<<http://www.aib.it/aib/contr/digirolamo3.htm>>

Di Girolamo, M. (2003). *Progetto per la realizzazione di un sistema informativo di misurazione e*

valutazione delle risorse elettroniche remote ad accesso riservato agli utenti dell'università di Milano Bicocca, Technical Report.

Dikeman, R.K. (1975). *Use of bibliometric techniques in serials management for libraries*, «Proceedings of the American Society for Information Science», 12, 55-56.

Dilevko, J. - Atkinson, E. (2002). *Evaluating academic journals without impact factors for collection management decisions*, «College & Research Libraries», 63(6), 562–657.

Duy, J. - Vaughan, L. (2003). *Usage data for electronic resources: a comparison between locally collected vendor-provided statistics*, «The Journal of academic librarianship», 29(1), 16-22.

Duy, J. - Vaughan, L. (2006). *Can electronic journal usage data replace citation data as a measure of journal use? An empirical examination*, «Journal of Academic Librarianship», 32(5), 512–517.

Eco, U. (2001). *Come si fa una tesi di laurea. Le materie umanistiche*, 12 ed., Milano, Bompiani.

Edge, D. O. (1977). *Why I am not a cocitationist*, «Society for the Social Studies of Science Newsletter», 2, 13-19.

Edwards, S. (1999). *Citation Analysis as a Collection Development Tool: a Bibliometric Study of Polymer Science Theses and Dissertations*, «Serials Review», 25(1), 11–21.

Egghe, L. (1995). *Extension of the general "success breeds success" principle to the case that items can have multiple sources*, in M.E.D. Koenig, A. Bookstein (Eds) *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*, (Rosary College, River Forest, Il., USA, 1995), 147–156, Learned Information, Inc, Medford, NJ.

Egghe, L. (2004a). *Positive reinforcement and 3-dimensional informetrics*, «Scientometrics», 60(3), 497–509.

Egghe, L. (2004b). *The source-item coverage of the Lotka function*, «Scientometrics», 61(1), 103–115.

Egghe, L. (2005a). *Expansion of the field of informetrics: Origins and consequences*, «Information Processing and Management», 41(6), 1311–1316.

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

Egghe, L. (2005c). *Zipfian and Lotkaian continuous concentration theory*, «Journal of the American Society for Information Science and Technology», 56(9), 935–945.

Egghe, L. (2006a). *How to improve the h-index*, «The Scientist», 20(3), 14.

Egghe, L. (2006b). *An improvement of the h-index: The g-index*, «ISSI Newsletter», 2(1), 8-9.
<http://pds4.egloos.com/pds/200703/08/11/g_index.pdf>

Egghe L. (2006c). *Theory and practise of the g-index*, «Scientometrics», 69(1), 131-152.
<<http://uhdSPACE.uhasselt.be/dSPACE/bitstream/1942/981/1/theory%20&%20practice.pdf>>

Egghe, L. - Ravichandra Rao, I. K. - Bhusan Sahoo, B (2006). *Proof of a conjecture of Moed and*

Garfield on authoritative references and extension to non-authoritative, «Scientometrics», 66(3), 537-549.

Egghe, L. - Rousseau, R. (1990). *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*, Amsterdam, Elsevier.
<<http://hdl.handle.net/1942/587>>

Fassoulaki, A.- Papilas, K. - Paraskeva, A. - Patris, N. (2002). Impact factor bias and proposed adjustments for its determination, «Acta Anaesthesiologica Scandinavica», 46(7), 902–905.

Fussler, H.H. (1949). *Characteristics of the research literature used by chemists and physicists in the United States*, «Library Quarterly», 19, 19-35.

Fassoulaki, A. - Paraskeva, A. - Papilas, K. - Karabinis, G. (2000). *Self-citations in six anaesthesia journals and their significance in determining the impact factor*, «British Journal of Anaesthesia», 84(2), 266–269.

Feeley, T. H. (2008). *A Bibliometric Analysis of Communication Journals from 2002 to 2005*, «Human Communication Research», 34(3), 505–520.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1111/j.1468-2958.2008.00330.x>>

Fenton, J. E. - Brazier, H. - De Souza, A. - Hughes, J. P. - McShane, D. P. (2000). *The accuracy of citation and quotation in otolaryngology/head and neck surgery journals*, «Clinical Otolaryngology and Allied Sciences», 25(1), 40–44.

Figa'-Talamanca, A. (2000). *L'Impact Factor nella valutazione della ricerca e nello sviluppo dell'editoria scientifica*. IV Seminario Sistema Informativo Nazionale Per La Matematica SINM 2000 : un modello di sistema informativo nazionale per aree disciplinari.
<<http://siba2.unile.it/sinm/4sinm/interventi/fig-talam.htm>>

Figà-Talamanca, A. (2004). *Uso e abuso delle banche dati nella valutazione delle riviste e della qualità scientifica: un problema di potere in Partecipare la scienza*, a cura di Adriana Valente e Daniela Luzi, Roma, Biblink.
<<http://www.irpps.cnr.it/it/system/files/FigaTalamancaUsoEAbuso.pdf>>

Fonseca, E. N. Da (1973). (In Portuguese). *Bibliografia Estatística e Bibliometria: Uma Reivindicacao de Prioridade*, «Ciencia da Informacao», 2(1), 5–7.
<<http://capim.ibict.br/index.php/ciinf/article/viewFile/1625/1234>>

Forgionne, G. A. - Kohli, R. (2001). *A multiple criteria assessment of decision technology system journal quality*. «Information & Management», 38(7), 421–435.

Franceschet, M. (2010). *A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar*, «Scientometrics», 83(1), 243-258.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1007/s11192-009-0021-2>>

Franceschini, F. - Maisano, D. (2010). *The Hirsch spectrum: A novel tool for analyzing scientific journals*, «Journal of Informetrics», 4(1), 64-73.

Frandsen, T. F. - Rousseau, R. (2005). *Article impact calculated over arbitrary periods*, «Journal of the American Society for Information Science and Technology», 56(1), 58–62.

- Frandsen, T. F. - Rousseau, R. - Rowlands, I. (2006). *Diffusion factors*, «Journal of Documentation», 62(1), 58–72.
- Frank, M. (2003). *Impact factors—Arbiter of excellence?* «Journal of the Medical Library Association», 91(1), 4–6.
- Fuhr, N. et al. (2001). *Digital libraries: a generic classification and evaluation scheme* [conference paper], *Proceedings di ECDL 2001*, in «LNCS», 2163, 187- 199.
- Fussler, H. H. - L. Simon, L. J. (1969). *Patterns in the use of books in large research libraries*, University of Chicago studies in library science, Chicago, University of Chicago Press.
- Galimberti, P. (2010b). *Archivi istituzionali e valutazione della ricerca: l'importanza dei dati e della loro validazione*, «Bollettino del CILEA», 116, 30-33.
- Galimberti, P. (2010a). *Verso un nuovo scenario per la valutazione della ricerca: potenzialità dell'Open Access e limiti imposti dal contesto*, «JLIS.it», 1 (1), 87-110.
<<http://leo.cilea.it/index.php/jlis/article/view/16/23>>
- Galluzzi, A. (2001). *Strumenti di valutazione per i servizi digitali. Quali strategie in un contesto ibrido?*, «Biblioteche oggi», 19(10), 6-15.
<<http://www.bibliotecheoggi.it/2001/20011000601.pdf>>
- Garretto, M. (2002). *Statistica: lezioni ed esercizi*, in Quaderni didattici del dipartimento di Matematica
<http://www.dm.unito.it/quadernididattici/garetto/quaderno_statistica.pdf>
- García-Pérez, M. A. (2010). *Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology*, «Journal of the American Society for Information Science and Technology», 61(10), 2070-2085.
<<http://www.whitelines.nl/html/google-page-rank.html>>
- Garfield, E. (1955a). *Citation Indexes – New Dimension in Documentation*, presentato al American Documentation Institute Meeting, atti del convegno tenutosi a Philadelphia, 2 – 4 Novembre, 1955.
- Garfield, E. (1955b). *Citation Indexes for Science: A new dimension in documentation through association of ideas*, «Science», 122(3159), 108-111.
- Garfield, E. (1965). *Can citation indexing be automated?* In M. E. Stevens, V. E. Giuliano, & L. B. Heilprin (Eds.), *Proceedings of the symposium on statistical association methods for mechanized documentation* (pp. 189–192). Washington, DC, National Bureau of Standards. (National Bureau of Standards Misc. Publication 269).
- Garfield, E. (1971). *The mystery of the transposed journal lists—wherein Bradford's Law of Scattering is generalized according to Garfield's Law of Concentration*, «Current Contents», 3(33), 5–6.
- Garfield, E. (1972). *Citation analysis as a tool in journal evaluation*, «Science», 178(4060), 471-

479.

<<http://www.garfield.library.upenn.edu/essays/V1p527y1962-73.pdf>>

Garfield, E. (1976a). *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>>.

Garfield, E. (1976b). *Permuterm Subject Index: Autobiographical review*, «Journal of the American Society for Information Science», 27(5-6), 288-291.

<<http://www.garfield.library.upenn.edu/essays/v3p070y1977-78.pdf>>

Garfield, E. (1979). *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*, New York, John Wiley & Sons.

Garfield, E. (1986). *Do Nobel Prize Winners Write Citation Classics?*, «Current Contents», 23(3-8), reprinted in Garfield, E., «Essays of an Information Scientist», 9, 182.

<<http://www.garfield.library.upenn.edu/essays/v9p182y1986.pdf>>

Garfield, E. (1994). *Using the impact factor*, «Current Contents», 29, 3-5.

Garfield, E. (1996). *Fortnightly review: how can impact factors be improved?*, «British Medical Journal», 313, 411-413.

Garfield, E. (2000a). *The use of JCR and JPI in measuring short and long term journal impact*, comunicazione al Council of Scientific Editors Annual Meeting, May 9, 2000.

Garfield, E. (2000b). *How Laboratory Explosions led me to document the Information Explosion*, «Chemical Heritage», 18(3), 43-45.

<<http://www.garfield.library.upenn.edu/papers/chemheritage18%283%29fall2000.html>>

Garfield, E. (2001). *From Sputnik to the World Wide Web -A retrospective view of Citation Indexing*, in *Quantum Leaps by Decade: Future Caching the Past - Forty years of creating new communities for science librarianship through collaboration*, ALA Annual Meeting, June 18, 2001.

<<http://garfield.library.upenn.edu/papers/acrl2001.html>>

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>>

Gaster, M. (1929). *Eliezer Crescas and his Beth Zebul: The Bible references in Talmud and Midrash*, «Hebrew Union College Annual», 6, 277–295.

Gerritsma, W. - Van Veller, M. - Van Zeist, C. - Van der Togt, P. - Leon, C. (2010). *Bibliometrics in the library, putting science into practice*, Book of Abstracts of the 11th International Conference on Science and Technology Indicators, Leiden.

Giesen, D. (2002). *Google PageRank calculation*, <<http://www.whitelines.nl/html/google-page-rank.html>>

Giglia, E. (2010). *I sistemi bibliotecari di Ateneo: realtà a confronto e prospettive future*, «AIDA informazioni», 28, 1-2.

<[http://www.aidainformazioni.it/pub/giglia\(2\)122010.html](http://www.aidainformazioni.it/pub/giglia(2)122010.html)>

Gilbert, N. G. (1977). *Referencing as persuasion*, «Social Studies of Science», 7(1), 113-122.

Glänzel, W. - Schubert, A. (1985). *Price distribution. An exact formulation of Price's square root law*, «Scientometrics», 7(3-6), 211-19.

Glänzel, W. - Schoepflin, U. (1999). *A bibliometric study of reference literature in the sciences and social sciences*, «Information Processing and Management», 35, 31-44.

Glänzel, W. (2003). *Bibliometrics As a Research Field. A Course on Theory and Application of Bibliometric Indicators*, Course Handouts.

<http://www.norslis.net/2004/Bib_Module_KUL.pdf>

Gnoli, C. (2000). *Le citazioni bibliografiche*, «AIB-WEB».

<<http://www.aib.it/aib/contr/gnoli2.htm>>

Godin, B. (2005). *Measurement and statistics on science and technology: 1920 to the present*, London, Routledge.

Godin, B. (2006). *On the origins of bibliometrics*. «Scientometrics», 68(1), 109-133.

Gonzalez-Pereira, B. - Guerrero, V. P. - Moyaanegón, B. F. (2009), *The SJR indicator: A new indicator of journals' scientific prestige*,

<<http://arxiv.org/ftp/arxiv/papers/0912/0912.4141.pdf>>

Gorraiz, J. - Wieland, M. (2009a). *Bibliometrie – eine neue Herausforderung für die Bibliotheken 30. Österreichischer Bibliothekartag: The next Generation*, Das Angebot der Bibliotheken, Graz, 17 September 2009.

<www.bibliothekartag.at/2009/nextgen/abstract20090916t03o01p01.htm>

Gorraiz, J. - Wieland, M. (2009b). *The end of barren disambiguation: introducing an international Standard Researcher Number (ISRN) and an International Standard Institution Number (ISIN)*.

<<http://phaidra.univie.ac.at/o:37899>>

Gorraiz, J. - Wieland, M. (2009c). *Multi-authored publications: their influence in the distribution of the financing costs in world licenses*, «Research Evaluation», 18(3), 215-220.

Gosnell, C.F. (1944). *Obsolescence of books in college libraries*, «College & Research Libraries», 5, 115-25.

Griffith, B. C. (1979). *Science literature-how faulty a mirror of science?* «Aslib Proceedings», 31, 381-391.

Griffith, B. C. - Servi, P. N. - Anker, A.L. - Drott, M. C. (1979). *The aging of scientific literature: A citation analysis*, «Journal of Documentation», 35, 179-196.

Gross, P. L. K. - Gross, E. M. (1927). *College libraries and chemical education*, «Science»,

66(1713), 385-89.

Guédon, J. C. (2004). *Per la pubblicità del sapere : i bibliotecari, i ricercatori, gli editori e il controllo dell'editoria scientifica*, traduzione dall'originale inglese di Maria Chiara Pievatolo, Brunella Casalini, Francesca Di Donato, Pisa, Plus-Pisa University Press.

Gumpenberger, C. - Wieland, M. - Gorraiz, J. (2012), *Bibliometric practices and activities at the University of Vienna*, «Library Management», 33(3), 174-183.
<<http://dx.doi.org/10.1108/01435121211217199>>

Haest, F. (2010). *Evaluation of Research Performance. Is there a role for libraries?*, WissKom2010, Jülich, 8.-10. November 2010.

Hamaker, C. - Spry, B. (2005). *Google Scholar*, «Serials», 18(2), 70-72.

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>>

Harnad, S. - Brody, T. - Vallieres, F. - Carr, L. - Hitchcock, S. - Gingras, Y., et al. (2004). *The access/impact problem and the green and gold roads to open access*, «Serials Review», 30(4), 310-314.
<<http://eprints.ecs.soton.ac.uk/15852/>>

Harzing, A. W. (2008). *Google Scholar-a new data source for citation analysis*, University of Melbourne.
<<http://perso.telecom-paristech.fr/~maitre/LTCI/bibliometrie/harzing2.pdf>>

Harzing, A.W. (2010). *The Publish or Perish Book*, Melbourne, Tarma Software Research.

Harzing, A. W. - Van der Wal, R. (2008). *Google Scholar as a new source for citation analysis?*, «Ethics in Science and Environmental Politics», 8(1), 61-73.

Harzing, A.W. - Wal, R. van der (2007). *Google Scholar: the democratization of citation analysis Ethics*, «Science and Environmental Politics», 8(1), 61-73.
<<http://www.harzing.com/download/gsdemo.pdf>>

He, Q.(1999). *Knowledge discovery through co-word analysis*, «Library Trends», 48(1), 133-159.

Henkle, H.H. (1938). *The periodical literature of biochemistry*, «Bulletin of the Medical Library Association», 27, 139-147.

Herther, N. K. (2009). *Research evaluation and citation analysis: key issues and implications*, «The Electronic Library», 27(3), 361-375.

Hertz, D. H. (1987). *History of the development of ideas in bibliometrics*, in A. Kent, (Ed.), *Encyclopedia of library and information sciences*, 42 (7), New York, Marcel Dekker, 144-219.

Hicks, D. (1987). *Limitations of co-citation analysis as a tool for science policy*, «Social Studies of Science», 17, 295-316.

Hicks, D. (2004). *The Four Literatures of Social Science*, in H. Moed, H. - Glänzel, W. - U. Schmoch (a cura di), *Handbook of Quantitative Science and Technology Research*, Dordrecht, Kluwer Academic Publishers, 473-496.

<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.10.500&rep=rep1&type=pdf>>

Hirsch, J. E. (2005). *An index to quantify an individual's scientific research output*, «Proceedings of the National Academy of Sciences of the United States of America», 102(46), 16569–16572.

<<http://arxiv.org/abs/physics/0508025>>

Holton, G. (1979). *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, 39-68.

Hönekopp, J. - Kleber, J. (2008). *Sometimes the impact factor outshines the H index*, «Retrovirology», 5, 1-3.

<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1186/1742-4690-5-88>>

Hood, W. W., & Wilson, C. S. (2001a). *The literature of bibliometrics, scientometrics, and informetrics*, «Scientometrics», 52(2), 291–314.

Hood, W. W. - Wilson, C. S. (2001b). *The scatter of documents over databases in different subject domains—How many databases are needed?*, «Journal of the American Society for Information Science and Technology», 52(14), 1242–1254.

Howard, D. (2004). *White: Reward, persuasion, and the Sokal Hoax: A study in citation identities*, «Scientometrics», 60(1), 93-120.

Hulme, E. W. (1923). *Statistical Bibliography in Relation to the Growth of Modern Civilization*, London, Grafton.

Hyman, A. (1979). *Torah hakethubah vehamessurah: A reference book of the scriptural passages quoted in Talmudic, Midrashic and early Rabbinic literature*, 2nd ed., Tel-Aviv: Dvir. (In ebraico).

Ingwersen, P. (1998). *The calculation of Web Impact Factors*, «Journal of Documentation», 54(2), 236–243.

Isidro, F. A. (2011). *Is Google Scholar useful for bibliometrics? A webometric analysis*, «Scientometrics», 91(2), 343-351.

Jascó, P. (2004a). *Web of Science citation indexes*, «Gale-Reference Review».

<<http://www.gale.com/servlet/HTMLFileServletimprint=9999@ion=7&fileName=/reference/archiv e/200408/webscience.html>>

Jascó, P. (2004b). *Savvy searching. Citation searching*, «Online Information Review», 28(6), 454-60, p. 454.

Jascó, P. (2004b). *Peter's Digital Reference Shelf- Google Scholar Beta*.
<<http://googlescholar.notlong.com/>>

Jascó, P. (2004c). *Citation searching*, «Online Information Review», 28(6), 454-460.

Jascó, P. (2005a). *Google scholar—The pros and the cons*, «Online Information Review», 29(2), 208–214.

<<http://library.gmu.edu/mudge/Reading/googlescholararticle.pdf>>

Jascó, P. (2005b). *Peter's Digital Reference Shelf- Google Scholar (Redux)*.

<<http://googlescholar2.notlong.com/>>

Jascó, P. (2005c). *As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced database*, «Current Science», 89 (9), 1537-1547.

Jascó, P. (2005d). *Comparison and Analysis of Citedness Scores in Web of Science and Google Scholar*, in Fox Edward A. et al. (editors), *Digital Libraries: Implementing Strategies and Sharing Experiences*, Berlino, Springer, 360-369.

Jascó, P. (2008a). *Google Scholar revisited*, «Online Information Review», 32(1), 102-114.

Jascó, P. (2008b). *Peter's Digital Reference Shelf- Scirus*.

<<http://www.gale.cengage.com/reference/peter/200806/scirus.htm>>

Jascó, P. (2008c). *Testing the Calculation of a Realistic h-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster*, «Library Trends», 56(4), 784-815.

Jascó, P. (2008d). *The pros and cons of computing the h-index using Google Scholar*, «Online Information Review», 32(3), 437-452.

Jascó, P. (2008e). *The pros and cons of computing the h-index using Web of Science*, «Online Information Review», 32(5), 673-688.

Jennings, C. (2001). *Citation Data: The Wrong Impact?*, «Cortex», 37(4), 585-589.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1016/S0010-9452%2808%2970601-4>>

Jewett, C. C. (1849). *Report of the Assistant Secretary relative to the library*, presented December 13, 1848, in Third annual report of the Board of Regents of the Smithsonian Institution to the Senate and House of Representatives, Washington, Tippin and Streeper.

Jiménez-Contreras, E. - Bailón-Moreno, R. - Ruiz-Baños, R., et al. (2006). *A Bibliometric Model for Journal Discarding Policy at Academic Libraries*, «Journal of the American Society for Information Science and Technology», 57(2), 198-207.

Jin, B. H. (2006). *H-index: An evaluation indicator proposed by scientist*, «Science Focus», 1(1), 8-9.

Katsaros, D. - Sidiropoulos, A. - Manolopoulos, Y. (2007). *Age Decaying H-Index for Social Network of Citations*, «Proceedings of Workshop on SocialAspects of the Web», Poznan, Poland.

Kelland, J. (1989). *The Impact of monographs in vertebrate zoology on the scientific literature: a citation analysis*, «Collection Management», 11(1–2), 77–95.

- Kelland, J. - Yonge, A. (1994). *Citation as a form of library use*, «Collection Management», 19(1-2), 81-100.
- Kent, A. et al. (1979). *Use of Library Materials: The University of Pittsburgh Study*, New York, Marcel Dekker.
- Kessler, M. M. (1963a). *Bibliographic coupling between scientific papers*, «American Documentation», 14(1), 10-25.
- Kessler, M. M. (1963b). *An experimental study of bibliographic coupling between technical papers*, «IEEE Transaction on Information Theory», 9(1), 49.
- Kim, P. J. - Lee, J. Y. - Park, J.-H. (2009). *Developing a new collection-evaluation method: Mapping and the user-side h-index*, «Journal of the American Society for Information Science and Technology», 60, 2366–2377.
- King, D. A. (2004). *The scientific impact of nations*, «Nature», 430(6997), 311–316.
- Knieval, J.E. - Kellsey, C. (2005). *Citation analysis for collection development: a comparative study of eight humanities fields*, «The Library Quarterly», 75(2), 142-68.
- Koenig, M. (1978). *Citation analysis for the arts and humanities as a collection management tool*, «Collection Management», 2(3), 247-261.
- Kousha, K. - Thelwall, M. (2007). *Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis*, «Journal of the American Society for Information Science and Technology», 58(7), 1055-1065.
- Kousha, K. - Thelwall, M. (2008). *Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines*, «Scientometrics», 74(2), 273-294.
- 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.
- Kumar, R. - Raghauam, P. - Rajagopalan, S. et al. (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>>
- Kushkowski, J. - Parson, K. - Wiese, W. (2003). *Master's Doctoral Theses Citations: Analysis and Trends of a Longitudinal Study*, «Libraries and the Academy», 3, 459–479.
- Lancaster, F. W. (1988). *If you want to evaluate your library*, Urbana-Champaign, University of Illinois, Graduate School of Library and Information Science.
- Lancaster, F. W. (1993). *If you want to evaluate your library*, 2nd ed., Urbana-Champaign, University of Illinois, Graduate School of Library and Information Science.
- Larkey, S. V. (1953). *The Army Medical Library Research Project at the Welch Medical Library*, «Bull Med Library Association», 41(1), 32–40.
<<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC199596/>>

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, edited by S. Hardin, Medford, Information Today Inc.

Latour, B. (1987). *Science In Action: How to Follow Scientists and Engineers Through Society*, Cambridge Mass., Harvard University Press.
<www.shi.uo-mo.com/images/misc/200704211104130.pdf>

Leiding, R. (2005). *Using citation checking of undergraduate honors thesis bibliographies to evaluate library collections*, «College & Research Libraries», 66, 417–429.

Levine-Clark, M. - Gil, E. L. (2009). *A comparative citation analysis of Web of Science, Scopus, and Google Scholar*, «Journal of Business and Finance Librarianship», 14(1), 32-46.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1080/08963560802176348>>

Lewis, G. (2001). *The quantity and quality of female researchers—A bibliometric study of Iceland*, «Scientometrics», 52(1), 29–43.

Leydesdorff, L. (2009). *How are new citation-based journal indicators adding to the bibliometric toolbox?*, «Journal of the American Society for Information Science and Technology», 60(7), 1327-1336.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1002/asi.21024>>

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>>

Leydesdorff, L. - Opthof, T. (2010). *Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations*, «Journal of the American Society for Information Science and Technology», 61(11), 2365-2369.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1002/asi.21371>>

Lin, X. (1998). *Map displays for information retrieval*, «Journal of the American Society for Information», 48(1), 40–54.

Li, X. (2003). *A review of the development and application of the Web impact factor*, «Online Information Review», 27(6), 407-417.
<<http://proquest.umi.com.eres.library.manoa.hawaii.edu/pqdweb?did=534325541&sid=1&Fmt=4&clientId=23440&RQT=309&VName=PQD>>

Line, M. (1978). *Rank lists based on citations and library uses as indicators of journal usage at individual libraries*, «Collection Management», 2, 13–116.

Liu, N. C. - Cheng, Y. (2005a). *The Academic Ranking of world universities—Methodologies and problems*, «Higher Education in Europe», 30(2), 127–136.

Liu, N. C. - Cheng, Y. (2005b). *Academic ranking of world universities using scientometrics—A*

comment to the "Fatal Attraction", «Scientometrics», 64(1), 101–109.

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

Lundberg, J. D. (2003). *The "omnipotent" Science Citation Index impact factor*, «Medical Journal of Australia», 178(6), 253–254.

Luther, J. (2001). *White Paper on Electronic Journal Usage Statistics*, Council on Library and information resources, 2a ed.

<<http://www.clir.org/pubs/abstract/pub94abst.html>>

MacColl, J. (2010). *Library roles in university research assessment*, «LIBER Quarterly», 20(2), 152-68.

<<http://liber.library.uu.nl/publish/articles/000498/article.pdf>>

Magrill, R. - St. Clair, G. (1990). *Undergraduate term paper citation patterns by disciplines and level of course*, «Collection Management», 12(3-4), 25–56.

Marchionini, G. (2000). *Evaluating digital libraries: a longitudinal and multifaceted view*, «Library Trends», 49(2), 304- 333.

Marshakova, I. V. (1973). *A system of document connection based on references*, «Scientific and Technical Information Serial of VINITI», 6(2), 3-8.

Martin, B. R. - Irvine, J. (1983). *Assessing basic research. Some partial indicators of scientific progress in radio astronomy*, «Research Policy», 12, 61-90.

McCain, K. W. - Bobbick, J. E. (1981). *Patterns of journal use in a department library: a citation analysis*, «Journal American Society Information Science», 32(49), 257-267.

Meadows, J. A. (1974), *Communication in science*, London, Butterworths.

Meho, L.I. - Yang, K. (2007). *A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar*, «Journal of the American Society for Information Science and Technology», 58(13), 1-21.

Menczer, F. (2004). *Lexical and semantic clustering by web links*, «Journal of the American Society for Information Science and Technology», 55(14), 1261-1269.

Merton, R. (1942). *La struttura normativa della scienza*, in *La sociologia della scienza. Indagini teoriche ed empiriche*, Milano, Franco Angeli, 1981, 349-62.

Merton, R. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*, Chigago, University of Chicago Press.

Merton, R. (1988). *The matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property*, «Isis», 79, 606-623.

<<http://www.garfield.library.upenn.edu/merton/matthewii.pdf>>.

Metitieri, F. - Ridi, R. (2008). *Biblioteche in rete: istruzioni per l'uso*, 3a edizione, Roma-Bari, Laterza, 2005.

Disponibile online (con aggiornamento a giugno 2008):

<<http://www.laterza.it/bibliotecheinrete/default.asp>>

Ming -Yueh, T. (1998). *Library journal use and citation half-life in medical science*, «Journal of the American Society for Information Science», 49(4), 1283-1292.

Ministero dell'Università e della ricerca scientifica (2009), Decreto Ministeriale 28 luglio 2009 prot. n. 89/2009, *Valutazione dei titoli e delle pubblicazioni scientifiche*.

<<http://attiministeriali.miur.it/anno-2009/luglio/dm-28072009-n-89.aspx>>

Moed, H. F. (2000). *Bibliometric indicators reflect publication and management strategies*, «Scientometrics», 47(2), 323–346.

Moed, H. F. (2002a). *Measuring China's research performance using the Science Citation Index*, «Scientometrics», 53(3), 281–296.

Moed, H. F. (2002b). *The impact-factors debate—The ISI's uses and limits*, «Nature», 415, 731–732.

Moed, H. F. (2005). *Citation analysis in research evaluation*, Dordrecht, The Netherlands, Springer.

Moed, H. F. (2008). *UK Research Assessment Exercises: Informed judgments on research quality or quantity?*, «Scientometrics», 74(1), 153-161.

<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1007/s11192-008-0108-1>>

Moed, H. F. (2009). *Measuring contextual citation impact of scientific journals*, preprint pubblicato su *ArXiv*:

<<http://arxiv.org/ftp/arxiv/papers/0911/0911.2632.pdf>>

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, 131-149.

Moed, H.F. - Garfield, E. (2004). *In basic science the percentage of authoritative references decreases as bibliographies become shorter*, «Scientometrics», 60(3), 295-303.

<<http://garfield.library.upenn.edu/papers/scientomv60i3p295y2004.pdf>>

Moed, H. F. - Glänzel, W. - Schmoch, U. (Eds.). (2004). *Handbook of quantitative science and technology research: The use of publication and patent statistics*, in *Studies of S&T Systems*, Dordrecht, London, Kluwer Academic.

Morriello, R. (2007). *L'indice di Hirsch (h-index) e altri indici citazionali dopo l'impact factor: uso nella valutazione della ricerca scientifica e nelle politiche documentarie delle biblioteche*, «Biblioteche oggi», 25(1), 23-32.

<<http://www.bibliotecheoggi.it/2007/20070102301.pdf>>

Morriello, R. - Ortigari, A. (2006). *Statistiche d'uso delle risorse elettroniche per la valutazione delle raccolte in un contesto di cooperazione*, in *Gestione delle raccolte e cooperazione nella biblioteca ibrida*, a cura di Nicola Benvenuti e Rossana Morriello, Firenze, Firenze University

Press, 37-56.

Morris, H. - Harvey, C. - Kelly, A. (2009). *Journal rankings and the ABS journal quality guide*, «Management Decision», 47(9), 1441-1451.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1108/00251740910995648>>

Morse, P. M. - Leimkuhler, F. F. (1979). *Technical Note—Exact Solution for the Bradford Distribution and Its Use in Modeling Informational Data*, «Operations Research», 27:187-198.

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

<http://www.conceptsymbols.com/narin/narin_1975_evalbibliometrics_images.pdf>

Narin, F. - Moll, J. K. (1977). *Bibliometrics*, «Annual Review of Information Science and Technology», 12, 35-58.

Nederhof, A. J. - Van Raan, A. F. J. (1987). *Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics*, «Scientometrics», 11(5-6), 333-350.

Nicholls, P. T. (1988). *Price's square root law: empirical validity and relation to Lotka's law*, «Information Processing & Management», 24(4), 469-477.

Nicholls, P. T. (1989). *Bibliometric modeling processes and the empirical validity of Lotka's law*, «Journal of the American Society for Information Science», 40(6), 379-385.

Nicolaisen, J. (2007). *Citation Analysis*, «Annual Review of Information Science and Technology», 41, 609-41.

Nisonger, T. (1994). *A Methodological Issue Concerning the Use of Social Sciences Citation Index Journal Citation Reports Impact Factor Data for Journal Ranking*, «Library Acquisitions – Practice and Theory», 18(4), 447-458.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1016/0364-6408%2894%2990052-3>>

Nisonger, T. (1998). *Management of Serials in Libraries*, Englewood, Libraries Unlimited.

Nisonger, T. E. (2004). *The benefits and drawbacks of impact factor for journal collection management in libraries*, «The Serials Librarian», 47(1-2), 57-75.

Norris, M. - C. Oppenheim (2010). *Peer review and the h-index: Two studies*, «Journal of Informetrics», 4(3), 221-232.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1016/j.joi.2009.11.001>>

Noruzi, A. (2005a). *Google Scholar: The New Generation of Citation Indexes*, «LIBRI», 55(4), 170-180.

Noruzi, A. (2005b). *Web impact factors for Iranian Universities*, «Webology», 2(1).

<<http://www.webology.org/2005/v2n1/a11.html>>

Noruzi, A. (2006). *The Web Impact Factor: A Critical Review*, «The Electronic Library», 24(4), 490-500.

<http://eprints.rclis.org/bitstream/10760/7161/1/Web_Impact_Factors,_A_critical_review.pdf>

Noyons, E. C. M. (2001). *Bibliometric mapping of science in a science policy context*, «Scientometrics», 50(1), 83–98.

Noyons, E. C. M. (2004). *Science maps within a science policy context*, in Moed, H., Glänzel, W., Schmoch, U., *Handbook of Quantitative Science and Technology Research, The use of Publication and Patent Statistics*, in *Studies of S&T Systems*, Kluwer Academic Publishers, Dordrecht, 237–255.

Noyons, E. C. M. - Luwel, M. - Moed, H. F. (1999). *Combining mapping and citation analysis for evaluative bibliometric purposes*, «Journal of the American Society for Information Science», 50(3), 115–131.

Oppenheim, C. (2008). *Out With the Old and in With the New: the RAE, Bibliometrics and the New REF*, «Journal of Librarianship and Information Science», 40(3), 47-149.
<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1177/0961000608092550>>

Organisation for Economic Co-operation and Development (OECD)(1963). *The Measurement of Scientific and Technical Activities: 'Frascati Manual'*, Paris, OECD.

Osareh, F. (1996). *Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature I*, «Libri», 46(3), 149-158.

Osburn, C. B. (1983). *Toward a re-conceptualization of collection development*, in G. McCabe, & B. Kreissman, *Advances in library administration and organization*, Greenwich, JAI Press, 175-198.

Otlet, P. (1934), *Traité de documentation*, disponibile online:
<http://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf>

Pan, E. (1978). *Journal citation as a predictor of journal usage in libraries*, «Collection Management», 2, 29–35.

Parise, S. (2008). *La formazione delle raccolte nelle biblioteche pubbliche: dall'analisi dei bisogni allo sviluppo delle collezioni*, Milano, Bibliografica.

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>>

Parker Ladwig, J. - Sommese, A. J. (2005). *Using cited half-life to adjust download statistics*, «College & Research Libraries», 66(6), 527-542.

Peritz, B. C. - Bar-Ilan, J. (2002). *The sources used by bibliometrics–scientometrics as reflected in references*, «Scientometrics», 54(2), 269–284.

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), 235-255.

Peters, H.P.F. - Van Raan, A.F.J. (1994). *On determinants of citation scores – a case study in chemical engineering*, «Journal of the American Society for Information Science», 45(1), 39-49.

- Piazzini, T. (2010). *Gli indicatori bibliometrici: riflessioni sparse per un uso attento e consapevole*, «JLIS.it», 1(1), 63-86.
<<http://leo.cilea.it/index.php/jlis/article/viewFile/24/38>>
- Pichappan, P. - Sarasvady, S. (2002). *The influence of author self-citations on bibliometric macro indicators*, «Scientometrics», 54(2), 285–290.
- Pierce, S. J. (1992). *On the origin and meaning of bibliometric indicators: journals in the social sciences, 1886-1985*, «Journal of the American Society for Information Science», 43(7), 477–487.
- Pitkin, R. M. - Branagan, M. A. - Burmeister, L. (1999). *Accuracy of data in abstracts of published research articles*, «Journal of the American Medical Association», 281(12), 1110-1111.
<<http://psychrights.org/research/Digest/Science4Sale/AbstractInnaccuracy%28PitkinJAMA1999%29.pdf>>
- Polanco, X. - Boudourides, M. A. - Besagni, D. et al. (2001), *Clustering and mapping web sites. For displaying implicit associations and visualizing networks*, Working paper.
<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.81.9795&rep=rep1&type=pdf>>
- Pomerantz, J. - Mon, L. - McClure, C. R. (2008). *Evaluating remote reference service: a practical guide to problems and solutions*, «Portal: libraries and the academy», 8(1), 15- 30.
- Popov, S. B. (2005). *A parameter to quantify dynamics of a researcher's scientific activity*, pubblicato su *ArXiv*: physics/0508113, <<http://arxiv.org/abs/physics/0508113>>.
- Portal, S. G. (1995), *Principios teoricos y metodologicos de los estudios metricos*, «Investigacion Bibliotecologica: Archivonomia, Bibliotecologia, e Informacion», 8, 23–32.
- Price, D. J. d. S. (1951). *Quantitative measures of the development of science*, «Archives Internationales d'Histoire des Sciences», 14, 85-93.
<<http://garfield.library.upenn.edu/price/pricequantitativemeasures1951.pdf>>
- Price, D. J. d. S. (1963). *Little Science, Big Science*, New York, Columbia University Press, traduzione italiana di questa prima edizione: Price D.J.d.S. (1967). *Sociologia della creatività scientifica*, Milano, Bompiani, 1967.
- Price, D. J. d. S. (1965). *The exponential curve of science*, «Discovery», 6, 240 – 243.
- Price, D. J. d. S. (1965). *Networks of scientific papers*, «Science», 149, 510-515.
<<http://garfield.library.upenn.edu/papers/pricenetworks1965.pdf>>
- 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(5), 292-306.
- Prime, C. - Bassecoulard, E. - Zitt, M. (2002). *Co-citations and co-sitations—A cautionary view on an analogy*, «Scientometrics», 54(2), 291–308.
- Pritchard, A. (1969). *Statistical bibliography or bibliometrics?*, «Journal of Documentation», 25(4), 348-49.

Rabinowitz, R. N. (1952). *Ma'amar D al hadpasat ha-Talmud: Toldot hadpasat ha-Talmud*. Edited, with corrections, addenda, and indexes by A. M. Habermann, Jerusalem, Mosad ha-Rav Kuk. (In ebraico).

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

Ravichandra Rao, I. K. (1983). *Quantitative Methods for Library and Information Science*, New Delhi, Wiley Eastern Ltd.

Reale, E. (2008). *La valutazione della ricerca pubblica: un'analisi della valutazione triennale della ricerca*, a cura di Emanuela Reale, Milano, Franco Angeli.

Regolamento concernente la struttura ed il funzionamento dell'Agenzia nazionale di valutazione del sistema universitario e della ricerca (ANVUR) (2010), Decreto del Presidente della Repubblica 1 febbraio 2010, n.76.

<<http://anvur.miur.it/documenti/Gazzetta-ANVUR.pdf>>

Revelli, C. (2010). *Citazione bibliografica*, Roma, Associazione Italiana Biblioteche.

Rice, B. A. (1983). *Selection and evaluation of chemistry periodicals. Role of Serials in Sci-Tech Libraries*, New York, 43-59.

Rider, F. (1944). *The scholar and the future of the research library*, New York, Hadham Press.

Ridi, R. (1995). *Citare Internet*, «Bollettino AIB», 35(2), 211-220.

<<http://www.aib.it/aib/boll/1995/95-2-211.htm>>

Ridi, R. (1998). *Ricerca e selezione delle fonti di informazione*, Roma, Ministero per i beni culturali e ambientali - RAI - Italia lavoro – AIB.

<<http://www.educational.rai.it/mediateche/mediateca.asp>>

Ridi, R. (2006). *Citare Internet: tradizioni da confermare e miti da sfatare*, «Bollettino AIB», 46(3), 247-253.

<<http://www.aib.it/aib/boll/2006/0603247.htm>>

Ridi, R. (2007). *La biblioteca come ipertesto: verso l'integrazione dei servizi e dei documenti*, Milano, Editrice bibliografica.

Ridi, R. (2010). *Bibliometria: una introduzione*, «Bibliotime», 12(1).

<<http://spbo.unibo.it/bibliotime/num-xiii-1/ridi.htm>>

Riolo, M. (2001). *Le risorse elettroniche di un sistema bibliotecario: analisi e monitoraggio del loro utilizzo*, Tesi di Diploma Universitario, Facoltà di Economia, Università degli Studi di Milano-Bicocca.

Roberts, M. - MacRoberts, B. (1989). *Problems of citations analysis: a critical review*, «Journal of the American Society for Information Science», 40(5), 342–349.

Rodríguez i Gairín, J.M (1997). *Valorando el impacto de la información en Internet: AltaVista, el*

"Citation Index" de la Red, «Revista Española De Documentación Científica», 20(2), 175-181.

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), 949-958.

Roy, R. P. M. (1980). *Citation analysis: a new tool for the modern librarian*, «IASLIC Bulletin», 25, 109–116.

Saad, G. (2006). *Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively*, «Scientometrics», 69(1), 117-120.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1007/s11192-006-0142-9>>

Saha, S. - Saint, S. - Christakis, D. A. (2003). *Impact factor—A valid measure of journal quality?*, «Journal of the Medical Library Association», 91(1), 41–46.

Saracevic, T. (2000). *Digital library evaluation: toward an evolution of concepts*, «Library Trends», 49(3), 350- 369.

Schmidmaier, D. (1984). *Zur Geschichte der Bibliometrie*, «Zentralblatt für Bibliothekswesen», 98, 404–406.

Schneider, J. W. (2004). *Verification of bibliometric methods' applicability for thesaurus construction*, Aalborg, Department of Information Studies, Royal School of Library and Information Science.

<<http://biblis.db.dk/uhtbin/hyperion.exe/db.jessch04>>

Schneider, J. W. - Borlund, P. (2002). *Preliminary study of the potentiality of bibliometric methods for the construction of thesauri*, In H. Bruce, R. Fidel, P. Ingwersen & P. Vakkari, *Proceedings of the Fourth International Conference on Conceptions of Library and Information Science*, Seattle, Greenwood Village, Libraries Unlimited.

Schneider, J. W. - Borlund, P. (2004). *Introduction to bibliometrics for construction and maintenance of thesauri: methodical considerations*, «Journal of Documentation», 60(5), 524-549.

Schoepflin, U. - Glänzel, W. (2001). *Two decades of "Scientometrics"—An interdisciplinary field represented by its leading journal*, «Scientometrics», 50(2), 301–312.

Schrader, A. M. (1984). *In search of a name: information science and its conceptual antecedents*, «Library and Information Science Research», 6, 227–271.

Schultz, C. K. (1968). *H[ans] P[eter] Luhn: pioneer of information science; selected works*, Spartan Books, New York.

Scimago Grupo (2007), *SCImago journal & country rank: un nuevo portal, dos nuevos rankings*, «El profesional de la Información», 16(6), 645–646.

<<http://www.elprofesionaldeinformacion.com/contenidos/2007/noviembre/11.pdf>>

Seglen, O. (1997). *Why the Impact Factor of Journals should not be used for evaluating Research*, «British Medical Journal», 314, 498-502.

<<http://www.bmj.com.eres.library.manoa.hawaii.edu/content/314/7079/497.1.full?sid=b6d7eb50-42e0-49f6-b859-4b3e56be1fad>>

Serrai, A. (1984). *Dai "loci communes" alla bibliometria*, Roma, Bulzoni.

Serrai, A. (1985). *Validità e utilità delle descrizioni bibliometriche: la regola di Trueswell*, (Schegge; 21), «Il bibliotecario», 6, 95-99.

Serrai, A. (2010). *La citazione bibliografica, in Natura elementi e origine della bibliografia in quanto mappa del sapere e delle lettere*, Bulzoni, 31-33.

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), 337-339.

Shaw, R. R. (1951). *Machines and the Bibliographical Problems of the Twentieth Century*, Urbana, University of Illinois Press.

Shepard, F. (1948). *How to use Shepard's Citations*, Colorado Springs, CO: Shepard's Citations.

Shera, J. H. - Egan, M. E. (1951). *Bibliographic organization*, Chicago, University of Chicago Press.

Sidiropoulos, A. - Katsaros, D. - Manolopoulos, Y. (2006). *Generalized Hirsch h-index for disclosing latent facts in citation networks*, «Scientometrics», 72(2), 253-280.

<<http://arxiv.org/pdf/cs/0607066v1.pdf>>

Simkin, M. V. - Roychowdhury, V. P. (2003). *Read before you cite!*, «Complex Systems», 14(3), 269-274. <<http://www.complex-systems.com/pdf/14-3-5.pdf>>

Singh, G. - Rekha, M. - Moin, A. (2007). *A bibliometric study of literature on digital libraries*, «The Electronic Library», 25(3), 342-348.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1108/02640470710754841>>

Slote, S. J. (1989). *Weeding the collection III*, 3rd rev., Englewood, Libraries Unlimited.

Small, H. (1973). *Co-citation in the scientific literature: A new measurement of the relationship between two documents*, «Journal of the American Society of Information Science», 24(4), 265-269.

Small, H. (1978). *Cited documents as concept symbols*, «Social Studies of Science», 8(3), 327-40.

<<http://www.jstor.org/stable/284908>>

Small, H. (1999). *Visualizing science by citation mapping*, «Journal of the American Society for Information Science», 50 (9), 799-813.

Small, H. (2003). *Paradigms, citations, and maps of science: A personal history*, «Journal of the American Society for Information Science and Technology», 54(5), 394-399.

Small, H. (2004). *On the shoulders of Robert Merton: towards a normative theory of citation*, «Scientometrics», 60(1), 71-79.

Small, H. - Griffith, B. C. (1974). *The structure of scientific literature. I: Identifying and graphing specialties*, «Science Studies», 4, 17-40.

Smith, L. C. (1981). *Citation Analysis*, «Library Trends», 30(1), 83–106.

Spinak, E. (1998). *Indicadores cientometricos*, «Ciência da informação», 27(2), 141-48.

Stankus, T. (1992). *Making Sense of Journals of the Life Sciences: From Speciality Origins to Contemporary Assortment*, New York, Haworth Press.

Stankus, T. - Mills, C.V. (1992). *Which Life Science Journals Will Constitute the Locally Sustainable Core Collection of the 1990s and Which Will Become 'Fax-Access' Only? Predictions Based on Citation and Price Patterns, 1979–1989*, «Science and Technology Libraries», 13, 73–114.

Sylvia, M. J. (1998). *Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies*, «Collection Building», 17(1), 20 – 28.

Száva-Kováts, E. (2002). *Unfounded attribution of the "half-life" index-number of literature obsolescence to Burton and Kebler: a literature science study*, «Journal of the American Society for Information Science and Technology», 53(13), 1098-1105.

Tague-Sutcliffe, J. M. (1994). *Quantitative methods in documentation*, in B. C. Vickery, (Ed.), *Fifty years of information progress: a journal of documentation review*, «ASLIB», London, England, 147–188.

Tague-Sutcliffe, J. - Beheshti, L. - Reespotter, L. (1981). *The law of exponential growth. Evidence, implications and forecasts*, «Library Trends», 30(1), 125-145.

Tammaro, A. M. (2000). *Misurazione e valutazione della biblioteca digitale*, «Biblioteche Oggi», 18(1), 66-70.

<<http://www.bibliotecheoggi.it/2000/20000106601.pdf>>

Tammaro, A. M. (2001a). *Qualità della comunicazione scientifica. 1: Gli inganni dell'impact factor e l'alternativa della biblioteca digitale*, «Biblioteche oggi», 19(7), 104-107.

<<http://www.bibliotecheoggi.it/2001/20010710401.pdf>>

Tammaro, A. M. (2001b). *L'alternativa all'impact factor*, «Biblioteche oggi», 19(8), 74-78.

<<http://www.bibliotecheoggi.it/2001/20010807401.pdf>>

Tammaro, A. M. (2004). *Indicatori di qualità delle pubblicazioni scientifiche e open access*, in *Partecipare la scienza*, a cura di Adriana Valente e Daniela Luzi, Roma, Biblink, 51-88.

Tarantino, E. (2006). *Troppo o troppo poco? Web of science, Scopus, Google scholar: tre database a confronto (un caso di studio)*, «Bollettino AIB», 46(1-2), 23-32.

<<http://www.aib.it/aib/boll/2006/0601023.htm>>

Tenopir, C. (2005). *Google in the Academic Library*, «Library Journal», 130(2), 32.

<<http://www.libraryjournal.com/article/CA498868.html>>

Thelwall, M. (2000). *Web impact factors and search engine coverage*, «Journal of Documentation», 56(2), 185–189.

Thelwall, M. (2001a). *Extracting macroscopic information from Web links*, «Journal of the American Society for Information Science and Technology», 52(13), 1157–1168.

Thelwall, M. (2001b). *A web crawler design for data mining*, «Journal of Information Science», 27(5), 319–325.

Thelwall, M. (2001c). *Results from a web impact factor crawler*, «Journal of Documentation», 57(2), 177–191.

Thelwall, M. (2001d). *The responsiveness of search engine indexes*, «Cybermetrics», 5(1), paper 1, Retrieved July 14, 2007.

<http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html>

Thelwall, M. (2001e). *Exploring the link structure of the Web with network diagrams*, «Journal of Information Science», 27(6), 393–401.

Thelwall, M. (2002a). *A comparison of sources of links for academic Web impact factor calculations*, «Journal of Documentation», 58(1), 66–78.

Thelwall, M. (2002b). *Conceptualizing 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.

Thelwall, M. (2002c). *Methodologies for crawler based Web surveys*, «Internet Research—Electronic Networking Applications and Policy», 12(2), 124–138.

Thelwall, M. (2002d). *Evidence for the existence of geographic trends in university Web site interlinking*, «Journal of Documentation», 58(5), 563–574.

Thelwall, M. (2002e). *An initial exploration of the link relationship between UK university Web sites*, «ASLIB Proceedings», 54(2), 118–126.

Thelwall, M. (2002f). *The top 100 linked-to pages on UK university web sites—High inlink counts are not usually associated with quality scholarly content*, «Journal of Information Science», 28(6), 483–491.

Thelwall, M. (2004). *Link Analysis. An Information Science Approach*, Amsterdam, Elsevier.

Thompson, J. W. (2002). *The death of the scholarly monograph in the humanities? Citation patterns in literary scholarship*, «LIBRI», 52(3), 121–136.

Tonta, Y. - Yurdagül, Ü. (2005). *Scatter of journals and literature obsolescence reflected in document delivery requests*, «Journal of the American Society for Information Science and Technology», 56(1), 84–94.

Tonta, Y. - Yurdagül, Ü. (2010). *Does Urquhart's Law hold for consortial use of electronic journals?*, «Scientometrics», 83, 793–808.

Torres-Salinas, D. - Moed, H. F. (2009). *Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics*, «Journal of informetrics», 3(1), 9-26.

<http://ec3.ugr.es/publicaciones/TorresSalinas_D_y_Moed_HF_Library_Catalog_Analysis_as_a_tool_in_studies_of_social_sciences_and_humanities_An_exploratory_study_of_published_book_titles_in_Economics.pdf>

Tosi P. (1999). *Come valutare la ricerca Giudizio dei pari o indicatori numerici? Quali sono i parametri più idonei per valutare "i prodotti intellettuali"*, in Working paper Unitn, n. 10, Università di Trento.

Trayhurn, G. (2002). *Citations and 'impact factor'—The Holy Grail*, «British Journal of Nutrition», 88(1), 1-2.

Trueswell, R. L. (1969). *Some behaviorial patterns of library users: the 80/20 rule*, «Wilson Library Bulletin», 43(5), 458-461.

Ugolini, D. - Casilli, C. (2003). *The visibility of Italian journals*, «Scientometrics», 56(3), 345-355.

Urquhart, D. J. (1959). *Use of scientific periodicals*, in *Proceedings of the international conference on scientific information*, Washington, DC, November 16-21, 1958 (Vol. I, pp. 287-300), Washington, National Academy of Sciences, National Research Council. Retrieved.

<http://books.nap.edu/openbook.php?record_id=10866&page=287>

Urquhart, D.J. (1977). *Urquhart's Law*, «Journal of Documentation», 33, 149.

Urquhart, D. J. - Bunn, R. M. (1959). *A national loan policy for scientific serials*, «Journal of Documentation», 15, 21-37.

Urquhart, J. A. - Urquhart, N. C. (1976). *Relegation and stock control in libraries*, Stocksfield, Oriel Press.

Usberti, M.(2002). *La citazione bibliografica delle risorse elettroniche remote*, Università degli studi di Parma.

<<http://www.burioni.it/forum/usb-cit0.htm>>

Usberti, M. (2010). *Suggerimenti per l'impaginazione della tesi e per la citazione delle opere*, Università degli studi di Parma, 2010.

<<http://www.bibliotecapagnegro.unipr.it/help-desk/Bibliografiatesi.html>>

Vaidya, J. S. (2005). *V-index: A fairer index to quantify an individual's research output capacity*, «British Medical Journal», 331, 1339-c-1340-c.

Valente, A. (2004). *Partecipare la scienza*, a cura di Adriana Valente e Daniela Luzi, Roma, Biblink.

<http://www.irpps.cnr.it/it/system/files/Partecipare_la_scienza.pdf>

Vlachy, J. (1978). *Frequency distributions of scientific performance: A bibliography of Lotka's law and related phenomena*, «Scientometrics», 1, 109-30.

- Vanclay, J. K. (2009). *Bias in the journal impact factor*, «Scientometrics», 78(1), 3-12.
<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1007/s11192-008-1778-4>>
- Van Leeuwen, T. N. - Moed, H. (2001). Development and Applications of New Journal Impact Measures, «Cortex», 37(4), 607-610.
<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1016/S0010-9452%2808%2970607-5>>
- Van Leeuwen, T. N. - Moed, H. F. (2002). *Development and application of journal impact measures in the Dutch science system*, «Scientometrics», 53(2), 249–266.
- Van Leeuwen, T. N. - Visser, M. S. - Moed, H. F. - Nederhof, T. J. - van Raan, A. F. J. (2003), *The Holy Grail of science policy: Exploring and combining bibliometric tools in search of scientific excellence*, «Scientometrics», 57(2), 257-280.
<<http://www.cwts.nl/TvR/documents/AvR-HolyGrail.pdf>>
- Van Raan, A. F. J. (1988). *Handbook of Quantitative Studies of Science and Technology*, Amsterdam, Elsevier.
- Van Raan, A. F. J. (2000). *The Pandora's box of citation analysis: Measuring scientific excellence, the last evil?*, in B. Cronin, H. B. Atkins, *The Web of Knowledge. A Festschrift in Honor of Eugene Garfield*. Ch., 15, 301–319. Medford, ASIS Monograph Series.
- Van Raan, A. F. J. (2004a). *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, 19-50.
- Van Raan, A. F. J. (2004b). *Sleeping Beauties in science*, «Scientometrics», 59(3), 461–466.
<<http://www.cwts.nl/tvr/documents/avr-slbeau-scientom.pdf>>
- Van Raan, A. F. J. (2005). *Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods*, «Scientometrics», 62(1), 133-143.
<<http://www.cwts.nl/TvR/documents/AvR-RankingScientom.pdf>>
- Vaughan, L. (2004a). *New measurements for search engine evaluation proposed and tested*, «Information Processing and Management», 40(4), 677–691.
- Vaughan, L. (2004b). *Web hyperlinks reflect business performance—A study of US and Chinese IT companies*, «Canadian Journal of Information and Library Science», 28(1), 17–31.
- Vaughan, L. (2004c). *Exploring website features for business information*, «Scientometrics», 61(3), 467–477.
- Vaughan, L. (2006). *Visualizing linguistic and cultural differences using web co-link data*, «Journal of the American Society for Information Science and Technology», 57(9), 1178–1193.
- Vaughan, L. - Hysen, K. (2002). *Relationship between links to journal Web sites and impact factors*, «ASLIB Proceedings», 54(6), 356–361.
- Vaughan, L. - Gao, Y. - Kipp, M. (2006). *Why are hyperlinks to business Websites created? A content analysis*, «Scientometrics», 67(2), 291–300.

- 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), 1313–1322.
- Vaughan, L. - Shaw, D. (2005). *Web citation data for impact assessment—A comparison of four science disciplines*, «Journal of the American Society for Information Science and Technology», 56(10), 1075–1087.
- Vaughan, L. - Thelwall, M. (2003). *Scholarly use of the Web—What are the key inducers of links to journal Web sites?*, «Journal of the American Society for Information Science and Technology», 54(1), 29–38.
- Vaughan, L. - Thelwall, M. (2004). *Search engine coverage bias—Evidence and possible causes*, «Information Processing and Management», 40(4), 693–707.
- Vaughan, L. - Thelwall, M. (2005). *A modeling approach to uncover hyperlink patterns—The case of Canadian universities*, «Information Processing and Management», 41(2), 347–359.
- Vaughan, L. - Wu, G. Z. (2004). *Links to commercial websites as a source of business information*, «Scientometrics», 60(3), 487–496.
- Vaughan, L. - You, J. (2006). *Comparing business competition positions based on Web co-link data—The global market vs. the Chinese market*, «Scientometrics», 68(3), 611–628.
- Vianello, N. (1970). *La citazione di opere a stampa e manoscritti*, Olschki.
- Vinkler, P. (2001). *An attempt for defining some basic categories of scientometrics and classifying the indicators of evaluative scientometrics*, «Scientometrics», 50(3), 539–544.
- Vinkler, P. (2002). *Subfield problems in applying the Garfield (Impact) Factors in practice*, «Scientometrics», 53(2), 267–279.
- Vinkler, P. (2004). *Characterization of the impact of sets of scientific papers—The Garfield (Impact) Factor*, «Journal of the American Society for Information Science and Technology», 55(5), 431–435.
- Vinkler, P. (2010a). *The Evaluation of Research by Scientometric Indicators*, New York, Neal Schuman.
- Vinkler, P. (2010b). *Indicators are the essence of scientometrics and bibliometrics*, «Scientometrics», 85(3), 861–866.
<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1007/s11192-010-0159-y>>
- Walter, G. - Bloch, S. - Hunt, G. - Fisher, K. (2003). *Counting on citations—A flawed way to measure quality*, «Medical Journal of Australia», 178(6), 280–281.
- Wan, J. - Hua, P. - Rousseau, R. (2007). *The pure h-index: calculating an author's h-index by taking co-authors into account*, «Collnet Journal of Scientometrics and Information Management», 1(2), 1–5.
<<http://eprints.rclis.org/11401/>>

Weinberg, B. H. (1974). *Bibliographic coupling: a review*, «Information Storage & Retrieval», 10, 189-96.

Weinberg, B. H. (1997). *The earliest Hebrew citation indexes*, «Journal of the American Society for Information Science», 48(4), 318–330.

Weinberg, B. H. (1999). *Indexes and religion: Reflections on research in the history of indexes*, «The Indexer», 21(3), 111–118.

<http://www.theindexer.org/files/21-3/21-3_111.pdf>

Weinreich, M. (1980). *Shlomo Noble (Trans.). History of the Yiddish language*, Chicago, University of Chicago Press.

Weinstock, M. (1971). *Citation indexes*, in *Encyclopaedia of Library and Information Science*, 5, New York, Marcel Dekker, 16-40.

West, J. D. (2010a). *Eigenfactor: ranking and mapping scientific knowledge*, PhD Dissertation, University of Washington.

<http://octavia.zoology.washington.edu/people/jevin/Documents/Dissertation_JevinWest.pdf>

West, J. D. (2010b). *How to improve the use of metrics: Learn from Game Theory*, «Nature», 465, 871-872.

<<http://octavia.zoology.washington.edu/people/jevin/Documents/West2010Nature.pdf>>

West, J. D. - Bergstrom, T. C. - Bergstrom, C. T. (2010a). *The Eigenfactor metrics TM: a network approach to assessing scholarly journals*, «College and Research Libraries», 71(3), 236-244,

<<http://crl.acrl.org/content/71/3.toc>>

West, J. D. - Bergstrom, T. C. - Bergstrom, C. T. (2010b). *Big Macs and Eigenfactor Scores: Don't Let Correlation Coefficients Fool You*, «Journal of the American Society for Information Science & Technology», 61(9), 1800-1807

<<http://octavia.zoology.washington.edu/people/jevin/Documents/West2010JASIST.pdf>>

West, J. D. - Stefaner, M. - Bergstrom, C. T. (2009). *The Eigenfactor™ Metrics: How does the Journal of Biological Chemistry stack up?*, «The American Society for Biochemistry and Molecular Biology», 20-21.

<<http://octavia.zoology.washington.edu/people/jevin/Documents/ASBMB2009.pdf>>

West, J. D. - Jensen, M. C. - Dandrea, R. J. - Gordon, G. J. - Bergstrom, C. T. (2011). *Author-Level Eigenfactor Metrics: Evaluating the Influence of Authors, Institutions and Countries Within the SSRN Community*, Harvard Business School NOM Unit Working Paper, 11-001.

<<http://www.eigenfactor.org/href=>>>

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), 423–434.

White, H. D. (2004). *Citation analysis and discourse analysis revisited*, «Applied Linguistics», 25(1), 89–116.

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. (1989). *Bibliometrics*, «Annual Review of Information Science and Technology», 24, 119-65.

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), 327-355.

Whitehouse, G. H. (2001). *Citation rates and impact factors—Should they matter?*, «British Journal of Radiology», 74(877), 1-3.

Wilson, C. S. (1995), *The formation of subject literature collections for bibliometric analysis: the case of the topic of Bradford's Law of Scattering*, Ph.D. dissertation, The University of New South Wales, Sydney.

<<http://handle.unsw.edu.au/1959.4/17213>>

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

Wittig, G. R. (1978). *Documentation note: statistical bibliography - a historical footnote*, «Journal of Documentation», 34, 240-241.

Wleklinski, J. M. (2005). *Studying google scholar—Wall to wall coverage?*, «Online», 29(3), 22-26.

Wuff, J. - Nixon, N. D. (2004). *Quality Markers and Use of Electronic Journals in Academic Health Sciences*, «Journal of Medical Library Association», 92(3), 315-322.

<<http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=442173&blobtype=pdf>>

Xiao, N. (2010). *The Evaluation of Research by Scientometric Indicators*, «Journal of Academic Librarianship», 36(6), 554-555.

<<http://dx.doi.org.eres.library.manoa.hawaii.edu/10.1016/j.acalib.2010.08.016>>

Young, T. (1845). *Course of Lectures on Natural Philosophy and the Mechanical Arts*, London, Taylor and Walton.

Yu, G. - Wang, X. H. - Yu, D. R. (2005). *The influence of publication delays on impact factors*, «Scientometrics», 64(20), 235-246.

Yu, L. - Apps, A. (2000). *Studying e-journals user behavior using log files: the experience of SuperJournal*, «Library and information science research», 22(3), 311-338.

Zani, M. (2005). *La rivincita delle citazioni*, «Bibliotime», 8(2).

<<http://didattica.spbo.unibo.it/bibliotime/num-viii-2/zani.htm>>

Zhang, C. T. (2009). *The e-Index, Complementing the h-Index for Excess Citations*, «Plos One», 4(5).

Zhang, Y. (2010). *Developing an holistic model for digital library evaluation*, «Journal of American Society for Information Science», 61(1), 88-110.

<<http://dx.doi.org/eres.library.manoa.hawaii.edu/10.1371/journal.pone.0005429>>

Zhao, D. Z. (2005). *Challenges of scholarly publications on the Web to the evaluation of science—A comparison of author visibility on the Web and in print journals*, «Information Processing and Management», 41(6), 1403–1418.

Zhao, D. Z. (2006). *Towards all-author co-citation analysis*, «Information Processing and Management», 42(6), 1578–1591.

Zhao, D., & Logan, E. (2002). *Citation analysis using scientific publications on the Web as data source—A case study in the XML research area*, «Scientometrics», 54(3), 449–472.

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

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*, Cambridge, Addison-Wesley.

Gli URL sono stati verificati l'ultima volta il 28 settembre 2012.

Le traduzioni, salvo indicazioni diverse, sono della candidata.