

**UNIVERSITA' DEGLI STUDI DI PADOVA**



**FACOLTA' DI SCIENZE STATISTICHE  
CORSO DI LAUREA IN  
STATISTICA E GESTIONE DELLE IMPRESE**



**Gli indicatori bibliometrici per la  
valutazione delle riviste scientifiche**

**RELATORE: DOTT. LISI FRANCESCO**

**LAUREANDA: ILARIA CALISTI  
MATRICOLA: 572273-GEI**

**ANNO ACCADEMICO 2009/2010**

# *INDICE*

<b>1</b>	<b>La bibliometria e la valutazione delle riviste scientifiche</b>	<b>1</b>
<b>2</b>	<b>Impact Factor</b>	<b>5</b>
2.1	Origine dell'Impact Factor	5
2.2	Calcolo dell'Impact factor	7
2.3	Punti di forza	8
2.4	Punti di debolezza	9
2.5	Metodi alternativi all'Impact Fcator	12
<b>3</b>	<b>H-index</b>	<b>15</b>
3.1	Origine dell'H-index	15
3.2	Calcolo dell'H-index	17
3.3	Esempio dell'H-index	20
3.4	Punti di forza dell'H-index	23
3.5	Punti di debolezza dell'H-index	24
3.6	Indici alternativi	26
<b>4</b>	<b>Usage Factor</b>	<b>27</b>
4.1	Modello Open Access	27
4.2	Calcolo dell'Usage Fcator	29
4.2.1	Analisi quantitativa	29
4.2.2	Analisi qualitativa	30
4.3	Punti di forza dell'Usage Factor	31
4.4	Punti di debolezza dell'Usage Factor	32
4.5	Indici alternativi	32
<b>5</b>	<b>Eigenfactor</b>	<b>33</b>
5.1	Metodo di ricerca	34
5.2	Metodo di calcolo	38
5.2.1	Data Input	38
5.2.2	Creare la matrice di croce-citazione	38
5.2.3	Modificare la matrice di croce-citazione	39
5.2.4	Identificazione dei nodi	41
5.2.5	Calcolare il vettore d'influenza	41
5.2.6	Calcolare l'indice Eigenfactor	43
5.2.7	Emettere i risultati	43
5.3	Punti di forza	44
5.4	Metodi alternativi	45

<b>6</b>	<b>Altri aspetti della bibliometria</b>	<b>46</b>
6.1	L'impact factor delle riviste statistiche	46
6.1.1	Dati	48
6.1.2	Metodo di studio	50
6.1.3	Tempo al raggiungimento del picco T*	51
6.1.4	Percentuale delle citazioni nei primi anni	52
6.2	Il mondo dell'Open Access	55
6.3	Publish or Perish	56
6.3.1	Che cos'è Publish or Perish	56
6.3.2	Come funziona Publish or Perish	57

# *Capitolo 1 - La bibliometria e la valutazione delle riviste scientifiche*

La bibliometria è un'antica disciplina applicata alle varie conoscenze scientifiche per studiare la loro propagazione attraverso la pubblicazione. Questa scienza risale alla fine dell'800, ma solamente negli ultimi vent'anni si è sviluppata l'analisi bibliometrica per mezzo dei database di grandi dimensioni resi utilizzabili on-line.

Tale materia è un ramo della scientometria, ovvero una scienza che valuta e misura la produzione scientifica, nata verso la fine del XX secolo.

La bibliometria adopera tecniche matematiche e statistiche per la valutazione della ricerca e per analizzare l'impatto delle pubblicazioni sulle comunità scientifiche. Lo strumento per fare tutto ciò è l'analisi citazionale, cioè l'uso delle citazioni nei lavori scientifici per stabilire legami con altre pubblicazioni o altri ricercatori. L'elenco di tutte queste citazioni si può trovare nel database ISI (Institute for Scientific Information), divenuto nel 1997 WoS (Web of Science) a causa dello spostamento nel web della banca dati, meno conosciuto però con questo nome.

I criteri per la valutazione della ricerca scientifica sono due:

1. INDICATORI BIBLIOMETRICI: sono un metodo quantitativo basato sul numero delle citazioni ricevute dalla pubblicazione. Più elevato è il numero delle citazioni ricevute dalla pubblicazione, maggiore è il numero di autori che hanno utilizzato nel loro lavoro il contenuto della pubblicazione in

oggetto, quindi maggiore è l'impatto di questa pubblicazione sulla comunità accademica.

2. PEER REVIEW: metodo qualitativo basato sul giudizio di esperti. La recensione di esperti, traduzione della peer review in italiano, consiste in un ristretto numero di studiosi specializzati nell'ambito della pubblicazione oggetto di studio, con il compito di analizzare e giudicare la produzione scientifica.

La possibilità di utilizzare uno dei due procedimenti non esclude l'altro, anzi proprio per il fatto che non esiste una regola standard per la valutazione, è legittimo usufruire della cooperazione degli indicatori bibliometrici e della peer-review.

Come tutti i criteri anche essi hanno i loro vantaggi e limiti. Gli indicatori bibliometrici hanno come punto a favore il fatto di essere oggettivi nel senso che il giudizio citazionale viene ricavato da tutta la comunità di studiosi e non da un ristretto numero di essi. Inoltre gli indici godono della peculiarità di essere economici perché sono disponibili in rete. D'altra parte però essi hanno due grandi svantaggi: non esiste un metodo bibliometrico standard e non hanno una copertura totale dei dati per tutte le discipline accademiche.

Per quanto riguarda la recensione di esperti, invece, i ricercatori si chiedono se la valutazione delle pubblicazioni affidata a due o tre individui sia sufficiente, a differenza degli indici che vengono analizzati dall'intera comunità. Oltre tutto i costi e i tempi nell'applicare questo metodo sono sicuramente superiori rispetto all'altro. Infatti i pochi studiosi impiegano in media non meno di due anni per esaminare tutte le pubblicazioni.

Dato che la valutazione delle riviste si basa sulle citazioni che ricevono le pubblicazioni, è giusto elencare i fattori che influenzano gli autori citanti lavori precedenti: qualità del contenuto, sesso dell'autore (perché la totalità degli studiosi è prevalentemente composta da individui di sesso maschile, di conseguenza le citazioni provenienti da scienziati maschi sono superiori rispetto a quelle fatte da figure femminili), numero degli autori della pubblicazione, fonte, citazioni precedenti se sono presenti, temi trattati, area di studio, assimilazione delle tematiche.

Nel caso in cui i lavori citati siano contemporanei o siano ricerche in corso, le caratteristiche da considerare sono: affinità e area del contenuto, se il lavoro citato è a favore della tesi dell'autore oppure se l'autore citato è importante, facilità di accesso alla pubblicazione, presenza di autocitazioni, distanza temporale (il lavoro citato è stato pubblicato da poco rispetto al lavoro citante).

Dopo questa breve introduzione, i principali indicatori utilizzati nell'analisi bibliometrica sono:

- IMPACT FACTOR: è la frequenza con cui un articolo medio di una rivista viene citato dalle altre riviste in uno specifico anno o intervallo di tempo.
  - CITED-HALF-LIFE: misura la validità nel tempo degli articoli citati o la durata delle citazioni nel tempo.
  - IMMEDIACY INDEX: misura il successo del lavoro nell'anno di pubblicazione in relazione alla quantità e velocità delle citazioni ricevute nell'anno in corso.
  - EURO FACTOR: misura l'impatto delle pubblicazioni prevalentemente scritte in lingua non inglese e pubblicate dagli editori non americani.
  - WEB IMPACT FACTOR: misura le citazioni presenti nei siti internet, mediante l'analisi dei link.
- H-INDEX: misura l'impatto del lavoro degli scienziati basandosi sul numero delle loro pubblicazioni e delle citazioni ricevute.
  - H-B-INDEX: definito come l'h-index, ma basato sull'analisi di uno specifico argomento.
  - G-INDEX: definito come l'h-index, con la sola differenza che esso considera le pubblicazioni più citate.
  - A-INDEX: media del numero di citazioni ricevute da pubblicazioni provenienti da quelle dell'h-index.
- USAGE FACTOR: è il fattore di utilizzo delle risorse elettroniche per misurare la qualità delle riviste.
  - WEB IMPACT FACTOR: indice che studia l'analisi dei link per misurare le citazioni ricevute da una rivista in rete.

- EIGENFACTOR: misura l'importanza della totalità delle riviste sulla comunità scientifica.
  - ARTICLE INFLUENCE: misura l'influenza media per articolo di una rivista, 5 anni dopo la sua pubblicazione.

Nel capitolo 1 è illustrato l'indice bibliometrico dell'Impact Factor utilizzato per valutare la qualità delle riviste. Nel capitolo 2, invece, è spiegato l'indice di Hirsch relativo alla valutazione dei lavori dei singoli autori. Mentre nei capitoli 3 e 4 sono delucidati gli indici dell'Usage Factor e dell'Eigenfactor rispettivamente, i quali si riferiscono alla valutazione delle riviste.

## *Capitolo 2 - Impact Factor*

L'impact factor (IF), chiamato anche indice d'impatto, è l'indicatore numerico - statistico più conosciuto e utilizzato, mediante il quale la comunità scientifica misura la frequenza con cui un articolo medio di una rivista viene citato dalle altre riviste in uno specifico anno o intervallo temporale.

### **2.1 ORIGINE DELL'IMPACT FACTOR**

Questo indice fu ideato nel 1955 da Eugene Garfield, un chimico americano, con lo scopo di categorizzare, valutare, comparare e ordinare le riviste. In origine, l'indice di impatto nasce con il compito di essere utile alle biblioteche nella scelta delle riviste da acquistare ed eliminarne altre, aiutare gli autori nello scegliere la rivista alla quale inviare i propri lavori e mostrare ai lettori il prestigio delle varie riviste. Solo in seguito esso verrà usato per la valutazione dei lavori scientifici di singoli autori o gruppi di ricercatori, basando tale valutazione sul numero delle citazioni ricevute per ogni lavoro. Garfield (1955) è anche il fondatore dell'Institute of Scientific Information (ISI), una delle più grandi aziende d'informazione scientifica nel mondo, nella quale possiamo trovare il calcolo dell'IF. All'inizio la banca dati dell'ISI faceva parte della Thomson, di recente acquistata dalla Reuters e quindi oggi

l'IF è di proprietà della Thomson Reuters. La banca dati dell'ISI è formata da un'enorme quantità di dati relativi alle citazioni scientifiche, impiegate per calcolare l'indice, rappresentate da una lista di articoli estratti da un grande numero di riviste scientifiche nel mondo. Nel database le citazioni vengono poi ordinate per poter osservare quante volte ogni citazione viene citata in un certo periodo di tempo e da quale autore o gruppo di ricercatori. Successivamente i dati vengono pubblicati annualmente nello Science Citation Index (SCI), un archivio elettronico delle citazioni scientifiche sempre aggiornato.

Per mezzo della banca dati elettronica si possono fare più facilmente, rispetto che nel passato, ricerche bibliografiche. È sempre stato possibile fare ricerche bibliografiche a partire da un lavoro importante e andando indietro nel tempo. Ciò avviene prendendo in considerazione gli articoli presenti in bibliografia di un'opera, poi si cercano i lavori citati da questi e così via, finché si arrivava all'origine della problematica. Oggi si possono ricercare le citazioni incrociate molto più facilmente rispetto a qualche anno fa. È possibile, inoltre, identificare dei lavori che citano un'importante opera del passato, poi si cercano i lavori che li citano a loro volta e così via, in modo da analizzare gli sviluppi di una problematica. Ciò ha determinato la creazione dell'archivio elettronico ISI il quale deve contenere una grande quantità di articoli e riviste, con le citazioni correlate, che fanno parte dei lavori scientifici più importanti nel mondo. La banca dati è organizzata innanzitutto scegliendo l'insieme delle riviste che la costituiscono, da queste, in un secondo momento, si estraggono gli articoli con l'indicazione dei nomi degli autori e le citazioni di articoli, libri, relazioni a congresso, eccetera. Infine tutti questi dati vengono inseriti in una banca dati elettronica. Eugene Garfield (1955) sostiene che l'ISI sia completa perché è composta da tutte le riviste presenti nelle maggiori biblioteche americane, ovvero quelle riviste pubblicate negli Stati Uniti e dalle grandi case editrici nord-americane,

dalle riviste pubblicate nei principali paesi di lingua inglese e inoltre contiene anche qualche rivista “nazionale” dei principali paesi europei.

Un'altra importante caratteristica da non dimenticare è che l'ISI include riviste con pubblicazioni a cadenza regolare. Le riviste con annate mancanti o in ritardo nella pubblicazione comportano costi aggiuntivi in questa banca dati elettronica e perciò essa preferisce cancellare queste riviste, preferendone altre più ordinarie. Infatti ci sono alcune riviste che impiegano molto tempo per la valutazione degli articoli, ritardando così la loro pubblicazione.

## 2.2 CALCOLO DELL'IMPACT FACTOR

Innanzitutto per calcolare l'indice di impact factor di una specifica rivista è necessario che essa sia inserita nelle banche dati dell'ISI e poi occorre un minimo di 2 anni dalla data di pubblicazione per ottenere le citazioni. Infatti molte riviste di recente pubblicazione potrebbero non essere presenti nelle banche dati elettroniche della Thomson Reuters e di conseguenza per esse non viene calcolato l'indice. Inoltre i lavori pubblicati all'inizio dell'anno hanno a disposizione più tempo per essere citati rispetto a quelli pubblicati alla fine dell'anno.

La definizione teorica dell'impact factor è:

*L'IF della rivista X nell'anno t, è il rapporto tra il numero di citazioni rilevate nell'anno t (C<sub>t</sub>) sulle riviste incluse nella banca dati, di articoli pubblicati negli anni t-1 e t-2 sulla rivista X, diviso per il numero totale degli articoli pubblicati (P<sub>t</sub>) negli anni t-1 e t-2 sulla rivista X.*

In formule:

$$IF = \frac{\sum_{i=1}^2 C_{t-i}}{\sum_{i=1}^2 P_{t-i}}$$

Dalla definizione aggiungiamo che il punteggio delle citazioni di una rivista scientifica può essere ottenuto dalla media delle citazioni di tutti gli articoli contenuti nella rivista.

Oltre a ciò, c'è da precisare che il numero totale degli articoli pubblicati è composto da articoli, note e rassegne, non rientrano invece editoriali, lettere, notizie, correzioni, database, rassegne di libri, software e altro materiale editoriale.

Esempio:

2500 = numero delle citazioni della rivista X nel 2005, relativamente agli articoli pubblicati nell'intervallo temporale 2004-2003.

500 = numero di articoli pubblicati dalla rivista X nei periodi 2004-2003.

Allora  $IF = 2500/500 = 5$

Quindi ogni articolo pubblicato sulla rivista X, nell'intervallo di tempo compreso nel biennio 2004-2003, è stato citato nel 2005 in media 5 volte.

Più alto risulta l'IF, migliore è la qualità della rivista sulla quale l'indice è stato calcolato, perché vuol dire che la rivista in questione è maggiormente citata dai ricercatori.

## **2.3 PUNTI DI FORZA**

Il principale punto di forza dell'IF è la sua facilità e semplicità di utilizzo per la valutazione della ricerca scientifica. Esso è un indice bibliometrico rapidamente accessibile perché si trova in una banca dati elettronica, ben fornita, ed è anche poco costoso. Il calcolo dell'impact factor è in grado di determinare in modo valido e corretto la qualità scientifica degli articoli nei quali esso viene calcolato. Per i suoi inventori l'indice è un fondamentale strumento per decidere se acquistare o mantenere l'abbonamento di una rivista. Per esempio se una rivista è poco citata, in relazione agli articoli che pubblica, il database ISI suggerisce di cancellare tale

rivista perché il costo e il tempo per esaminare gli articoli e le sue citazioni risulterebbe inutile. La maggior parte delle riviste presenti nel database elettronico sono in lingua inglese, questo favorisce la loro maggiore divulgazione rispetto a quelle scritte in altre lingue.

Seglen (1997) spiega che l'IF dipende dai campi di ricerca della rivista. Per esempio i campi di ricerca dinamici, come la biochimica e la biologia molecolare, che pubblicano articoli molto rapidamente, hanno un elevato numero di citazioni e di conseguenza un alto impact factor, determinato proprio dalla grande quantità di materiale messo a disposizione di altri scienziati. Le riviste del campo matematico, i cui lavori impiegano più tempo per l'accettazione e la revisione, hanno generalmente un basso valore dell'indice. Inoltre gli articoli delle riviste appartenenti a campi accademici innovativi e in continua espansione ricevono un numero maggiore di citazioni relativamente all'ammontare del materiale citabile. Ciò non avviene per le riviste che trattano temi tradizionali, nelle quali riscontriamo una quantità di citazioni inferiore. Inoltre Seglen (1997) afferma che un campo contiene riviste con più alto IF se ha la capacità di essere citato dai campi di ricerca adiacenti ad esso. Per esempio la medicina di base e quella clinica hanno dei punti in comune, ma è importante specificare che la medicina clinica si fonda su quella di base e da essa si specializza in tanti altri sottocampi, ma non vale il contrario. Da qui Seglen (1997) dichiara che il numero delle citazioni ricevute da una rivista è proporzionale alla grandezza del suo campo di ricerca, quindi le riviste specializzate in un sottocampo avranno un IF più basso.

## **2.4 PUNTI DI DEBOLEZZA**

L'impact factor, nonostante sia l'indice più usato per la valutazione delle riviste, è criticato in diversi aspetti.

Alessandro Figà Talamanca (2000) sostiene che l'IF può essere maneggiato in vari modi, creando così distorsioni sul reale valore di una rivista. Per prima cosa, egli sostiene che i ricercatori o gruppi di ricercatori possono essere i primi a manipolare l'indice perché prima di pubblicare i loro articoli su una specifica rivista, prendono in

considerazione la lista delle riviste ordinate in base ai valori dell'IF. Dopo aver analizzato questa lista scelgono una rivista, tra quelle con più alto IF, nella quale pubblicare i loro lavori. Seguendo tale tattica essi sperano di ricevere il maggior numero di citazioni. L'unica condizione per applicare questa strategia è che gli scienziati abbiano a disposizione l'elenco delle riviste ed il loro IF, che si trovano nel database ISI. Talamanca (2000) fa queste critiche perché egli è convinto che l'azienda dell'ISI abbia come scopo principale quello di guadagnare, attraverso gli scienziati o le biblioteche che acquistano le riviste dal database elettronico. In questo modo l'Institute of Scientific Information agisce come tutte le aziende private. In particolar modo, Talamanca (2000) si chiede come mai l'ISI non si pone come obiettivo principale quello di promuovere l'informazione scientifica. I ricercatori poi riempiono i loro articoli di citazioni relative alla rivista nella quale li vogliono pubblicare. Inoltre se essi appartengono ad un gruppo di scienziati favorevoli alle citazioni reciproche è palese che queste citazioni non sono sincere come chi le fa per richiamare contenuti di opere famose, esprimendo così la devozione per un particolare articolo o scrittore. Il numero delle citazioni è significativo per misurare la qualità delle riviste. Oggi, purtroppo, Talamanca (2000) ritiene che la pratica della citazione viene usata in modo errato, nel senso che una volta si citava per ringraziare gli ispiratori della ricerca, oggi invece la tecnica citazionale è strettamente legata alla carriera lavorativa, all'accesso ai finanziamenti e alla diffusione delle riviste. Oltre alle citazioni di gruppo, sono gli stessi comitati di redazione che chiedono ai ricercatori di citare lavori della rivista nella quale vogliono pubblicare i loro articoli, come condizione necessaria per la pubblicazione. Per questi motivi aumenta anche il numero delle autocitazioni, cioè citazioni dirette a lavori pubblicati nella stessa rivista nella quale si vogliono pubblicare i propri articoli. Tutto questo comporta un aumento dell'IF della rivista e con esso la qualità e la diffusione. Proprio per questo motivo, spiega Talamanca (2000), alcuni scrittori decidono di fare citazioni solo perché a loro volta verranno citati. Con questa teoria sfortunatamente l'indice perde la sua obiettività e in particolar modo la sua veridicità.

Anche Seglen (1997) ritiene che l'impact factor abbia alcuni problemi di natura tecnica e di fondamento. Per cominciare, egli sottolinea che c'è disparità tra numeratore e denominatore perché quest'ultimo include articoli, note e rassegne, ma al numeratore le citazioni vengono fatte su tutti i tipi di documenti. Questo fatto potrebbe causare una distorsione nel calcolo dell'indice nel caso in cui una rivista

decida di non accettare lavori che hanno poca probabilità di essere citati. Un ingrossamento dell'IF è stato individuato anche per le riviste di rassegna, le quali sono sottoposte a giudizio numerose volte da vari ricercatori; lo stesso vale per quelle riviste che pubblicano molte rassegne, le quali ricevono un numero più elevato di citazioni rispetto agli articoli ordinari. Per quanto riguarda l'aumento mirato dell'indice, infine, Seglen (1997) segnala che nella somma delle citazioni non si tiene presente la distinzione tra quelle provenienti da bibliografie e le autocitazioni. Queste ultime infatti dovrebbero essere cancellate in quanto non ritenute valide per valutare la qualità di una rivista perché è ovvio che uno scienziato ritiene la sua rivista un buon lavoro. Seglen (1997) sostiene anche che il numero totale delle citazioni relative ad un articolo è approssimativamente proporzionale alla lunghezza dello stesso, quindi gli articoli estesi accrescono l'IF, ma un numero elevato di citazioni non esprime di per sé la qualità di una rivista. In aggiunta a ciò, secondo Seglen (1997), sarebbe opportuno distinguere le citazioni che esprimono giudizi positivi da quelle di critica, perché esse contribuiscono in modo opposto nel calcolo dell'IF. D'altra parte però l'inclusione nell'indice di autocitazioni e citazioni negative possono rappresentare un vantaggio perché aumentano il valore dell'IF, attribuendo in questo modo un maggiore prestigio alla rivista.

Inoltre Seglen (1997) afferma che l'intervallo di tempo di due anni, condizione necessaria per calcolare l'indice, non è sufficiente per tutte le riviste. Infatti esistono riviste alle quali occorre un tempo più lungo per la rettifica e l'accettazione, causata dal fatto che esse sono revisionate e controllate in modo più severo e perciò questo lavoro di rassegna richiede più tempo.

Dopo Talamanca (2000) e Seglen (1997), il professore Gianluca Castelnuovo enuncia alcuni limiti e difetti dell'impact factor. Oltre a quanto detto sopra, egli aggiunge che l'IF, in relazione all'intervallo temporale, è più alto per i lavori pubblicati all'inizio dell'anno, perché essi hanno un tempo più lungo per ricevere citazioni, rispetto a un lavoro pubblicato a fine anno. Un'altra critica è che l'indice non sembra essere adeguato per la valutazione della ricerca scientifica, in quanto esso si basa sulla media aritmetica del numero delle citazioni e viene calcolato solo su una parte degli articoli pubblicati su una rivista. Infine Castelnuovo attacca alti punteggi d'IF legati a riviste con diffusione a livello internazionale e appartenenti a campi di ricerca multidisciplinari. Questa critica è spiegata dal fatto che alcuni settori

studiano temi di carattere locale, ma certo non meno importanti, e perciò hanno una diffusione limitata e quindi un basso valore dell'indice.

In conclusione questi tre scienziati concordano che l'ISI è un database contenente riviste esclusivamente in lingua inglese, andando a discapito delle riviste locali o nazionali che per la maggior parte fanno pubblicazioni nella loro lingua d'origine. Da un lato questa caratteristica è positiva perché la lingua inglese è la lingua divulgativa più utilizzata, dall'altra parte però ostacola i lavori originari di una terra che parla una lingua diversa dall'inglese. Cioè la pubblicazione in lingua inglese non è sufficiente per provare la qualità di una rivista.

## **2.5 METODI ALTERNATIVI ALL'IMPACT FACTOR**

I ricercatori scientifici hanno elaborato altri indici per risolvere i problemi e i limiti dell'IF.

Tali indici alternativi sono sotto elencati.

### **CITED HALF – LIFE**

Questo indice nasce per risolvere il problema dell'IF legato alla condizione temporale dell'uscita delle pubblicazioni limitate a due anni, cercando di includere quelle riviste che impiegano più tempo per i loro lavori scientifici. Il cited half-life (CHAL) è definito come l'età mediana degli articoli, contenuti nel database ISI, e citati ogni anno, cioè la durata delle citazioni nel tempo. In altre parole questo indice misura il numero di anni impiegati per arrivare a raggiungere il 50% del totale delle citazioni nell'anno in corso, riferite alla rivista considerata.

In formule:  $CHAL = C_i + (C_{i+1} - C_i) \frac{0.5 - F_i}{F_{i+1} - F_i}$

dove  $C_i$  rappresenta la i-esima citazione e  $F_i$  è la frequenza cumulata della i-esima citazione.

Per esempio un valore CHAL = 5, per una rivista pubblicata nell'anno 2009, vuol dire che la metà del totale delle citazioni registrate nell'anno 2009 sono presenti in articoli pubblicati negli ultimi cinque anni, l'altra metà degli articoli pubblicati prima, invece, registra la vita media degli articoli citati.

## **IMMEDIACY INDEX**

L'immediacy index (II) misura il numero delle citazioni di una rivista nell'anno della sua pubblicazione. Questo indice è stato creato per essere di aiuto alle biblioteche sui tagli dei periodici a cui esse sono abbonate, perché ne valuta l'immediatezza, ovvero la prossimità delle citazioni rispetto alle pubblicazioni degli articoli.

Il seguente esempio mostra come il II viene calcolato.

2006 = anno della pubblicazione degli articoli per la rivista X

500 = articoli pubblicati dalla rivista X nell'anno 2006.

3000 = numero delle citazioni

$$\begin{aligned}\text{Allora il II} &= \frac{(\text{numero citazioni ricevuti dalla rivista X nell'anno Y})}{(\text{numero totale di articoli pubblicati nell'anno Y})} \\ &= 3000/500 = 6\end{aligned}$$

## **EURO-FACTOR**

L'euro-factor (EF) è un indice nato per risolvere il limite dell'IF che considera prevalentemente lavori scientifici in lingua inglese o pubblicati dalle maggiori case editrici nord-americane, escludendo così ricerche scientifiche scritte in altre lingue o pubblicate in altri paesi.

L'algoritmo creato da Roland Hofbauer (2002) analizza le relazioni bibliometriche tra le riviste europee. La formula per questo indice è data dal rapporto tra il numero delle citazioni (A) e il coefficiente dell'euro factor posto uguale a 10, moltiplicato per la radice quadrata del numero delle citazioni sommate al numero di articoli (B).

$$EF = \frac{A}{EF\text{-coefficiente} * \sqrt{A+B}}$$

Sfortunatamente VICER, database dell'euro-factor, non provvede ad alcuna spiegazione dettagliata dell'algoritmo al di fuori della semplice formula, nella quale arbitrariamente si pone il coefficiente EF pari a 10. Ogni anno vengono calcolati i nuovi punteggi dell'indice e vengono pubblicati ogni gennaio nel database.

## **WEB IMPACT FACTOR**

Il web impact factor è un indice che prende in considerazione le citazioni presenti in siti al di fuori della banca elettronica dell'ISI.

Oltre all'uso di indici alternativi, si possono anche proporre semplici soluzioni ai problemi dell'IF, come per esempio: scartare le auto-citazioni che gli autori fanno esclusivamente per aumentare il valore dell'IF; valutare la qualità delle riviste in ogni campo tematico e non fra tutte le riviste presenti; considerare per la valutazione della produttività scientifica non solo le riviste e gli articoli ma anche libri, CD-ROM, video, convegni, ect.

# *Capitolo 3 - H-Index*

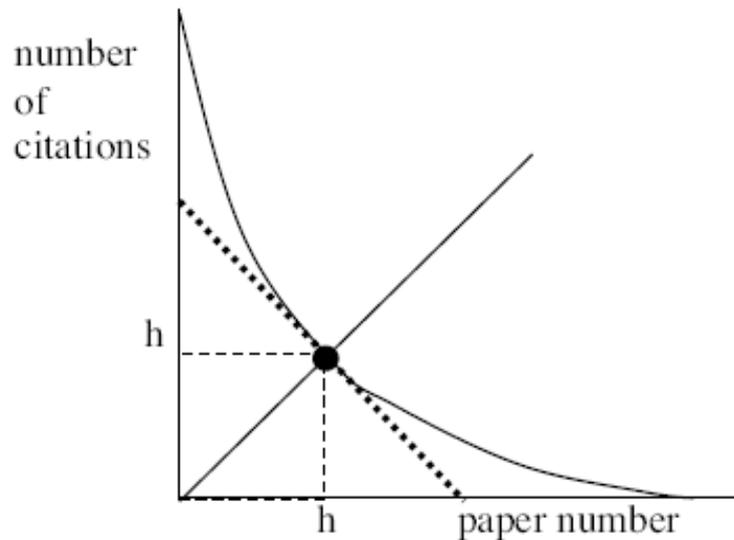
## **3.1 ORIGINE DELL'H-INDEX**

Dopo l'Impact Factor, l'indice più diffuso e conosciuto è l'h-index, chiamato anche indice di Hirsch. Infatti fu ideato da Hirsch (2005), all'università della California a San Diego.

Il suo creatore aveva come obiettivo quello di ridurre i numerosi problemi dell'IF, costruendo così un indice valido e oggettivo, in grado di misurare con efficacia la valutazione dei lavori scientifici dei ricercatori o gruppi di ricercatori. L'h-index, di fatto, valuta gli autori appartenenti alla medesima disciplina, a differenza dell'IF quindi non fa confronti tra autori di campi accademici che non hanno nessun punto in comune. Esso inoltre tiene conto dell'età del ricercatore e del numero di anni durante i quali egli ha pubblicato i suoi lavori. L'h-index, infatti, è in grado di misurare, in base al numero delle pubblicazioni emesse e delle citazioni ricevute, l'impatto del ricercatore scientifico all'interno del suo campo disciplinare.

L'h-index si definisce nel seguente modo:

*Uno scienziato possiede un indice  $h$ , se  $h$  dei suoi  $N_p$  lavori hanno almeno  $h$  citazioni ciascuno ed i rimanenti  $(N_p-h)$  lavori hanno ognuno al più  $h-1$  citazioni.*



La figura sopra rappresenta la curva dell'h-index. Nell'asse delle ascisse troviamo il numero di pubblicazioni, mentre nell'asse delle ordinate il numero delle citazioni. L'incontro tra la bisettrice del primo e terzo quadrante con la curva identifica il valore dell'indice h. Inoltre l'area sotto la curva è l'insieme delle citazioni.

Un h-index pari a zero riconosce che lo scienziato, corrispondente a tale valore, è inattivo.

Per esempio uno scienziato che ha pubblicato 185 lavori, ordinati in modo decrescente per quanto riguarda le citazioni, ha un h-index pari a 29 se 29 dei suoi 185 lavori hanno ciascuno almeno 29 citazioni e gli altri 154 (185-29) hanno ciascuno meno di 29 citazioni.

Posizione Articolo	1	2	3	...	>	28	29	30	...	>	184	185
Numero citazioni	192	141	103	...	>	29	29	28	...	>	0	0

## 3.2 CALCOLO DELL'INDICE H

L'indice si può quantificare per mezzo di Publish or Perish (PoP), un programma sviluppato da Harzing (2006) che fornisce le principali variabili bibliometriche in tutti i maggiori campi di ricerca. Questo software funziona grazie a Google Scholar, un motore di ricerca che copre le risorse del web, non fornisce però la lista completa degli articoli pubblicati da ogni ricercatore, ma solo una frazione di essi compresa tra il 40% e il 70%. Questo database è gratuito e si avvale del marchio di qualità legato agli algoritmi di Google. Inoltre esso applica gli stessi criteri per tutti, quindi si possono fare confronti all'interno dello stesso settore scientifico. Oltre a Google Scholar, ci sono altri database per il calcolo dell'H-index, come ad esempio l'ISI (Institute of Scientific Information) già conosciuto per l'impact factor, e Scopus, una banca dati bibliografica citazionale multidisciplinare.

Per utilizzarli è sufficiente avere una connessione internet. Per esempio PoP mediante le queries di Google Scholar offre informazioni sulle citazioni delle riviste, sugli articoli, su uno scrittore o un gruppo di ricercatori.

Per fare un'analisi sull'impatto di un autore bisogna inserire il nome dell'autore nel campo author's name e premere invio. Google Scholar, una volta che ha recuperato le citazioni, mostrerà l'elenco di esse e calcolerà le metriche citazionali nel pannello results.

Per ricercare opere scritte da più autori basta semplicemente inserire i loro nomi sempre all'interno del campo author's name. Inoltre usando l'operatore booleano AND otteniamo i lavori prodotti dalla collaborazione di diversi scienziati, mentre con OR otteniamo i singoli lavori di ogni autore.

Nel caso in cui si vuole fare la ricerca in un limitato intervallo di tempo è necessario inserire gli anni di inizio e di fine pubblicazione nel campo years of publication between...and...

Se si vuole fare la ricerca per aree tematiche è fondamentale conoscere le tipologie di soggetti di specifiche pubblicazioni e si osservano i box corrispondenti a quelle aree di soggetti.

Nel calcolo dell'h-index ci sono tre regole fondamentali da non trascurare:

1. Individuare un campo accademico nel quale fare confronti tra i diversi ricercatori presenti all'interno di esso. Questa operazione si attua mediante i box corrispondenti alle tipologie di soggetti presenti nel database Google Scholar. Infatti non è molto logico comparare autori di campi disciplinari diversi perché ogni settore ha le proprie peculiarità. Inoltre è opportuno dividere lo stesso settore in sezioni più piccole in modo tale da trovare un valore efficace dell'indice.
2. Una volta scelta la banca dati per il calcolo dell'indice si deve rimanere all'interno di essa per fare i paragoni tra i diversi scienziati. Tale scelta viene effettuata dall'individuo che ha deciso di fare la ricerca. Difatti, ogni database ha i suoi elementi distintivi, quindi il valore dell'indice proveniente da banche dati diverse non risulterebbe veritiero per la valutazione dei ricercatori. Per esempio Google Scholar è adatto per le citazioni in lingua inglese e per i settori umanistici, mentre il database ISI Web of Knowledge è maggiormente usato per le materie scientifiche.
3. Fissare un intervallo temporale di riferimento per l'analisi valutativa. In questo punto sarebbe più appropriato introdurre altri indici più validi per comparare le pubblicazioni annuali, l'età degli scienziati e per osservare se essi hanno lavorato maggiormente nel passato o nel presente. È possibile fare questo mediante un apposito campo presente nel database di ricerca. Ovviamente i lavori più recenti, e quindi di interesse comune, avranno un numero di citazioni più elevato rispetto a quelli del passato.

Nel semplice calcolo dell'h-index però non sono sufficienti queste tre regole, ci sono altri aspetti da non sottovalutare o comunque da spiegare.

Per quanto riguarda le pubblicazioni emesse da ogni scienziato, l'indice si focalizza su quelle che hanno avuto maggior diffusione, non solo sul loro numero effettivo. Per questo fatto l'h-index si scosta notevolmente dall'impact factor il quale considera tutte le pubblicazioni e citazioni, anche quelle negative. Oltre tutto l'indice di Hirsch non è condizionato dalle riviste che hanno riscontrato più popolarità, perché questo indice vuole appunto definire l'impatto dello scienziato nel suo settore accademico.

Come detto nel punto 1, si possono analizzare i lavori dei diversi scienziati basandosi sul campo al quale essi appartengono. Per esempio nell'ambito della matematica un

ricercatore avrà un valore dell'indice pari al numero di anni di lavoro, mentre un medico avrà un valore di sicuro più elevato, in quanto quest'ultimo osserverà un numero di citazioni superiore al numero delle pubblicazioni. Ciò avveniva analogamente per l'IF perché il campo della medicina è più citato e conosciuto, ma soprattutto è collegato ad un numero elevato di altri settori, rispetto alla fisica e alla matematica.

Specificando quanto detto al punto 3, il calcolo dell'indice considera l'epoca durante la quale gli articoli delle riviste sono stati scritti, volendo affermare che gli autori di pubblicazioni del passato saranno meno citati rispetto ad altri più recenti, perché trattano temi ormai obsoleti e di scarso interesse. Oltre a ciò Hirsch (2005) ha anche voluto esaminare gli anni concreti di lavoro dei vari scienziati. In conseguenza di ciò gli autori che hanno avuto una vita lavorativa breve non hanno riportato contributi evidenti nelle banche dati, a differenza di scienziati ben noti e con una carriera lunga e ben cospicua.

A differenza dell'impact factor, l'indice H non prende in esame per il suo calcolo le autocitazioni, per il fatto che ritiene che esse influenzino notevolmente il valore dell'indice. Infatti, un autore che produce molti lavori scientifici nel proprio campo, citerà con alta probabilità i suoi articoli precedenti, incrementando così l'indice.

Studiando più in dettaglio l'indice, verificiamo che esso presenta una crescita lineare nel corso del tempo dovuta al fatto che ogni anno ciascun scienziato produce e pubblica nuovi articoli e riviste e perciò è più soggetto ad essere citato. Ciò avviene con un tasso di crescita fissato.

Questa crescita si interrompe solo se l'autore sospende la sua attività di ricerca oppure diminuisce le pubblicazioni.

Nel caso in cui la produttività del ricercatore rimanga costante avremmo una situazione del genere:

$$h \sim mn$$

n = numero di anni calcolati dalla prima pubblicazione

m = linea di crescita dello scienziato, considera anche l'anzianità scientifica.

Hirsch (2005) individua i seguenti esempi per stabilire la capacità produttiva di un ricercatore:

- $m = 1$  e  $n = 20$  ottengo  $h = 20$ , cioè con un tasso di crescita costante e un indice pari a 20 identifica uno studioso di successo. Esso è stato in grado di pubblicare 20 articoli, citati almeno 20 volte in un intervallo di tempo di venti anni.
- $m = 2$  e  $n = 20$  consegua  $h = 40$ , in questo caso si parla di uno studioso di alto livello, il quale dopo venti anni di attività lavorativa ha prodotto 40 articoli, i quali sono stati citati almeno 40 volte.
- $m = 3$  e  $n = 20$  raggiunga  $h = 60$  identificando così uno studioso geniale che è riuscito a creare 60 articoli, citati almeno 60 volte, in venti anni di carriera di ricercatore.

### 3.3 ESEMPIO DELL'INDICE H

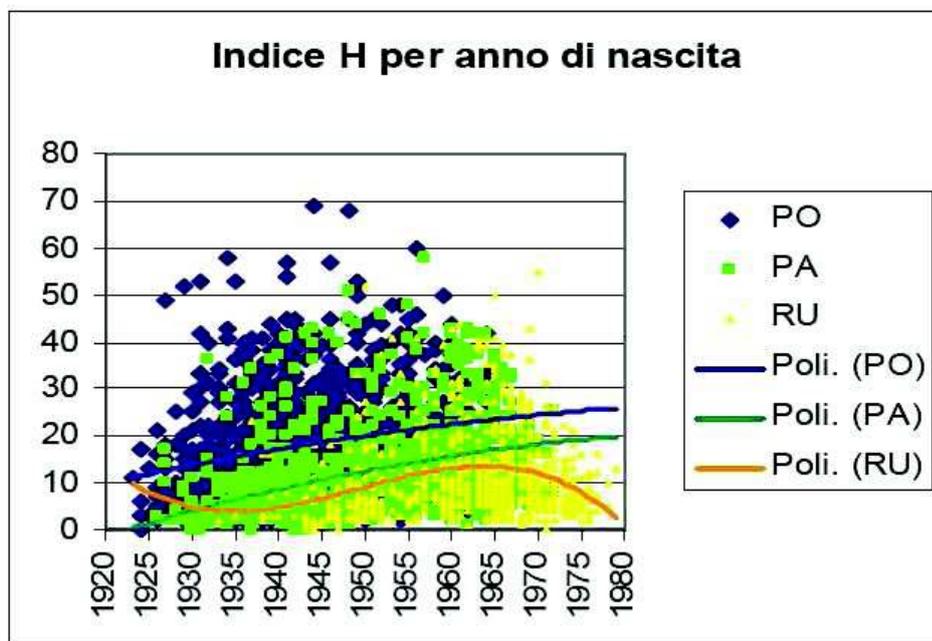
UN ESEMPIO SULL'AREA FISICA

Paolo Rossi (2008) con questo esempio ha voluto analizzare la produttività scientifica dei fisici italiani mediante l'uso degli indicatori bibliometrici e in particolar modo dell'h-index.



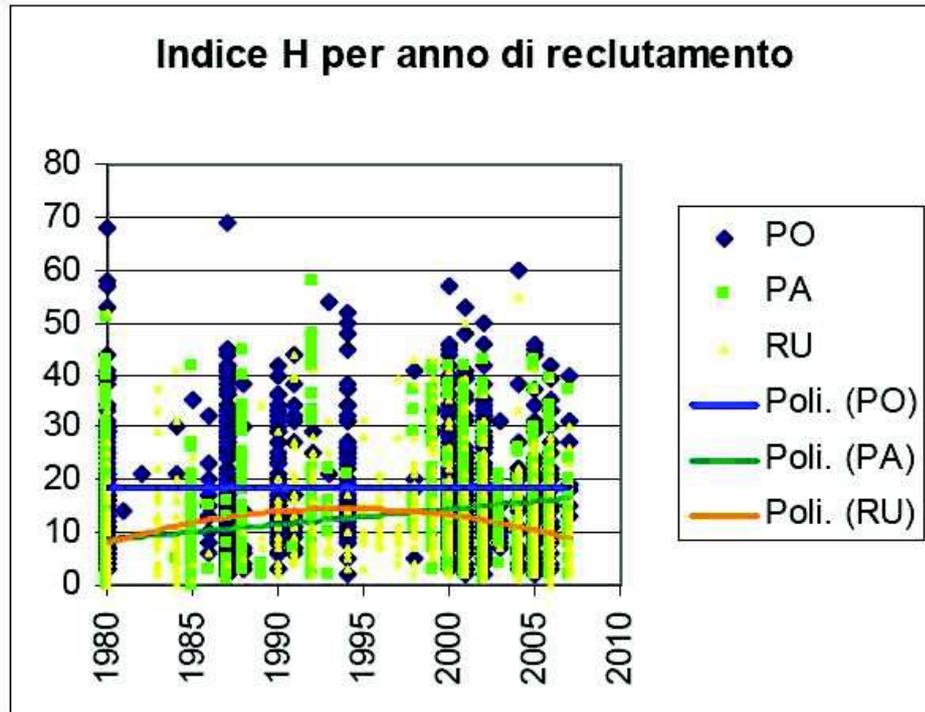
Dal grafico si può osservare che la curva ha una distribuzione di tipo gamma. Il valore medio è circa 14, mentre il valore più probabile è 6.5, dunque nella curva si nota una asimmetria verso sinistra. Nella parte destra invece è molto evidente un calo esponenziale per i valori alti dell'indice h.

Sarebbe più opportuno fare questo studio in base all'anno di nascita degli scienziati in quanto l'indice h tiene conto della loro produttività in base agli anni lavorativi, infatti, basti pensare che un ricercatore con anni di esperienza alle spalle avrà prodotto un numero maggiore di lavori rispetto a un neo-ricercatore e perciò il primo avrà anche più citazioni rispetto al secondo.



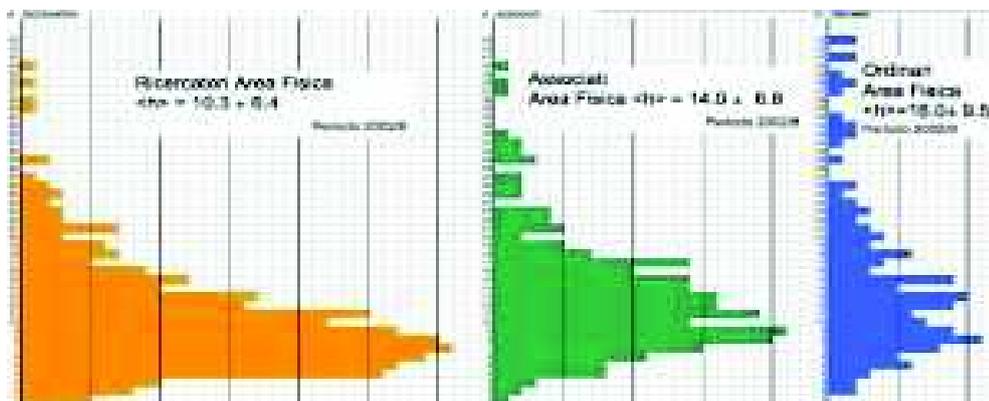
Da questo grafico, che illustra i valori dell'indice h in base all'anno di nascita degli scienziati, si denota una crescita dei valori dell'indice per quei ricercatori scientifici nati attorno agli anni 30, i quali nel corso della loro vita hanno prodotto un numero cospicuo di pubblicazioni e perciò ricevuto un numero sempre maggiore di citazioni. Mentre coloro che sono nati successivamente riscontrano un declino dei valori dell'h-index a causa del loro limitato periodo di produttività scientifica. Si osserva, infatti, che gli ordinari hanno valori maggiori dell'h-index perché sono nati prima dei neo-ricercatori, che invece hanno valori minori.

L'anno di reclutamento inoltre mette in luce i valori medi dell'indice per i ricercatori, per gli ordinari e gli associati.



Dal grafico qui sopra si osserva che la linea blu, che rappresenta gli ordinari, mostra una stabilità per questo gruppo di individui con un valore medio di H pari a 20. La linea verde, che identifica gli associati, denota una crescita tendenziale con un valore medio pari a  $H = 15$ , ciò accade perché negli anni '80 ci sono stati pochi reclutati. Infine per quanto riguarda i ricercatori, linea arancione, c'è una crescita fino all'anno '95 e poi una successiva diminuzione con il valore medio del h-index pari a 10; il calo è probabilmente determinato dalla nascita recente di essi e quindi da una vita lavorativa breve (visibile nel grafico precedente).

Infine le distribuzioni per i tre gruppi di individui studiati in questo esempio sono le seguenti:



La parte del grafico colorata in arancione individua i ricercatori, i cui valori di h sono compresi tra 10.3 e 6.4. il valore più basso per i neo-ricercatori parte da  $h=1$ , cioè essi hanno prodotto un articolo citato almeno una volta. In verde invece troviamo gli associati con indice di Hirsch compreso tra il valore 14 e 8.8. anche in questa categoria il valore minimo corrisponde ad 1. Infine per gli ordinari, in blu, l'intervallo del h-index è tra 18 e 9.5. in questo gruppo di individui però il valore più basso è 2, ciò significa che essi hanno prodotto almeno 2 articoli, i quali sono stati citati due volte.

Paolo Rossi (2008) con questo esempio ha voluto sottolineare valori relativamente bassi dell'h-index se confrontati con i rispettivi valori medi per ogni gruppo di persone analizzate sopra. Probabilmente questo aspetto è influenzato dalle caratteristiche del campo accademico della fisica, all'interno del quale gli scienziati impiegano molto tempo per le loro ricerche ed inoltre è altrettanto difficile fare nuove scoperte.

### **3.4 PUNTI DI FORZA DELL'H-INDEX**

L'indice è facile da calcolare e da trovare, grazie ai database elettronici, tra essi Google Scholar è gratuito.

Hirsch (2005) afferma inoltre che l'h-index è un valido indicatore bibliometrico perché si basa su fondamenti statistico-matematici.

L'indice viene calcolato in base all'importanza dell'impatto delle riviste o degli articoli e non in base alla quantità di essi. Infatti uno studioso potrebbe aver prodotto una miriade di lavori, senza che questi abbiano riscontrato popolarità nei lettori. A differenza di uno scrittore che ha fatto poche pubblicazioni, ma ha ottenuto grande successo.

Le citazioni vengono valutate in base alla tipologia dell'opera. Infatti Hirsch (2005) aveva diagnosticato per tempo che ci sono alcune riviste maggiormente citabili

rispetto ad altre. Ciò dipende dall'area tematica nella quale esse rientrano perché ci sono discipline connesse a molte altre che godono di conseguenza di maggior notorietà semplicemente perché sono più diffuse. Per esempio i lavori in ambito medico avranno sicuramente una diffusione e un impatto maggiore se confrontati con qualsiasi altro campo accademico, perché essi hanno un'importanza maggiore sull'intera comunità.

L'indice non include le auto-citazioni nel calcolo. Infatti Hirsch (2005) ha voluto eliminare le auto-citazioni perché pensava che esse influenzassero il valore dell'h-index, in quanto gli stessi scrittori hanno la possibilità di citare i propri lavori.

Le citazioni variano con l'età del ricercatore, perché gli scienziati più anziani avranno prodotto, nella loro vita, un numero di articoli superiore a confronto dei neo-ricercatori. E l'h-index valuta i lavori dei singoli ricercatori considerando l'età lavorativa di ognuno di essi.

L'indice di Hirsch è in grado di confrontare l'operato di scienziati appartenenti a diversi campi accademici, anche se è più opportuno fare ricerche all'interno di uno stesso settore per garantire l'omogeneità nei dati.

L'h-index si può calcolare sia per il singolo ricercatore sia per un gruppo di ricercatori, grazie alla struttura delle banche dati.

Infine, dato che esso è in grado di contestualizzare la carriera degli autori, l'h-index serve ai redattori e agli editori di periodici a trovare critici qualificati nella valutazione delle riviste.

### **3.5 PUNTI DI DEBOLEZZA DELL'H-INDEX**

L'indice è diffuso particolarmente nelle scienze matematiche e fisiche e meno in altri campi accademici.

Così come l'IF, anche l'h-index viene calcolato soprattutto per le riviste in lingua inglese o pubblicate dalle grandi case editrici nord-americane. Non considerando così lavori importanti pubblicati in altre lingue meno divulgative.

I ricercatori che hanno avuto una carriera lavorativa breve sono penalizzati perché l'h-index non tiene conto dei loro lavori, in quanto essi sono un numero molto basso, di conseguenza non si valuta nemmeno l'effetto che questi scienziati hanno determinato.

Oltre a ciò l'indice non esamina il contesto della citazione, nel senso che alcune citazioni vengono fatte solamente per agevolare un'introduzione. Questo determina un aumento dei valori dell'h-index, escludendo l'importanza e il significato delle citazioni.

L'h-index non distingue gli articoli pubblicati da un singolo autore e quelli scritti da un gruppo di ricercatori. Per questo motivo, nei lavori prodotti da un insieme di persone, coloro che hanno collaborato in minima parte vengono valutati allo stesso modo di chi invece ha contribuito maggiormente alla ricerca. Per esempio i professori universitari, che scrivono articoli ed emettono pubblicazioni, sono citati abbondantemente dai loro studenti, ottenendo pertanto un elevato valore dell'indice, di gran lunga superiore al numero delle loro pubblicazioni.

Per quanto riguarda le grandi collaborazioni, basate non solo sulle pubblicazioni ma anche su conferenze e incontri, conseguono un indice h alto perché gli autori degli articoli sono ben noti e importanti e di conseguenza molto citabili dagli altri ricercatori.

Secondo Egghe (2006) l'indice calcolato per le donne scienziate è distorto perché esse sono in numero inferiore rispetto alla controparte e quindi la quantità delle loro pubblicazioni è più bassa. L'indice di Hirsch è distorto anche per quanto riguarda la zona geografica di residenza perché m (tasso di crescita dello scienziato) è influenzato dalla zona nella quale vivono gli autori. Infine Leo Hegghe (2006) sottolinea il fatto che l'indice non tiene conto delle pubblicazioni che hanno ricevuto tante citazioni nel passato, ma si conteggiano solo quelle del periodo preso in esame, pertanto l'autore in questione perde la fama e la gloria che aveva ottenuto precedentemente.

## 3.6 INDICI ALTERNATIVI

### H-B-INDEX

Questo indice esposto da Banks (2009) valuta le diversità tra gli studiosi che trattano temi maggiormente visibili, perciò più citabili, e quelli che invece producono lavori in filoni meno visibili.

L'h-b-index è l'applicazione dell'h-index, ma basato su uno specifico argomento. Inoltre questo indicatore bibliometrico è definito allo stesso modo dell'indice di Hirsch.

### G-INDEX

Questo indice è stato presentato con lo scopo di superare il limite dell'h-index che non considera gli articoli più citati. Il g-index invece annovera nel calcolo anche gli articoli più citati.

Esso viene definito in questo modo:

*Un insieme di articoli ha indice g se gli articoli più citati g hanno complessivamente un numero di citazioni pari a  $g^2$ , con g maggiore o uguale ad h.*

In formule: 
$$g\text{-index} = \left(\frac{\alpha-1}{\alpha-2}\right)^{\frac{\alpha-1}{\alpha}} h$$

con  $\alpha > 2$  altrimenti non avrebbe senso la formula.

### A-INDEX

L'a-index corrisponde alla media del numero di citazioni ricevute da lavori provenienti dalle pubblicazioni dell'h-index.

La formula di questo indicatore è: 
$$a\text{-index} = \frac{1}{h} \sum_{j=1}^h cit$$

dove cit = numero delle citazioni.

Infine se uniamo l'indice h e g otteniamo il numero di esiti positivi di un ricercatore relativamente alle sue pubblicazioni e citazioni.

## *Capitolo 4 - Usage Factor*

L'usage factor (UF), chiamato anche fattore di utilizzo, è un indicatore del tutto nuovo, originato in seguito al fenomeno dell'open access. Esso ha lo scopo di misurare la qualità dei periodici scientifici in base all'uso delle risorse elettroniche, per questo motivo viene definito fattore di utilizzo. Quindi si può dire che l'UF è generato dai lettori che utilizzano un motore di ricerca elettronico, in quanto si valuta il numero di download per una determinata rivista. Questo nuovo metodo è molto diverso dal conteggio delle citazioni per indici bibliometrici quali l'IF o l'H-index, perciò si afferma che l'UF è complementare ad essi, ma non sostitutivo.

### **4.1 MODELLO OPEN ACCESS**

L'open access (OA), nominato sopra, è un motore di ricerca mediante il quale si vuole liberalizzare la letteratura scientifica mettendola a disposizione di tutti. Infatti, attraverso l'uso dell'OA si può scaricare, leggere, copiare, ricercare o creare link; con le uniche limitazioni che venga mantenuto il controllo dell'autore sull'opera e che venga riconosciuto il diritto d'autore.

Per mezzo della tecnica dell'Open Access i ricercatori hanno l'opportunità di fare analisi bibliometriche per valutare l'impatto della ricerca su uno scienziato o un

gruppo di ricercatori, con assoluta trasparenza nel processo di valutazione perchè i dati sono disponibili a tutti i lettori del web. Questo processo viene effettuato nei seguenti modi: analisi citazionale, che calcola il numero delle citazioni espresse da altri ricercatori, analisi dei download, che conteggia il numero di volte che viene scaricato un articolo, e infine valutazione tramite blogs.

Numerosi sono i vantaggi provenienti dal modello di accesso aperto:

- permette la massima diffusione e quindi il massimo impatto degli articoli
- grazie all'uso del web l'autore ha meno costi di pubblicazione
- attraverso internet i lettori di ogni genere hanno una maggiore facilità e rapidità di accesso agli articoli
- spinge gli scienziati a pubblicare nuovi articoli perché si divulgano prima e quindi hanno maggiore probabilità di ottenere più citazioni/download
- citazioni più immediate legate all'istantaneità di internet
- partecipa alla valutazione della ricerca scientifica, attraverso nuovi indicatori tra i quali l'usage factor.

Però anche l'OA ha i suoi limiti:

- difficoltà nel definire e misurare l'impatto di un articolo di un ricercatore scientifico
- complicazioni nel confrontare il modello di accesso aperto con altri metodi
- criticità nel ottenere set di dati comparabili con altri indici
- problemi legati alle differenze disciplinari perché il modello non suddivide i diversi campi accademici, ma li valuta tutti insieme.

Per concludere, alcuni studi di Norris-Oppeheim-Rowland (2008) hanno registrato che le riviste con alto IF hanno un maggior numero di articoli presenti nel mondo dell'OA, che le riviste elaborate da un gruppo di scienziati hanno maggior probabilità di essere disponibili nell'OA e infine che un articolo o una rivista con un più grande numero di citazioni ha più facilità di essere un articolo o una rivista OA.

## 4.2 CALCOLO DELL'USAGE FACTOR

### 4.2.1 ANALISI QUANTITATIVA

Un'analisi quantitativa ha cercato di studiare in che modo la disponibilità di documentazione elettronica ha aumentato la produzione scientifica, ponendo in relazione il numero di risorse a disposizione, l'utilizzo di esse e il numero di articoli prodotti. Più esattamente nel modello OA si relazionano il numero di riviste presenti nel web, il numero di articoli scaricati e la quantità dei lavori scientifici. Cioè si confronta il numero di articoli scaricati con il numero di articoli pubblicati e in questo modo si valuta la ricerca scientifica e la sua disponibilità nel mondo di internet.

La definizione dell'indice, quindi, è la seguente:

$$UF = \frac{\text{numero articoli scaricati}}{\text{numero articoli pubblicati on-line}}$$

E' necessario che gli articoli siano compatibili, cioè pubblicati con lo stesso metodo standard (in questo caso ci si riferisce a COUNTER, un programma per pubblicare i propri lavori), e li si osserva in uno specifico intervallo temporale.

In questo metodo di analisi, però, ci sono delle problematiche legate all'uso delle risorse perché i dati potrebbero non essere compatibili tra di loro. Infatti, i vari ricercatori usano diversi tipi di programmi per la pubblicazione dei loro articoli e inoltre le pubblicazioni nel modello OA non possiedono dati statistici in grado di misurare il numero dei download dei lettori.

Le difficoltà sono riscontrabili pure nella produzione scientifica per il fatto che le pubblicazioni sono di diverse tipologie e perciò non si possono confrontare con altre di diverso prototipo, e ancora le complessità legate al gap temporale tra il download e le pubblicazioni dell'articolo.

## 4.2.2 ANALISI QUALITATIVA

Oltre all'analisi quantitativa, c'è un'analisi qualitativa con lo scopo di esaminare se migliorando la disponibilità e la facilità di utilizzo dei dati messi a disposizione del ricercatore, aumenta la quantità degli articoli.

Questo studio si svolge recuperando i dati, relativi alle citazioni, dalle biblioteche. A volte purtroppo si verificano delle difficoltà perché i dati non sono trattati in modo automatico e anche per il fatto che molto spesso il ricercatore non indica la natura della fonte che ha adoperato, perciò mancano dei dati importanti. Per risolvere questi problemi si può svolgere l'analisi su un campione di dati e da questo estendere i risultati alla totalità del materiale.

L'analisi qualitativa, con l'aiuto dell'impatto delle citazioni, è in grado di controllare se migliorando l'offerta aumenta la quantità degli articoli pubblicati, le citazioni degli autori e il relativo h-index. Le uniche problematiche connesse a questo studio sono: complicazioni nel separare ambiti disciplinari diversi, presenza di autocitazioni che causano distorsioni nella valutazione dell'autore e infine assenza di metodi capaci di distinguere le citazioni positive da quelle negative, le quali determinano ulteriori alterazioni nella stima della qualità dell'autore.

L'influenza della produzione scientifica nel modello OA si determina tenendo in considerazione i seguenti punti:

1. analisi delle bibliografie citate nei lavori;
2. osservazione delle lingue maggiormente citate (inglese);
3. studio delle caratteristiche formali delle risorse elettroniche:
  - i) titolo degli articoli, infatti un titolo corto e chiaro attrae il lettore molto più di uno lungo e difficile da capire, in quanto richiede più attenzione;
  - ii) impaginazione e qualità degli articoli, infatti i testi scritti in modo corretto, ben strutturati e con le adeguate spaziature invogliano il lettore a soffermarsi alla lettura;
4. valutazione della qualità delle fonti on-line: qualità della rivista, calcolo dell'IF degli articoli, conoscenza dell'autore e dell'editore, indicizzazione, ect;

5. previsione del plagio, perchè nel web è molto facile eseguire il “copia-incolla” dei lavori scientifici.

### **4.3 PUNTI DI FORZA DELL'USAGE FACTOR**

L'usage factor è un indice valido per la scelta della cancellazione o dell'acquisto di una rivista, infatti dai download i ricercatori sono in grado di capire se una determinata rivista è interessante e quindi subirà altrettanti scaricamenti dal motore di ricerca o il contrario.

L'usage factor ha una copertura molto più ampia rispetto all'IF, grazie alla sua presenza nel OA. Infatti l'usage factor è disponibile non solo agli scienziati che sono iscritti nel database ISI, ma a tutti i lettori che possiedono una connessione a internet.

L'indice di utilizzo essendo complementare all'impact factor si può dire che si pone in bilanciamento ad esso perché analizza giornali sottovalutati dall'IF e inoltre fa le sue ricerche in uno scenario diverso e molto più ampio, introducendo così nuovi tipi di articoli e riviste.

L'UF ha come peculiarità una facilità e rapidità di utilizzo, favorita dagli archivi open access.

Esso valuta in modo veritiero la produzione scientifica, perchè l'indice può essere visto da tutti in internet.

L'UF contribuisce al cambiamento della comunicazione scientifica, nel senso che si vuole renderla disponibile a tutti.

## 4.5 PUNTI DI DEBOLEZZA DELL'USAGE FACTOR

Un punto critico dell'usage factor è la difficoltà nel calcolarlo e in modo particolare nel fare paragoni con gli altri indici, perché i vari scienziati usano diversi programmi per le pubblicazioni. Perciò manca la caratteristica dell'omogeneità per fare i confronti tra gli autori delle riviste, a causa dell'incompatibilità dei vari sistemi di pubblicazione.

Complicazioni nel determinare il range temporale per l'analisi, si propone 2 anni, ma sarebbe più appropriato considerare un intervallo temporale di 5 anni con il fine di analizzare i picchi generati dal momento della pubblicazione dell'articolo e l'inizio delle citazioni.

Con il sistema OA gli autori sono favorevoli a pubblicare le loro riviste con questo metodo, ma d'altra parte c'è lo scontento degli editori, causato dal fatto che l'open access non ha scopi di lucro.

Anche in questo indice bibliometrico, come nell'IF o nell'h-index, c'è la presenza di autocitazioni che potrebbero aumentarne erroneamente il calcolo.

I confronti sono difficili da fare, non solo a causa dei vari metodi di pubblicazione delle riviste, ma anche per la scarsa suddivisione degli ambiti disciplinari.

Infine non c'è un modo per distinguere le citazioni positive da quelle negative, le quali contribuiscono ad alzare in modo sbagliato l'indice, così come avviene per l'impact factor.

## 4.5 INDICI ALTERNATIVI

### WEB IMPACT FACTOR

L'indice WIF è stato creato da T. C. Almind e P. Ingwersen (2004) per approfondire la produzione scientifica nel web. Il calcolo dell'indice si costruisce mediante l'analisi dei link, esso si definisce come segue:

$$\text{WIF} = \boxed{\text{(numero contatti ricevuti da un documento pubblicato in rete)}} * \boxed{\text{(numero pagine del documento)}}$$

## *Capitolo 5 – Eigenfactor*

L'Eigenfactor, abbreviato in EI, è un indicatore bibliometrico in grado di misurare il numero di volte che gli articoli pubblicati su una rivista negli ultimi 5 anni sono stati citati nell'anno considerato.

L'EI è un progetto di ricerca accademica non commerciale, sponsorizzato dal laboratorio del professor Carl Bergstrom (2007) nel dipartimento di biologia presso l'università di Washington. Bergstrom (2007), con la collaborazione del suo team, ha utilizzato i recenti progressi della rete di analisi e teoria dell'informazione per lo sviluppo di un nuovo metodo per la valutazione dell'influenza delle riviste accademiche e per il mapping della struttura della ricerca accademica. Tale indice è disponibile nel sito web [www.eigenfactor.org](http://www.eigenfactor.org) che rileva le informazioni all'interno della vasta rete di pubblicazioni creata dalla letteratura accademica, le quali sono collegate tra loro da citazioni bibliografiche.

L'Eigenfactor misura l'influenza delle riviste allo stesso modo dell'algoritmo PageRank di Google sui ranghi d'influenza delle pagine web. Questo metodo attribuisce un giudizio d'importanza ad ogni pagina in base alla rete dei collegamenti ipertestuali tra pagine web. In questo modo una pagina è importante se a sua volta viene citata da altre pagine importanti. Analogamente per l'EI, le riviste sono considerate influenti se sono spesso citate da altre riviste influenti.

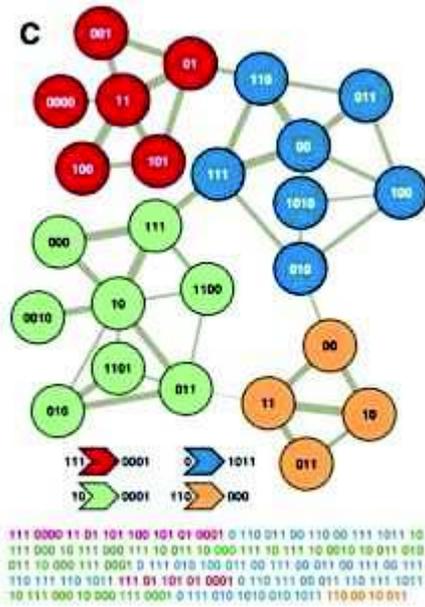
## 5.1 METODO DI RICERCA

L'eigenfactor è una misura dell'importanza della totalità delle riviste sulla comunità scientifica. Il punteggio EI di una rivista viene definito come la stima della percentuale di tempo nel quale gli utenti di una libreria spendono con quella rivista. L'algoritmo EI corrisponde ad un semplice modello di ricerca nel quale i lettori cercano le citazioni come se si stessero muovendo da una rivista all'altra. Infatti questo algoritmo calcola la traiettoria di un'ipotetica camminata di un ricercatore che inizialmente va in una libreria e seleziona, in modo casuale, un articolo di una rivista. Dopo aver letto l'articolo lo scienziato sceglie una citazione, sempre in modo casuale, e va in cerca della rivista citata. Da questa nuova rivista egli seleziona un altro articolo a caso, dal quale poi sceglie un'altra citazione per la prossima rivista. Il ricercatore fa questo procedimento, meglio definito come camminata, all'infinito. L'algoritmo EI si avvale della matematica per simulare questo procedimento. Inoltre la frequenza con la quale il ricercatore visita ogni rivista, durante la sua camminata, ci fornisce una misura dell'importanza della rivista all'interno della rete delle citazioni accademiche.

Il mapping, nominato sopra, consiste proprio nel disegnare i legami tra le varie riviste, esplicitando le relazioni esistenti, quindi rappresentando tutti i collegamenti citazionali. Questi legami sono visibili per mezzo di nodi e connessioni a moduli.

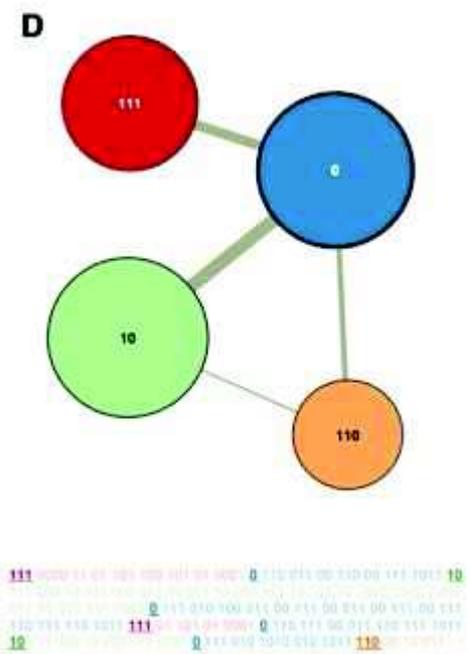


FIG C:



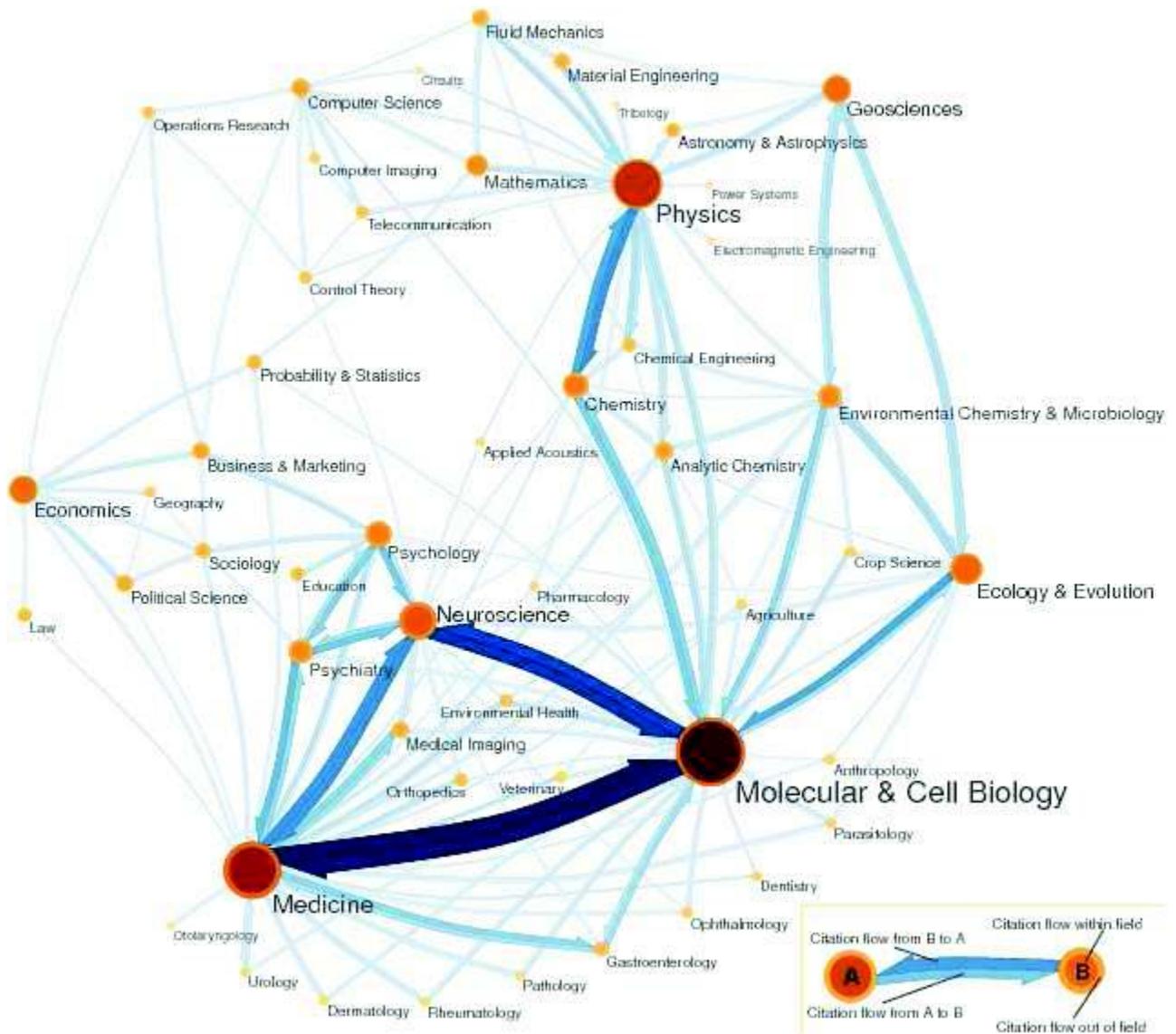
La figura C descrive la camminata sulla rete a due livelli, dove i grandi gruppi ricevono nomi univoci, ma i nomi dei nodi nei gruppi non sono univoci, infatti essi vengono riutilizzati. Così si risparmia il 32% della descrizione della rete.

FIG D:



Nella figura D vengono riportati solo i nomi dei moduli e non le posizioni all'interno del modulo, ciò offre la descrizione modulare della rete e serve da mappa.

## Mappa delle scienze



I cerchi arancioni rappresentano i campi accademici, i cerchi più grandi e più scuri sono la dimensione del campo misurata dall' algoritmo EI, le frecce blu riproducono il flusso di citazione tra i campi, infine le frecce più grandi e più scure descrivono il più elevato volume di citazioni.

## 5.2 METODO PER IL CALCOLO

Innanzitutto per il calcolo dell'indice Eigenfactor si considera il database ISI e i dati del JCR a partire dal 2006.

Le tappe per il conteggio dell'EI sono spiegate nei prossimi paragrafi.

### 5.2.1 DATA INPUT

Abbiamo bisogno di 4 inputs, di cui 2 archivi e 2 costanti:

\_ ARCHIVIO DELLE RIVISTE: viene utilizzato il CD-ROM della versione del Thomson Reuters dove troviamo una lista delle riviste ISI e una del JCR che contiene solo le scienze naturali e le scienze sociali, si combinano queste due liste. Poi si contano le citazioni date nel periodo considerato, cioè il 2006, a riviste pubblicate 5 anni prima, ovvero nel periodo 2001-2005.

\_ ARCHIVIO DEGLI ARTICOLI: archivio che contiene il numero di articoli prodotti da ogni rivista nei 5 anni prima aver ricevuto citazioni. A differenza degli articoli di Thomson Reuters che considera come target-window solamente 2 anni. Dato che Thomson perfeziona continuamente i suoi dati, è sempre meglio utilizzare il CD più recente il quale contiene misure più vicine alla realtà odierna.

\_  $\alpha$ : costante posta pari a 0.85.

\_  $\epsilon$ : costante pari a 0.00001.

### 5.2.2 CREARE LA MATRICE DI CROCE-CITAZIONE

La rete delle citazioni delle riviste può essere rappresentata con una matrice, denominata  $Z$ , dove  $Z_{ij}$  rappresenta il numero di volte che gli articoli pubblicati nella rivista  $j$  durante un certo periodo, citano articoli nella rivista  $i$  pubblicati 5 anni prima. La matrice  $Z$  è una matrice quadrata di dimensione  $n \times n$ , dove  $n$  è il numero delle sole riviste estratte dal database ISI.

Esempio: date le riviste A, B, C, la matrice Z è la seguente.

	A	B	C
A	2	0	3
B	4	1	1
C	0	2	7

Da sinistra verso destra si leggono il numero delle citazioni ricevute, mentre dall'alto verso il basso troviamo il numero di citazioni fatte. Dunque la rivista A cita 2 volte se stessa, cita 4 volte la rivista B e non cita C. Inoltre A ha ricevuto 3 citazioni dalla rivista C e nemmeno una da B. Analogamente per le altre riviste.

### 5.2.3 MODIFICARE LA MATRICE DI CROCE-CITAZIONE

Una volta creata la matrice di croce-citazione Z, bisogna fare alcune modifiche a quest'ultima al fine di calcolare l'eigenfactor. Inizialmente occorre porre tutti zeri nella diagonale ( $Z_{ii}=0$ ) così si escludono le auto-citazioni. È importante che queste non siano presenti perché si evita il gonfiamento della rivista determinata da comportamenti opportunistici, inoltre perché non ci sono informazioni sulle auto-citazioni per riviste non appartenenti al JCR.

Successivamente si devono normalizzare le colonne della matrice Z, dividendo ciascuna entrata di ogni colonna con la somma di tutte le colonne, così da stimare la somma delle colonne per ogni colonna j, come  $Z_j = \sum_i Z_{ij}$ . Poi si dividono le entrate di ogni colonna con la corrispondente somma delle colonne e infine si mettono tutte queste entrate in una nuova matrice H, dove  $H_{ij} = Z_{ij} / Z_j$ .

Esempio: sia  $Z$  la seguente matrice, dove i numeri dentro le parentesi rappresentano il numero di libri che ogni rivista ha pubblicato.

	A	B	C	D	E	F
A(3)	1	0	2	0	4	3
B(2)	3	0	1	1	0	0
C(5)	2	0	4	0	1	0
D(1)	0	0	1	0	0	1
E(2)	8	0	3	0	5	2
F(1)	0	0	0	0	0	0

Ora viene modificata la matrice  $Z$ , ponendo tutti zero sulla diagonale principale.

	A	B	C	D	E	F
A(3)	0	0	2	0	4	3
B(2)	3	0	1	1	0	0
C(5)	2	0	0	0	1	0
D(1)	0	0	1	0	0	1
E(2)	8	0	3	0	0	2
F(1)	0	0	0	0	0	0

Infine, si ottiene la matrice  $H$  che corrisponde alla matrice  $Z$ , ma con le colonne normalizzate.

	A	B	C	D	E	F
A(3)	0	0	$2/7$	0	$4/5$	$1/2$
B(2)	$3/13$	0	$1/7$	1	0	0
C(5)	$2/13$	0	0	0	$1/5$	0
D(1)	0	0	$1/7$	0	0	$1/6$
E(2)	$8/13$	0	$3/7$	0	0	$1/3$
F(1)	0	0	0	0	0	0

## 5.2.4 IDENTIFICAZIONE DEI NODI

Alcune riviste non citano le altre riviste. Esse sono chiamate “dangling nodes”, cioè nodi penzolanti, e possono essere identificate osservando la matrice Z. Le colonne che contengono tutti zeri sono esattamente i nodi. Queste colonne devono essere scritte come un vettore riga composto da un solo 1 e tutti gli altri elementi 0. Questo vettore lo chiameremo  $d$ . L’elemento 1 sta a significare che una rivista è un nodo, mentre gli 0 rappresentano nodi non penzolanti.

Esempio.

	A	B	C	D	E	F
$d_i$	0	1	0	0	0	0

## 5.2.5 CALCOLARE IL VETTORE D’INFLUENZA

A questo punto è necessario costruire una matrice P di transizione e da essa calcolare il suo miglior eigenvector normalizzato  $\pi^*$ , in modo tale che i suoi componenti diano somma 1. Questo vettore ci fornisce il peso delle riviste utilizzate nel calcolo dell’EI. Per calcolarlo esigiamo dei seguenti elementi:

$\alpha$ : è un vettore colonna che rappresenta il numero degli articoli pubblicati in ogni rivista nei 5 anni di target-window, inoltre è un vettore normalizzato. Definiamo anche  $A_{tot}$  come il numero totale di articoli pubblicati da tutte le riviste.

Esempio: (sommiamo i numeri che nell’esempio di prima si trovavano tra parentesi)

$$A_{tot} = 3+2+5+1+2+1 = 14$$

	A	B	C	D	E	F
$\alpha^T_i$	3/14	2/14	5/14	1/14	2/14	1/14

Allora

$\underline{\pi}^{(0)}$  : è un vettore colonna applicato in ripetizione al vettore d'influenza. Ogni elemento del vettore viene posto uguale a  $1/n$ , dove  $n$  identifica il numero delle riviste.

	A	B	C	D	E	F
$\pi_i^{(0)T}$	1/6	1/6	1/6	1/6	1/6	1/6

Esempio:

$\underline{\pi}^*$ : è il miglior eigenvector normalizzato della matrice  $P$  e definisce l'influenza di una rivista, in altre parole esso corrisponde alla frazione di tempo spesa per ogni rivista in  $P$ .

La matrice di transizione  $P$  è così definita:  $P = \alpha H' + (1-\alpha)\alpha e^T$  dove  $e^T$  è un vettore riga di tutti 1,  $\alpha e^T$  è una matrice con colonne identiche ad  $\alpha$  e  $H'$  è la matrice  $H$  sopra descritta, con tutte le colonne equivalenti ai nodi e sostituite dal vettore  $\alpha$ .

Esempio:

	A	B	C	D	E	F
A	0	3/14	2/7	0	4/5	3/6
B	3/13	2/14	1/7	1	0	0
C	2/13	5/14	0	0	1/5	0
D	0	1/14	1/7	0	0	1/6
E	8/13	2/14	3/7	0	0	2/6
F	0	1/14	0	0	0	0

Calcolare il leader eigenvector dalla matrice stocastica  $P$  è un procedimento troppo lungo, perché ci sono troppe moltiplicazioni matriciali con matrici di grandi dimensioni. Un metodo alternativo, molto più facile e sbrigativo, è quello di operare sulla matrice  $H$ . Per stimare il vettore d'influenza si adopera la seguente equazione:  $\pi^{(k+1)} = \alpha H \pi^{(k)} + [\alpha d \pi^{(k)} + (1-\alpha)]\alpha$ .

Questa interazione convergerà all'eigenvector di  $P$ . Per trovare il vettore dell'EI si ripetono varie iterazioni e dopo ognuna di queste si osserva se il residuo  $\tau = \pi^{(k+1)} - \pi^{(k)}$  è minore di  $\epsilon$  ( con  $\epsilon=0.00001$ ). Se è così, allora  $\pi^* \approx \pi^{(k+1)}$ .

Usufruento dell'esempio di prima, otteniamo che il vettore stazionario converge dopo 16 iterazioni al precedente vettore con  $\alpha = 0.85$  e  $\varepsilon = 0.00001$ .

	A	B	C	D	E	F
$\pi_i^*$	0,3040	0,1636	0,1898	0,0466	0,2753	0,0206

### 5.2.6 CALCOLARE L'INDICE EIGENFACTOR

Il valore dell'eigenfactor per ogni rivista, è dato dal prodotto della matrice H con il vettore  $\pi^*$  normalizzato, poi si moltiplica per 100 per calcolarne la percentuale.

In formule:  $EI = 100 \frac{H\pi^*}{\sum_i(H\pi^*)}$

Il punteggio  $EI_i$  della rivista  $i$  è definito come la percentuale del totale delle citazioni pesate, che la rivista  $i$  riceve dalle 7611 fonti presenti nel database.

Esempio:

	A	B	C	D	E	F
$EI_i$	34,05	17,20	12,18	3,65	32,92	0,00

### 5.2.7 EMETTERE I RISULTATI

Infine i risultati ottenuti vengono sistemati in Excell, mettendo i seguenti parametri: anno della pubblicazione della rivista, nome corto dell'autore, nome lungo dell'autore, gruppo (scienze naturali o scienze sociali), campo accademico, EI, AI, IF, numero totale degli articoli pubblicati in 5 anni, numero totale delle citazioni ricevute in 5 anni.

## 5.3 PUNTI DI FORZA

Prima di tutto gli algoritmi utilizzati per il calcolo dell'eigenfactor, valutano l'importanza delle riviste mediante la struttura dell'intera rete, la quale è molto vasta. Questo indice è in grado di fornire informazioni sul prezzo e sul valore di migliaia di riviste accademiche, così che le biblioteche o i ricercatori hanno la possibilità di scegliere in modo più sicuro le riviste da acquistare.

Il sito web [www.Eigenfactor.org](http://www.Eigenfactor.org) contiene non solo elenchi di riviste accademiche, ma anche tesi, riviste popolari, giornali in genere, ect, in questo modo si valuta più facilmente il divario tra le riviste delle scienze sociali e naturali, dalle quali poi si calcola l'EI. Precisamente l'indice tiene presente le differenze di prestigio tra le riviste citanti e si adatta per i vari modelli di citazione tra le discipline. Infatti le diverse discipline hanno vari standard e scale di tempo in cui valutare le citazioni. Per esempio l'articolo di una rivista di biologia cellulare potrebbe ricevere 30 citazioni entro 2 anni, mentre l'articolo di una rivista leader di matematica potrebbe ricevere solamente 2 citazioni nello stesso periodo. L'algoritmo dell'eigenfactor nota subito queste differenze e automaticamente permette un miglior confronto tra le differenti aree di ricerca.

Un grande vantaggio dell'EI è osservare le citazioni della rivista in un intervallo temporale di 5 anni, non ci si limita ai soli 2 anni dell'impact factor. Questo è un punto di forza fondamentale perché in molti campi accademici, le riviste o gli articoli non vengono citati fino a parecchi anni dopo la loro pubblicazione.

Infine è importante sottolineare che il sito dal quale si calcola l'indice è gratuito, facile da usare e da scaricare.

## 5.4 METODI ALTERNATIVI

L'indice alternativo all'EI è l'ARTICLE INFLUENCE (AI), ovvero l'influenza di un articolo, che appunto misura l'influenza media delle citazioni per articolo della rivista  $i$ , 5 anni dopo la sua pubblicazione.

Esso è composto dal punteggio dell'EI fratto 100, diviso dalla frazione di tutti gli articoli che ogni rivista ha pubblicato.

In formule:  $AI_i = 0.01 \frac{EI_i}{\alpha_i}$  dove  $EI_i$  è il punteggio dell'eigenfactor della rivista  $i$ , e  $\alpha_i$  è il vettore articolo normalizzato.

Questo indice nasce perché, mentre l'EI misura l'influenza totale di una rivista, l'AI misura l'influenza per articolo di una rivista. Quindi per confrontare questi due indici è necessario dividere l'influenza della rivista per il numero di articoli pubblicati e in questo modo si ottiene proprio l'AI.

Esempio: (in riferimento all'esempio dell'EI)

	A	B	C	D	E	F
$AI_i$	1,59	1,20	0,34	0,51	2,30	0,00

# *Capitolo 6 - Altri aspetti della bibliometria*

## **6.1 L'IF DELLE RIVISTE STATISTICHE**

Dopo aver illustrato gli indici bibliometrici utilizzati per la valutazione delle riviste, è importante soffermarsi sulla diffusione delle citazioni delle riviste nei vari campi accademici. In altre parole è opportuno analizzare il picco massimo delle citazioni per ogni rivista, contenuta ovviamente nel database ISI (Institute of Scientific Information). Oltre a ciò, si può anche calcolare la percentuale delle citazioni che uno specifico giornale ha ricevuto dopo 2 o 3 anni dalla sua pubblicazione.

E' noto che le riviste statistiche hanno bassi valori di impact factor e inoltre impiegano maggior tempo a raggiungere il picco del numero massimo di citazioni, cioè hanno uno sviluppo più lento nella diffusione delle citazioni rispetto alle riviste di altre discipline.

Nella letteratura accademica l'impact factor è l'indice bibliometrico maggiormente usato per determinare la qualità delle riviste. Nonostante ciò è molto criticato dagli scrittori scientifici perché può essere facilmente manipolato. Per esempio Hernán (2008), Reedijk e Moed (2008), Kostoff (1998), Seglen (1997) e Talamanca (2000) affermano che l'IF può essere gonfiato mediante la pratica dell'auto-citazione. Inoltre le citazioni possono esprimere giudizi sia positivi che negativi, perciò ricevere tante citazioni non sempre implica una maggiore stima dal mondo scientifico, anzi nel caso di citazioni negative può voler dire esattamente il contrario. L'impact factor però non possiede intrinsecamente un criterio per dividere i tipi di

citazioni e di conseguenza anche le citazioni negative possono determinare una distorsione nel calcolo dell'indice.

Tabella 1. 2007 impact factors per riviste statistiche.

classifica IF	Riviste Statistiche	IF
1600	Journal of the Royal Statistical Society Series B-Statistical Methodology	2,21
1741	Journal of the American Statistical Association	2,086
1906	Annals of Statistics	1,944
2326	Journal of the Statistical Society Series A-Statistics in Society	1,654
2827	Journal of the Royal Statistical Society Series C-Applied Statistics	1,362

La tabella 1, presa dal Journal Citation Reports dell'ISI, mostra i più alti punteggi per le riviste di statistica. Il miglior giornale, Journal of the Royal statistical Society Series B-Statistical Methodology, arriva ad un picco massimo di 2.21, posizionandosi al posto 1600 tra tutte le riviste presenti nel database. Però, tra le migliori riviste, nella top 10 gli IFs variano da 69.03, detenuto da A Cancer Journal for Clinicians, ad un 28.75 di Nature. Dunque, da queste classifiche, si può affermare che le riviste di statistica hanno valori di impact factor decisamente molto bassi.

Dalla definizione di impact factor (numero di citazioni nell'anno in corso di articoli pubblicati in una rivista nei due anni precedenti, diviso il numero totale di articoli pubblicati negli stessi anni nella medesima rivista) sembra che una rivista per arrivare a raggiungere alti livelli di IF debba avere i suoi articoli citati spesso e molto velocemente, in particolar modo subito dopo la loro pubblicazione. Questa condizione può essere spiegata dal fatto che un articolo quando viene pubblicato di solito tratta di temi di comune interesse in quel specifico momento. Perciò, dopo un certo periodo di tempo esso potrebbe risultare ormai vecchio e noioso, per cui gli scienziati non sono più interessati a citarlo, ma si dirigono verso altri articoli trattanti nuovi temi interessanti.

Inoltre, Fok e Franses (2007) dichiarano che la velocità di diffusione delle citazioni dipende da alcuni fattori legati alle caratteristiche degli articoli e delle riviste, come per esempio il numero di autori del lavoro che potrebbero aumentare le autocitazioni, il numero di pagine in quanto un articolo più lungo ha maggior probabilità di contenere citazioni rispetto ad uno più corto.

Altri non si sono fermati solamente a commentare, ma Contreras, Edwards e Wizala (2006) hanno addirittura stimato un modello logistico di diffusione delle citazioni per le discipline accademiche. Tale modello prende il nome di CURRENT IMPACT FACTOR (CIF) e stimano anche un impact factor di lungo termine, che rappresenta il valore al quale l'impatto dovrebbe convergere (M). Il rapporto M/CIF misura la diversità dei movimenti (lenti o rapidi) nelle diverse discipline. Purtroppo il modello in questione non include il campo della statistica, perciò non si possono trarre conclusioni sul problema dell'enorme diversità dei valori dell'indice e della diffusione delle citazioni.

Un modello migliore per comparare la statistica con altri campi accademici è il modello di Bass (Bass 1969), semplicemente perché esso contiene la disciplina messa in discussione.

### **6.1.1 DATI**

Per studiare le differenze degli IFs nei vari campi si considerano i dati dell'ISI del 2007 e si includono le seguenti riviste: Journal of the Royal Statistical Society Series B-Statistical Methodology, Journal of the American Statistical Association e Annals of Statistics. Phenal (1999) afferma che l'analisi citazionale è più attendibile quando i dati distribuiti sono più citati, allora si considerano i migliori 20 articoli per ognuna di queste riviste, così da avere un numero considerevole di citazioni per stimare il modello. Si esamina anche la top 25 degli articoli statistici più citati, contenuti in riviste di diverse discipline. Oltre ad esse si includono nello studio i 20 migliori articoli di altre materie. Il numero medio delle citazioni varia di molto tra la top 20 di statistica e quella degli altri settori. Più in dettaglio si osserva che, per gli articoli statistici classificati tra la 11-esima e 20-esima posizione, il numero delle citazioni arriva a 1000, mentre per le tre riviste menzionate sopra questo numero sale a 4000.

Si vuole dimostrare che le differenze trovate non sono causate dalla discrepanza del numero assoluto di citazioni. Alcuni scienziati pensano che questa diversità sia determinata dal modo con il quale si effettuano le citazioni, che varia da rivista a rivista.

La tabella 2, visibile qui sotto, mostra le riviste incluse nell'analisi, il valore degli IFs delle riviste considerate, il numero di libri presi in esame per ognuna di esse, la loro classifica in base al numero delle citazioni ed infine il numero medio di citazioni per i libri studiati.

Tabella 2. Riviste incluse nell'analisi.

Journal	IF	numero di libri	posizione dei libri	numero medio delle citazioni
<i>Statistica</i>				
1-Journal of the Royal Statistical Society Series B-Statistical Methodology	2,21	20	1-20	2462
2-Journal of the American Statistical Association	2,086	20	1-20	3196
3-Annals of Statistics	1,944	20	1-20	701
4-Top 25 dei libri di statistica più citati		23	1-25	7158
<i>Altre discipline</i>				
5-Journal of Biological Chemistry	5,581	20	1-20	1637
6-Nature	28,751	10+10	1-10; 491-500	7152;1008
7-Science	26,372	10+10	1-10; 491-500	8597;1057
8-Water Resources Research	2,154	20	1-20	611
9-Physical Review Letters	6,944	10+10	1-10;91-100	4184;875
10-Enviromental Modelling & Software	2,099	10	1-10	92
Totale		193		2839

## 6.1.2 METODO DI STUDIO

L'approccio all'analisi sopra descritta è di stimare un modello di Bass per ogni serie di momenti del numero di citazioni di un particolare libro. Cioè, il numero di citazioni che un libro ottiene nell'anno  $t$  dalle osservazioni  $Y_t$ , e il numero cumulativo di citazioni nell'anno  $t$  dal totale delle citazioni  $N_t$ .

Il modello di Bass è il seguente:

$$Y_t = mp + (q - p)N_t - \frac{q}{mN_{t-1}^2}$$

Dove  $p$  = parametro innovativo (influenze interne)

$q$  = parametro imitativo (influenze esterne)

$m$  = potenziale numero totale di citazioni, cioè livello di saturazione

In particolare è fondamentale calcolare il tempo al raggiungimento del picco massimo delle citazioni, ovvero  $T^* = \frac{\log(q/p)}{(p+q)}$ .

Chiamando  $N_t = m^*$ , si ridefiniscono i parametri  $p$  e  $q$  come segue:

$$q = -\frac{\log(1-2m^*/m)}{2T^*(1-m^*/m)} \quad \text{e} \quad p = -\left(\frac{2m^*}{m} - 1\right)q$$

quindi il modello diventa il seguente:

$$Y_t = \frac{m(2m^*/m - 1)\log(1 - 2m^*/m)}{2T^*(1 - m^*/m)} - \frac{2m^*/m \log(1 - 2m^*/m)}{2T^*(1 - m^*/m)} N_{t-1} + \frac{\log(1 - m^*/m)}{2mT^*(1 - m^*/m)} N_{t-1}^2$$

Oltre al periodo medio  $T^*$ , che differisce tra le discipline, è opportuno calcolare anche la percentuale delle citazioni che un libro riceve dopo 2 anni dalla pubblicazione. Per ogni anno  $t$  la percentuale è stimata con la funzione  $f(t) = \frac{N_t}{m}$ , cioè il numero cumulativo di citazioni nell'anno  $t$  diviso il numero potenziale di citazioni. Per l'analisi presa in esame, si è interessati a calcolare  $f(2)$  e per completezza  $f(3)$ .

### 6.1.3 TEMPO AL RAGGIUNGIMENTO DEL PICCO T\*

Tabella 3. Valori del raggiungimento al picco.

Journal	n	T* medio	SD T*	P-value	
				T*=2	T*=3
1-Journal of the Royal Statistical Society Series B	15	29,10	13,82	0,000	0,000
2-Journal of the American Statistical Association	18	22,17	10,63	0,000	0,000
3-Annals of Statistics	15	16,98	8,03	0,000	0,000
4-Top 25 dei libri statistici più citati	18	23,70	8,5	0,000	0,000
5-Journal of Biological Chemistry	17	5,16	1,73	0,000	0,000
6-Nature	13	7,00	4,32	0,001	0,006
7-Science	7	8,36	10,24	0,151	0,215
8-Water Resources Research	17	16,11	6,03	0,000	0,000
9-Physical Review Letters	11	11,57	4,14	0,000	0,000
10-Enviromental Modelling & Software	2	5,13	0,52	0,074	0,109

La tabella 3 riporta i valori stimati per T\*, per le riviste considerate nell'analisi e i p-value per un semplice t-test che confronta le medie dei picchi per t = 2 e t = 3. Il test evidenzia che tutte le riviste di statistica hanno picchi significativi dopo 3 anni. La rivista Enviromental Modelling & Software ha tempi di picchi che non sono significativamente diversi da 2 anni. Le riviste classificate nelle ultime 3 posizioni differiscono significativamente da una sola rivista di statistica, cioè Journal of the Royal Statistical Society B.

I libri che si trovano nella top 25 hanno un picco piuttosto tardi (attorno a 24 anni). Alcuni studiosi ritengono che questa lista non sia affidabile perché è composta da libri ben citati, la maggior parte di loro sono relativamente vecchi, infatti l'anno di pubblicazione è compreso in un intervallo temporale che varia tra il 1949 e il 1986. Sicuramente ci saranno libri molto più nuovi che trattano gli stessi argomenti con le opportune modifiche e aggiunte, dovute alle scoperte innovative e alle odierne conoscenze. Dalla tabella appare che le riviste abbiano ricevuto citazioni in un tempo rapido, avendo come conseguenza un alto valore di T\*. Inoltre è sorprendente osservare che la rivista classificata alla decima posizione, che consta di soli 2 articoli nel database ISI, impiega poco tempo al raggiungimento del picco. È appropriato chiedersi come essa faccia a ricevere tutte queste citazioni così velocemente.

### 6.1.4 PERCENTUALE DELLE CITAZIONI NEI PRIMI ANNI

Tabella 4. Percentuale del numero di citazioni nei primi 2 anni.

Journal	Top 10	Altre 10
1-Journal of the Royal Statistical Society Series B-Statistical Methodology	1,15%	0,93%
2-Journal of the American Statistical Association	0,72%	1,41%
3-Annals of Statistics	1,93%	2,05%
4-Top 25 dei libri statistici più citati	0,96%	0,69%
5-Journal of Biological Chemistry	5,33%	7,03%
6-Nature	5,41%	2,82%
7-Science	7,49%	6,68%
8-Water Resources Research	1,66%	2,72%
9-Physical Review Letters	2,27%	3,22%
10-Enviromental Modelling & Software	2,82%	
Totale	2,64%	2,83%

La tabella 4 espone le percentuali del numero delle citazioni che le riviste ricevono dopo i primi 2 anni dal momento della loro pubblicazione. Da essa si può osservare che le riviste Journal of Biological Chemistry, Nature e Science ricevono una percentuale più alta di citazioni nei primi 2 anni, rispetto alle altre riviste considerate nella top 10. Le riviste di statistica, classificate nelle prime 4 posizioni, ricevono una percentuale decisamente inferiore, anche se per gli articoli catalogati tra la 11-esima e 20-esima postazione le percentuali sono leggermente più alte.

Tabella 5. Percentuale del numero di citazioni nei primi 3 anni.

Journal	Top 10	Altre 10
1-Journal of the Royal Statistical Society Series B-Statistical Methodology	2,10%	1,44%
2-Journal of the American Statistical Association	1,16%	2,28%
3-Annals of Statistics	3,15%	3,39%
4-Top 25 dei libri statistici più citati	1,54%	1,12%
5-Journal of Biological Chemistry	9,57%	12,27%
6-Nature	9,55%	5,30%
7-Science	13,92%	11,59%
8-Water Resources Research	2,70%	4,37%
9-Physical Review Letters	3,67%	5,10%
10-Enviromental Modelling & Software	5,09%	
Totale	4,61%	4,80%

La tabella 5, invece, riporta i risultati delle percentuali delle citazioni ricevute dalle riviste nei successivi 3 anni dalla pubblicazione. Allo stesso modo, si nota un'alta percentuale per le medesime riviste nel caso di  $t = 2$ . Ovviamente, stimando i dati per 3 anni i valori salgono perché le riviste hanno a disposizione un tempo maggiore per essere viste e citate.

Dall'analisi sopra spiegata e dalle tabelle riportate, che offrono dati concreti, si può affermare che le riviste statistiche hanno una diffusione più lenta rispetto alle riviste di altre discipline. Per questo motivo si potrebbe pensare ad una modifica nel calcolo degli impact factors. Per esempio, Fok e Franses (2007) dichiarano che il tempo di vetrina usato per il calcolo degli IFs dovrebbe variare in base alle riviste, più precisamente esso dovrebbe dipendere dal tempo necessario per una rivista al raggiungimento del picco delle citazioni. In altre parole, il tempo di vetrina di 2 anni è estendibile a  $T^*$  anni. In questo modo le differenze tra i diversi campi accademici non esisterebbero più. Infatti, ogni disciplina ha il suo stile citazionale. In generale gli scienziati hanno notato che le riviste in ambito medico citano in modo immediato ogni altra rivista, ciò è causato dal fatto che il tema della salute è di fondamentale importanza e perciò deve diffondersi in modo esteso e rapido.

Per analizzare gli effetti che portano al raggiungimento del picco basta utilizzare il rapporto  $m^*/m$ , cioè la percentuale del numero potenziale di citazioni quando le riviste raggiungono il momento del picco.

Tabella 6. Percentuale di citazioni al raggiungimento del picco.

Journal	$m^*/m$
1-Journal of the Royal Statistical Society Series B-Statistical Methodology	47%
2-Journal of the American Statistical Association	47%
3-Annals of Statistics	44%
4-Top 25 dei libri statistici più citati	47%
5-Journal of Biological Chemistry	38%
6-Nature	41%
7-Science	33%
8-Water Resources Research	45%
9-Physical Review Letters	41%
10-Enviromental Modelling & Software	47%

Nella tabella 6 sono riportate le percentuali di citazioni al raggiungimento del picco per le riviste presenti nel data base relative allo studio. Le percentuali risultano essere più alte per i giornali statistici (47%) rispetto alle altre discipline (40%). Quindi utilizzando un tempo di vetrina pari a  $T^*$  per la stima degli IFs, le riviste di statistica ricevono un miglior punteggio rispetto alle altre discipline.

È rilevante soffermarsi anche sul fenomeno delle auto-citazioni, cioè articoli in un giornale che rimando ad altri articoli nello stesso giornale. Esse potrebbero variare da rivista a rivista, provocando stime distorte degli IFs.

In conclusione, dopo aver mostrato questi dati e dopo aver fatto alcune considerazioni, è apparso che le riviste statistiche hanno una lenta diffusione che causa un numero minore di citazioni. Ciò è spiegato dal fatto che il campo della statistica ha bisogno di un tempo più lungo di vetrina temporale.

## 6.2 IL MONDO DELL'OPEN ACCESS

Per mezzo delle nuove tecnologie informatiche, ovvero al mondo elettronico del World Wide Web, la letteratura scientifica è entrata a far parte anch'essa della rete elettronica. Ciò prende il nome di Open Access, in altre parole corrisponde alla possibilità di trovare in rete la letteratura scientifica e non solo, ogni utente gode dell'opportunità di poter scaricare e utilizzare a suo piacimento le riviste elettroniche.

La tecnica dell'Open Access funziona tramite due strategie: l'auto-archiviazione, mediante la quale gli autori depositano i loro lavori, rispettando il diritto d'autore, oppure tramite la pubblicazione degli articoli in riviste ad accesso aperto.

In questo nuovo scenario le riviste vengono valutate per mezzo: dell'**analisi citazionale**, cioè vengono registrate le citazioni fatte dai ricercatori accademici, dell'**analisi dei download** ovvero attraverso il conteggio dei scaricamenti di uno specifico articolo o di una determinata rivista, oppure **valutazione tramite blogs**.

Gli OA favoriscono la rapidità e la facilità di accesso alle riviste, grazie al fatto che internet, nel mondo odierno, è una realtà alla portata di tutti. A differenza del contesto del database ISI, l'OA facilita la diffusione delle riviste riducendone i costi e massimizzandone l'impatto. Infatti l'archivio dell'ISI è un'enorme biblioteca che contiene specifiche riviste, le quali sono disponibili solamente ai scienziati e tra l'altro a prezzi onerosi, inoltre esse devono rispettare determinate condizioni come l'esposizione di 2 anni dopo la pubblicazione. Mentre l'OA impone come unica regola di mantenere i diritti dell'autore. Infatti, i principali vantaggi di questo metodo sono il massimo impatto degli articoli e la massima disseminazione, intesa nel senso di propagazione di conoscenza per raccogliere conoscenza, determinando così una crescita (Antonella de Robbio, 2008). C'è anche più trasparenza nel processo di valutazione perché questo è visibile a tutti. Inoltre, dall'ambiente dell'impact factor con elevati costi delle riviste, si passa all'open acces che porta ad una riduzione dei prezzi proprio perché è necessario pagare solamente la connessione ad internet. Per di più, alcuni studi hanno verificato che le riviste OA ricevono un numero maggiore di citazioni, per il fatto che esse sono disponibili a tutti, e in modo immediato. Infatti, mentre lo scenario dell'IF necessitava di una vetrina temporale di 2 anni dalla pubblicazione per ricevere citazioni, l'OA non ha bisogno di alcun intervallo di

tempo. Si superano così i lunghi tempi di pubblicazione detenuti da quelle riviste che impiegano anche diversi anni per il processo di revisione.

In conclusione, grazie al modello dell'accesso aperto, la conoscenza della letteratura scientifica è diventata un'opportunità per tutti.

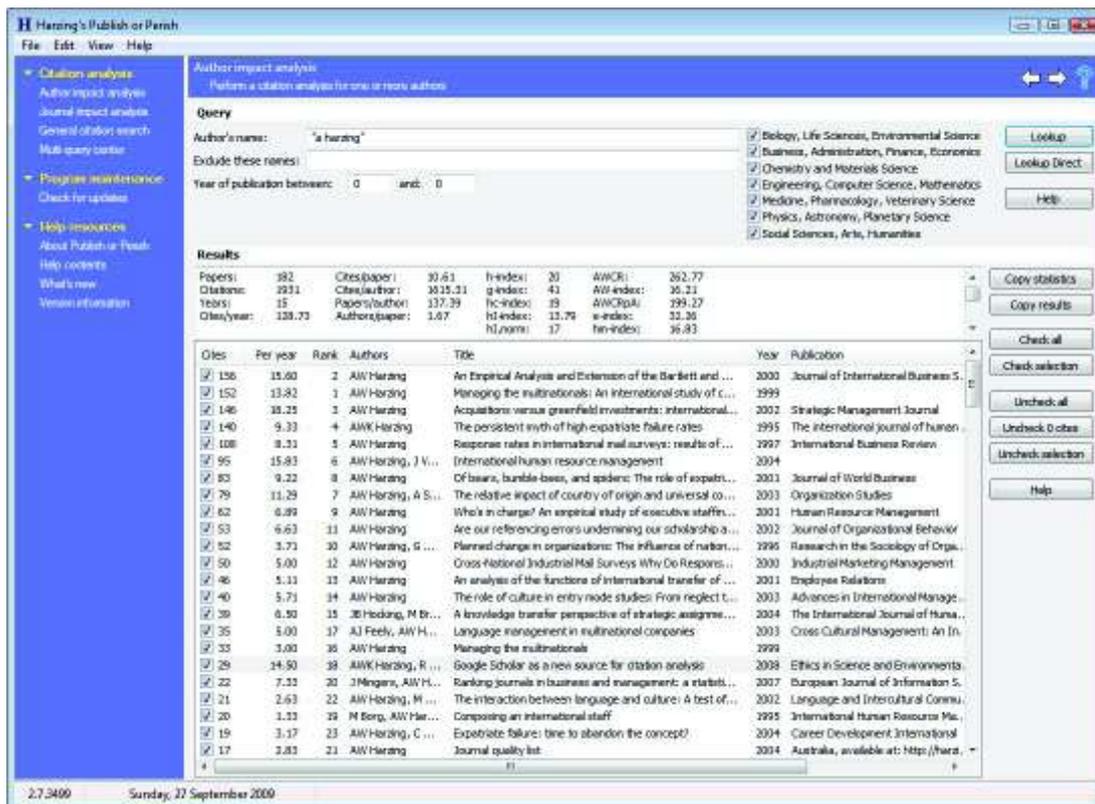
## **6.3 PUBLISH OR PERISH**

### **6.3.1 CHE COS'È PUBLISH OR PERISH**

Nel 2006 è apparso in rete Publish or Perish (PoP) un software gratuito sviluppato da Anne Will Harzing e disponibile sul sito internet [www.Harzing.com](http://www.Harzing.com). Questo programma recupera e analizza le citazioni accademiche per mezzo di Google Scholar. In pochi secondi PoP fornisce le principali variabili bibliometriche in tutti i principali campi di ricerca e le seguenti statistiche: numero totale di libri, numero totale di citazioni, numero medio di citazioni per articolo, numero medio di citazioni per autore, numero medio di articoli per autore, numero medio di citazioni per anno, tasso delle citazioni ponderato per l'età e un'analisi del numero di autori per ogni libro.

Questo programma viene utilizzato per studiare l'impatto dei lavori di un autore sulla comunità scientifica. Infatti Publish or Perish viene adoperato per il calcolo dell'h-index.

Una possibile schermata di un'analisi svolta mediante tale programma è riportata qui sotto:



La figura che si trova sopra è un esempio di una ricerca con PoP. A lato della finestra si osserva anche un Help, infatti Publish or Perish offre un aiuto scheda di dettaglio con la ricerca di consigli e di informazioni supplementari sulle metriche di citazione.

Il programma si può scaricare dal link:

- <http://www.harzing.com/download/PoPSetup.exe> (per la versione Windows)
- <http://www.harzing.com/download/pop-linux.tgz> (per la versione di Linux)

Attualmente non c'è alcun file della Guida per la versione di Linux.

### 6.3.2 COME FUNZIONA PUBLISH OR PERISH

Pop utilizza le queries di Google Scholar per ottenere informazioni sulle citazioni, le quali successivamente vengono trasformate in numeri e inserite nelle statistiche del programma.

La pagina “Analisi dell’impatto degli autori” contiene i seguenti pannelli:

- Il pannello di query sull’autore
- Il pannello dei risultati

Per le informazioni di base, si possono consultare i seguenti pannelli:

- Metriche citazionali
- Accuratezza dei risultati
- Ancora sull'analisi citazionale.

Per effettuare una semplice analisi dell'impatto di un autore:

1. Inserire il nome dell'autore nel campo **Author's name**
2. Cliccare su **Look up** o premere Invio

Dopo aver fatto questi comandi, Google Scholar recupererà le citazioni ed emetterà l'elenco di esse per poi calcolare le principali statistiche, visibili nel pannello **Results**.

Generalmente, l'elenco dei risultati ottenuti contiene opere di autori diversi da quello che è stato ricercato. Si può migliorare la ricerca consultando la sezione **Accuracy of the results** per ulteriori approfondimenti e suggerimenti. Si può inserire il nome dell'autore più specifico, inserendo le iniziali. Per esempio la ricerca di *Harzing* può essere raffinata ulteriormente se si ricerca *A Harzing* o (*Harzing A* che è lo stesso). Il campo **Author's name** è case-insensitive quindi scrivere *harzing*, *Harzing* o *HARZING* è la stessa cosa. Google Scholar combina di default il nome e le iniziali in qualsiasi punto della lista degli autori. Quindi per combinare le iniziali di un autore solo con il suo cognome e per fare in modo che il sistema ricerchi la combinazione per intero, si deve digitare il nome tra virgolette. Per esempio, "*AW Harzing*".

Per ricercare gli articoli scritti da più autori, basta inserire i loro nomi nel campo **Author's name**. È possibile avvalersi degli operatori booleani AND o OR: il primo fornirà solo gli articoli scritti da entrambi gli autori, mentre il secondo produrrà come risultati tutti gli articoli scritti separatamente dai due autori o elaborati dai due autori in collaborazione con altri autori. Invece per escludere il nome di un determinato autore, bisogna inserirlo all'interno del campo **Exclude these names**.

Nel caso in cui un autore ha pubblicato solo da un certo anno in poi (o prima di un terminato anno), si possono inserire gli anni di inizio e fine nel campo **Years of publication between ... and ...**. Si utilizza questo campo anche quando si vogliono prendere in considerazione le pubblicazioni di un autore in un determinato intervallo temporale.

Nel caso in cui l'elenco dei risultati è misero, si possono includere o escludere manualmente delle citazioni dall'analisi riempiendo o rassettando i box nell'elenco dei risultati. Per esempio, il bottone **Check all** permette di controllare e modificare tutti i box, il bottone **Uncheck all** ripulisce tutti i box, quando si utilizzano i tasti della tastiera per scorrere su e giù nella lista dei risultati, si può usare la barra spaziatrice per muovere il cursore all'interno e/o all'esterno della casella selezionata.

Infine la selezione delle pubblicazioni più rilevanti risulta più agevole se si ordinano i risultati in base al numero di citazioni (**Cites**), all'autore (**Authors**), al titolo (**Title**), all'anno (**Year**), al numero di pubblicazioni (**Publication**) e all'editore (**Publisher**). La classificazione in base a queste tipologie si effettua cliccando sull'intestazione della colonna corrispondente.

## BIBLIOGRAFIA

1. Abramo G. (2010), Ci vuole metodo per valutare la ricerca, *La voce*, pag. 1001454
2. Battiston R. (2008), La valutazione della rivista al tempo di Google, da *Meritocrazia.com*
3. Battiston R., Rossi P., Miglietta F. (2009), Publico Ergo Sum dal sito [www.PublicoErgoSum.org](http://www.PublicoErgoSum.org)
4. Bergstrom C. (2008), Pseudocode for calculating Eigenfactor™Score and Article Influence™Score using data from Thomson Reuters Journal Citation Reports, dal sito [www.Eigenfactor.org](http://www.Eigenfactor.org)
5. Borman L. e Daniel H.D. (2007), What do we know about the h-index, *Journal of the American Society for Information Science and Technology*, pag. 1381-1385
6. Contreras C., Edward G. e Mizala A. (2006), The current impact factor and the long term impact of Scientific Journal by discipline: a logistic diffusion model estimation, *Scientometrics* 69, pag. 689-695
7. De Castro P. e Toni F. (2008), utilizzo di risorse elettroniche e produzione scientifica in un ente di ricerca: esiste una relazione?, *Seminario: La biblioteca digitale tra accesso e produzione di conoscenza: un percorso tra le statistiche d'uso e citazionali. A cura del Consorzio iteruniversitario per le Applicazioni del Supercalcolo per Università e Ricerca.*
8. Egghe L. (2006), Theory and practise of the g-index, *Scientometrics* 69, n. 1, pag. 131-152
9. Figà 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. Lecce*

10. Fok D. e Franses P.H. (2007), Modelling the diffusion of Scientific publications, *Journal of Econometrics* 139, pag. 376-390
11. Franceschet M. (2010), Istruzioni per l'uso della bibliometria, *La voce*, pag 1001545
12. Garfield E. (1972), Citation analysis as tool in journal evaluation, *Science*, 178-471
13. Gargiulo P. (2009), Open Access, diritto d'autore, valutazione della ricerca e vantaggi citazionali, *Convegno: Accesso aperto alla conoscenza e valutazione della ricerca. Lecce*
14. Hirsch J. (2005), An index to quantify an individual's scientific research output, da arXiv: *Physics* 0508025v5
15. Moed H.F., Burger W.J.M. , Frankfort J.G. e Van Raan a.F.J. (1985), The application of bibliometric indicators: important field and time dependent factor to be considered, *Scientometrics* vol. 8, pag. 177-203
16. Moriello R. (2007), L'indice di Hirsch e altri indici citazionali dopo l'impact factor, *Biblioteche oggi*, pag. 23-32
17. Nierop E. (2009), Why do Statistics Journals have low impact factors?, *Statistica Neerlandica* vol. 69, n. 1, pag. 52-62
18. Seglen O. (1997), Why the impact factor of journals should not be used for evaluating research, *British Medical Journal*, pag. 314-497

Link di riferimento:

- [www.eigenfactor.org](http://www.eigenfactor.org)
- [www.harzing.com](http://www.harzing.com)
- [www.publicoergosum.org](http://www.publicoergosum.org)
- [www.scopus.it](http://www.scopus.it)