

Considerazione sui nuovi indicatori proposti nelle procedure ASN

È stata pubblicata la bozza di decreto sulle nuove procedure per l'ASN. La bozza sarà oggetto di parere obbligatorio da parte del CUN a giorni. Dalla bozza risulta quanto segue (Allegato C, punto 2, relativo agli SC bibliometrici):

“Gli indicatori bibliometrici da utilizzare nelle procedure di abilitazione a professore di prima e seconda fascia sono i seguenti:

- a) *il numero di articoli pubblicati nei 10 anni consecutivi precedenti la data stabilita nel bando candidati su riviste scientifiche contenute nelle banche dati internazionali “Scopus” e “Web of Science” pesato per l'indicatore di impatto della rivista scientifica;*
- b) *l'indice h di Hirsch, rilevato nelle banche dati internazionali “Scopus” e “Web of Science” e calcolato con riferimento alle pubblicazioni edite nei 10 anni consecutivi precedenti la data stabilita nel bando candidati.”*

Il presente documento mette in luce numerosi, gravi problemi relativi al primo dei due indicatori bibliometrici e l'impatto negativo per la maggior parte dei SSD dell'area 09

Nella seconda parte del documento si fa esplicito riferimento all'impatto che tale indicatore ha sull'SSD di Ingegneria Informatica (ING-INF/05, SC 09/H1). Questo settore è di grandi dimensioni, comprende una pluralità di sotto-discipline, ed è caratterizzato da una forte produzione interdisciplinare, frutto di collaborazioni con la maggior parte delle aree ERC di PE (Physical and Engineering Science) e LS (Life Sciences). Queste considerazioni sono comunque comuni a molti settori dell'area dell'informazione.

Sull'uso del “fattore di impatto” nelle nuove procedure ASN

Sono cambiati gli indicatori bibliometrici per l'abilitazione. I nuovi criteri sono *obbligatori*¹: non si può ottenere l'abilitazione se non si passano le soglie degli indicatori. Non si può non osservare che, d'ora in avanti, la *prima preoccupazione* di chi aspira a far carriera sarà, quindi, attrezzarsi a superare le soglie degli indicatori.

Il primo indicatore dovrebbe stimare la *produttività scientifica* di un candidato negli ultimi 10 anni. Per calcolarlo, le riviste Scopus/ISI sono *pesate per il fattore di impatto*. È lecito chiedersi cosa sia in realtà il “fattore di impatto”. Pare logico identificarlo *con l'impact factor* (IF) della rivista. Ci sono in realtà diversi enti che calcolano un numero chiamato IF. Uno è Thomson Reuters (lo stesso del DB ISI-WoS). Visto che il DB ISI-WoS è uno di quelli menzionati dal decreto, nel resto del documento i conti sono fatti prendendo a riferimento il più recente valore di IF presente in questo database, relativo al 2015.²

È ben noto **che l'IF di una rivista non è indicativo della sua qualità scientifica**: si veda ad esempio la voce “impact factor” su Wikipedia (https://en.wikipedia.org/wiki/Impact_factor) e le

¹ Si evince dalla lettura congiunta dell'Allegato C e dell'art. 4.

² Come verifica, i conti che seguono sono stati ripetuti usando:

a) L'IF su 5 anni, sempre fornito da Thomson Reuters

b) l'IF calcolato su *citefactor* (<http://www.citefactor.org/journal-impact-factor-list-2014.html>), relativo al 2014.

I risultati che si ottengono sono qualitativamente identici.

fonti citate in essa. I problemi citati nel documento linkato sopra sono infatti tra quelli che hanno portato ad abbandonare l'ipotesi di usare l'IF nelle precedenti tornate dell'ASN.

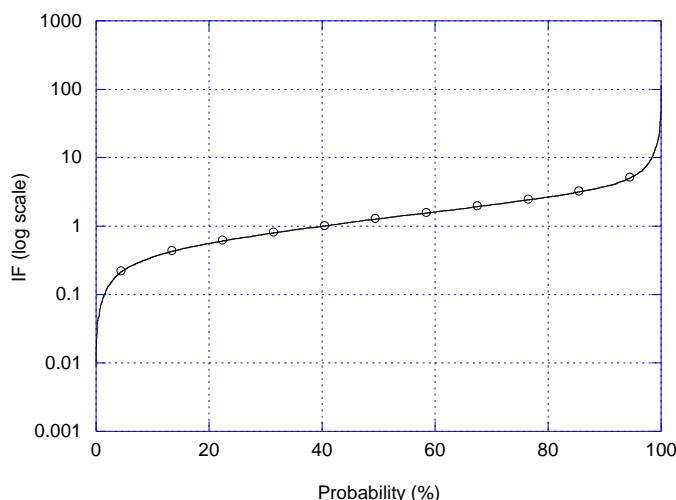
Esistono alcuni problemi parziali di applicabilità e alcuni problemi sostanziali per i quali il valore delle mediane potrebbe essere del tutto falsato.

1) Riguardo all'applicabilità dell'IF in modo prescrittivo in una procedura come l'ASN, esistono numerosi problemi pratici.

Il primo è che *l'IF cambia (significativamente) di anno in anno, e non è prevedibile a priori*: non si può sapere quale sarà l'effetto di una pubblicazione quando la si sottometta, visto che in genere passa più di un anno dalla sottomissione alla pubblicazione.

Il secondo è che *l'IF non esiste* per tutte le riviste Scopus/ISI. Ad esempio, l'elenco delle riviste del GEV ING-INF/05 conta 1126 titoli, e solo 675 di questi hanno un IF (circa il 60%). Tra le riviste mancanti ce ne sono alcune di notevole prestigio. Questo vale anche per le riviste di molti altri SSD.

Il terzo è che il range di variazione dell'IF è di *ordini di grandezza* (da qualche decimo di punto a decine di punti). Di seguito un grafico della distribuzione (si noti che la scala dell'asse y è logaritmica³) ed alcune statistiche riassuntive:



<i>Journal Impact Factor</i>	
Numero campioni	11109
Minimo	0.00
Massimo	115.84
Media	1.96
Mediana	1.28
Deviazione standard	2.96
Varianza campionaria	8.78
Curtosi	253.56
Asimmetria	10.76

È cosa nota che sommare tra loro numeri che variano di ordini di grandezza porta a risultati *privi di senso*: le code della distribuzione assumono infatti un peso preponderante rispetto alla massa della medesima. Per questo motivo, ordinare sulla base dell'IF aggregato significa ordinare sull'incidenza di outlier molto prima che sulla *produttività scientifica*.

Il fatto che il range di variazione degli IF occupi vari ordini di grandezza ha delle conseguenze non banali, in considerazione del fatto che la *distribuzione degli IF è disomogenea tra aree scientifiche* diverse: ad esempio, medici, biologi e fisici hanno accesso a riviste con IF elevato (anche maggiore di 10), mentre per molti ingegneri le riviste considerate di alto prestigio hanno IF più bassi (si veda a questo proposito l'analisi dettagliata presentata nella seconda parte del documento). Ciò significa che chi, tra questi ultimi, collabora con aree applicative "ad alto impact factor" (magari apportando contributi marginali alla propria disciplina) ha accesso a riviste a cui chi si occupa del progresso

³ Per ottenere un grafico in scala logaritmica sono stati omessi i pochi campioni nulli di IF, 58 in tutto.

della propria disciplina non può accedere. Si noti che il fatto che le commissioni valutino la pertinenza e rilevanza della rivista rispetto al settore non risolve il problema: il calcolo delle soglie verrà infatti svolto dal CINECA, che non può entrare nel merito della coerenza tra riviste e tematiche del settore.

Capita, inoltre, che la stessa rivista ospiti contributi provenienti da settori differenti. In questo caso, è del tutto legittimo che, per i diversi settori, tale rivista rivesta un prestigio differente. L'utilizzo di un valore unico (l'IF, appunto) obbliga un settore ad accettare supinamente "l'opinione di maggioranza".

2) Il problema sostanziale è che le considerazioni di cui sopra potrebbero rendere inutile l'indicatore: le soglie degli indicatori bibliometrici dovrebbero servire a tagliar fuori chi non ha i titoli per andare ad una valutazione di merito, evitando quindi lavoro inutile alle commissioni. Con una distribuzione di IF come quella sopra indicata, si corre il rischio che questo criterio possa essere implementato *soltanto* mettendo delle soglie molto basse, che quindi finiscono per essere inefficaci.

Proposte alternative

Di seguito si formulano alcune proposte alternative all'uso dell'IF per definire un indicatore di produttività scientifica, elencandone i possibili vantaggi e/o problemi:

- a) Attribuire *peso unitario* a ciascuna pubblicazione su rivista Scopus/ISI (come nelle precedenti edizioni dell'ASN).
 - Il fatto di limitarsi alle pubblicazioni Scopus/ISI offre già sufficiente garanzia di escludere pubblicazioni su riviste di dubbia scientificità e scarsa diffusione o continuità.
 - *Tutte* le riviste Scopus/ISI entrano a far parte del conteggio dell'indicatore.
 - Il sistema di calcolo di questo indicatore è già funzionante ed ampiamente testato (il che limita l'incidenza di ricorsi dovuti a problemi di calcolo imprevisti ed imprevedibili).
 - Tutti i ricercatori hanno impostato il loro lavoro negli ultimi anni sulla base di *questo* criterio, e quindi mantenerlo consente loro di raccogliere i frutti di quanto hanno seminato.
- b) Attribuire a ciascuna pubblicazione su rivista Scopus/ISI un peso calcolato sulla base della classificazione dei GEV delle riviste di quel settore, con un'escursione *numericamente limitata* (ad esempio quella usata per la VQR: 1.0 per i prodotti *eccellenti*, 0.8 per quelli *buoni*, etc.).
 - Si recupera il ruolo delle comunità scientifiche nel definire l'importanza di una rivista.
 - Si mantiene comunque, per i GEV che lo riterranno opportuno, la possibilità di classificare le riviste sulla base dell'IF.
 - Si mantiene l'incentivazione a pubblicare "bene", cioè su riviste considerate top-level dalla comunità, ma si evitano le sproporzioni numeriche macroscopiche (e.g., un articolo sulla rivista X ne vale 20 o 100 sulla rivista Y) dovute all'uso dell'IF non quantizzato. Tali sproporzioni rischiano invece di incentivare comportamenti opportunistici nelle comunità scientifiche.
 - Si mantiene una classificazione coerente tra VQR e ASN (è irragionevole tenere due indicatori diversi per misurare la qualità delle stesse riviste).
 - La classificazione è comunque modificabile nel tempo, ma con una periodicità scelta dai GEV; d'altro canto l'IF viene ricalcolato anno per anno, troppo velocemente rispetto alle dinamiche di questo sistema.
 - *Non tutte* le riviste Scopus/ISI entrano a far parte del conteggio dell'indicatore. Si rende quindi necessario introdurre una clausola di default per le riviste fuori dalla classificazione del GEV di un SC/SSD: ad esse va attribuito comunque un peso (e.g., zero, maggiore di zero, pari alla media dei pesi dati dagli altri GEV, etc.).

Analisi degli IF relativa al SSD ING-INF/05

Qui di seguito, alcune considerazioni valide per l'SSD ING-INF/05, probabilmente estendibili ad altri raggruppamenti grandi e plurali.

Come sopra richiamato, l'SSD ING-INF/05 è caratterizzato da un forte tasso di interdisciplinarietà, che comporta collaborazioni praticamente con tutti i settori ERC di PE e LS, e dalla presenza di numerose sottodiscipline (e.g., intelligenza artificiale, valutazione delle prestazioni, architetture dei calcolatori, networking, image processing, basi di dati, sistemi real-time, e molte altre), ciascuna delle quali ha le proprie riviste di riferimento.

Il GEV ING-INF/05 ha classificato le riviste di interesse per il settore⁴, dividendole in quattro categorie (in classe 1 ci sono le riviste considerate migliori). Sono state censite oltre 1100 riviste. Le riviste che *non hanno* IF sono così distribuite:

Classe GEV	# riviste GEV	# riviste senza IF	% mancanti
1	226	17	7.5%
2	223	33	14.8%
3	113	39	34.5%
4	564	361	64%
Totale	1126	451	40%

Il 40% delle riviste di interesse per il settore non ha IF. Quasi tutte quelle di classe 1 ne hanno uno, ma tra quelle che non ce l'hanno ce ne sono alcune tra le più importanti, ad esempio:

- ACM Transactions on Accessible Computing,
- ACM Transactions on Computation Theory,
- ACM Transactions on Speech and Language Processing,
- Communications in Numerical Methods in Engineering,
- IEEE Intelligent Systems and Their Applications,
- IEEE Transactions on Industrial Informatics,
- IEEE Transactions on Neural Networks,
- IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans,
- Journal of Algorithms,
- Journal of Computer Security.

Anche limitando l'analisi alle riviste classificate dal GEV, si osservano disparità marcate. Per le sole riviste di classe 1, gli IF vanno da 10.872 a 0.357. La rivista con IF=10.872 è "Molecular Systems Biology" (dove scrivono *anche*, ma non *principalmente* – e raramente da soli - ingegneri informatici), mentre quella con IF=0.357 è "Logical Methods in Computer Science". Quindi, chi pubblica 30 articoli sulla seconda ha meno probabilità di superare la soglia di chi ne pubblica uno solo sulla prima, *coeteris paribus*⁵. Il numero di pubblicazioni necessario ad eguagliare il peso di un articolo su Molecular Systems Biology (ai fini del superamento della soglia) è indicato nella tabella seguente, approssimato all'intero più vicino, per alcune riviste ben note di classe 1:

⁴ Il documento del GEV si trova qui: <http://tinyurl.com/ojwupd>

⁵ La seguente obiezione si sente di frequente: chi, negli ultimi 10 anni, ha pubblicato *un solo articolo* su Molecular Systems Biology verrà comunque bocciato dalla commissione non appena questa entrerà nel merito. È vero, ma ciò non cambia il fatto che *l'altro* suo collega alla fase di merito potrebbe non arrivarci nemmeno.

Journal name	# pubblicazioni necessarie
COMMUNICATIONS OF THE ACM	3
IEEE TRANSACTIONS ON MOBILE COMPUTING	4
IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	5
IEEE-ACM TRANSACTIONS ON NETWORKING	6
IEEE TRANSACTIONS ON COMPUTERS	7
IEEE TRANSACTIONS ON SOFTWARE ENGINEERING	7
JOURNAL OF THE ACM	8
PERFORMANCE EVALUATION	9
ACM TRANSACTIONS ON INFORMATION SYSTEMS	11
REAL-TIME SYSTEMS	11
ACM TRANSACTIONS ON DATABASE SYSTEMS	16
ACM TRANSACTIONS ON COMPUTER SYSTEMS	18
ACM Transactions on Embedded Computing Systems	23

Per fare un altro esempio, prendiamo due riviste unanimemente considerate *top-level* per il settore:

- IEEE Transactions on Fuzzy Systems: IF=8.746
- IEEE/ACM Transactions on Networking: IF=1.811

Queste sono le riviste sulle quali aspira a pubblicare chi si occupa di fuzzy systems e di networking, rispettivamente. Chi fa ricerca sul networking, però, *non ha accesso* a riviste con IF prossimo a 8.74, perché non ce ne sono. È irragionevole che - per avere identica possibilità di accesso alla fase di merito della ASN - quest'ultimo sia condannato in eterno a scrivere 5 articoli per ognuno di quelli scritti da un suo collega che si occupa di fuzzy systems. Analoghe disparità si riscontrano prendendo altre coppie di sottodiscipline pertinenti al settore.

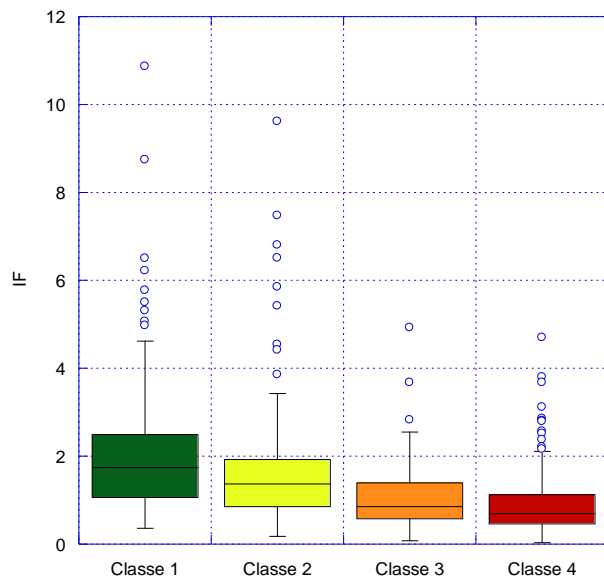
In questo modo il primo indicatore bibliometrico dell'ASN seleziona sulla base dell'*ambito di ricerca*, non della produttività. Già nel breve termine, la possibilità di accedere alla fase "di merito" dell'abilitazione verrà *distorta* a favore di alcune sottodiscipline a dispetto di altre. Nel medio termine, questo si rifletterà sulla composizione del settore, e del suo vertice in particolare.

Inoltre, la classificazione delle riviste fatta per IF è sostanzialmente incompatibile con quella svolta dal GEV. La correlazione tra la classe GEV e l'IF è di -0.374, negativa in quanto a classe più piccola corrisponde rivista più autorevole, ma troppo *piccola* in modulo perché si possa dire che le due misure si assomigliano. Si notano, tra mille altre, le seguenti peculiarità, che contribuiscono invece a renderle irreconciliabili:

- Il 20% delle riviste di classe 2, il 10% di quelle di classe 3, e l'8% di quelle di classe 4 hanno un IF maggiore di quello del 62% delle riviste di classe 1.
- La rivista di classe 4 con massimo IF (Integrated Computer-aided Engineering) ha IF 4.698 (circa 3 volte l'IF di IEEE Transactions on Computers, rivista di classe 1 e di riferimento per il settore).
- Tra le riviste del GEV che sono sopra il 90mo percentile dell'IF, il 38% *non* è in classe 1. Il 10.5% è addirittura in classe 4.
- Tra le prime 10 riviste per IF, 5 sono di classe 2.

Di seguito viene mostrato un box plot degli IF divisi per classe di rivista del nostro GEV. Si vede piuttosto bene che:

- gli outlier (di qualunque classe) sono numerosi e ben al di là della massa della distribuzione delle riviste di classe 1;
- esistono intersezioni ampie non solo tra classi adiacenti, ma anche tra classi distanti due unità.



Non è ragionevole che una rivista che va bene per la VQR non vada bene (o vada meno bene) per l'ASN e viceversa. Su quali riviste dovranno pubblicare i nostri colleghi d'ora in avanti, su quelle di classe 1 (per massimizzare la quota premiale di fondi per la propria istituzione di appartenenza), o su quelle ad alto IF (per superare le soglie dell'ASN)?

È lecito chiedersi come cambierà il comportamento dei ricercatori del settore qualora questo sistema venga messo in funzione. È interamente possibile che si diffondano pratiche quali:

- Abbandonare il proprio filone di ricerca, se non ha sbocco su riviste a alto IF.
- Disinteressarsi del progresso della *propria* disciplina, e mettersi a svolgere lavoro da tecnico informatico per gruppi di ricerca di altre aree (medici, biologi, etc.), in modo da poter accedere a riviste con IF che un ingegnere informatico non potrebbe permettersi altrimenti.

Pratiche del genere possono soltanto impoverire il settore, provocando danni gravi in tempi brevi.

Conclusioni:

- L'IF è una misura a copertura parziale e significativamente variabile nel breve termine.
- E' una misura non uniforme non solo tra SSD diversi, ma anche tra aree di ricerca appartenenti allo stesso SSD
- E' una misura che varia per ordini di grandezza, e quindi non può essere usata in modo additivo (a meno di non voler dare peso agli outlier prima che alla massa della distribuzione).
- In un settore grande, fatto di molte sottodiscipline ed aree applicative, l'IF discrimina sulla base della sottodisciplina e dell'area applicativa, molto prima che sulla produttività scientifica.
- L'IF è inconciliabile con la valutazione del GEV, cioè con l'idea che ciascun settore ha della qualità delle riviste scientifiche.