

# Un metodo *fair* per la valutazione di singoli, gruppi e dipartimenti con la VQR

Gianluigi Greco e Francesco Scarcello  
Università della Calabria

## Premessa.

Chiariamo subito: questo non è l'ennesimo intervento sulla correttezza o meno delle modalità di valutazione usate dall'ANVUR per la valutazione della ricerca VQR 2004-2010 (numero di lavori da sottomettere, peer-review VS dati bibliometrici, classificazione riviste etc.). Intendiamo invece muoverci nell'ambito delle regole definite nella VQR per analizzare in modo rigoroso il *problema di come la valutazione di una struttura possa essere ribaltata in modo fair su sottostrutture e (gruppi di) individui*. Per semplicità parleremo nel seguito di università e di dipartimenti, ma ciò che si dirà vale ovviamente per le altre strutture di ricerca e per le loro sottostrutture coinvolte nella VQR.

Il contributo qui riportato sintetizza informalmente il contenuto dei seguenti lavori di ricerca, entrambi disponibili in modalità open access dai siti delle rispettive case editrici:

- [“Fair division rules for funds distribution: The case of the Italian Research Assessment Program \(VQR 2004–2010\)”](#), G. Greco and F. Scarcello, *Intelligenza Artificiale*, 7(1):45-56, IOS Press, 2013.
- [“Mechanisms for Fair Allocation Problems: No-Punishment Payment Rules in Verifiable Settings”](#), G. Greco and F. Scarcello, *Journal of Artificial Intelligence Research (JAIR)*, Vol. 49: 403-449, 2014.

## Il problema.

Come è stato ribadito da più parti, e finanche dall'ANVUR stessa nel rapporto finale sulla VQR, è palesemente scorretto valutare un singolo ricercatore in base ai 3 prodotti *a lui assegnati dalla propria università* e conferiti per la valutazione. Infatti, la scelta di assegnare o meno un prodotto ad un ricercatore A è stata guidata dalla massimizzazione del risultato globale per la struttura di ricerca (o almeno così doveva essere), sicché in presenza di coautori deboli o poco attivi, si poteva e doveva “togliere” qualche lavoro buono ad A attribuendolo ad un coautore B che altrimenti poteva ritrovarsi costretto a sottomettere lavori insufficienti o a non sottometterne nessuno!

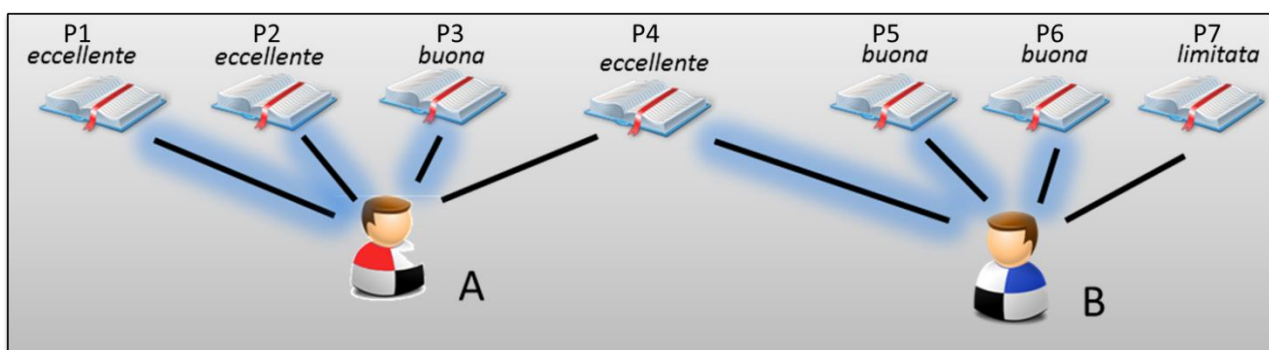
D'altra parte, laddove le università non abbiano centralizzato la selezione dei prodotti (e forse questa è la maggioranza dei casi), l'attribuzione degli stessi ad uno dei coautori potrebbe essere stata persino più arbitraria. Ad esempio, la scelta dell'attribuzione di una pubblicazione eccellente a B invece che ad A potrebbe essere stata determinata da questioni di “gerarchia” (o comunque “politica” in senso lato), che a maggior ragione non dovrebbero influire in alcun modo né sullo svantaggiato A né sull'ingiustamente premiato B.

Per i motivi sopra esposti è apparso dunque del tutto ragionevole che la VQR non dovesse essere utilizzata per valutazioni individuali. Era tuttavia ampiamente prevedibile, anche visto lo sforzo economico sopportato in tale operazione, che tale limite fosse poi forzato sia dal ministero sia dagli atenei. Infatti il primo effettua nell'ombra valutazioni individuali allorché usa già la VQR per dare pagelle sul reclutamento dei ricercatori (nonché parti del Fondo di Finanziamento Ordinario delle Università in base a queste!) e per valutare i membri dei collegi di dottorato di ricerca, mentre i secondi vorrebbero usare internamente le valutazioni “certificate” della VQR (ad esempio) per distribuire incentivi ai ricercatori o per capire cosa hanno eventualmente sbagliato nel reclutamento, o per definire in modo appropriato i collegi

dei dottorati di ricerca (è chissà cos'altro in futuro, in base alle decisioni ministeriali che saranno basate sulla VQR).

Ciò su cui si è inoltre poco riflettuto è che l'obiezione principale sulla non correttezza dell'uso della VQR per gli individui *mantiene tutta la sua validità anche rispetto al suo uso per i dipartimenti*. Ciò si concretizza laddove vi siano coautori appartenenti a più dipartimenti (ad esempio gli autori di questo contributo afferiscono a dipartimenti diversi dell'Università della Calabria) ed è tanto più rilevante quanto più le attività di ricerca di una università hanno respiro interdisciplinare (ad esempio, a Roma "La Sapienza" vi sono prodotti di ricerca i cui coautori afferiscono addirittura a cinque diversi dipartimenti).

Per esemplificare il problema, consideriamo i due ricercatori A e B le cui pubblicazioni sono riportate in figura ed assumiamo che essi appartengano a dipartimenti diversi.



I due ricercatori hanno scritto congiuntamente una pubblicazione *eccellente* (P4). Inoltre, A ha scritto altre 2 pubblicazioni eccellenti (P1 e P2) ed una buona (P3), mentre B ha scritto altre 2 pubblicazioni buone (P5 e P6) ed una di qualità limitata (P7). Entrambi hanno dovuto sottoporre a valutazione 3 pubblicazioni. Dunque, per massimizzare le performances dell'Università, A dovrebbe aver "ceduto" la pubblicazione congiunta a B, evitando la sottomissione della pubblicazione limitata. L'allocazione ottima per l'Ateneo è quindi quella illustrata in figura (gli archi autore-prodotto evidenziati indicano l'associazione sottoposta alla VQR). Se le valutazioni dell'ANVUR sono quelle attese dagli autori in base ai criteri pubblicati, l'università otterrà i seguenti punteggi per l'insieme dei prodotti P1—P6 selezionati:

Publicazione sottomessa	Valutazione ANVUR	Autori
P1	eccellente (valore 1)	A
P2	eccellente (valore 1)	A
P3	buona (valore 0,8)	A
P4	eccellente (valore 1)	A+B
P5	buona (valore 0,8)	B
P6	buona (valore 0,8)	B

Si noti che il valore complessivo dell'università risulta pari a  $1+1+0,8+1+0,8+0,8=5,4$ , mentre il valore delle pubblicazioni complessivamente assegnate ad A e B è, rispettivamente, pari a 2,8 (P1+P2+P3) e 2,6 (P4+P5+P6). In particolare, rispetto il caso in cui A "tiene" per se la pubblicazione eccellente congiunta (e forza B a sottometterne una limitata di valore 0), nell'allocazione sopra riportata A perde uno 0,2 (ora A deve accontentarsi di P3 al posto di P4), ma la sua università guadagna ben 0,8 (ora dispone di P3 al posto P7).

Ma allora ci chiediamo:

- È corretto valutare A e B sulla base dei prodotti a loro attribuiti in fase di sottomissione?
- Può A (e dunque il suo dipartimento) essere penalizzato per aver “ceduto” una pubblicazione eccellente al fine di massimizzare le performance dell’Università?

Vi sono poi esempi (reali) ben più conflittuali: A e B potrebbero avere un’unica pubblicazione eccellente congiunta (e magari tutte le altre limitate e a nome singolo), per cui l’assegnazione della stessa potrebbe generare tensioni ed attriti. In qualunque modo tale assegnazione sia effettuata (dall’Università in modo centralizzato ovvero dal confronto tra A e B), il ricercatore (diciamo A) a cui tale eccellenza sia stata “sottratta” avrebbe ben da lamentarsi se nella sua valutazione tale eccellenza non fosse proprio tenuta in conto! Ancor di più avrebbe da lamentarsi la sua Università, se A fosse un neo-assunto per il quale l’eccellenza avrebbe contribuito a migliorare le performance rispetto *la qualità del reclutamento* e quindi rispetto la ripartizione del finanziamento FFO!

Di fronte queste problematiche, emerge evidente che *ribaltare la valutazione aggregata dell’Università sulle singole componenti dell’aggregazione (siano essi individui, gruppi di individui, o dipartimenti) richiede molta cautela* ed approcci più sofisticati di quanto possano essere semplici formule basate su medie e normalizzazioni. In effetti *tutti i meccanismi di valutazione dipartimentali proposti sino ad ora sono minati alla loro base dal vizio di ignorare gli aspetti che concernono la corretta attribuzione delle pubblicazioni interdipartimentali* (lo stesso vale per le valutazioni delle aree e dei settori scientifico disciplinari, che ignorano gli aspetti relativi alle pubblicazioni inter-area ed inter-settore).

## Desiderata.

In questo articolo affrontiamo il problema sopra esposto in modo sistematico, partendo dalla postulazione delle seguenti proprietà che un qualunque *meccanismo di distribuzione della valutazione complessiva* dovrebbe soddisfare:

1. “*Indipendenza dall’assegnazione dei prodotti in fase di sottomissione*”: l’attribuzione ad un ricercatore di uno dei prodotti selezionati per la sottomissione non preclude l’uso di tale prodotto per la valutazione dei suoi coautori. Abbiamo visto nel precedente esempio quanto ciò sia rilevante.
2. “*Distribuzione perfetta*”: la somma delle valutazioni dei ricercatori deve coincidere con la valutazione complessiva dell’università. In particolare, non si può distribuire agli individui più di ciò che si è ricevuto. Si noti che il criterio, se fosse considerato in modo indipendente dalle altre proprietà, potrebbe essere risolto facilmente con una normalizzazione finale, a valle di qualunque modalità di distribuzione (come spesso si fa, senza troppe cautele).
3. “*Imparzialità*”: due ricercatori indistinguibili, cioè con prodotti con le stesse valutazioni ANVUR e coautori con le medesime caratteristiche devono ottenere la stessa valutazione.
4. “*Fairness*”: ogni gruppo di ricercatori C (eventualmente costituito da un singolo ricercatore o da un intero dipartimento) deve ricevere una valutazione **almeno** pari al proprio *contributo marginale* alla valutazione dell’università. Il contributo marginale di C all’università U non è altro che la differenza tra la valutazione complessiva di U nella VQR e quella che U avrebbe ottenuto se il gruppo C non fosse stato presente. Ad esempio, pubblicazioni di C con un coautore X esterno al gruppo C non rientrano necessariamente nel computo del contributo marginale di C, in quanto U potrebbe ancora disporre grazie alla presenza di X, qualora i ricercatori in C non fossero presenti. Dunque, il contributo marginale è dovuto unicamente alla presenza di C, ed il criterio richiede naturalmente che a C sia attribuito almeno questo contributo (si tratta di fairness, perché sarebbe ingiusto che tale contributo dovuto solo a C fosse in toto o in parte attribuito ad altri).

5. “*Indipendenza dalla selezione dell’insieme di prodotti da sottomettere alla VQR*”: la valutazione assegnata ad un individuo deve essere la stessa a parità di scelta *ottima* dei prodotti da parte dell’università (cioè qualunque scelta che massimizzi la valutazione totale per l’università). In altri termini, nessun ricercatore deve poter trarre beneficio (né detrimento) dal fatto che l’università si sia orientata verso una particolare selezione di prodotti (in cui magari molti suoi lavori siano stati sottomessi) piuttosto che verso una selezione alternativa di pari valore complessivo. Si noti come questo requisito metta a nudo il forte contrasto tra la struttura e l’individuo: dal punto di vista della struttura ogni allocazione ottima è equivalente alle altre. Gli individui (e con essi i dipartimenti) potrebbero invece voler spingere verso specifiche selezioni dei prodotti, finendo in contrasto gli uni con gli altri. Quando la proprietà in esame è soddisfatta, sappiamo invece che tale contrasto non ha ragion d’essere, poiché qualunque selezione di prodotti globalmente ottima produrrà esattamente la stessa valutazione per ogni individuo.

Per illustrare meglio quest’ultimo requisito consideriamo una proposta di cui si sente spesso parlare: valutare un ricercatore in base alla *media* di tutti i prodotti di cui è autore (che sono stati sottomessi e quindi valutati dall’ANVUR). Osserviamo come tale approccio non soddisfi il requisito di indipendenza dalla selezione con un semplice scenario di esempio. Assumiamo che un ricercatore X, oltre a 3 lavori eccellenti a nome singolo, abbia scritto un lavoro “limitato” insieme ad Y. Inoltre assumiamo che Y abbia scritto anche un lavoro “limitato” insieme a Z ed abbia altri due lavori a nome singolo. Assumiamo infine che Z, come X, abbia anche scritto 3 lavori eccellenti da solo. Dovendo Y sottomettere 3 lavori per la VQR, sarà indifferente per l’università selezionare il lavoro limitato scritto da Y insieme ad X oppure il lavoro limitato scritto da Y insieme a Z. Tuttavia, se la valutazione dei ricercatori si basa sulla media dei prodotti, la sottomissione del lavoro scritto insieme a X determinerebbe l’abbassamento della valutazione media di quest’ultimo. Al contrario, la valutazione media dei prodotti di Z sarebbe “eccellente”, risultando dunque in una palese ingiustizia a danno di X. Noi vogliamo invece che la selezione iniziale dell’università non infici la valutazione di X, che in questo esempio dovrebbe avere la stessa valutazione eccellente di Z.

Più in generale, tutti i criteri basati su medie e normalizzazioni di cui si è finora discusso (pubblicamente) non soddisfano i requisiti di cui sopra, poiché non è stato affrontato in modo formale il problema della gestione delle “reti di coautori”. Ad esempio, la soluzione banale di sfruttare solo l’allocazione dei prodotti agli autori usata per la sottomissione viola apertamente il primo requisito, come già discusso sin dall’inizio di questo articolo.

Un’altra soluzione estremamente “unfair” è quella attualmente proposta dall’ANVUR per la valutazione dei collegi di Dottorato, secondo la quale dovranno essere considerati *i migliori 3 lavori* per ogni autore. Ciò violerebbe in primo luogo il criterio della distribuzione perfetta, poiché il totale delle valutazioni supererebbe in generale quanto ottenuto dall’università (favorendo magari gruppi di autori che hanno scritto pochissimi lavori buoni tutti insieme, che potrebbero essere utilizzati ripetutamente da ciascuno dei componenti, portando ad una evidente supervalutazione di tali autori e quindi del collegio a cui appartengono). Ovviamente normalizzare non risolverebbe il problema, in quanto i gruppi sopra descritti sarebbero comunque privilegiati (e si può dimostrare formalmente che la proprietà di *fairness* risulterebbe violata, svantaggiando coloro che sono esterni a tali gruppi).

### **La soluzione.**

Alla luce del fallimento di metodi che molti ricercatori potrebbero considerare “naturali”, è lecito chiedersi: E’ possibile definire un meccanismo di ribaltamento delle valutazioni che soddisfi tutte le proprietà sopra elencate, o forse il problema che stiamo affrontando non ammette proprio soluzione?

Ciò che abbiamo osservato è che una soluzione in realtà esiste, ed essa è peraltro estremamente elegante. La soluzione si basa sulla teoria dei giochi cooperativi e, in particolare, sull'applicazione al caso della VQR del cosiddetto *Valore di Shapley*, definito nel 1953 da Lloyd S. Shapley (premio Nobel per l'economia nel 2012). Tale valore è spesso usato in letteratura per la redistribuzione "fair" di risorse, in quanto "rispettoso del contributo di ciascuno dei giocatori" (qui ricercatori). Di seguito discuteremo brevemente tale approccio, premettendo che l'esposizione sarà mantenuta informale al fine di renderla accessibile al più vasto pubblico possibile.

L'idea è di rappresentare il problema come un *gioco a coalizioni (coalitional game)*. Questi giochi permettono di modellare matematicamente molti problemi nell'ambito dell'economia, della politica e dell'informatica, ad esempio per distribuzione di risorse, sistemi di votazione o agenti artificiali. Ad ogni insieme  $S$  di individui (chiamato coalizione) si associa una valutazione  $v(S)$  che corrisponde al contributo che essi sono in grado di fornire. Nel nostro caso, possiamo associare ad  $S$  il massimo valore che si può ottenere allocando i prodotti di cui i ricercatori in  $S$  sono autori, secondo le regole della VQR ed in base ai voti effettivamente ottenuti nella valutazione ANVUR. Dunque,  $v(S)$  non è altro che la valutazione che l'università avrebbe ottenuto nella VQR se i ricercatori fossero solo quelli in  $S$ . Di seguito, riprendiamo il nostro esempio, evidenziando le possibili coalizioni ( $S_1, S_2$ , e  $S_3$ ) con le rispettive allocazioni.



Una volta definita la funzione "v" di valutazione, è possibile calcolare lo *Shapley value* (rispetto a tale funzione "v") di ogni ricercatore  $r$ , che indichiamo di seguito come  $svalue(r)$  e che rappresenterà per noi la sua valutazione nella VQR. Tale valore è calcolato considerando i contributi marginali in base alla funzione  $v(\cdot)$  di *tutte* le coalizioni (di ricercatori) che includono  $r$ . In particolare, detto  $N$  l'insieme di tutti i ricercatori della data università, la formula di valutazione dello Shapley value del ricercatore  $r$  è la seguente:

$$svalue(r) = \sum_{S \subseteq N \setminus \{r\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} (v(S \cup \{r\}) - v(S))$$

Intuitivamente, si considera la media dei contributi marginali di  $r$  considerando tutti i possibili modi in cui questo può partecipare a coalizioni di giocatori/ricercatori.

Ad esempio, ritornando ai nostri due ricercatori A e B (con la scala di valori della VQR, cioè eccellente=1, buona=0,8, accettabile=0,5 e limitata=0), i seguenti valori di Shapley value possono essere ottenuti con semplici calcoli algebrici:

$$svalue(A) = 2 \times 1 + (1 + 0,8) / 2 = 2,90 \quad \text{e} \quad svalue(B) = 2 \times 0,8 + (1 + 0,8) / 2 = 2,50$$

In pratica A e B ottengono una valutazione “base” corrispondente, rispettivamente, alle due eccellenze (2x1) ed alle due pubblicazioni buone (2x0,8). In aggiunta, i valori della pubblicazione buona di A e di quella eccellente congiunta vengono mediate, distribuendo per esse ad entrambi i ricercatori un valore aggiuntivo di 0,9.

Per contrasto, il criterio della media dei prodotti valutati dall'ANVUR assegnerebbe 2,85 al ricercatore A e 2,60 al ricercatore B, poiché il “contributo” della pubblicazione non eccellente di A verrebbe spalmato dalla media su tutti i suoi prodotti (inclusi quelli a firma unica). Questo ricercatore sarebbe quindi penalizzato da tale soluzione, mentre B sarebbe avvantaggiato, poiché otterrebbe il contributo intero della pubblicazione eccellente congiunta. Si noti inoltre che la somma di tali valori (5,45) supererebbe il totale dell'università (5,40). Ciò richiederebbe una successiva normalizzazione che peggiorerebbe ancora la situazione per A, portandolo a 2,82 (mentre B avrebbe 2,58).

Si noti che la valutazione basata sullo Shapley value può essere estesa in modo naturale a gruppi/coalizioni di ricercatori, aggregandone semplicemente i valori. In particolare, la valutazione di un dipartimento (o di un SSD) R, visto come insieme dei suoi ricercatori, è data dalla somma di tutti i valori  $svalue(r)$  calcolata su tutti i ricercatori  $r$  che appartengono ad R.

Ciò che abbiamo dimostrato nei lavori sopra citati è che è tutte *le proprietà desiderate per la VQR sono soddisfatte da questa valutazione basata sullo Shapley value.*

Inoltre, abbiamo dimostrato che la soluzione proposta soddisfa ulteriori proprietà non strettamente legate alla fairness, ma comunque di rilevante utilità pratica. Ad esempio, la soluzione è robusta rispetto a tentativi di *manipolazione strategica*. Ciò vuol dire che in future edizioni VQR (in cui si dichiara a priori che la valutazione di singoli e di dipartimenti userà tale criterio) sarà sempre conveniente per ogni ricercatore collaborare in modo sincero all'autovalutazione preliminare alla scelta dei prodotti. Senza necessità di minacciare alcuna sanzione, risulterà infatti svantaggioso per un ricercatore (razionale) qualunque tentativo di influenzare la selezione dei prodotti e/o la loro allocazione per fini che siano diversi dall'ottenimento dell'ottimo globale per l'università.

Come ultima nota, dal punto di vista computazionale, può essere utile osservare che è in corso l'implementazione della tecnica presso la nostra università, e che in essa sono sfruttate anche proprietà strutturali del grafo delle collaborazioni di ricerca in modo da rendere il calcolo efficiente anche su grandi numeri.

## **Conclusioni.**

Abbiamo proposto una tecnica per distribuire in modo corretto (*fair*) la valutazione di una struttura ai singoli e quindi ai gruppi ed alle sottostrutture in modo da rispettare l'effettivo contributo di ciascuno alla performance della propria struttura. La tecnica garantisce infatti che qualunque gruppo di ricercatori otterrà una valutazione almeno pari al proprio contributo marginale (indipendentemente dalla selezione e dall'allocazione dei prodotti sottomessi). Insieme alla proprietà di distribuzione perfetta, ciò implica che, ad esempio, qualunque gruppo di ricercatori senza collaborazioni con altri potrà suddividere tra i propri membri esattamente la somma dei valori dei propri prodotti (senza regalare né togliere nulla agli altri).

Tali proprietà implicano che, anche in presenza di GEV che abbiano utilizzato metri differenti, *la soluzione proposta permette almeno di rapportare correttamente le performance dei singoli ricercatori rispetto alle performance dei colleghi valutati dallo stesso GEV.*

Inoltre la nostra tecnica è immune da distorsioni dovute a ricercatori inattivi o parzialmente inattivi, che vengono correttamente gestiti (a differenza di alcune soluzioni improvvisate, in cui i ricercatori inattivi vengono esclusi arbitrariamente dai conteggi, rischiando di sopravvalutare alcuni gruppi a scapito di altri che hanno invece ricercatori parzialmente inattivi, ma che si sono comunque sforzati di produrre qualcosa!).

E' importante notare che l'approccio proposto potrà tornare ancora più utile per la prossima edizione della VQR. Dichiarando infatti sin dall'inizio che qualunque valutazione dei singoli godrà delle proprietà di indipendenza della selezione dei prodotti e della loro allocazione si potranno evitare fenomeni spiacevoli come quelli che hanno visto alcuni professori tenere per sé i prodotti migliori, spesso penalizzando il risultato complessivo della propria università e magari anche le valutazioni sul reclutamento di qualche giovane ricercatore (quindi danneggiando doppiamente il finanziamento della propria università!). Infine, chiarendo che le regole del gioco porteranno ad una corretta valutazione degli autori di prodotti interdipartimentali, collaborazioni tra dipartimenti e ricercatori di aree diverse sarebbero certamente incentivate. Ad oggi, invece, il timore di una valutazione ancora focalizzata sulla rigida assegnazione dei prodotti nella fase di sottomissione sta già creando barriere tra colleghi e dipartimenti indesiderate e dannose.