

## Index ke kvantifikaci vědeckovýzkumného výkonu jednotlivce (An index to quantify an individual's scientific research output)

Katedra fyziky, University of California at San Diego, La Jolla, CA 92093-0319

**Navrhuji index  $h$ , definovaný jako počet prací s počtem citací  $\geq h$ , jako index, použitelný k charakterisaci vědeckého výkonu výzkumného pracovníka.**

Pro nové vědecké pracovníky, kteří se snaží dosáhnout na Nobelovu cenu, je impakt a relevance jejich výzkumu záležitostí nespornou. U ostatních je otázkou, jak kvantifikovat kumulativní impakt a relevanci výkonu vědeckého výzkumu? Ve světě omezených zdrojů je taková kvantifikace (i když se možná leckomu přičí) často potřebná pro účely hodnocení a porovnání (např. univerzitní náborů na fakulty a povýšení, udělování grantů apod.).

Záznamy o publikační činnosti jednotlivce a citační záznamy obsahují nesporně užitečné informace. Tyto informace obsahují počet ( $N_p$ ) separátů, publikovaných během  $n$  let a počet citací ( $N_c$ ) každého separátu, údaje o periodikách, kde byly práce publikovány, jejich impaktový parametr, atd. Tato spousta informací je z různých hledisek hodnocena, přičemž hodnocení různými hodnotiteli se liší. Rád bych zde navrhl jediné číslo, index „ $h$ “, jako zvláště jednoduchý, užitečný způsob k vyjádření vědeckého výkonu výzkumného pracovníka.

Vědátor má index  $h$ , pokud jeho nebo její  $N_p$  práce (každé) má alespoň  $h$  citací a každá z ostatních prací ( $N_p - h$ ) má  $\leq h$  citací.

Zde uvedený výzkum se soustředil na fyziky; domnívám se ovšem, že  $h$  index by měl být použitelný i pro jiné vědecké disciplíny. (Na konci práce rozebírám některá pozorování  $h$  indexu v biologických vědách). Nejvyšší  $h$  u fyziků se zdá být E. Wittenovo  $h$ , které činí 110. To znamená, že Witten napsal 110 prací, z nichž každá (jednotlivě) měla 110 citací. To dává hraniční celkový počet citací Wittenových prací při  $h^2 = 12\,100$ . Ovšem, celkový počet citací ( $N_{c,tot}$ ), bude obvykle mnohem vyšší než  $h^2$ , neboť  $h^2$  podceňuje jak celkový počet citací  $h$  nejcitovanějších prací tak ignoruje práce s  $< h$  počtem citací. Vztahy mezi  $N_{c,tot}$  a  $h$  budou záviset na detailní formě konkrétní distribuce (1) a je užitečné definovat proporcionalitu konstanty  $a$  jako

$$N_{c,tot} = ah^2. \quad (1)$$

Empiricky jsem došel k tomu, že  $a$  má rozsah 3 až 5.

Další prominentní fyzikové s vysokými  $h$  jsou A.J.Heeger ( $h=107$ ), M.L.Cohen ( $h=94$ ), P.W.Anderson ( $h=91$ ), S.Weinberg ( $h=88$ ), M.E.Fisher ( $h=88$ ), M.Cardona ( $h=86$ ), P.G.deGennes ( $h=79$ ), J.N.Bahcall ( $h=77$ ), Z.Fisk ( $h=75$ ), D.J.Scalapino ( $h=75$ ), G.Parisi ( $h=73$ ), S.G.Louie ( $h=70$ ), R.Jackiw ( $h=69$ ), F.Wilczek ( $h=68$ ), C.Vafa ( $h=66$ ), M.B.Maple ( $h=66$ ), D.J.Gross ( $h=66$ ), M.S.Dresselhaus ( $h=62$ ), a S.W.Hawking ( $h=62$ ). Tvrdím, že  $h$  je vhodnější než jiná jednočíslová kritéria, běžně používaná k hodnocení vědeckého výkonu výzkumníků, viz následující:

- (i) Celkový počet prací ( $N_p$ ). Výhoda: měří produktivitu. Nevýhoda: neměří důležitost impaktu prací.
- (ii) Celkový počet citací ( $N_{c,tot}$ ). Výhoda: měří celkový impakt. Nevýhoda: těžko se získává a může být nafouklý malým počtem „velkých hitů“, které nemusí být pro jednotlivce reprezentativní, pokud je (on nebo ona) jedním z mnoha spoluautorů těch prací. V takových případech vztah v rovnici 1 vede k velmi netypické hodnotě

$a > 5$ . Jinou nevýhodou je, že  $N_{c,tot}$  přisuzuje větší váhu značně citovaným přehledovým článkům oproti původním vědeckým příspěvkům.

- (iii) Citace na separát (to je poměr  $N_{c,tot}$  k  $N_p$ ). Výhoda: umožňuje porovnání vědátorů různého věku. Nevýhoda: těžko se údaje získávají, cení více malou produktivitu, trestá vysoký výkon.
- (iv) Počet „významných prací“, definovaný jako počet sdělení s  $>y$  citací (např.  $y=50$ ). Výhoda: vyloučí nevýhody kritérií i, ii, a iii a poskytne obraz o širokém a únosném impaktu. Nevýhoda:  $y$  je libovolné a náhodně jednotlivcům prospěje nebo uškodí, dále  $y$  musí být přizpůsobeno různé věkové úrovni.
- (v) Počet citací každé z  $q$  nejcitovanějších prací (např.  $q = 5$ ). Výhoda: vyloučí mnoho nevýhod výše uvedených kritérií. Nevýhoda: Není to jediné číslo, což znesnadňuje závěr a porovnání  $q$  je také libovolné a náhodně zvýhodní či znevýhodní jednotlivce.

Místo toho, navrhovaný index  $h$  měří rozsáhlý impakt práce jednotlivce, vylučuje všechny nevýhody výše zmíněných kritérií, obvykle je jednoduše zjistitelný objednaním údajů o citacích „za stanovené období“ podle databáze Thomson ISI Web of Science (<http://isiknowledge.com>) a poskytuje přibližný odhad celkového počtu citací (rovnice 1).

Tím tvrdím, že dva jedinci s podobnými  $h$  jsou v rámci jejich celkového impaktu srovnatelné, i když jejich celkový počet citací se značně liší. Naopak, při porovnání dvou jedinců (stejného vědeckého stáří) s podobným počtem veškerých prací nebo všech citací a velmi odlišných hodnot  $h$ , ten s vyšším  $h$  je pravděpodobně dokonalejším vědcem. Pro daného jednotlivce se dá očekávat, že  $h$  by se mělo v závislosti na čase zvyšovat přibližně lineárně. V nejjednodušším možném modelu předpokládejme, že výzkumník publikuje  $p$  prací za rok a že každá práce docílí  $c$  nových citací za rok za každé následné roční období. Celkový počet citací po  $n+1$  let pak je

$$N_{c,tot} = \sum_{j=1}^n pcj^r = \frac{pcn(n+1)}{2} \quad (2)$$

Za předpokladu, že všechny práce až do roku  $y$  přispějí k tvorbě indexu  $h$ , bude

$$(n-y)c = h \quad (3a)$$

$$py = h \quad (3b)$$

Levá strana rovnice 3a je počtem citací nejnovějších prací, přispívajících k  $h$ ; levá strana rovnice (3b) je celkový počet, tvořící index  $h$ . Odtud, z rovnice 3,

$$h = \frac{c}{1+c/p} h^2 \quad (4)$$

Celkový počet citací (pro nepřilíš malé  $n$ ) je pak přibližně

$$N_{c,tot} \sim \frac{(1+c/p)}{2c/p} h^2 \quad (5)$$

formy rovnice 1. Koeficient  $a$  závisí na počtu citací na jednu práci dosažený za rok, jak dáno rovnicí 5. Jak dříve uvedeno, empiricky dochází k závěru, že  $a \approx 3-5$  je typickou hodnotou. Lineární vztah

$$h \sim mn \quad (6)$$

by měl platit obecně pro vědce, kteří tvoří práce stejné kvality stálou intenzitou během celé jejich kariéry; ovšem,  $m$  se bude mezi různými vědci hodně lišit. V jednoduchém lineárním

modelu,  $m$  se vztahuje k  $c$  a  $p$  je dáno rovnicí 4. Zcela obecně, sklon  $h$  oproti  $n$ , parametru  $m$ , by měl zajistit použitelné měřítko (standard) k porovnání vědátorů různého věku.

V lineárním modelu, minimální hodnota  $a$  v rovnici 1 je  $a = 2$ , pro případ  $c=p$ , kde práce s  $>h$  citací a ty s  $<h$  citací rovnoměrně přispívají k celkovému  $N_{c,tot}$ . Hodnota  $a$  bude větší jak pro  $c>p$  tak pro  $c<p$ . Pro  $c>p$ , většina příspěvků k celkovému počtu pochází od „silně citovaných prací“ ( $h$  prací, jež mají  $N_c > h$  prací, jež mají  $<h$  citací každá) jež poskytují nejvyšší přírůstek k  $N_{c,tot}$ . Shledáváme, že první situace obsahuje v sobě velkou většinu, pokud ne všechny případy. Pro lineární model, definovaný v tomto příkladě,  $a=4$  odpovídá  $c/p=5,83$  (další hodnota, která dává  $a=4$ ,  $c/p=0,17$ , není realistická).

Výše definovaný lineární model odpovídá distribuci

$$N_c(y) = N_0 \left( \frac{N_0}{h} - 1 \right) y \quad (7)$$

kde  $N_c(y)$  je počet citací yntého separátu (seřazeno od nejcitovanějšího k nejméně citovanému) a  $N_0$  je počet citací nejcitovanějších prací ( $N_0=cn$  ve výše uvedeném případě). Celkový počet prací  $y_m$  je dán  $N_c(y_m)=0$ ; odtud

$$y_m = \frac{N_0 h}{N_0 - h} \quad (8)$$

Můžeme psát  $N_0$  a  $y_m$  pomocí  $a$ , definovaného v rovnici 1 jako

$$N_0 = h \left[ a \pm \sqrt{a - 2a} \right] \quad (9a)$$

$$N_m = h \left[ a \pm \sqrt{a - 2a} \right] \quad (9b)$$

Pro  $a = 2$ ,  $N_0=y_m=2h$ . Pro větší  $a$ , vyšší znak v rovnici 9 odpovídá případu, kde vysoce citované práce dominují (případ více odpovídající skutečnosti), a nižší symbol odpovídá případu, kdy v počtu citací dominují méně často citované práce.

V realističtějším modelu,  $N_c(y)$  nebude funkce  $y$  lineární. Povšimněte si, že  $a=2$  bude moci bezpečně pokládáno za obecně na nižší úrovni, poněvadž nižší hodnoty  $a$  by mohly potřebovat druhou odvozeninu  $\partial^2 N_c / \partial y^2$ , která bude po velkých částech  $y$  negativní, což je nereálné. Celkový počet citací je dán plochou pod křivkou  $N_c(y)$ , která probíhá bodem  $N_c(h)=h$ . V lineárním modelu, nejnižší  $a=2$  odpovídá linii sklonu -1, jak ukázáno v obr.1.

Realističtější model bude exponenciálně rozložen formou

$$N_c(y) = N_0 e^{-\beta y} \quad (10)$$

Povšimněte si, že pro  $\beta \leq 1$ ,  $N_c''(y) > 0$  pro všechny  $y$ ; odtud  $a > 2$  je pravdivé. Lze psát, že distribuce v rámci  $h$  a  $a$  jako

$$N_c(y) = \frac{a}{2\beta I(\beta)} e^{-\beta y} \quad (11)$$

$$I(\beta) = \int_0^{\infty} dz e^{-\beta z} \quad (12)$$

s  $I(\beta)$  integrál

a  $\alpha$  stanovené rovnici

$$\alpha e^{\alpha} = \frac{c}{h(\beta)} \quad (13).$$

Maximálně citované práce mají citací dle (s nejméně jedinou citací) je stanoven pomocí  $N(y_m)=1$  podle

$$N(y_m) = \frac{c}{ah(\beta)} e. \quad (14),$$

$$y_m = h[1 + \alpha^{\beta} \ln(\frac{c}{h})]^{1/\beta}. \quad (15).$$

Daná distribuce výzkumníků může být modelována volbou většiny odpovídajících  $\beta$  a  $\alpha$  pro ten který případ. Např. pro  $\beta=1$ , je-li  $a=3$ ,  $\alpha=0,661$ ,  $N_0=4,54h$ , a  $y_m=h [1+,661nh]$ . S  $a=4$ ,  $\alpha=0,4644$ ,  $N_0=8,61h$ , a  $y_m=h[1 + 0,461n(h)]$ . Pro  $\beta=0,5$  nejnížší možná hodnota  $a$  je 3,70; v tom případě  $N_0=7,4h$  a  $y_m=h[1 + 0,51n(h)]^2$ . Větší hodnoty  $a$  zvýší  $N_0$  a zredukují  $y_m$ . Pro  $\beta=2/3$ , nejmenší možné  $a$  je  $a=3,24$ , pro kterýžto případ  $N_0=4,5h$  a  $y_m=h[1 + 0,661n(h)]^{3/2}$ .

Lineární vztah mezi  $h$  a  $n$  (rovnice 6) se ovšem zhoří, když výzkumník zpomalí produkci článků nebo zcela přestane publikovat. Mezi dvěma událostmi je časová prodleva. V lineárním modelu za předpokladu, že vědátor přestane publikovat po  $n_{stop}$  letech,  $h$  se stále

$$n_{stop} = \frac{h}{c} = \frac{1}{1 + c/p} n_{stop} \quad (16)$$

zvyšuje stejnou rychlostí po dobu a pak zůstane stálý, poněvadž nyní všechny publikované práce přispívají k  $h$ . V realističtějším modelu se  $h$  hladce ustálí během zvýšení  $n$  spíše než v závislosti na přerušené změně sklonu. Časová prodleva bude i tak větší u vědátorů, kteří publikovali po mnoho let, jak o tom svědčí rovnice 16.

Navíc, ve skutečnosti ne všechny práce musí přispět k  $h$ . Některé práce s nízkými citacemi nikdy výzkumníkům index  $h$  nepodpoří, zvláště když byly sepsány v pozdní části vědátorovy kariéry, když jeho  $h$  je již přijatelně vysoké. Jak uvádí Redner (3), většina prací nabývá svých citací během omezeného časového období popularity a pak již nejsou vůbec citovány. Odtud je možno odvodit, že půjde o případ, že práce, které přispěly k výzkumníkovu  $h$  v rané fázi jeho kariéry již nic k jeho vyššímu  $h$  nepřinesou ani v pozdější části kariéry jednotlivce. Nicméně je ovšem vždy pravda, že  $h$  se nemůže v průběhu času snižovat. Práce, které v kterémkoli daném čase mají přesně  $h$  citací jsou v riziku eliminace ze součtu jednotlivce poněvadž jsou přebity jinými pracemi, jež jsou více citovány. Je též možné, že práce „vypadnou“ a poté později přijdou zpět opět do součtu  $h$ , což se stává hlavně u prací, nazývaných „spící krásky“ („sleepy beauties“)(4).

U jednotlivých výzkumných pracovníků zmíněných dříve, shledávám  $n$  od doby, uběhlé od jejich prvního separátu do nynějška a nacházím tyto hodnoty pro směrnici křivky  $m$  definovanou v rovnici 6: Witten,  $m=3,89$ ; Heeger,  $m=2,38$ ; Cohen,  $m=2,24$ ; Gossard,  $m=2,09$ ; Anderson,  $m=1,88$ ; Weinberg,  $m=1,76$ ; Fisher,  $m=1,91$ ; Cardona,  $m=1,87$ ; deGennes,  $m=1,75$ ; Bahcall,  $m=1,75$ ; Fisk,  $m=2,14$ ; Scalapino,  $m=1,88$ ; Parisi,  $m=2,15$ ; Louie,  $m=2,33$ ; Jackiw,  $m=1,92$ ; Wilczek,  $m=2,19$ ; Vafa,  $m=$ ; Maple,  $m=1,94$ ; Maple,  $m=1,94$ ; Gross,  $m=1,69$ ; Dresselhaus,  $m=1,41$ ; a Hawking,  $m=1,59$ . Z prohlédnutí citačních záznamů mnoha fyziků docházím k tomuto závěru:

(i) hodnota  $n \approx 1$  (to je index 20 po 20 letech vědecké aktivity), charakterizuje úspěšného vědce.

(ii) hodnota  $m \approx 2$  (to je  $h$  index 40 po 20 letech vědecké činnosti) charakterizuje vynikající vědce, pravděpodobně se bude vyskytovat jen na předních univerzitách nebo v hlavních výzkumných laboratořích.

(iii) hodnota  $m \approx 3$  nebo vyšší (to je  $h$  index 60 po 20 letech, nebo 90 po 30 letech) charakterizuje skutečně unikátní jedince.

Parametr  $m$  přestává být užitečný, pokud vědátor si neudrží svou úroveň produktivity, zatímco parametr  $h$  si ponechává tuto vlastnost jako míru kumulativního dosažení, které se může neustále zvyšovat a to i dlouho poté, co vědec přestal publikovat.

Založeno na nalezených typických  $h$  a  $m$  hodnotách, domnívám se (s velkými zábrany proti chybám), že v případě fakult a velkých výzkumných univerzit,  $h \approx 12$  by mohlo být hodnotou, typickou pro postup do funkce plnohodnotného profesora. Členství v American Physical Society by typicky mohlo nastat při  $h \approx 15-20$ . Členství v National Academy of Sciences of the United States of America by mohlo být spojeno s  $h \approx 45$  a vyšším, až na výjimečné okolnosti. Povšimněte si, že tyto odhady zhruba odpovídají typickému počtu let podporovaného výzkumu přičemž se předpokládá, že hodnota  $m \approx 1$ ; časové úseky budou ovšem kratší v případě výzkumníků s vyššími hodnotami  $m$ . Zaznamenejte, že časové odhady jsou odvozeny od publikování první práce, k čemuž běžně dochází několik let před dosažením Ph.D.

Existuje, ovšem, řada námitek, které by měly být uvažovány. Jediná hodnota zřejmě nikdy neposkytne více než hrubou aproximaci mnohoznačného profilu jednotlivce, a nemělo by být uvažováno mnoho dalších činitelů v kombinaci s hodnocením jednotlivce. Navíc, skutečnost, že vždy se mohou u pravidel vyskytnout výjimky, zvláště v případě životních událostí jako udělování grantů, neplnění doby řešení atd. Mohou nastat rozdíly v typických hodnotách  $h$  různých vědních oblastí (přestože, u první aproximace v rozsáhlém oboru je více vědců, kteří sdílejí větší počet citací, takže typické hodnoty  $h$  by nutně nemusely být větší). Vědci, pracující mimo nosné obory nedosáhnou tak velkých hodnot  $h$  jako nejuspěšnější část těch, kdo pracují v oborech o něž je nejvyšší zájem. Přestože argumentují, že vysoké  $h$  je spolehlivým indikátorem vysokého výkonu, opak není vždy pravdou. V rámci daného podoboru je značná variace ve zkreslení distribuce citací a u autorů s relativně nízkým  $h$ , kteří mají několik málo původních prací s mimořádně vysokými citacemi,  $h$  index nebude plně odrážet stav autorových schopností. Obráceně, vědátor s vyšším  $h$ , dosaženým převážně prostřednictvím prací s velkým počtem spoluautorů by byl díky svému vysokému  $h$  příliš přeceňován. Podobory s typicky rozsáhlými spolupracemi (např. pokusy s vysokou energií) budou vykazovat vyšší hodnoty  $h$ , a domnívám se, že v případech velkých rozdílů v počtu spoluautorů by mohlo být užitečné porovnání různých jednotlivců, čímž by bylo možno dosáhnout normalizace  $h$  činitelem, který obráží průměrný počet spoluautorů. Ke stanovení vědeckého „stáří“ při vypočtení  $m$ , vůbec první práce nemusí být někdy vhodná jako výchozí bod, pokud představuje relativně malý raný příspěvek (k problematice) než autor dosáhne běžného výkonu.

Konečně, v jakémkoli systému citací, by měly být vyloučeny samocitace. Přestože samocitace mohou skutečně zvýšit index  $h$  vědeckého pracovníka, jejich dopad na výši  $h$  je mnohem menší než na celkový počet citací. Za prvé, všechny samocitace prací s  $<h$  citací jsou irelevantní, stejně jako samocitace prací, které mají více než  $h$  citací. K opravě  $h$  od samocitací bychom měli vzít v úvahu práce s počtem citací právě  $>h$  a spočítat samocitace v každé z těchto prací. Pokud práce s  $h+n$  citacemi má  $>n$  samocitací, mohla by se z  $h$  vypustit, a  $h$  by se o 1 snížilo. Obvykle by se jednalo o několik málo pokud vůbec o nějaké vypuštěné práce. Druhá strana této mince je, že vědečtí pracovníci, zamýšlející si samocitacemi zvýšit svůj  $h$  index, by se zaměřili na práce s počtem citací právě  $<h$ .

Při volbě zajímavé vzorové populace jsem se zaměřil při výpočtu  $h$  a  $m$  na fyziky, kteří získali Nobelovy ceny v posledních dvaceti letech (pro výpočet  $m$  jsem použil druhého ročního údaje z dat o počtu prvně publikovaných prací za rok nebo údaje z roku 1955, který byl prvním rokem databáze ISI). Ovšem, soubor byl dále omezen zahrnutím pouze těch jmen, která identifikovala vědce v citačním indexu ISI, který omezil náš soubor na 76% celkového objemu. Je to ovšem stále férový nástroj odhadu, neboť častost jména by neměla být v korelaci s  $h$  a  $m$ . Index  $h$  indikuje rozsah od 22 do 79, a  $m$  indikuje rozsah od 0,47 do 2,19. Průměry a standardní odchylky jsou  $(h)=41$ ,  $\delta_h=15$  a  $(m)=1,14$ ,  $\delta_m=0,47$ . Distribuce  $h$  indexu jeobrazena obr.2; medián je při  $h_m=35$ , níže než průměr následkem výše koncových hodnot pro vysoké hodnoty  $h$ . Je zajímavé, že respondenti Nobelových cen mají podstatné  $h$  indexy (84% mělo  $h$  nejméně 30), což svědčí o tom, že Nobelovy ceny nejsou udělovány těm, které na poli vědecké práce potkalo jednorázové štěstí. Zjevně, zjištěné hodnoty  $m$  často nebývají v porovnání s ostatními úspěšnými vědátory vysoké (49% z našeho vzorového souboru mělo  $m < 1$ ), zřejmě proto, že Nobelovy ceny jsou často udělovány dlouho po období nejvyšší produktivity výzkumných pracovníků.

Jiný vzorek, nově jmenovaní členové National Academy of Sciences za fyziku a astronomii v roce 2005, měli  $(h)=44$ ,  $\delta_h=14$ , nejvyšší  $h=71$ , nejnižší  $h=20$ , a členové Academy of Sciences za fyziku, podskupina posledních jmen, začínajících „A“ a „B“ má  $(h)=38$ ,  $\delta_h=10$  a  $h_m=37$ . Tyto příklady dále svědčí o tom, že index  $h$  je stabilní a důsledný odhadce vědeckého úspěchu.

Fascinující myšlenkou je rozšíření konceptu  $h$ -indexu na skupiny jednotlivců (což bylo poprvé zavedeno databází SPIRES). Databáze fyziky vysokých energií Spire (www.slac.stanford.edu/spires/hep ) nyní včlenila  $h$ -index do svých citačních sumářů, a též umožňuje skupinám vědců vypočtení  $h$ -indexu. Celkový  $h$  index skupiny obecně bude větší než jednotlivců téže skupiny, ale menší než součet  $h$ -indexů jednotlivců, jelikož některé z prací, které přispívají k  $h$ -indexu každého jednotlivce již nebudou přispívat k  $h$  skupiny. Na příklad, celkový  $h$ -index skupiny zhutnění hmoty na University of California at San Diego, katedře fyziky činí  $h=118$ , ze které nejvyšší příspěvek jednotlivce činí 25; nejvyšší individuální index je 66 a součet individuálních  $h$ -indexů je  $>300$ . Příspěvek každého jednotlivce k celkovému  $h$  skupiny nemusí být proporcionální individuálnímu  $h$ , a nejvyšší hodnota přispěvatele skupinového  $h$  nemusí být jednotlivec s nejvyšším  $h$ . Pro perspektivního graduovaného studenta s ohledem na rozdílné doktorandské programy, zařazení skupin nebo kateder v jeho nebo její oblasti zájmu vzhledem k jejich celkovému  $h$ -indexu by mělo být též zajímavé analyzovat pro administrátory jichž se tyto záležitosti týkají, to samé se týká zařazení kateder i celé instituce.

Závěrem diskutuji některá pozorování v oblasti biologických a biomedicínských věd . Ze seznamu nejcitovanějších vědců v období 1983-2005, sestaveného Christopherem Kingem od Thomson ISI jsem zjistil, že  $h$ -indexy prvních deseti osob z tohoto seznamu, všech pracujících v biologických vědách, kteří jsou, podle klesajícího  $h$ , zařazení takto: S.H.Snyder,  $h=191$ ; D.Baltimore,  $h=160$ ; R.C.Gallo,  $h=154$ ; P.Chambon,  $h=153$ ; B.Vogelstein,  $h=151$ ; S.Moncada,  $h=143$ ; C.A.Dinarello,  $h=138$ ; T.Kishimoto; R.Evans,  $h=127$ ; A.Ullrich,  $h=120$ . Z toho je vidět, že nepřekvapivě, všichni z tohoto souboru vysoce citovaných výzkumných pracovníků mají též vysoké indexy  $h$  a že vysoké  $h$ -indexy jsou v biologických vědách mnohem vyšší než ve fyzice. Mezi 36 novými vědeckými členy v National Academy of Science v biologických a biomedicínských vědách v roce 2005 shledávám  $(h)=57$ ,  $\delta_h=22$ , nejvyšší  $h=18$  a medián  $h_m=57$ . Tyto pozdější výsledky mají sklon být vyšší než ve fyzice; ovšem, také nasvědčují tomu, že rozdíl se zdá být mnohem vyšší u špiček než u průměru. Zřejmě by bylo zajímavé provést další průzkum k porozumění podobnosti a rozdílů v  $h$ -indexu u různých vědních oborů.

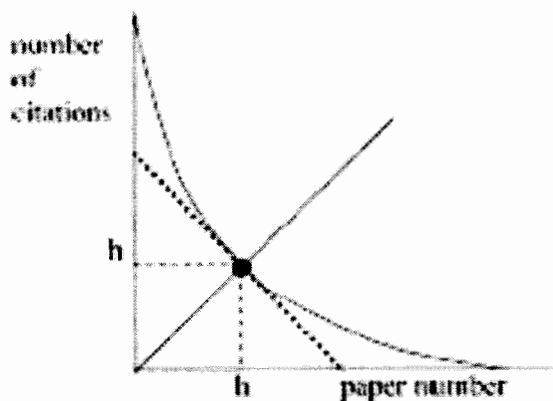
Shrnuto, navrhl jsem jednoduše vypočitatelný index  $h$ , který poskytuje odhad důležitosti, významnosti a širokého impaktu vědeckých kumulativních sdělení. Domnívám se, že tento index může zajistit užitečné měřítko, pomocí kterého lze nezaujatě porovnávat různé jednotlivce, soutěžící o stejný zdroj, kdy důležitým hodnotícím kritériem je vědecký úspěch.

#### Poděkování

Jsem zavázán mnoha kolegům z University of California v San Diego, ve skupině Condensed Matter, zvláště Ivanovi Schullerovi za stimulační diskuse na tato témata a dodání odvahy tyto nápady publikovat. Také děkuji mnoha čtenářům, kteří odepisovali s mnoha zajímavými komentáři od té doby, kdy byla tato práce poprvé zveřejněna v arXiv.org (6); dále recenzentům, kteří měli konstruktivní připomínky, které vedly ke zlepšení obsahu práce; a Travin Brookes a administraci databáze SPIRES za jejich rychlé včlenění  $h$ -indexu do databáze.

#### Legenda k obr. 1

Schematická křivka počtu citací versus počtu prací, s pracemi číslovanými podle počtu citací sestupně. Průsečík přímkou v úhlu  $45^\circ$  s křivkou udává  $h$ . Celkový počet citací je plocha pod křivkou. Za předpokladu, že druhá derivace nikde není negativní, minimální plocha je dána distribucí představovanou tečkovanou přímkou, v rovnici 1 pak vztahem  $a=2$ .



#### Legenda k obr. 2

Histogram, znázorňující počet nositelů Nobelovy ceny ve fyzice za posledních 20 let oproti  $h$ -indexu v rozmezí hodnot 35 a 39.

