

A tudományos kutatás információs hálózatainak hatékonysága

SCHUBERT ANDRÁS
schuba@iif.hu

MTAK TTO, Budapest



A projektek az Európai Unió
támogatásával valósulnak meg.

A hálózati szemlélet fontossága

"...the isolated man does not develop any intellectual power. It is necessary for him to be immersed in an environment of other men, whose techniques he absorbs during the first twenty years of his life. He may then perhaps do a little research of his own and make a very few discoveries which are passed on to other men. From this point of view [research] must be regarded as carried out by the human community as a whole, rather than by individuals." (Alan Turing)

A tudományos aktivitás mennyiségi vizsgálatának és értékelésének – vagyis a tudományometriának – ezért az egyének helyett egyre inkább a közöttük fennálló kapcsolatok felé kell fordítania a figyelmét.

Hatásfok és hatékonyság

Az angol „efficiency” kifejezésnek a magyarban két megfelelője is van: *hatásfok* és *hatékonyság*.

A *hatásfokot* a tényleges teljesítménynek az optimális teljesítményhez mért értékével határozhatjuk meg.

A *hatékonyságot* az eredmények és a ráfordítások arányaként szokás értelmezni.

Egy adott rendszerben rendkívül sokféle dolgot tekinthetünk teljesítménynek, eredménynek és ráfordításnak. Ezeket különféle módon mérhetjük, és többféleképpen értelmezhetjük az elérhető (elérendő) optimumot is. Ennek megfelelően ugyanannak a rendszernek a hatásfoka és a hatékonysága is különböző szempontok szerint egészen különböző lehet.

A hatásfok és a hatékonyság értelmezése és mérése mindig csak egy jól meghatározott cél ismeretében, annak szempontjából lehetséges.

Ez a kitétel egyébként minden értékelési tevékenységnél kulcsfontosságú!

Az információ hatékonysága

Az információ hatékonyságát napjainkban főként a pénzügyi (tőzsdei) és a fogadási piacok vonatkozásában szokás elemezni. Itt a hatékonyságot többnyire azzal a sebességgel mérik, amellyel a releváns információk eljutnak a piacok releváns szereplőjéhez. Ez a szemlélet felhasználható a tudományos információ (vagy pl. az információs szolgáltatások) hatékonyságvizsgálatának során is, de itt többnyire nem a sebességet tekintik a legmeghatározóbb attribútumnak.

A hálózatok hatékonysága

A hálózatok hatékonyságának vizsgálatának szakirodalmi példáit főként a telekommunikációs hálózatok területén találhatjuk, ahol általában a hálózat elemei (csomópontjai) közötti kapcsolati sebesség, az energiafelhasználás vagy a költséget képezik a hatékonyság számításának alapját.

A kutatási hálózatok esetében ezek kevésbé releváns szempontok. Felmerült viszont pl. a hálózat elemei közötti legrövidebb átlagos úthossz, mint hálózati hatékonysági mutatószám.* Ennek az érdekessége az, hogy míg bizonyos hálózatok esetében ez a felhasználók által közvetlenül érzékelhető minőségi tényező (pl. közlekedési hálózatok – átszállások száma), más esetekben (pl. digitális adatátvitel – e-mailek továbbítása) teljességgel rejtve marad.

*Bart Verspagen: Small Worlds and Technology Networks: The Case of European Research Collaboration, KNOW Conference, 2001, Athens.

Információs hálózatok hatékonysága

A szakirodalomban ilyen címszó alatt főleg az az igény fogalmazódik meg, hogy milyen fontos lenne kidolgozni az információs hálózatok hatékonyságának objektív mutatószámait és mérési módszertanát. Bár a témából már nemzetközi konferenciákat is rendeztek, egyelőre nem látszik körvonalazódni általánosan használható fogalmi keret vagy pláne konkrét módszertan.

A tudományos kutatás információs hálózatainak hatékonyságának vizsgálata során tehát nemigen támaszkodhatunk általános eredményekre, hanem egyes esetekre kidolgozott specifikus eredményekből kell építkeznünk.

A tudományos kutatás információs hálózatai

A tudományos kutatás információs hálózatainak példái szerteágazóak és változatosak. Ide tartoznak a kutatók és kutatási intézmények közötti formális és informális kapcsolatok, a közös pályázatok és a hozzájuk kapcsolódó virtuális és valóságos műhelytalálkozók (workshop) és konferenciák, stb. Ezek jelentőségét nem kisebbítve a tudománymetria szokásos és többé-kevésbé bevált szemlélete azoknak a kapcsolati hálózatoknak ad elsőbbséget, amelyek a tudomány formális, hivatalos és ellenőrzött fórumai – mindenekelőtt az elsőközlő tudományos folyóiratok – által publikált információkban jelennek meg. Ezek legjellemzőbb példái a társzerzőségi és az idézettségi kapcsolatok.

A társszerzői hálózat hatékonysága

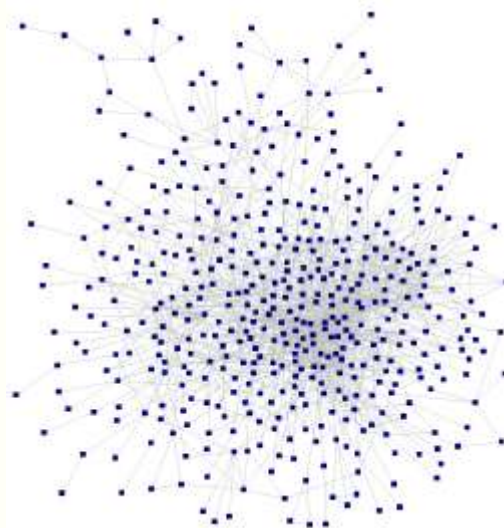
Széles körben bizonyított tény, hogy a tudományos publikációk társszerzőinek átlagos száma gyakorlatilag minden szakterületen folyamatosan növekszik, és általánosan elfogadott vélekedés, hogy a társszerzők számának ésszerű határokon belüli növekedése mind a cikkek létrehozása, mind a cikkben foglalt eredmények minél szélesebb körben való megismertetése szempontjából előnyös. Számos vizsgálat számol be a cikkek társszerzőinek száma és cikk idézettsége közötti pozitív korrelációról, bár egyes esetekben ellenpéldákat is találtak.

Az utóbbi időben egyre nagyobb figyelem irányul arra, hogy társszerzői hálózat eredményes működését nemcsak a társszerzői hálózat sűrűsége (a társszerzők átlagos száma), hanem a hálózat szerkezete is lényegesen befolyásolja. A hálózat szerkezetét jellemző mutatószámok között kereshetjük tehát a hatékonyság indikátorait is.

Az Erdős-szám, a strukturális–evaluatív tudományometriai mutatószám archetipusa

Definíció: Erdős Pál Erdős-száma 0. Egy szerző Erdős-száma n , ha társszerzőinek Erdős-száma közül a legkisebb érték $n-1$.

Erdős Pál társszerzőségi hálózata
(1416 cikk, 509 társszerző)



V. Krebs, <http://www.orgnet.com/Erdos.html>

Erdős-számon alapuló értékelés

↳ Kis Erdős-szám, mint értékmérő

A Nobel-díjasoknak és a tudományos közösség kiemelkedő személyiségeinek az Erdős-száma általában 2–4 között van (Einstein: 2, Neumann: 3, Pauling: 4, Eccles: 3, Gates: 4)

((Telcs: 2, Schubert: 3))

↳ Mérhető-e az "Erdősség"

Milyen mutatószámokkal jellemezhető a kutatóknak a társszerzői hálózatban elfoglalt helyzete?

Centralitás és centralizáltság

↪ Hálózati csomópontok centralitásának mérőszámai

- fokszám centralitás
- közelségi centralitás
- közbülsőségi centralitás
- sajátérték centralitás

↪ A hálózat egészének centralizáltsága az egyes csomópontok centralitásának egyenlőtlenségével mérhető

A centralizáltság mértéke teljes gráfokban minimális,
csillaggráfokban maximális

Erdős Pál centralitása társszerzői hálózatában kétségkívül kiemelkedő, de magának a hálózatnak a centralizáltsága nem kiemelkedően nagy, a hálózat struktúrája valamiféle működési optimum létezését sugallja.

Erdős társszerői hálózatának hatékonysága

“Experts have long argued about the optimal structure of a person's professional network. Some say that a dense, cohesive network brings more social capital, while others argue that a sparse, radial network, one that provides opportunities for innovation and entrepreneurial activity, equates to greater social capital. Erdős's network shows both patterns – a densely connected core along with loosely coupled radial branches reaching out from the core. According to Ron Burt – a leading expert on social capital – this structure may be the optimal pattern for success. According to Burt, radial ties reach diverse information and knowledge to create value, while the dense, trusted ties deliver value in discovered opportunities. For a complete analysis read Burt's seminal work *The Network Structure of Social Capital*.”

Valdis Krebs, <http://www.orgnet.com/Erdos.html>

Hirsch-típusú hálózati mutatószámok

A Hirsch-féle h-index definíciója: egy szerző h-indexe h , ha h cikke kapott legalább h idézetet és a többi nem kapott h -nál többet.

↪ Általánosítás hálózati csomópontokra: egy csomópont h-indexe h , ha h szomszédjának a fokszáma legalább h , és a többié nem nagyobb h -nál..

↪ Általánosítás hálózatokra: egy hálózat h-indexe h , ha h csomópontjának a fokszáma legalább h , és a többié nem nagyobb h -nál.

A csomópontok h-indexe a centralitás, a hálózatoké a centralizáltság egy lehetséges (nem normalizált) mérőszáma.

A. Korn, A. Schubert, A. Telcs: Lobby index in networks, *Physica A*, 388 (2009) 2221–2226

A h-index matematikai háttere

A h-index matematikai hátterül a Gumbel-féle karakterisztikus extrém értékek elmélete szolgál.

Egy $F(k)$ eloszlásfüggvényű valószínűségi változóból vett n elemű minta r -edik Gumbel-féle karakterisztikus extrém értéke $u_r := G^{-1}(r/n) = \max\{k: G(k) \geq r/n\}$, ahol $G(k) := 1 - F(k)$. Az elméleti h-index (H) definíciója ekkor $H := \max\{r: u_r \geq r\}$.

Kimutatható, hogy amennyiben az eloszlás Zipf tulajdonságú, vagyis $\lim_{k \rightarrow \infty} G(k) \cdot k^{-\alpha} = \text{konstans}$ és $n \gg 1$, akkor $H \approx c \cdot n^{1/(\alpha+1)}$, ahol c egy pozitív konstans.

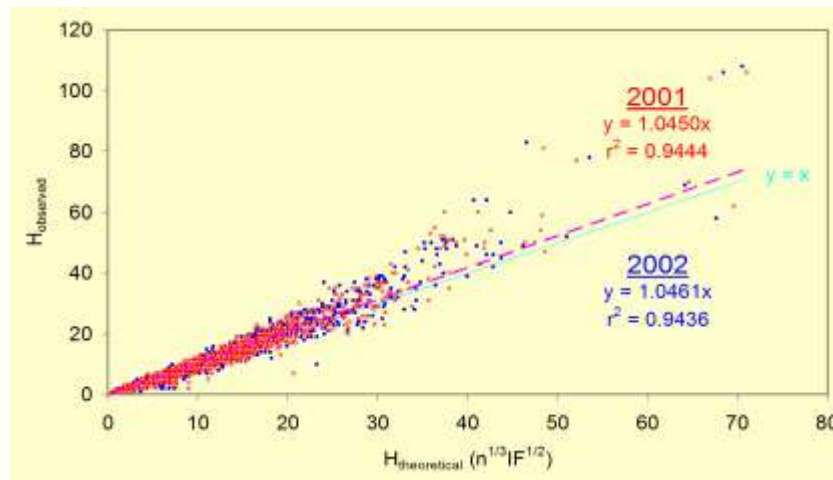
W. Glänzel: On the h-index: a mathematical approach to a new measure of publication activity and citation impact, *Scientometrics*, 67(2) (2006) 315–321.

A h-index függése a publikációk számától és az átlagos idézettségtől

A Glänzel-féle modellből a h-index, a publikációk száma (n) és az átlagos idézettség (x) között egy meglepően egyszerű összefüggés következik:

$$h \approx n^{1/3}x^{1/2}.$$

Az összefüggést nagyobb empirikus mintán először a folyóiratok h-indexe példáján igazoltuk, ahol az x az impakt faktor (IF).*



*A. Schubert, W. Glänzel, A systematic analysis of Hirsch-type indices for journals, *Journal of Informetrics*, 1(3) (2007) 179–184.

Folyóiratok szerzői közösségének „nyomatéka”

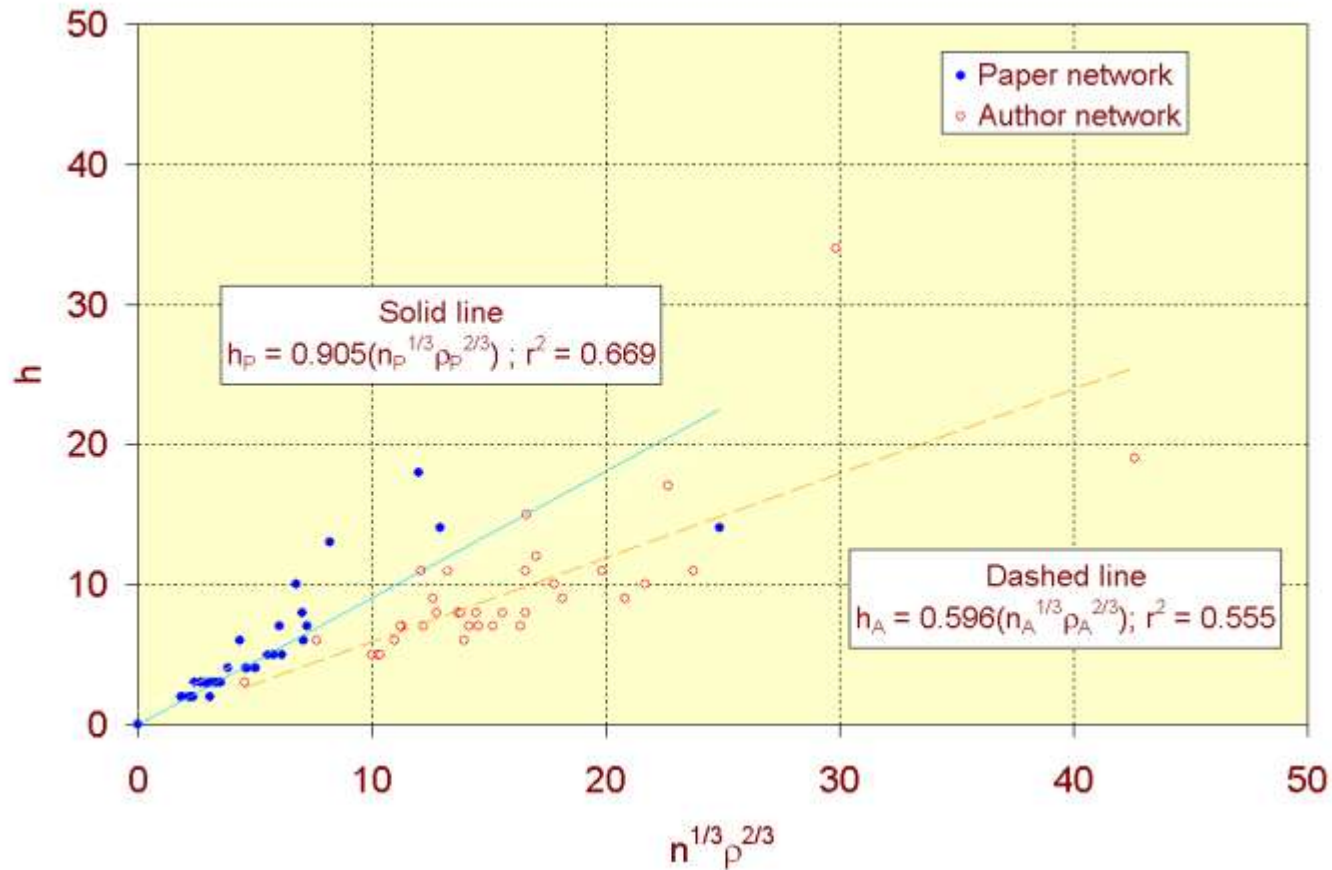
Egy közösségnek a mérete (tagjainak száma) és összetartó ereje (a tagok között fennálló kapcsolatok ereje) ad nyomatékot.

Tekintsük egy folyóirat két szerzőjét összekapcsoltnak, ha van legalább egy közös cikkük. Az így definiált szerzői hálózat h-indexe a közösség „nyomatékának” egy lehetséges mértéke.

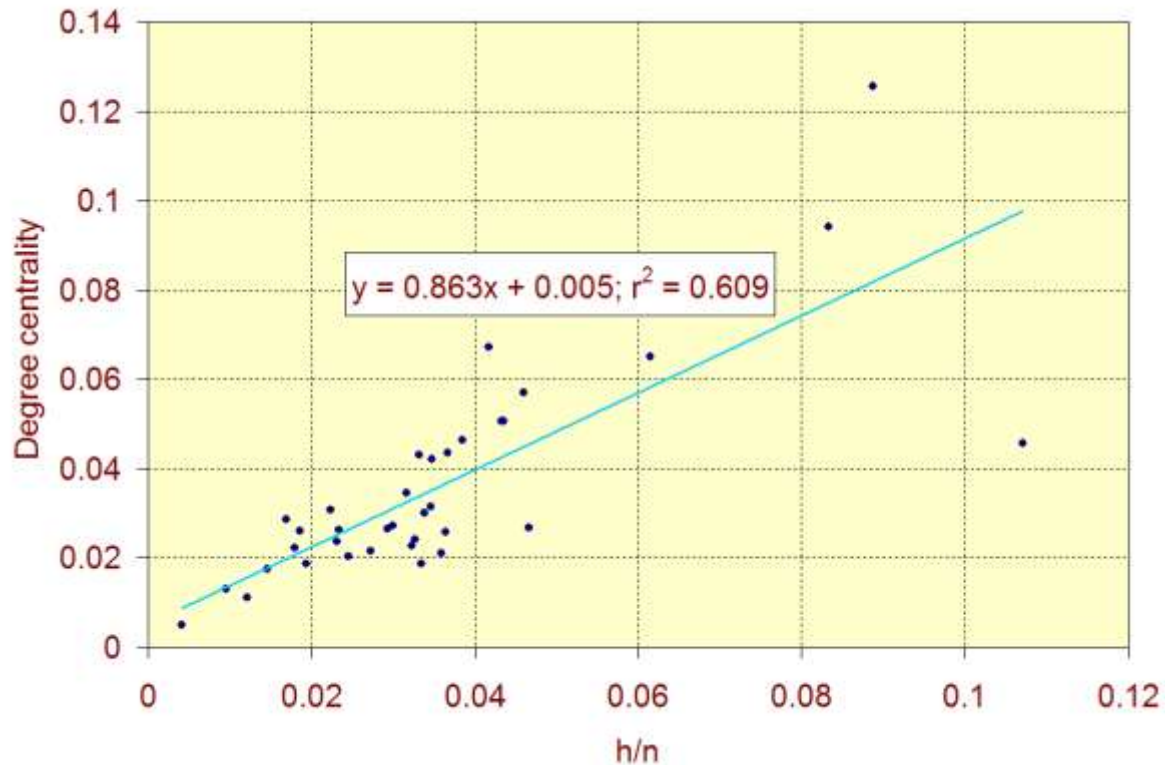
Empirikus vizsgálatunkban 36 fogorvosi folyóirat szerzői közösségét vizsgáltuk az 1999. évi publikációkban.*

*A. Schubert, A. Korn, A. Telcs: Hirsch-type indices for characterizing networks, *Scientometrics*, 78(2) (2009) 375–382.

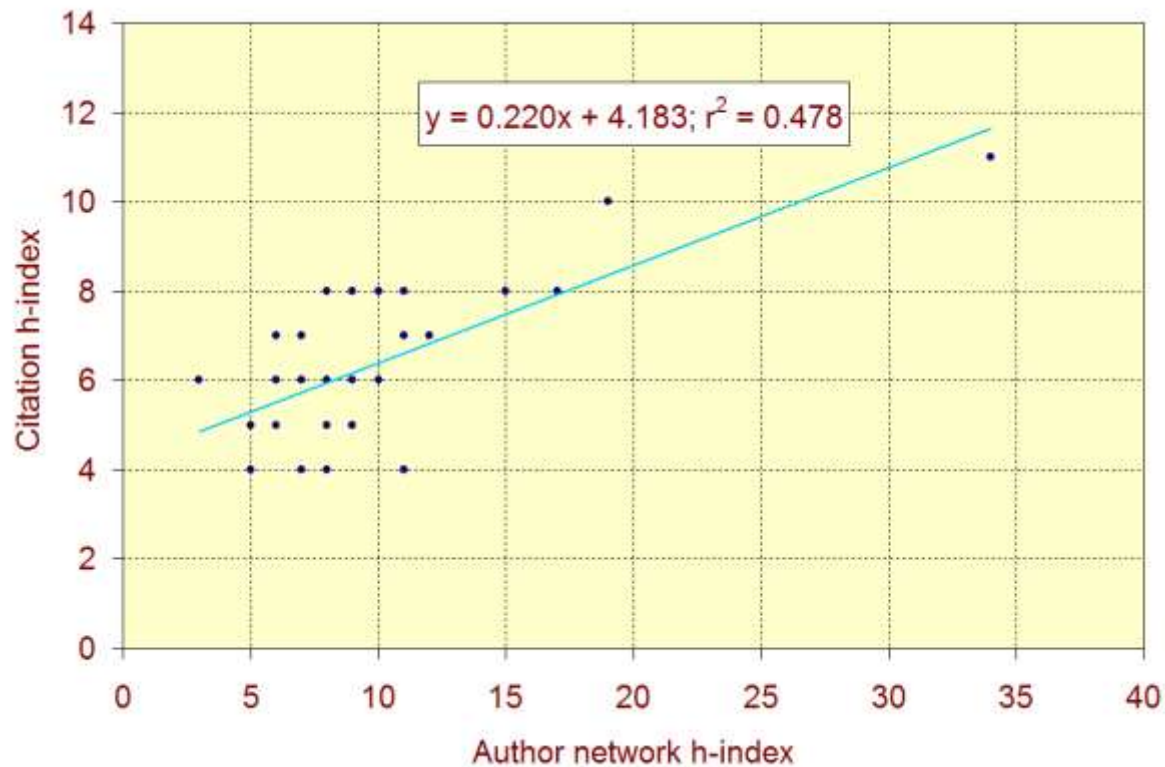
Megvizsgáltuk az egyes folyóiratokban a cikkek (közös szerzők szerinti) és a szerzők (közös cikkek szerinti) hálózatát is. A hálózati h-index mind a két esetben jól követte a Glänzel-féle modellt.



A „h-frakció” (vagyis a h-index és a szerzők számának hányadosa) erős korrelációt mutat a centralizáció egyik legegyszerűbb mérőszámával: a fokszám centralizációval.



Még meglepőbb, hogy a folyóiratok idézettségi h-indexe is határozottan korrelálni látszik a szerzői hálózathalozat h-indexével, vagyis a szerzői közösség „nyomatéka” a cikkek idézettségére is hatással van. (Természetesen ebben az önidézetek és kölcsönös idézetek hatása is szerepet játszik.)



Az idézettségi hálózat hatékonysága

Már az idéztelemzés kezdeti korszakában (az 1970-es években) történtek próbálkozások, hogy az idézeteket ne csak számuk szerint, hanem valamilyen súlyozott módon vegyék figyelembe. Narin és munkatársai a folyóiratok értékelésére vezette be az „influence weight” mutatószámot, amely az ÁKM módszertanát kísérelte meg a folyóiratok információforgalmára átültetni.

Az „influence weight” kései leszármazottjai a Thomson-Reuters Web of Knowledge Journal Citation Reports adatbázisában 2007 óta megtalálható Eigenfactor Score és Article Influence Score mutatószámok, valamint a SCImago (az Elsevier Scopus adatbázisa alapján szerkesztett tudományometriai adatbázis) SCImago Journal Rank (SJR) mutatószáma. Az SJR közvetlenül a Google PageRank mintájára készült.

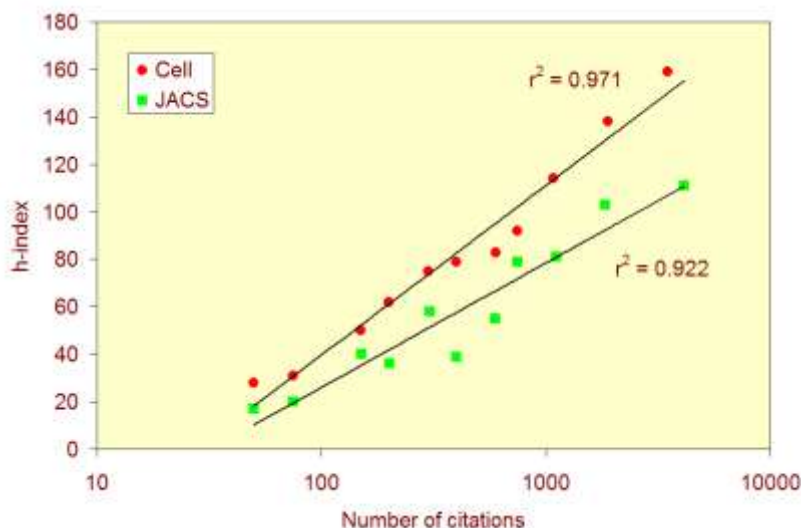
Mindezek a próbálkozások arra irányulnak, hogy az „értékesebb” forrásból származó idézetek nagyobb súlyt kapjanak, mint a „kevésbé értékesek”.

Bizonyos esetekben azonban ezt a célt úgy is elérhetjük, hogy nem használunk mesterséges súlyozást.

Erősen idézett cikkek közvetett idézettségi hatása

Közvetett idézettségi hatás: az idézet nemcsak dokumentálja, hanem generálja is az idézett cikkekre irányuló figyelmet.

Tekintsük egy cikk h-indexének az őt idéző cikkek halmazaának h-indexét.*



Az azonos idézettségű Cell és JACS cikkek közül a Cell cikkek h-indexe nagyobb, mert a Cell cikkeket idéző cikkeket általában többet idézik.

*A. Schubert: Using the h-index for assessing single publications, *Scientometrics*, 78 (3) (2009) 559–565.

A partnerkapcsolati index

Egy tetszőleges partnerkapcsolati hálózat valamely résztvevőjének a partnerkapcsolati indexe (PartnersHIP, PHI) φ , ha n partnere közül φ -vel volt legalább φ közös akciója, míg a többi $(n-\varphi)$ -vel nem volt φ -nél több.*

Társszerzői hálózatok esetében a „közös akció” természetesen a közösen írt cikkeket jelenti.

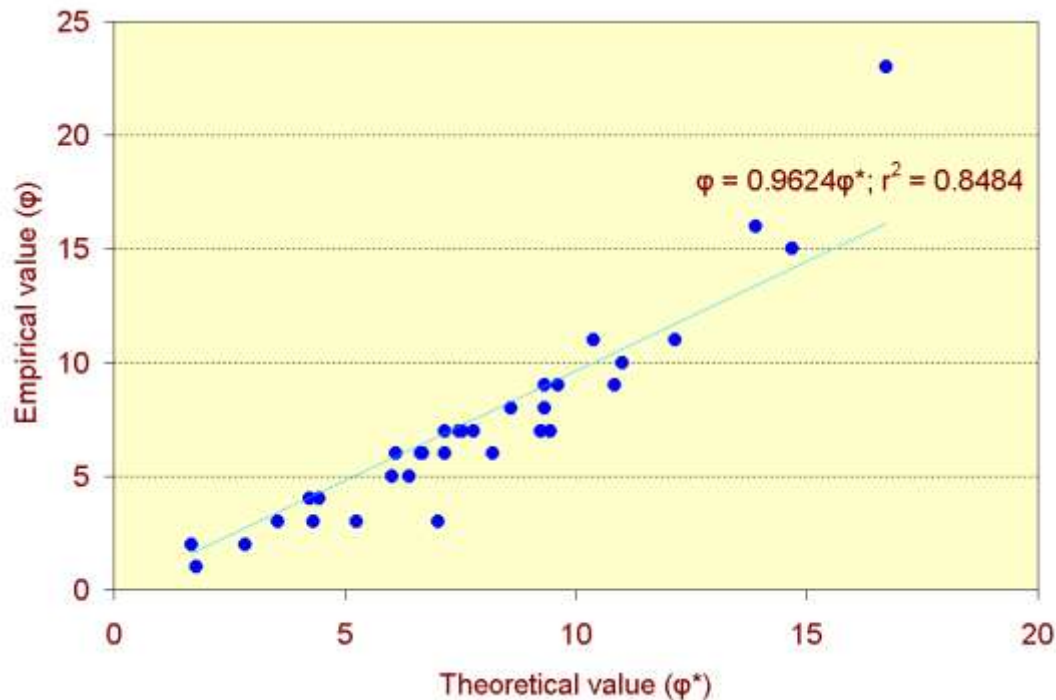
φ értéke akkor és csak akkor 0, ha a szerző csak önálló cikkeket írt.

φ értéke akkor 1, ha (a) a szerző tetszőleges számú kétszerzős cikket írt, mindet ugyanazzal a társszerzővel ÉS/VAGY (b) tetszőleges számú társszerzővel írt cikkeket, de mindegyikkel csakis egyet.

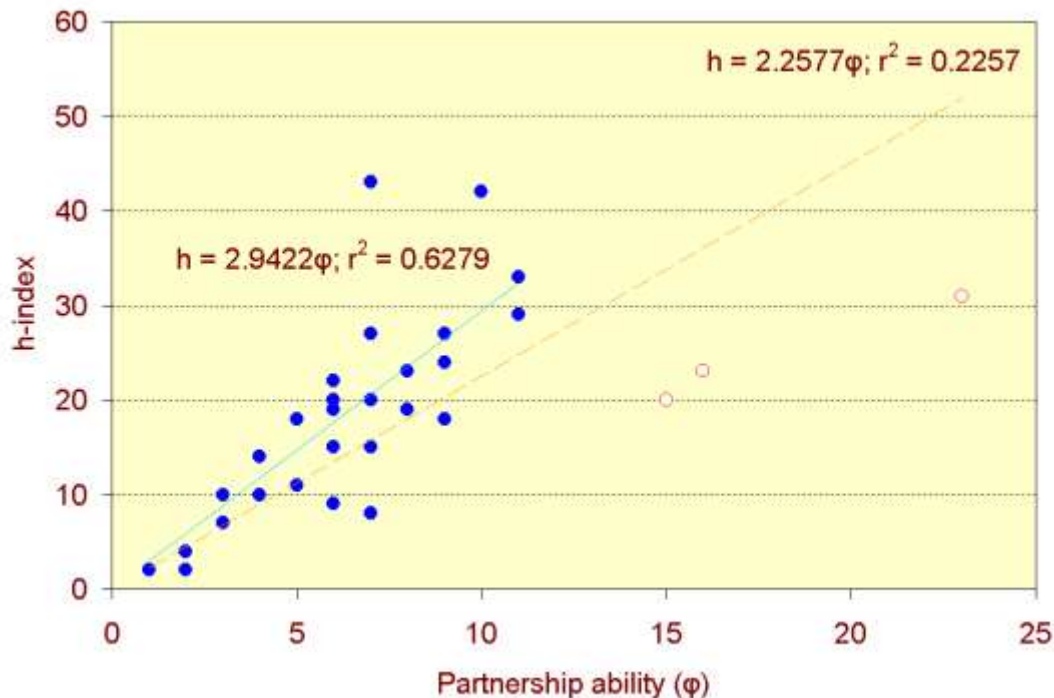
Minden más esetben φ értéke 1-nél nagyobb természetes szám.

*A. Schubert: A Hirsch-type index of co-author partnership ability, *Scientometrics*, 91(1) (2012) 303–308.

A társszerzői partnerkapcsolati index viselkedését egy jeles nemzetközi tudományos díj, a Hevesy Emlékérem díjazottjainak példáján vizsgáltuk meg. A 34 kutató esetében az elméleti modellel kiválóan egyező empirikus értékeket találtunk.



A példán az is beigazolódott, hogy a partnerkapcsolati index növekedése egy bizonyos határig az idézettségi h-indexszel mért „teljesítmény” növekedésével jár együtt. E határ fölött azonban ez az összefüggés megszűnni látszik. Az együttműködés „hatékonyságának” tehát – Ron Burt elképzelésének megfelelően – létezhet egy optimális szintje.



A partnerkapcsolati index egyezése az elméleti modellel egy valóban nagy mintán

Cabanac* a DBLP számítástudományi bibliográfiai adatbázis több, mint egymillió (!) szerzőt tartalmazó mintáján ellenőrizte a φ -index „Schubert–Glänzel modelljét”. Az eredmény látványosan meggyőző.

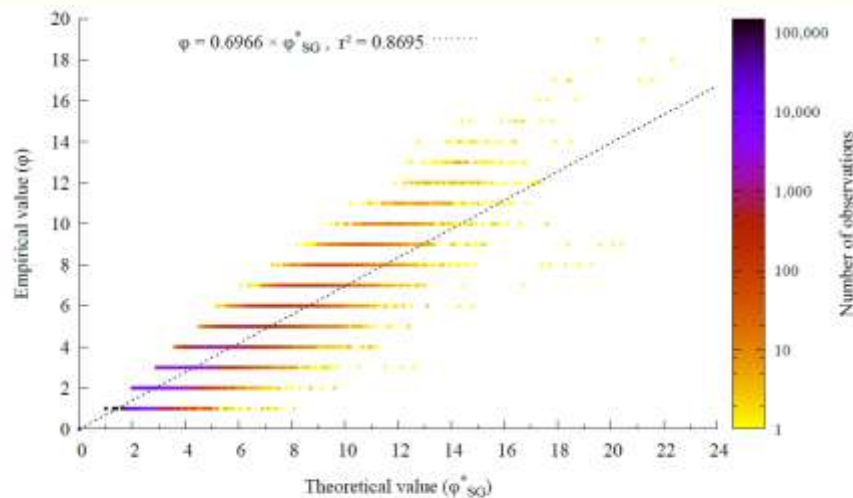
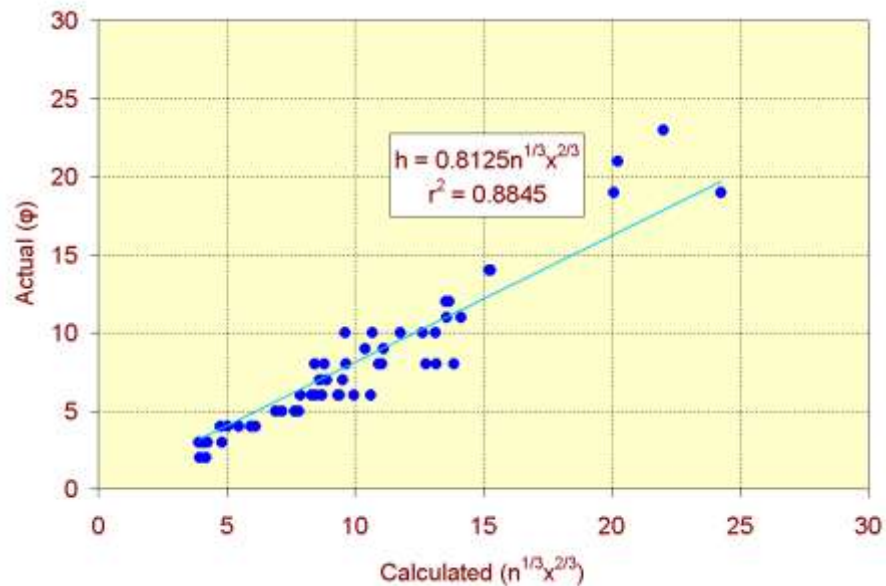


Fig. 2 Linear fit between the theoretical (φ^*_{SG}) and empirical (φ) values of the partnership ability index for the 1,072,213 authors of the DBLP_2012 dataset. Points are colored according to their density: light points show fewer observations than dark points, which show a larger number of observations

*G. Cabanac: Experimenting with the partnership ability φ -index on a million computer scientists, *Scientometrics*, to appear in 2013, DOI 10.1007/s11192-012-0862-y

A partnerkapcsolati index alkalmazása egy, a tudományometriától távolabbi területen

Jazz zenészek esetében a „közös akció” az együttes zenélés, illetve a közös lemezfelvétel készítése lehet. Ennek vizsgálata során ismét meggyőződhattünk az elméleti modell igen általános érvényességéről.*



*A. Schubert: Jazz discometrics – A network approach.
Journal of Informetrics, 6 (2012) 480–484.

Következtetések

- Az információs hálózatok hatékonyságának fogalma igen sokrétű és nehezen definiálható, bár definiálása és mérése rendkívül fontos lenne.
- A tudományos információs hálózatok hatékonyságának egy lehetséges megközelítése az idézettségben mérhető hatás mérése és viszonyítása.
- Az ilyen értelemben vett hatékonyság jellemzésében hasznosnak bizonyulhatnak a Hirsch-típusú mutatószámok.

Köszönettel tartozom

- a konferencia szervezőinek
megtisztelő meghívásukért,
- valamint a tisztelt hallgatóságnak
érdeklődésükért és türelmükért.