



MTA KIK Tudománypolitikai és
Tudományelemzési Osztály

A szakterületi kategorizáció és az ESI-rendszer alkalmazása az MTMT-ben (hatáselemzés)

Projektszám: TÁMOP-4.2.5.A-11/1-2012-0001

Projektcím: A Magyar Tudományos Művek Tára (MTMT) publikációs adatbázis szolgáltatások országos kiterjesztése

Teljesítés száma: 2.

Dátum: 2013. április

Szerzők: Soós Sándor, Schubert András, Horváth Dániel, Belső-Stefán Eszter

A szakterületi kategorizáció és az ESI-rendszer alkalmazása az MTMT-ben (hatáselemzés)

Soós Sándor, Schubert András, Horváth Dániel, Belső-Stefán Eszter

Tartalom

BEVEZETÉS	2
<i>A szakterületi kategorizáció funkciói az MTMT-ben.....</i>	<i>2</i>
<i>Sztenderd bibliometriai kategorizációs/osztályozási rendszerek és az Essential Science Indicators (ESI) kategóriái.....</i>	<i>2</i>
<i>Szerző- vs. közleményalapú kategorizáció.....</i>	<i>3</i>
AZ ESI-RENDSZER ALKALMAZÁSÁNAK VIZSGÁLATA	5
<i>Az MTMT-tartalom leképezése az ESI-rendszerbe.....</i>	<i>5</i>
<i>A szerzők ESI-osztályozásának vizsgálata I: a szerzői profil elemzése.....</i>	<i>5</i>
<i>A szerzők ESI-osztályozásának vizsgálata II: az ESI-kategóriák elemzése.....</i>	<i>8</i>
<i>ESI vs. MAB-osztályozás.....</i>	<i>9</i>
ÖSSZEĞZÉS ÉS AJÁNLÁSOK	15
MELLÉKLET	17

Bevezetés

Az alábbi jelentés az MTMT adatbázis fejlesztésének és felhasználási célrendszerének/lehetőségeinek egy alapvető összetevőjét, az adatbázisban megvalósítható *szakterületi kategorizáció* kérdéskörét vizsgálja. A bemutatott vizsgálatsorozat

- 1) az ESI kategóriarendszer MTMT-ben való alkalmazásának lehetőségét és feltételeit tárja fel, valamint
- 2) a – közleményalapú – ESI-osztályozás és a szerzőalapú besorolási rendszerek kapcsolatát elemzi, konkrétan az ESI-rendszer és a MAB-felosztás összehasonlító vizsgálata (az MTMT-re való párhuzamos probaalkalmazása) révén.

A szakterületi kategorizáció funkciói az MTMT-ben

Az MTMT sokrétű célrendszerének a kategorizációval összefüggő két nagy csoportja a hazai tudomány (publikációs) eredményeinek hozzáférhetővé tétele (mind bibliográfiai, mind pedig repozitóriumi szinten), valamint az adatbázisnak a hazai tudományértékelésben betöltött (hiánypótló) funkciója. A tartalom szakterületi besorolása – csakúgy, mint a hasonló célú nemzetközi, ún. citációs adatbázisokban (WoS-rendszer, Scopus stb.) ennek megfelelően két alapfunkciót szolgál:

- 1) Információ-visszakeresés, rendszerezés. Kézenfekvő felhasználási igény, hogy a hazai tudomány eredményei szakterületi rendszerben is kereshetők, lekérdezhetők legyenek.
- 2) Tudomány- (biblio-) metriai referenciahalmazok képzése. Kevésbé közismert, szakmai elvárás az adatbázis kutatásértékelési hasznosítását illetően, hogy az MTMT-ben olyan szakterületi alapú (közlemény- és/vagy felhasználó-) csoportok, ún. referenciahalmazok legyenek képezhetők, amelyekhez viszonyítva egy-egy szereplő (kutató, intézmény stb.) teljesítménye *az adott szakmában* megítélhető. Erre azért van szükség, mert a tudományometriai teljesítményértékelés egyik alaptétele a szakterület-függőség, amelynek értelmében a teljesítmények a szakterületi jellemzők (pl. idézési szokások és a vonatkozó idézettségi tendenciák) figyelembevételével hasonlítható csak össze. A gyakorlatban ez a szakterületi kategóriák mint tudományometriai szempontból konzisztens referenciahalmazok képzése (és a szereplők nyers mutatóinak a saját kategóriájukra való relativizálása) nyomán valósul meg.

Fontos megjegyezni, hogy a két alapcél egy-egy kategóriarendszer jellemzően eltérő mértékben szolgálja.

Sztenderd bibliometriai kategorizációs/osztályozási rendszerek és az Essential Science Indicators (ESI) kategóriái

A fenti két alapfunkcióból kiindulva célszerű azon osztályozási rendszerek, ill. szakterületi sémák MTMT-beli alkalmazását megfontolni, amelyeket szintén erre a kettős célra alakítottak ki. A rendelkezésre álló sztenderdek köréből ezért nem elsősorban a bibliográfiai leírást

szolgáló rendszerező teauruszok, hanem a nemzetközi citációs adatbázisokhoz, így elsősorban a Thomson *Web of Science*, illetve a *Scopus* szolgáltatásokhoz fejlesztett bibliometriai sémák látszanak alkalmasnak. E sémák közül tudományometriai szempontból kiemelt szerepet játszik a Thomson *Essential Science Indicators* nevű szolgáltatásának szakterületi kategorizációja (**ESI-kategorizáció**), amely egy 22 területi besorolásból álló, egyszintű (nem-hierarchikus) rendszer. Jelenlegi, főként pedig bibliometriai alkalmazásának előnyeit (szemben a további besorolási rendszerekkel) az alábbiakban összegezhetjük:

- Az ESI-rendszer az egyetlen olyan, citációs adatbázishoz kifejlesztett séma, amelynek besorolásai *egyértelműek* (minden közleményt – folyóiratot – egyetlen kategóriához rendel), és *mezoszintű* kategóriarendszert szolgáltat, vagyis viszonylag robusztus, de még informatív szakterületi felbontást képvisel. A rendelkezésre álló alternatív rendszerek mindegyike (így a WoS “Web of Science Categories”, ill. “Subject Categories” sémája, vagy a Scopus azonos nevű kategorizációja) jellemzően 200–300 (mikroszintű) szűkebb tématerületet tartalmaz, amelyek átfednek, vagyis a közlemények általában több tématerülethez tartoznak (a hozzárendelés többértelmű). Ezek a sémák, ha hierarchikusak is (Scopus), nem tartalmaznak olyan szintet, amely egyértelmű felosztást mutat. Határozott előnye ez utóbbi típusú felbontásnak és leképezésnek, hogy hívebben tükrözi a tudomány mindenkori szerveződését (amely jellemzően interdiszciplináris és életlen határu területek rendszereként írható le), ezért az információ-visszakereső funkciót hívebben szolgálhatja. A bibliometriai alkalmazás szempontjából viszont világos hátrány, hogy egyidejűleg több mikroszakterülethez rendelt közlemények alapján a teljesítménymérés alapjául szolgáló referencialalmaz-képzés problematikus, és szükségképpen (a kategorizáción túlmutató) szubjektív megfontolásokat igényel.
- A tudomány- és bibliometriai hasznosítás szempontjából további, és a fenténél is erősebb érv az ESI-rendszer mellett, hogy ennek ún. referenciaértékei az ESI-szolgáltatás révén (az EISZ-előfizetői kör számára hazai feltételek mellett is) nyilvánosan és legálisan hozzáférhetőek. A referenciaértékek az egyes ESI-területek mindenkori produktivitási és idézettségi adatait (normáit) tartalmazzák, és nélkülözhetetlenek a normalizált, a szakterület sajátosságait figyelembe vevő értékelésben.

Szerző- vs. közleményalapú kategorizáció

A nemzetközi citációs adatbázisok sémáinak az MTMT-ben való alkalmazása témakörében megkerülhetetlen kérdés, hogy az érintett szakterületi besorolások mindegyike az indexelt folyóiratokat kategorizálja. Ilyen módon a közlemények a lelőhelyük (a folyóirat) besorolása révén kapcsolódnak a szakterületekhez. Az osztályozás alapegysége tehát a közlemény (*közleményalapú kategorizáció*). Ez a besorolás (a szolgáltatók tájékoztatása szerint) részben empirikus (hivatkozáselemzésre épülő), részben szakértői alapú. Az MTMT fejlesztésében ugyanakkor egyértelműen megnyilvánuló igény mutatkozik az ún. *szezőalapú szakterületi felbontás*, így például a MAB képzési területeket tartalmazó (hierarchikus) rendszer alkalmazására. Ez utóbbi – legalábbis eredeti funkciójában és a legtermészetesebb módon – a szerzőket osztályozza képzési területek (jellemzően a fokozatszerzés területe) szerint. Alapvető ezért a kétféle, vagyis a szerzői és a közleményalapú szakterületi rendszerek – ez esetben konkrétan a MAB- és az ESI-rendszer – közötti viszony tisztázása annak megállapítása céljából, hogy a kettő egymást helyettesítő, avagy kiegészítő (párhuzamos) alkalmazása

milyen előnyökkel/hátrányokkal jár az MTMT használatát és hasznosítását illetően. A vizsgálat kiindulópontja a következőkben foglalható össze:

- A legtöbb esetben a szerzők életműve, ill. sokszor annak szűkebb szakaszai is multidiszciplinárisak. Ezért jogos várakozás, hogy az ESI-rendszer szigorú, vagyis közleményalapú alkalmazása nem eredményez egy-az-egyhez típusú leképezést a szerzők és az ESI-szakterületek között: A szerzők egy-egy szakterület helyett a mindenkori kutatási területüket (és a publikációs szokásaikat) jellemző ESI-profillal lesznek jellemezhetőek, mégpedig a közleményeik ESI-ben való eloszlása szerint. Ez azt is jelenti egyben, hogy a szerzők egy-egy kategóriába való (pl. önkéntes, adatközlő-alapú) besorolása, amilyen MAB-séma bevezetése, eltérő módon osztályozza az MTMT-tartalmat, mint az ESI ilyen jellegű adaptációja. A valódi kérdés azonban éppen az, hogy ez az eltérés milyen mértékű és jellegű.
- Az ESI következetes, és az eredeti alkalmazás sajátosságait megőrző alkalmazása az MTMT-ben egy további korláttal is jár. Az ESI szakterületi kategóriái ugyanis rögzített összetételű folyóirathalmazok: az Thomson *Essential Science Indicators* szolgáltatásának alapját adó folyóiratbázis felosztását adják. Ez a folyóiratbázis tízezres nagyságrendű, és a Web of Science bázisát jól reprezentáló, kimondottan kutatásértékelési célból válogatott „elit” lista. Az ESI-rendszer gyakorta használatos tudományometriai célokra, de a következetes alkalmazás a fentiekből adódóan azt jelentené, hogy az MTMT-tartalomnak csak az része kategorizálható, amely (1) nemzetközi folyóiratcikk és (2) leőhelye, a közlő folyóirat szerepel az ESI-listában. Becslésünk szerint ez a WoS-azonosítóval rendelkező (de legalábbis a WoS-ban indexelt) MTMT-tartalom döntő részét fedi. Ennek a korlátnak egy szakmailag adekvát kezelési módjára az Ajánlások c. szakaszban térünk vissza.

Az ESI-rendszer alkalmazásának vizsgálata

Az ESI-kategóriarendszer MTMT-beli próbaalkalmazását a előzőekben leírt megfontolások alapján három összefüggő kísérletben vizsgáltuk meg. Az alapkérdés mindhárom esetben az volt, hogy az ESI milyen mértékben alkalmas a szerzők szakterületi besorolására, azok tényleges, az MTMT-ben nyilvátartott publikációs portfóliója alapján. A jelen terminológia szerint tehát a kérdés úgy fogalmazható meg, hogy az ESI mint közleményalapú rendszer használható-e szerzőalapú osztályozásként.

Az MTMT-tartalom leképezése az ESI-rendszerbe

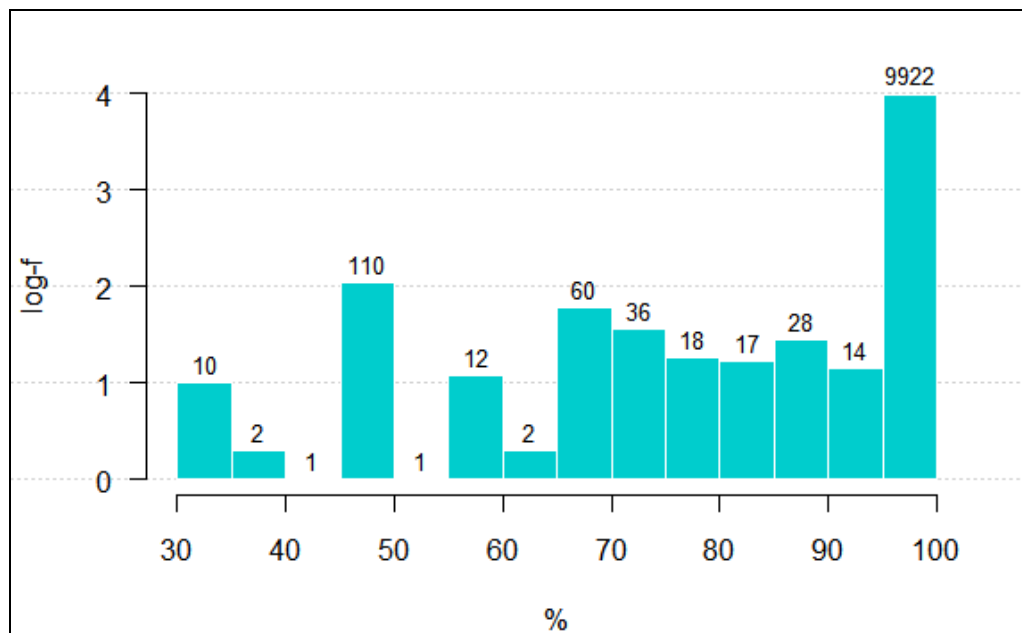
Mindhárom kísérlet alapja egy olyan leképezés létrehozása volt, amely kapcsolatot teremt az MTMT-szerzők és az ESI-kategóriák között, vagyis az ESI szerint kategorizálja a szerzők életművének abban megjeleníthető részét. Ehhez az előkészítő művelethez az MTMT-ben nyilvántartott, az ESI folyóiratbázisában 2000 és 2012 között megjelent, évszám, kötetszám és kezdő oldalszám szerint azonosítható 90 865 publikációt használtuk. Minden cikk minden szerzői azonosítóval (AUTHID) ellátott szerzőjéhez hozzárendeltük a közlő folyóirat ESI szerinti szakterületét.

A szerzők ESI-osztályozásának vizsgálata I: a szerzői profil elemzése

Az első kísérlet arra irányult, hogy képet kapjunk a szerzői profilok koherenciájának mértékéről, vagyis arról, hogy a szerzők publikációs rekordja az ESI szempontjából mennyire egyöntetű, ill. milyen mértékben reprezentálható egy-egy ESI-kategóriával. Az elemzés egy további fontos szempontja volt, hogy feltételeztük a szerzői besorolások időszakosságának lehetőségét az MTMT-ben, vagyis azt, hogy a kutatói életpálya alakulásának függvényében a szerzők (adott esetben ön-) besorolása természetes módon változhat. Ebből fakadóan nem a teljes életművek ESI-összetételét vizsgáltuk, hanem – a maximális rugalmasság, az évenkénti “mobilitás” jegyében – az életművek egyéves szakaszait. A vizsgálati egység ilyen módon egy szerzői év publikációs rekordja volt.

Az éves publikációs listák szakterületi szerkezetét az ún. *domináns kategória részesedésével* jellemeztük. A mutatót úgy határoztuk meg, mint a közlemények megoszlását tekintve legnagyobb részarányú kategória, az ún. *domináns kategória* részesedésének mértékét (százalékban kifejezve), vagyis azt, hogy a leginkább jellemző ESI-szakterület mekkora hányadát teszi ki a publikációs rekordnak – szerzőnként és évekre bontva. Az eredményeket az 1. ábrán tüntettük fel: a hisztogram az éves szerzői profilok eloszlását mutatja be a domináns kategória (százalékos) részesedése szerint.

1. ábra A domináns kategória részesedése (%) a szerzők ESI-profiljából évenként (logaritmusos skála)

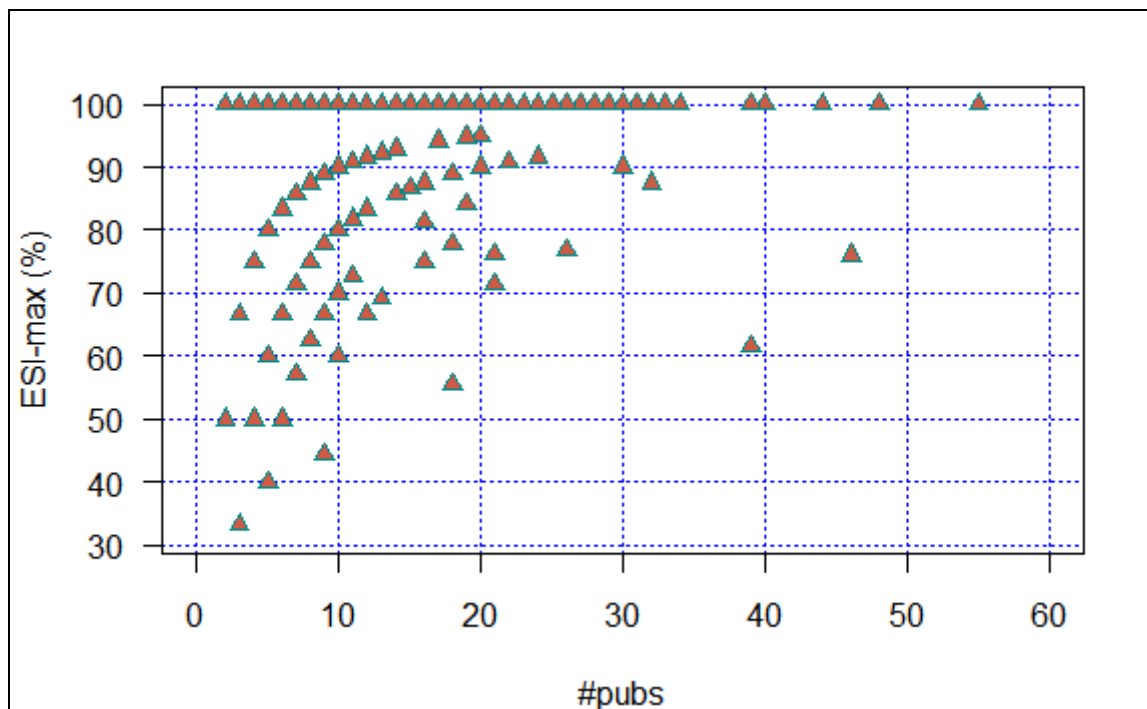


Az 1. ábra tanúsága szerint a szerzők életművének azon része, amely az ESI-ben megjelenik, szűk időablakokban mérve meglepően magas koherenciát mutat. A vizsgált esetek (évek) 97%-ában az évi kibocsátás egyetlen ESI-kategóriában koncentrálódik. A logaritmusos skála jól példázza, hogy (sokszoros) nagyságrendi különbség mérhető az egyöntetű, ill. a valamennyire heterogén profilú évek között: az egyöntetű időszakok tízezres köre mellett a kevésbé egyöntetű évek száma néhány száz.

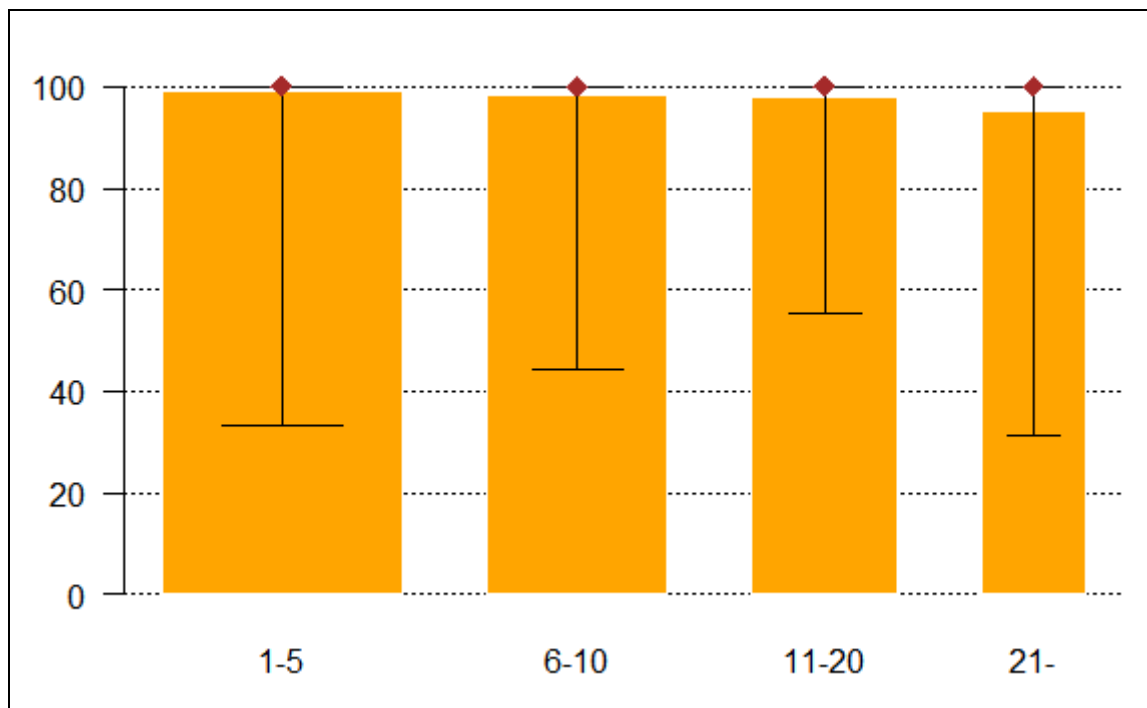
A koherencia ilyen magas mértéke láttán okkal feltételezhető, hogy a választott szűk időablak torzítja az eredményeket. Közelebbről, mivel egy-egy év termése relatíve kicsi (különösen, ha kizárólag az ESI-rendszerben megjelenő közleményekre szorítkozunk), számíthatunk rá, hogy a domináns kategória kizárólagossága sok esetben egyetlen cikket jelent (amely az egyértelmű besorolás okán szükségképpen egyetlen ESI-kategóriát képvisel). Ennek a torzításnak a kiküszöbölése céljából a domináns kategória mellett bevontuk az elemzésbe a közlemények számát is (ami az ESI-közleményeket illeti), vagyis figyelembe vettük, hogy a domináns kategória részesedése hány publikációra vonatkozik.

A közleményszám kezelése céljából a vizsgált szerzők körére bevezettük az ún. *referenciaév* fogalmát, amely egy-egy szerző életművéből azt az évet jelöli ki (az összes bennfoglalt évből), amelyben a domináns kategória részesedése maximális. Másképp fogalmazva, a szerzőket az életmű „legkoherensebb” évével jellemeztük, ilyen értelemben azt vizsgáltuk, hogy munkásságuk éves időszakokra bontva mennyire ESI-koherens a „legjobb esetben”. Ezt a mutatót, vagyis a referenciaév koherenciáját (a domináns kategória részesedését), valamint a referenciaév közleményszámát hasonlítja össze szerzőnként a 2. ábra. Ugyanezt az összehasonlítást élesíti tovább a 3. ábra, amely azt ábrázolja, hogy ezek a koherenciaértékek hogyan oszlanak el a közleményszám alapján képzett szerzőcsoportokban. A referenciaévek közleményszámának eloszlásából kiindulva a szerzőket négy csoportra bontottuk: a referenciaévben (1) maximum 5, a (2) 6–10, a (3) 11–20, ill. a (4) 20 feletti közleményszámmal jellemezhető szerzőkre. A négy csoportot a koherenciaérték átlagával, minimumával, maximumával, ill. mediánjával ábrázoltuk.

2. ábra A domináns kategória részesedése vs. a közlemények száma a referenciaévben



3. ábra A domináns kategória részesedése a referenciaév közleményszámától függően (oszlopok: csoportátlag; hibasávok: csoport minimuma/maximuma; jelölők: a csoport mediánja)



A kimutatások alapján megállapítható, hogy az MTMT-minta a mennyiségi torzításra vonatkozó feltevések ellenkezőjét mutatja: a 2. ábra szerint a domináns kategória magas (sőt, kizárólagos) részesedése egyrészt a közleményszámtól jelentős részben független, ugyanúgy jellemzi az alacsony, mint a magas produktivitású szerzőket, másrészt látható tendencia, hogy

a dokumentumszám növekedésével növekszik (nem pedig csökken) a koherencia. Az alacsonyabb produktivitási tartományokban (pl. [0,10]) ugyan nagyobb a szóródás, de további elemzések megerősítik, hogy az itteni populáció erősebben szóródó része valójában kiugró értékeket/eseteket képvisel.

Ezek az eredmények még letisztultabban jelennek meg a 3. ábrán, amely szerint a négy, rendre növekvő produktivitási szintet jegyző szerzői kör (méretcsoport) átlagos ESI-koherenciája csaknem megegyezik, és majdnem azonos a maximummal. Enyhe csökkenő tendencia mutatkozik ugyan a produktívabb szerzők felé haladva, illetve a szóródás is a 20 feletti közleményszámú méretcsoportban a legnagyobb (bár alig haladja meg a legkisebb méretcsoport értékét, ami ugyancsak ellentmond a korábbi feltevéseknek), de az eloszlás mediánja mindenhol a 100%-os részesedés (vagyis a teljes koherencia). Összegezve tehát megállapítható, hogy az MTMT-szerzők a referenciaév tekintetében igen jól leképezhetők egy-egy ESI-kategóriára.

A szerzők ESI-osztályozásának vizsgálata II: az ESI-kategóriák elemzése

Míg az első vizsgálat fókuszában a szerzői profilok koherenciája állt, a második kísérlet az ESI-kategóriák koherenciájára irányult. Csakúgy, mint az előző esetben, most is a szerzők életművéből, illetve annak ESI-összetételéből indultunk ki. A cél azonban ezúttal az egyes szakterületi kategóriák MTMT-beli egyöntetűségének feltérképezése volt, vagyis lényegében az első kísérlet fordítottja. Az alapkérdés most az adott kategóriához sorolható szerzők közleményeinek ESI-megoszlása volt.

A kategóriák jellemzéséhez az ún. *diszciplináris profil* mutatót vezettük be, amelynek egy további, szerzőszintű mutató, az adott kategóriára vonatkozó *szerzői diszciplináris* képezte az alapját. Egy szerzőnek valamely szakterületre vonatkozó diszciplinárisát úgy definiáltuk, mint a szerző összes cikke között az adott szakterülethez sorolt cikkek százalékos arányát.

A szakterületek szerzői populációjához soroltunk minden olyan szerzőt, akinek legalább egy publikációja az illető szakterülethez tartozott. Az egyes szakterületek diszciplináris profilját úgy készítettük el, hogy a szakterületbe sorolt szerzőket csökkenő diszciplináris szerint sorba rendeztük, és kumulatív gyakorisági diagramot készítettünk (Melléklet, 7. ábra); itt a függőleges tengelyen a diszciplináris, a vízszintes tengelyen a kumulatív gyakoriság található).

A szakterületek szerzői diszciplinárisát, ill. annak ellentettjeként a multidiszciplinárisát a diszciplináris profil különféle statisztikai mutatószámaival jellemeztük. A 1. tábla az átlagértéket, valamint a 100%-os, a legalább 90%-os és a legalább 50%-os diszciplinárisú szerzők arányát tünteti fel minden ESI-kategóriára vonatkozóan. Minél nagyobbak ezek a mutatószámok, annál nagyobb a szakterület diszciplinárisa, ill. annál kisebb a multidiszciplinárisa. A tábla a kategória ún. X-értéke szerint rangsorolja a területeket (ld. a következő szakaszt).

1. tábla Az ESI-területek diszciplináris mutatói az MTMT-ben

X-rank	field	#au	avg	100%	90%	50%	X
1	math	564	52.3%	37.9%	44.3%	69.9%	61.4%
2	soci	1040	18.7%	53.0%	53.2%	65.0%	58.4%
3	geos	615	31.5%	33.3%	37.2%	61.0%	55.1%

4	econ	254	22.4%	42.9%	43.7%	63.4%	53.5%
5	spac	196	39.9%	33.7%	41.9%	58.2%	53.1%
6	clin	3835	46.9%	23.0%	28.5%	58.0%	52.0%
7	plan	1709	35.2%	19.5%	22.5%	53.5%	50.0%
8	chem	2377	35.1%	16.7%	21.0%	49.0%	49.1%
9	agri	1134	22.9%	19.3%	30.6%	43.7%	44.5%
10	phys	1382	37.4%	15.2%	19.9%	42.3%	44.4%
11	comp	625	18.0%	23.4%	24.2%	43.2%	44.2%
12	neur	1826	22.7%	10.6%	11.2%	39.2%	42.8%
13	mate	913	24.2%	15.4%	16.4%	34.7%	40.0%
14	engi	1890	16.9%	17.3%	18.1%	35.0%	38.6%
15	psyc	495	17.8%	12.1%	12.5%	32.9%	38.4%
16	micr	788	17.9%	8.4%	8.9%	31.2%	37.4%
17	biol	3483	17.9%	7.3%	7.4%	24.2%	33.3%
18	envi	804	14.7%	8.2%	8.3%	20.5%	33.2%
19	mole	1754	10.7%	3.8%	3.8%	13.5%	27.3%
20	immu	797	10.5%	3.9%	4.0%	14.4%	26.9%
21	phar	1375	10.4%	4.6%	4.7%	14.0%	26.6%
22	mult	681	5.0%	1.9%	1.9%	6.3%	18.1%

Rövidítések: **Agri:** Agricultural Sciences; **Biol:** Biology & Biochemistry; **Chem:** Chemistry; **Clin:** Clinical Medicine; **Comp:** Computer Science; **Econ:** Economics & Business; **Engi:** Engineering; **Envi:** Environment/Ecology; **Geos:** Geosciences; **Immu:** Immunology; **Mate:** Materials Science; **Math:** Mathematics; **Micr:** Microbiology; **Mole:** Molecular Biology & Genetics; **Mult:** Multidisciplinary; **Neur:** Neuroscience & Behavior; **Phar:** Pharmacology & Toxicology; **Phys:** Physics; **Plan:** Plant & Animal Science; **Psyc:** Psychiatry/Psychology; **Soci:** Social Sciences, general; **Spac:** Space Science

Egy érdekes új lehetőség a táblázatban X-szel jelölt mutatószám használata. Az X értékét a diszciplinaritási profil görbéjének és a diagram diagonálisának (az ábrákon a piros szaggatott egyenes) metszéspontjaként határozhatjuk meg. Szóban kifejezve ez azt jelenti, hogy egy szakterület X-értéke X, ha a szerzők X százalékának a diszciplinaritása legalább X. A szakterületek X szerinti rangsorában – kevéssé meglepő módon – a matematika bizonyult a legdiszciplinarisabbnak és a multidiszciplináris tudományok a legmultidiszciplinárisabbnak. A társadalom- és a földtudományok, a közgazdaságtudomány, az űrkutatás, a klinikai tudományok és a növény- és állattudományok szerzői is 50% fölötti diszciplinaritást mutatnak. Az élettudományok legtöbb területe igen multidiszciplinárisnak mutatkozik, ami feltehetően abból is ered, hogy az ESI szakterületi besorolása nem mindig "természetes határokat" követ, a szerzők nagy részének kutatási-publikációs tevékenysége ezeken a határokon átível.

ESI vs. MAB-osztályozás

A harmadik vizsgálat során az ESI-rendszer alkalmazását a szerzőalapú kategorizáció egy olyan esetével szembesítettük, amelynek az MTMT-ben való bevezetése elvárás. A *Magyar Akkreditációs Bizottság* (többszintű) besorolási rendszere, a MAB-osztályozás a tudomány- és szakterületeket a *képzési területek* rendszerével azonosítja, így a MAB-kategóriák végzettség alapján természetes módon hozzárendelhetők a kutatókhoz, ill. ilyen módon a szerzőkhöz. A két, konceptuálisan jelentősen különböző rendszer MTMT-beli kapcsolatát ez esetben is a közös nevező, vagyis a szerzők jellemzői alapján vizsgáltuk meg. A kísérlet alaplépései a következők voltak:

- Az MTMT-ben több adatbázis együttes felhasználása révén feltártuk azt a szerzői kört, amelynek MAB szerinti besorolása azonosítható volt. Ennek feltételrendszereként az *Akadémiai Adattár* (AAT, a korábbi köztestületi adatbázis utódja) és az MTMT kapcsolata szolgált: (1) az MTMT-ben köztestületi azonosítóval (MTAID) ellátott, ilyen módon az AAT-ben nyilvántartott szerzőkhöz hozzárendeltük az utóbbi adatbázisban feltüntetett szakterületi besorolásukat, a *fokozatszerzés szakterületét*. Mivel a fokozatszerzés szakterületének megadásához szolgáló lista az AAT-ben lényegében megegyezik a MAB-kategorizáció legalsó szintjével, ez a leképezés úgy tekinthető, mint az érintett MTMT-szerzők MAB szerinti szakterületi besorolása, illetve annak igen jó közelítése. Az vizsgálat első fázisának eredményeként tehát az MTMT-mintát leképeztük a MAB-kategóriák (egyszintű, de nagyfelbontású) rendszerére.
- A második fázisban a érintett szerzők ESI-ben kategorizált közleményeihez (folyóiratcikkeihez) hozzárendeltük a illető szerző MAB-besorolását. Az ennek eredményeképpen létrejött kettős hozzárendelés az MTMT-mintához tartozó közleménylistát kétféle szakterületi kóddal ruházta fel: minden közleményt két szakterületi változóval, (1) annak ESI-besorolásával, ill. (2) a (köztestületi) szerző MAB-területével írtunk le.

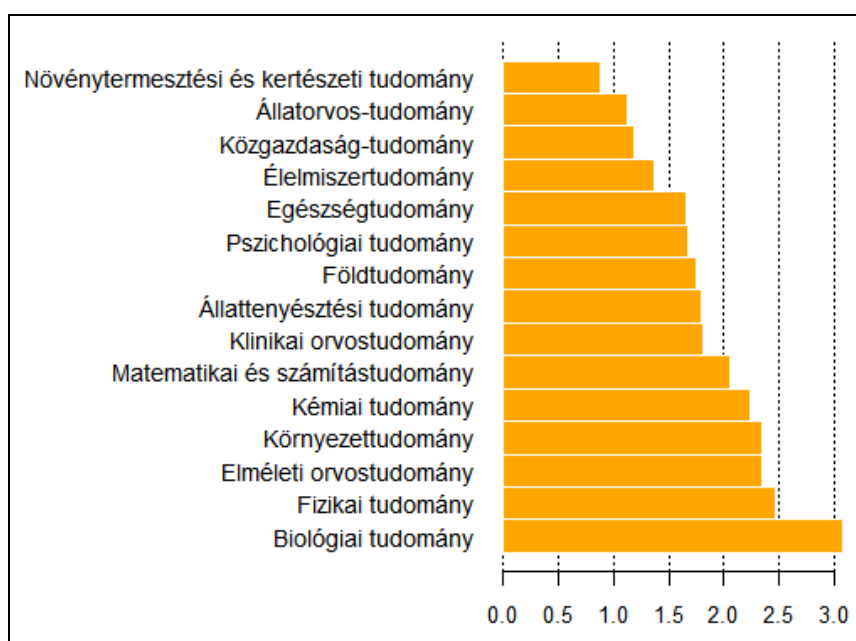
A kísérlet ezek után annak vizsgálatára irányult, hogy a mintabeli közlemények két változója, az ESI-, ill. a MAB-besorolás értékei milyen viszonyban állnak egymással. Az esetlegességek hatásának csökkentése céljából az MTMT-mintában nagyon alacsony közleményszámmal bíró MAB-területeket kihagytuk az analízisből (a legszűkebb, még számításba vett kategória így $n=70$ dokumentumot számlált) Az összehasonlításhoz két "tudományterképezési" és egy "bibliometriai" megközelítést alkalmaztunk:

- *A MAB-kategóriák heterogenitása (az ESI szerint)*. Elsőként a MAB-kategóriák ESI-összetételét vizsgáltuk, azaz a MAB-területek konzisztenciáját, heterogenitását, ill. általános leképezhetőségét az ESI-területek rendszerére. Ennek mérőszámaként – a MAB- és ESI-szakterületek együttjárása alapján – minden (mintabeli) MAB-kategóriát a Shannon-féle entrópiamutatóval jellemeztünk. E mutató értéke ebben az esetben azt fejezi ki, hogy a vonatkozó közlemények hány ESI-besorolás között oszlanak meg és mennyire egyenletes ez a megoszlás. Értéke 0, ha a közlemények egyetlen ESI-kategóriához tartoznak (egyöntetűség), és annál magasabb, minél nagyobb a kategóriák száma és az eloszlás egyenletessége, a heterogenitás. A vizsgált MAB-osztályok entrópiáját a 4. ábra szemlélteti: a sor elején a legkoherensebb (legkisebb entrópiájú), a végén pedig a legheterogénebb besorolások állnak.
- A kvantitatív kép részletezése céljából elkészítettük a MAB- és az ESI-területek kapcsolatrenderének térképét, a "kvalitatív képet", ugyancsak a két változó együttjárása alapján (5. ábra). A (kétdimenziós) hálózat csomópontjai egyfelől az ESI-területek (kör alakú csomópontok), másfelől a MAB-besorolások (négyzetek): az élek és az élsúlyok azt jelzik, hogy a MAB-besorolások, belső összetételüket tekintve, mekkora részarányban kapcsolódnak (járnak együtt) a vonatkozó ESI-kategóriákkal. Az áttekinthetőség kedvéért a 10%-ot (0.1) meghaladó kapcsolatokat tüntettük fel.
- A harmadik megközelítés szűkebb értelemben is a tudomány- (biblio-) metriai alkalmasság kérdését célozta, ami a MAB-rendszert illeti, pontosabban azt, hogy a (szerzőkön keresztül) egy-egy MAB-területtel azonosított közleményhalmaz érvényes referencialhalmaz lehet-e a bennfoglalt közlemények tudománymetriai megítélésében,

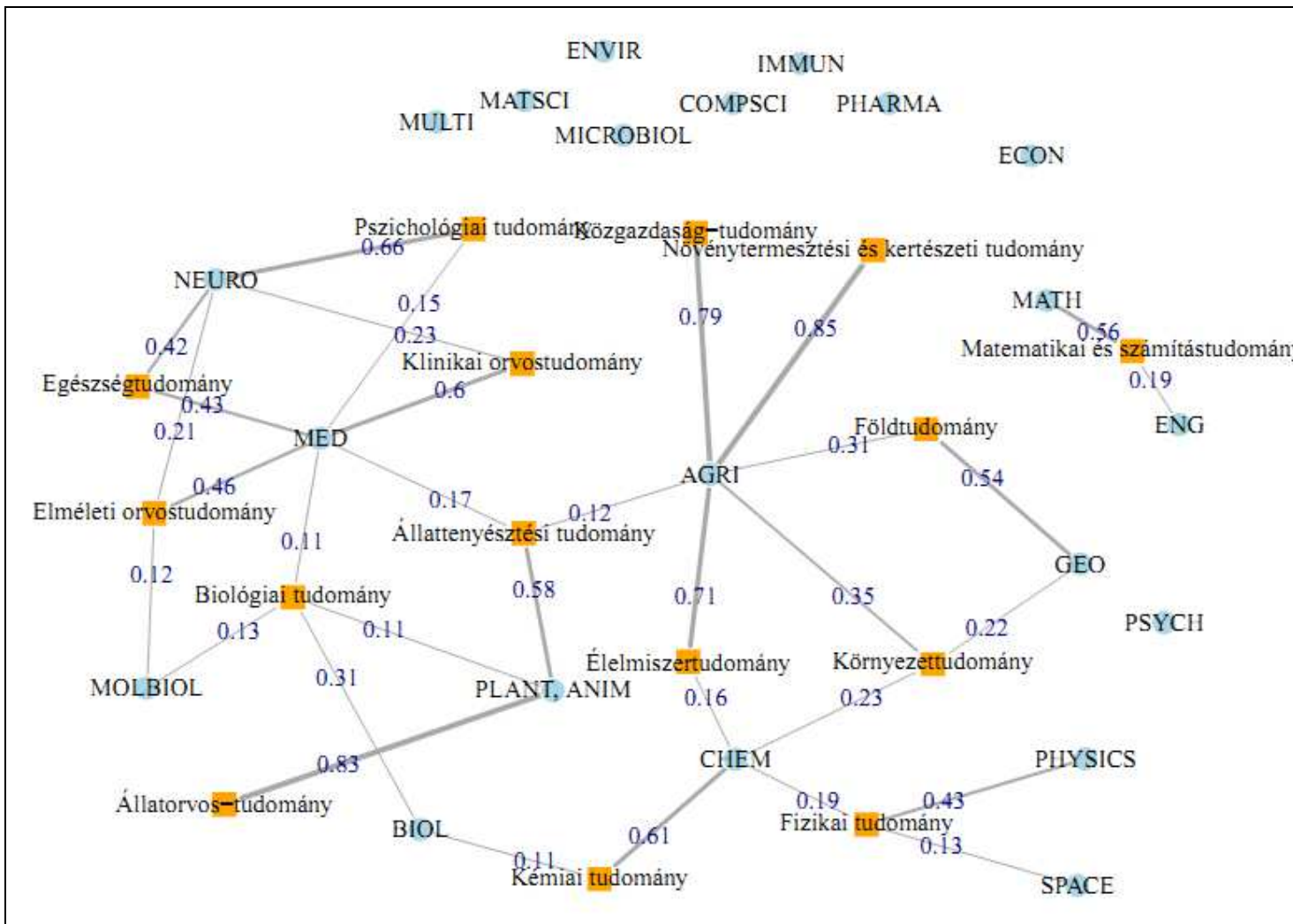
amennyiben az ESI-rendszerből (mint sztenderdből) indulunk ki. Ebben az elemzésben ezért a mintaközlemények ESI-besorolását az adott kategória idézettségi referenciaértékével helyettesítettük, a Thomson vonatkozó szolgáltatása segítségével. A referenciaérték az egy dokumentumra jutó v. átlagos idézetszám (citation per paper, cpp) tízéves átlaga mind a 22 szakterületre vonatkozóan. A mintabeli közlemények referenciaértéke alapján ezt követően előállítottuk a vizsgált MAB-területeken megnyilvánuló referenciaérték-eloszlást (6. ábra) annak megítéléséhez, hogy a terület "idézési viselkedése" mennyire konzisztens, és ilyen értelemben mennyire tekinthető jóldefiniált szakterületnek (referenciahalmaznak).

A három megközelítést célszerű a MAB-területek jellemzésén keresztül összefüggően tárgyalni (ld. a következő szakaszokat).

4. ábra A MAB-kategóriák belső heterogenitása (entrópia)

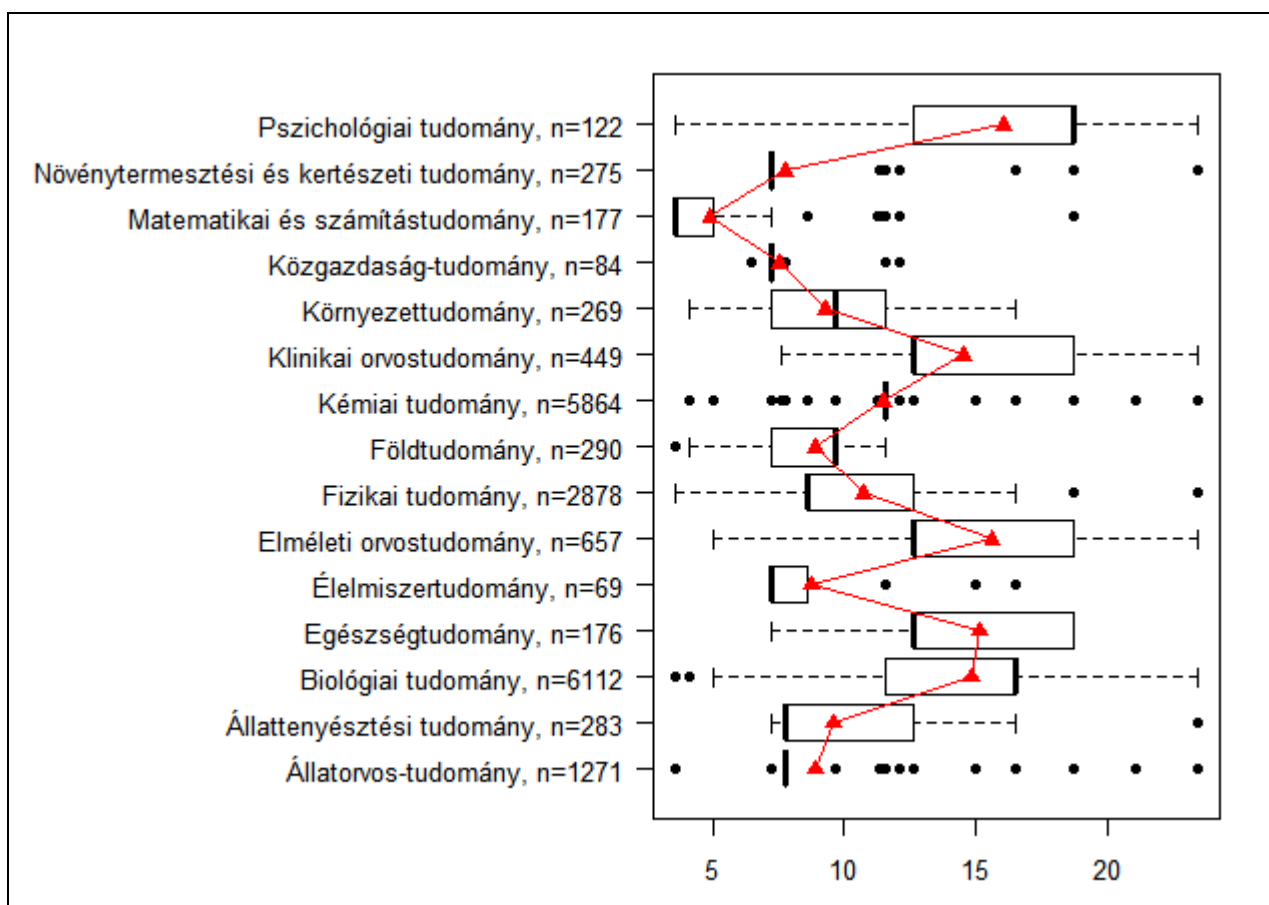


5. ábra A MAB- és az ESI-kategóriák kapcsolatrendszer



*

6. ábra Az ESI-referenciaértékek eloszlása a MAB-kategóriákon belül. A görbével összekapcsolt adatpontok az átlagokat jelzik.



2. tábla Az ESI-referenciaértékek eloszlásának statisztikái

	min	25%	median	75%	max
Állatorvos-tudomány, n=1271	7,76	7,76	7,76	7,76	7,76
Állattenyésztési tudomány, n=283	7,24	7,76	7,76	12,62	16,5
Biológiai tudomány, n=6112	5,02	11,56	16,5	16,5	23,49
Egészségtudomány, n=176	7,24	12,62	12,62	18,76	18,76
Élelmiszertudomány, n=69	7,24	7,24	7,24	8,56	8,56
Elméleti orvostudomány, n=657	5,02	12,62	12,62	18,76	23,49
Fizikai tudomány, n=2878	3,53	8,56	8,56	12,62	16,5
Földtudomány, n=290	4,07	7,24	9,68	9,68	11,56
Kémiai tudomány, n=5864	11,56	11,56	11,56	11,56	11,56
Klinikai orvostudomány, n=449	7,59	12,62	12,62	18,76	23,49
Környezettudomány, n=269	4,07	7,24	9,68	11,56	16,5
Közgazdaság-tudomány, n=84	7,24	7,24	7,24	7,24	7,24
Matematikai és számítástudomány, n=177	3,53	3,53	3,53	5,02	7,24
Növénytermesztési és kertészeti tudomány, n=275	7,24	7,24	7,24	7,24	7,24
Pszichológiai tudomány, n=122	3,53	12,62	18,76	18,76	23,49

A referenciaértékek kategóriánkénti eloszlásának tanúsága szerint több MAB-terület rendkívül erős konzisztenciát mutat ("növénytermesztési és kertészeti tudomány" "kémia tudomány", "állatorvos-tudomány"). Ezt a konzisztenciát az ESI-MAB-térkép jól magyarázza: az említett területek mindegyike jórészt egy ESI-kategóriához kapcsolódik (rendre 85%-ban az *Agricultural Sciences*, 61%-ban a *Chemistry*, ill. 83%-ban a *Plant and Animal Science* kategóriához). Hasonlóan kirívó a "közgazdaság-tudomány" konzisztenciája is, az előzőekhez képest ugyanakkor egy figyelemre méltó különbséggel, amely különbség a kvalitatív oldalon jelentkezik. A térkép alapján ez a MAB-terület is döntően egyetlen ESI-területtel mutat együttjárást, ez azonban az *Agricultural Sciences*, amely a várakozásokkal ellentétes. Sajátos módon az *Economics* ESI-besorolás üres (a térképen izolált elem).

A MAB-területek másik csoportjára jellemző, hogy kvantitatív szempontból igen heterogének. Kirívó példa a (relatív kis elemszámú) "Pszichológiai tudomány", és a (nagy méretű) "Biológiai tudomány", ahol a referenciaérték-intervallum – kiugró értékek nélkül is – csaknem az ábrázolt [1, 25] tartomány egészét fedi. Érdekes módon az előbbi kvalitatíve kevésbé heterogén, viszont elég polarizált: kb. 65%-ban a *Neuroscience and Behavior* (várható érték ≈ 19), 15%-ban pedig a *Clinical Medicine* (várható érték ≈ 12.5) kategóriájához kötődik. A fennmaradó csekély részt a "puhább", nem-természettudományi orientációjú részterületek alacsony, vagy az előbbieknél is idézés-intenzívebb részterületek igen magas értékei határozzák meg. A heterogenitás gyakorlati mértékét ebben az esetben a kis elemszám mérsékli, nevezetesen a 122 cikk csaknem 70%-ának egyöntetűsége. Más a helyzet a Biológiai tudomány esetében: itt hatezer közlemény szóródik ugyanilyen széles tartományban. A Biológiai tudomány kvalitatíve is a legváltozatosabb összetételű a MAB-területek között (ld. 5. ábra). A térkép szerint 31%-ban a – szintén gyűjtőkategóriának tekinthető – *Biology* köréhez rendelődik, de az ESI számos további élettudományi elemével van átfedésben (a min. 10% körüli részesedést a *Molecular Biology and Genetics*, *Plant and Animal Science*, *Clinical Medicine* szolgáltatja), amelyek között jelentős referenciaérték-beli különbségek adódnak (magyarázva ezzel a 6. ábrán látható eloszlást). Hasonlóan értelmezhető az *Elméleti orvostudomány* viselkedése, ami kvantitatíve nem sokban különbözik a *Klinikai orvostudománytól*. A (nagy méretű) Fizikai tudomány mint MAB-terület összetételében és a vonatkozó értékek eloszlásában ugyancsak relatíve ESI-inkonzisztens (bár több, mint 40%-a a *Physics* kategóriához kapcsolódik). Megjegyzendő, hogy a referenciaértékek eloszlásának tulajdonságai egy-egy esetben ellensúlyozzák a sokféleséget: a (kisméretű) Környezettudomány (MAB) ESI-profilja viszonylag széles, értéktartománya sem szűk, de az értékek eloszlása szabályos normáeloszlásnak látszik: ilyen módon az átlagérték a heterogenitás dacára is jól reprezentálja a területet. Az esetek többségében ez azonban nem, vagy kevésbé mondható el: a legtöbb "széthúzó" MAB-területen az átlagértékeket a ferde eloszlások elhúzzák.

A fentiekből a MAB és az ESI-kategorizáció (MTMT-beli) viszonyát illetően az alábbi, viszonylag általános megállapítások tehetők:

- A MAB-területek (és tetszőleges, a közlemény-alapútól eltérő osztályok) tudományometriai alkalmasságának vizsgálatában együttesen kell mérlegelni azok kvalitatív és kvantitatív összetételét, amit ebben az esetben az ESI-kategóriákra alapuló profil (ESI-összetétel) és a résztvevő kategóriák hatásmutatójának (cpp) eloszlása képvisel. A közgazdaságtudomány (mint MAB-kategória) példája jól mutatja, hogy bár a minta kvantitatíve lényegében egységes, a vonatkozó hatásmutató egy relatíve távoli (a térkép alapján sok egyéb MAB-kategóriában részt vevő) területről

származik (az agrártudomány ESI-területéről, amely az agrárközgazdaságtan uralkodó szerepére utal az MTMT-mintában).

- Azok a MAB-területek, amelyek ESI-összetétele változatos (kvalitatíve heterogén), kvantitatív szempontból sem mutatnak konzisztenciát, vagyis a hatásmutató-értékek széles tartományban mozognak. Másképp fogalmazva, az a várakozás, hogy a MAB-kategóriát felépítő ESI-területek “tudománymetriailag szomszédosak”, azaz a részt vevő területek hatásmutatói hasonlóak, ezért a heterogenitás ellenére a kategória – hatás mérési szempontból – konzisztens, többnyire nem teljesül. Kivételek akadnak, ilyenek látszik a sajátos belső eloszlású környezettudomány. (Fordítva a kapcsolat kevésbé tendenciózus: a szóródó értékek nem feltétlenül heterogenitást jeleznek. A kvantitatíve szóródó területek többször nem nagy számú ESI-területtel fednek át, hanem néhány, de – idézési szokásait tekintve – távoli területtel, l. a Pszichológiai tudomány példáját).
- A fentiekből adódóan elmondható, hogy a MAB-kategóriák – az ESI terminusaiban – különböző (aggregációs) szintű, és ezért jórészt átfedő területeket reprezentálnak. Viszonylag általános, hogy a kisebb, jobban definiált (“mezoszintű”) természettudományi MAB-besorolások jól leképeződnek az ESI-rendszerben, kvalitatív és/vagy kvantitatív szempontból egyaránt. Találunk azonban “makroszintű” diszciplínát is, (Kémiai tudomány) ahol a két rendszer megfelelése erős. A nagy tudományterületeket megnevező MAB-címkék azonban többnyire erősen inhomogének, mind tematikusan (kvalitatíve), mind a mérés szempontjából (kvantitatíve): a várakozásoknak megfelelően ilyen a “Fizikai tudomány” vagy a “Biológiai tudomány”. A társadalomtudományok közül a közgazdaságtudomány jelenik meg a mintában, amely azonban speciális viselkedést mutat, utalva arra, hogy ezeken a területeken a kutató és a közlemény ESI-szakterületi betagozódása meglehetősen távol kerülhetnek egymástól.

Összegzés és ajánlások

A fentiekben ismertetett vizsgálat sorozat viszonylag összetett képet rajzolt ki az ESI kategorizáció MTMT-beli alkalmazásának lehetőségeiről. Az elemzések szerint (1) a szerzők évenkénti szakterületi koherenciája igen magas, ugyanakkor (2) a szerzők teljes életművéből levezetett ESI-szakterületek – mint az MTMT-ben azonosított *szerzőalapú* referencialalmazok –, eltérő mértékben ugyan, de multidiszciplinárisak. Különösen az “átfogóbb” kategóriák (pl. a fizika és főleg az élettudományok) heterogének. Nagyon hasonló eredményekre vezet (3) a MAB-területek és az ESI-területek szerzőkön keresztüli összehasonlítása: egyes szűkebb, jellemzően természettudományi orientációjú MAB-területekre az ESI-rendszer jó közelítést ad mind az információ-visszakeresés, mind pedig a bibliometriai referencialalmaz funkcióját tekintve. A nagyléptékű MAB-kategóriákat, megintcsak főként az élettudományok terén viszont a szerzőalapú megfeleltetés kvalitatíve heterogén és bibliometriailag is széttartó szerzőosztályokként jeleníti meg, amennyiben a szerzők ESI-profiljából indulunk ki. (Ez utóbbinak nyilván része az is, hogy a MAB-osztályok meghatározása esetén a *fokozatszerzés* szakterületéből volt lehetőségünk kiindulni, míg az életmű ESI-besorolásai a mindenkori aktivitást regisztrálják.) Ez az áttekintés sajátos módon azt implikálja, hogy az MTMT-ben

- a (közlemények kategorizációja révén létrehozott) **időszakos szerzőalapú osztályozás**, vagyis a szerzők egy-egy ESI-kategóriához rendelése a

teljesítményértékelési funkciónak megfelelhet. Ennek alapja, hogy az egy-egy szakterületet képviselő referencialalmaz definíció szerint egyidőben publikált, praktikusán azonos évben megjelent közlemények csoportja: ha pedig a szerzők éves szinten egy-egy kategóriát reprezentálnak, az ugyanazon kategóriához tartozó szerzők aggregátuma (a referencialalmaz) szintén koherens lesz.

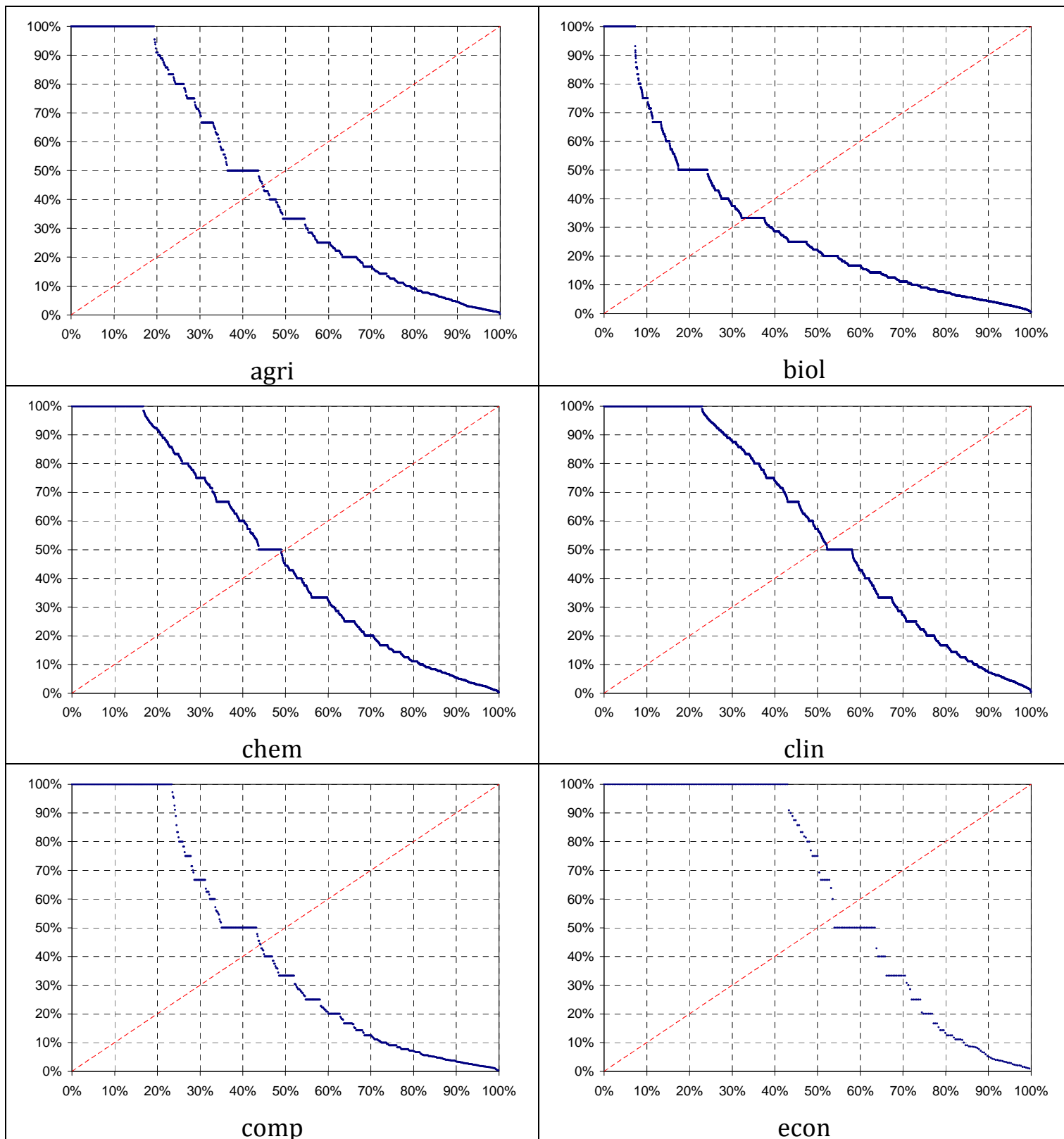
- A **teljes életművekre** alapozott ESI-kategorizáció eredményei, vagyis szerzőcsoportok mint ESI-kategóriák **belső heterogenitása** miatt az így alkalmazott ESI-rendszer információ-visszakeresési célokra kevésbé alkalmas.
- A **közlemények ESI-besorolásától független** szerzőalapú osztályozási rendszer, a MAB-felosztás mind információ-visszakeresési, mind pedig értékelési célra korlátozottan (pontosabban: **szelektíve**, egyes kategóriákban inkább, másokban kevésbé) alkalmas.

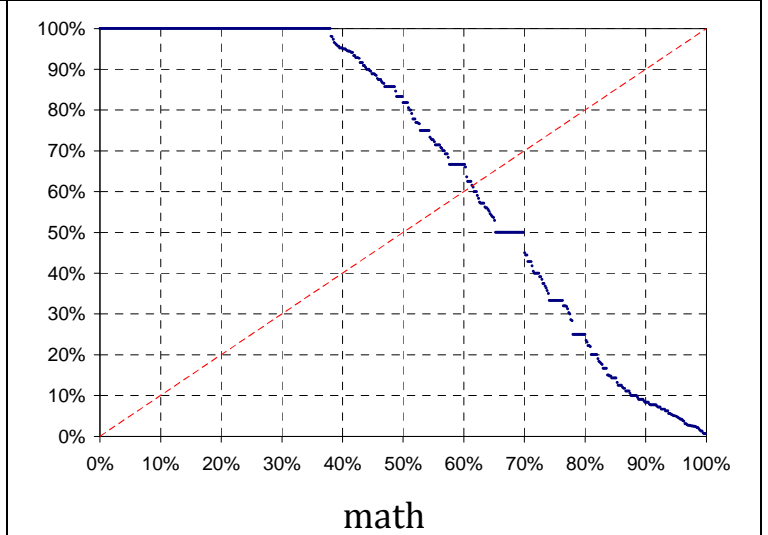
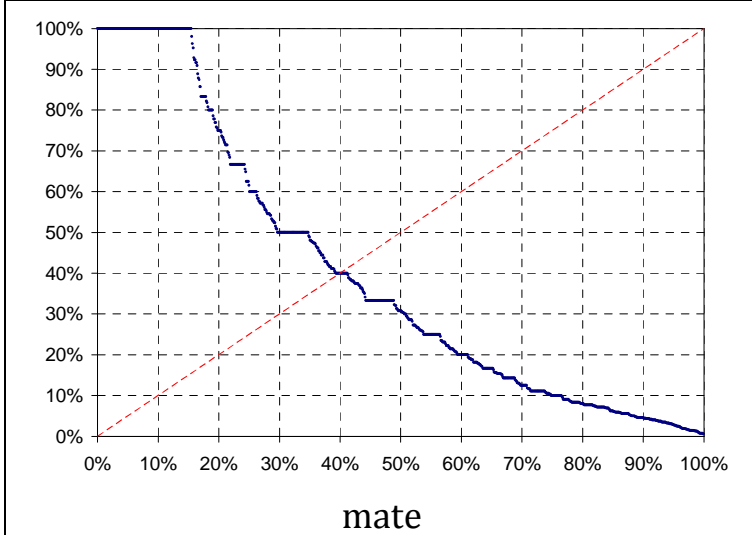
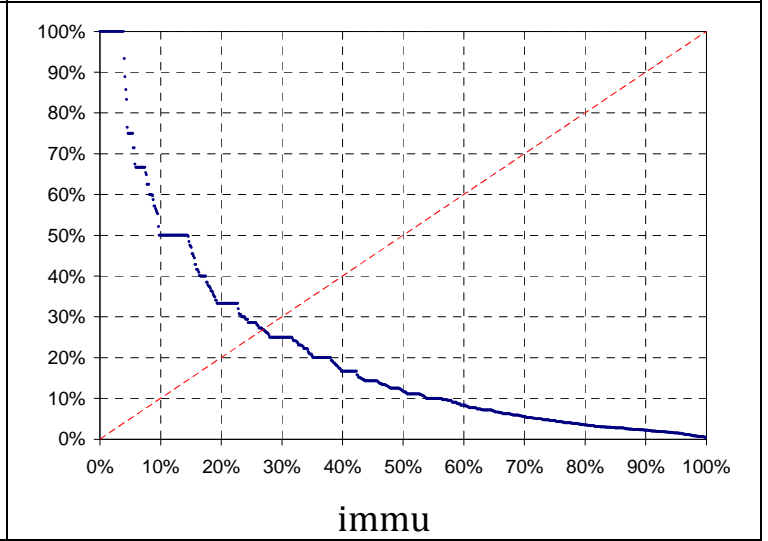
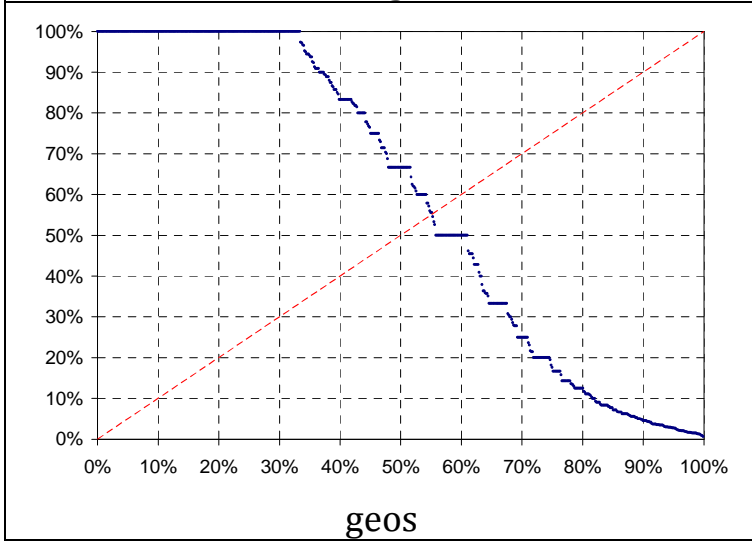
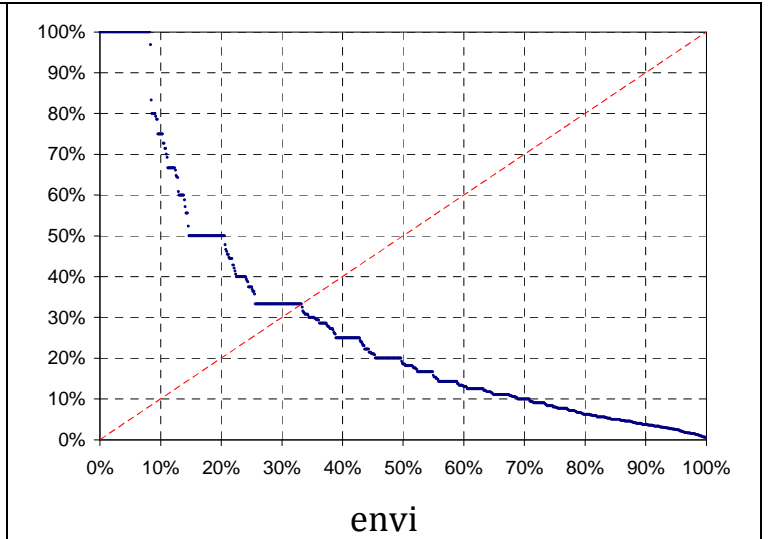
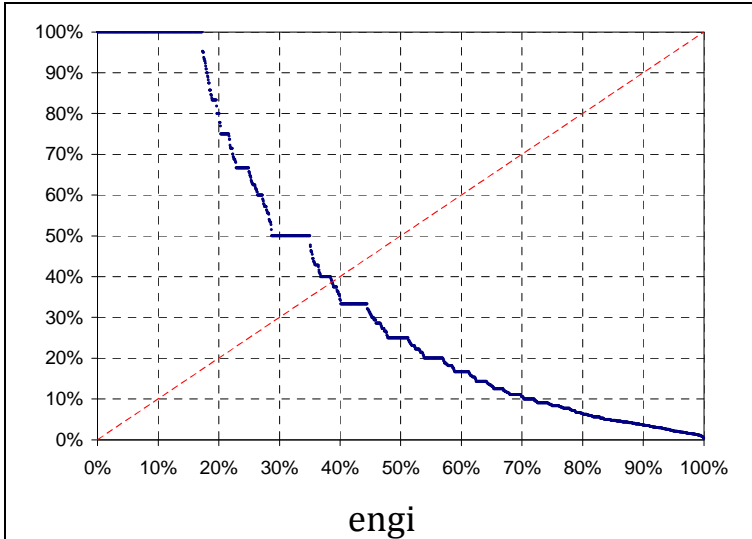
A fenti következtetésekből az ESI-rendszer (szerzőalapú) alkalmazásának korlátai mellett ugyanakkor egyértelműen megragadható egy konstruktív mozzanat is, amely az ESI adekvát és hasznos bevezetésének módját illeti. Mivel a támogatható megoldás is (vagyis az időszakos, értékelési célra történő szerzői besorolások rendszere) lényegében a közlemények kategorizációján alapszik, és mivel az ESI-rendszer a korábban tárgyalt természetéből adódóan többnyire az életmű egy részének (és nem egészének) kategorizációjára alkalmas,

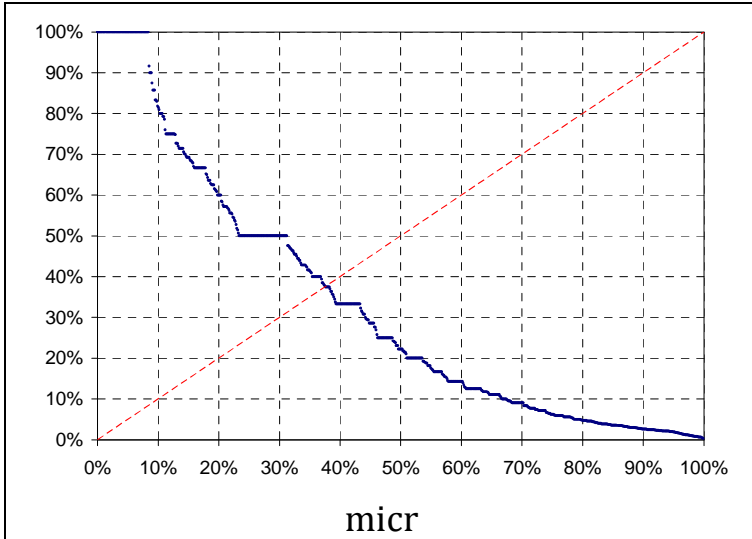
- a javasolt megoldás egy **részleges, közleményalapú ESI-kategorizáció**, amely alapján (1) az ESI-folyóiratbázishoz tartozó MTMT-tartalom minden közleménye jellemezhető ill. a vonatkozó referencialalmazon belül relatív mutatókkal mérhető, és (2) erre építve a szerzők (intézmények stb.) *nemzetközi teljesítménye* közleményeik relatív mutatóinak aggregációjával összehasonlítható és értékelhető.

Melléklet

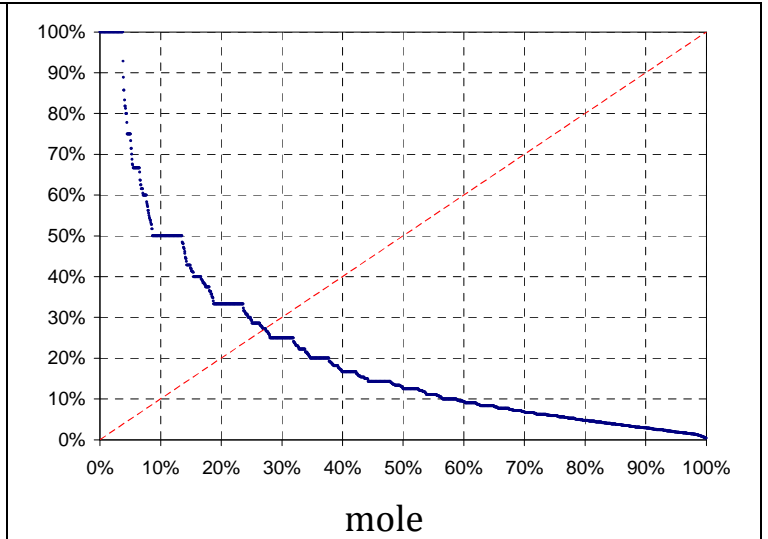
7. ábra Az egyes ESI-szakterületek X-értéke az MTMT-ben



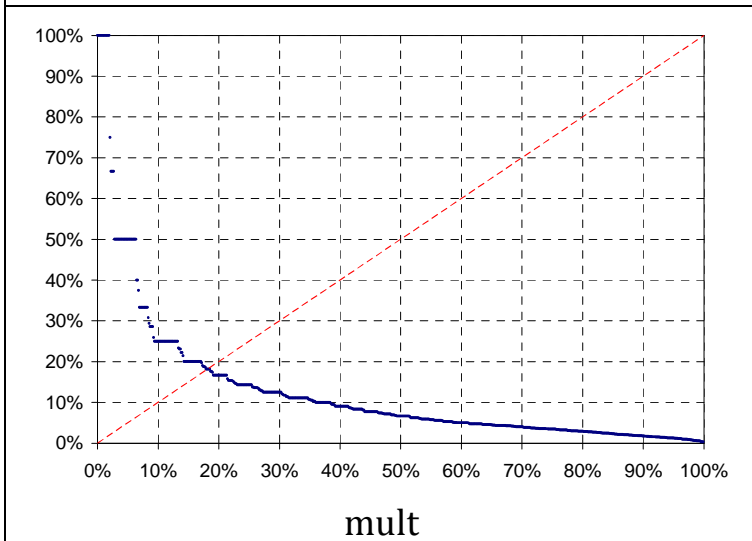




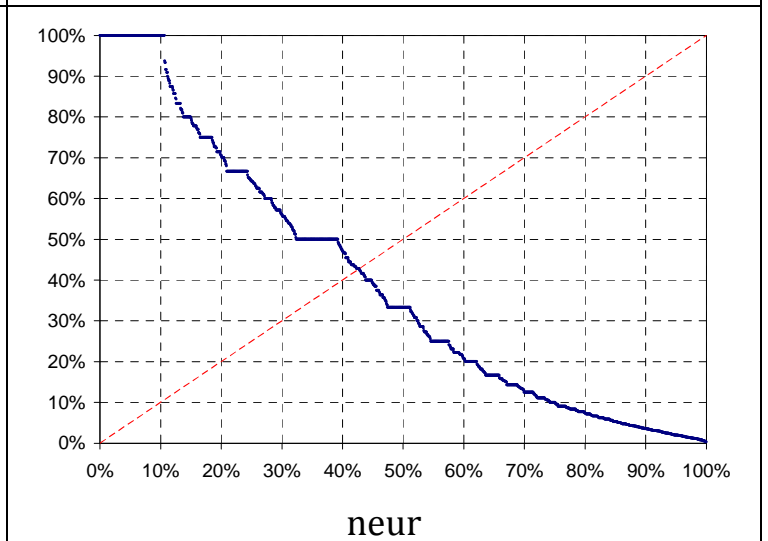
micr



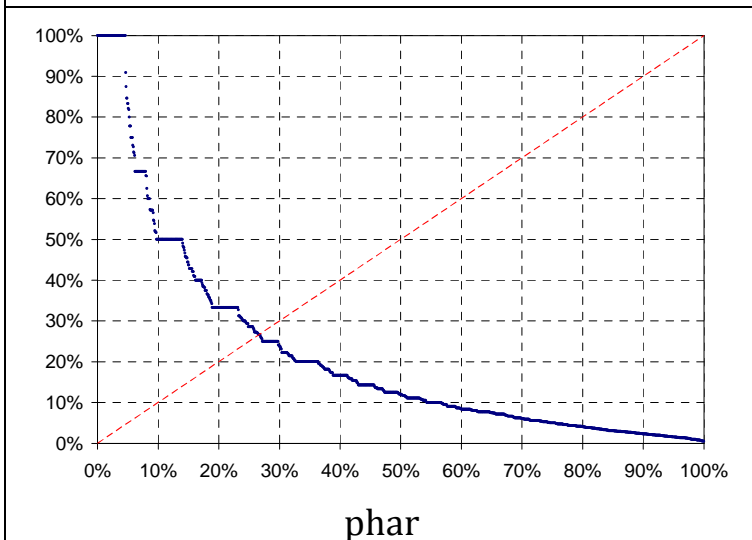
mole



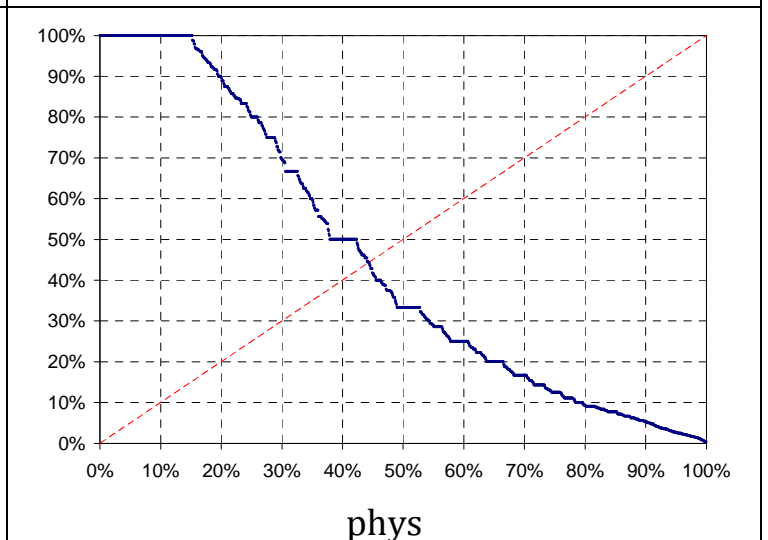
mult



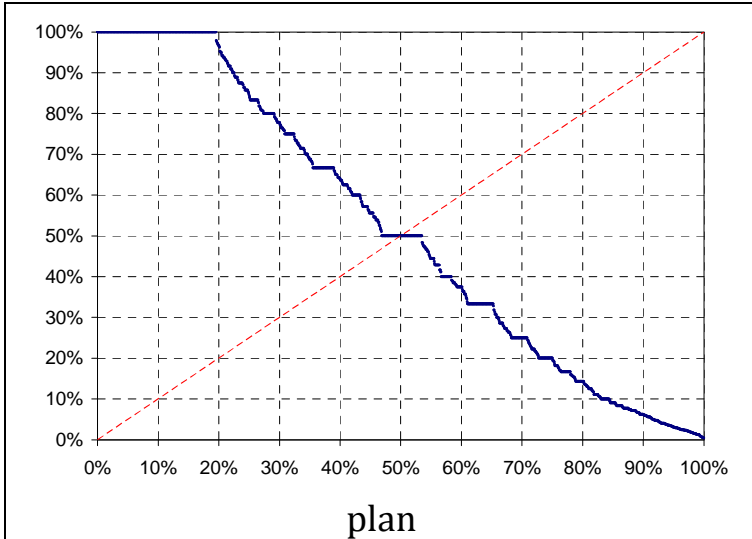
neur



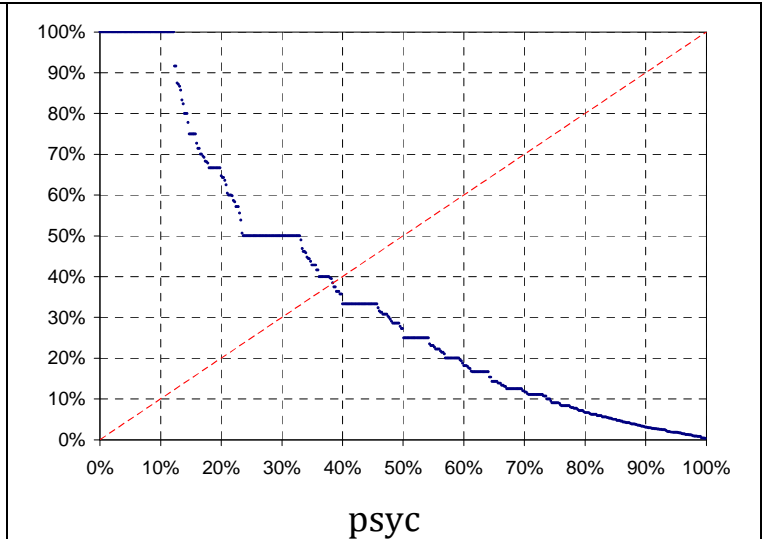
phar



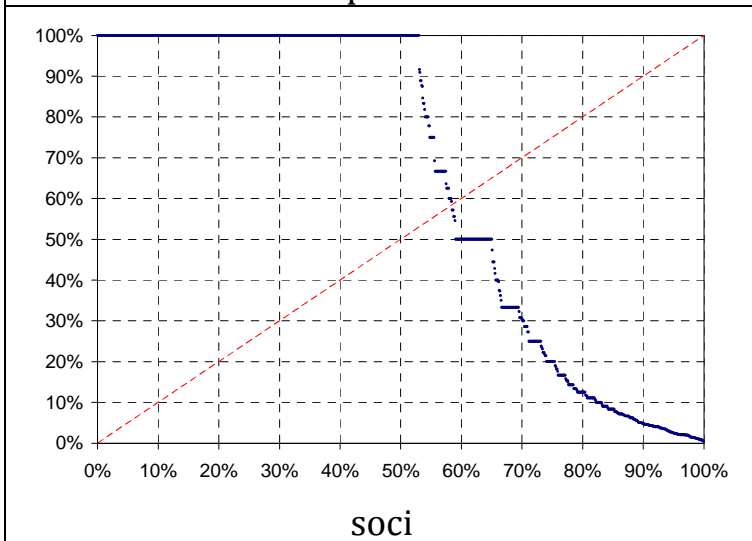
phys



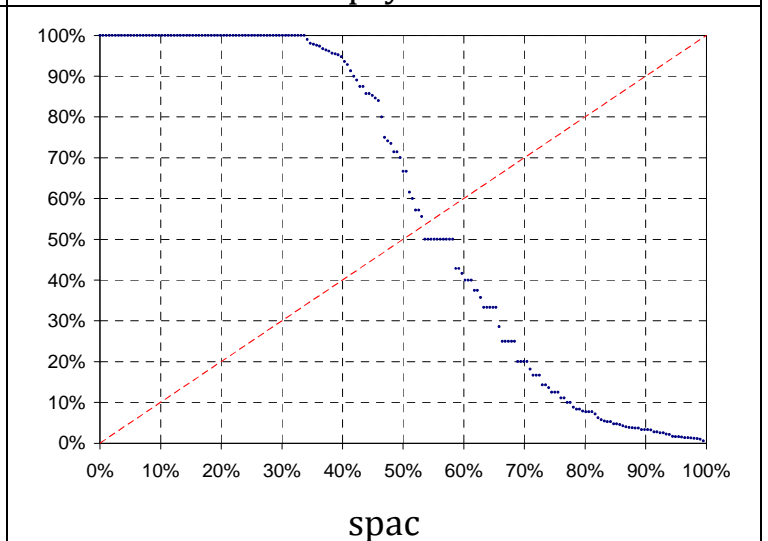
plan



psyc



soci



spac