

Integrált osztályozási módszer az MTA adattárainak tudományometriai alkalmazásához

Az elmúlt években a hazai tudományos adattárak, főként az MTA adatbázisai egyre erősödő szerephez jutottak a magyar tudomány teljesítményének értékelésében, értékelő tudományometriai kimutatások készítésében. A rangsorképzésnek ugyanakkor gyakori módszertani kérdése, hogy melyek azok a szakmai körök (szakterületi csoportok), amelyekben belül érvényes összehasonlítás tehető: vajon ezek megegyeznek-e a tudományos osztályokkal vagy éppen keresztülszelik-e azokat, avagy a finomabb felbontás-e a megoldás. Az alábbi tanulmány egy empirikus, kizárólag bibliográfiai adatokra támaszkodó, a bibliometriát és a szövegfeldolgozást ötvöző osztályozási eljárás bemutatásával igyekszik hozzájárulni a kérdés megválaszolásához.

Bevezetés

Az összehasonlító tudományometriai értékelések, rangsorok felállításának visszatérő kritikája, egyben fontos szempontja az összehasonlíthatóság kritériumaira vonatkozik. Gyakran emlegetett körülmény, hogy a különböző tudományterületek és álkörzetek eltérő publikációs és hivatkozási szokásrendszerrel rendelkeznek, ezért például tudományometriai rangsorok felállítása csak a gondosan körülhatárolt specialitásokon belül lehet informatív. Ez az érv egyike volt azoknak a kritikai reflexióknak is, amelyek a közelmúltban publikáltak és az *MTA Köztudományi Publikációs Adattára (KPA)* épülő rangsorok [1] vonatkozásában felmerültek. Az észrevétel konkrétan arra az eljárásra vonatkozott, hogy a szerzői rangsorok keretét az akadémiai osztálystruktúra szolgáltatta, vagyis a rangsorok felállítása az egyes osztályok szintjén történt. A kritika szerint az akadémiai osztályok nem tekinthetők szakterületi szempontból homogén, és így az összehasonlításnak alapot szolgáltató egységeknek:

„Az önálló szakterület ugyanis *nem akadémiai osztályokat jelent*, szűkebb szakmacsoportokat is csak nagyon óvatosan. A *Műszaki Osztályon* biztos igazságtalan ilyen számok alapján összehasonlítani egy anyagtechnológust egy áramlástanossal, vagy egy elektronikai technológust egy szabályozástechnológussal szakemberrel, vagy egy biológiai/villamos határterületen dolgozó szakembert egy villamosgépes tudóssal. Ezen valamit segíthetne a szakterületek megadása a táblázatokban, de ez sajnos tipikusan formális és semmitmondó

(„Műszaki tudomány”) ([2]; kiemelés az eredetiben).

A felvetésben megjelenő probléma, az MTA osztályok belső szerkezetének kérdése és tudományometriai elemezhetősége inspirálta az alábbiakban bemutatott tanulmányt. A munka fő célkitűzése egy olyan módszer kiválasztása–kidolgozása, valamint demonstrációja, amely kellően érzékeny módon képes feltárni az egyes osztályok szakterületi struktúráját, vagyis informatív és (belsőleg) összehasonlítható szakmai csoportokra tagolni azokat. A célkitűzés része, hogy ez a tagolás az ismert célok miatt a legkevésbé legyen *a priori*, tentatív vagy önkényes, ehelyett a lehető legnagyobb mértékben empirikus vagy „adatvezérelt”. Ebből adódik, hogy a javasolt módszer kizárólagos információforrása célszerűen maga a KPA, pontosabban az osztályokhoz tartozó publikációs listák. A megközelítés ilyen módon igyekszik megoldani, hogy a tagolást az adatokban – a publikációk korpuszában – meg nyilvánuló szakterületi struktúra alapján végezzük el. A módszer tehát nem támaszkodik más adatbázisra, így például az osztályok tagságát nyilván tartó Köztudományi Tagok adatbázisára, részben az említett adatbányászati, „bottom-up” felfogás miatt, részben pedig azért, mert a rendszer felhasználói által készített „szabadszavas” besorolások – a részletezni kívánt bizottsági hovatartozást leszámítva – sokszor egyéniek (a szakterületi leírás eltérő szintjét jelölik meg – l. a fent idézett észrevételt), nem mérhető össze.

Röviden tehát, a javaslat lényege a KPA-ban szereplő és az egyes tudományos osztályokhoz tarto-

zó publikációk szerzőinek bibliometriai klaszterezése. A szerzők szakmai kapcsolatrendszerének feltárásához az elméleti keretet a bibliometriai tudománytérképezés (science and technology mapping) és a hálózatelemzés (network analysis) szolgáltatja.

Anyag és módszerek

A megközelítésmód kidolgozása és szemléltetése céljából egy élettudományi osztályt, a *Biológiai Tudományok Osztályát* választottuk ki. A vizsgált anyagát ennek megfelelően a VIII. osztályhoz tartozó publikációk egy részhalmaza szolgáltatja. A mintát a KPA nyilvános keresője segítségével gyűjtöttük ki, és a (KPA-ból ilyen módon) kinyerhető folyóiratcikk típusú közleményekre szűkítettük. Ez a gyűjtemény 1011 szerző összesen 416 közleményét tartalmazta. Fontos hangsúlyozni, hogy a felhasznált anyag – mint mintaadatbázis – a VIII. osztályhoz tartozó személyeknek és publikációknak csupán egy töredékét képviseli és nem tekinthető reprezentatívnak; jelen célunkhoz mérten ugyanakkor elegendő alapanyag, amennyiben vállalásunk a módszer jellemzése és demonstrációja, nem pedig a VIII. osztály egészének empirikus vizsgálata.*

A szerzők klaszterezéséhez a publikációk két olyan jellemzőjét választottuk ki, amely viszonylag közvetlen módon tükrözi szerzője szakterületét. Abból a feltevésből kiindulva, hogy egy közlemény címléírásában a vonatkozó szakterület indikátorai között elsősorban a lelőhely, illetve a cím, az abban hivatkozott tematika és fogalmak sorolhatók fel, az egyik információforrás a közlemény lelőhelye volt, amely a vizsgált korpuszban folyóiratcikkekről lévén szó, a publikációt közlő folyóiratnak felel meg.

A strukturáláshoz használt másik információtípust a cím szövegének feldolgozásából állítottuk elő. A publikáció szakterületét a címben szereplő fogalmak, ebben az értelemben kulcsszavak reprezentálták.

Az alkalmazott klaszterezési módszer ennek a két aspektusnak a kombinációjára épült. A módszer az alábbi lépésekben írható le.

- Első lépésben a szerzők–folyóiratok, illetve a szerzők–kulcsszavak kereszttáblájából meghatároztuk a szerzők lelőhely, illetve fogalomhasználat szerinti hasonlóságának mértékét (technikai-

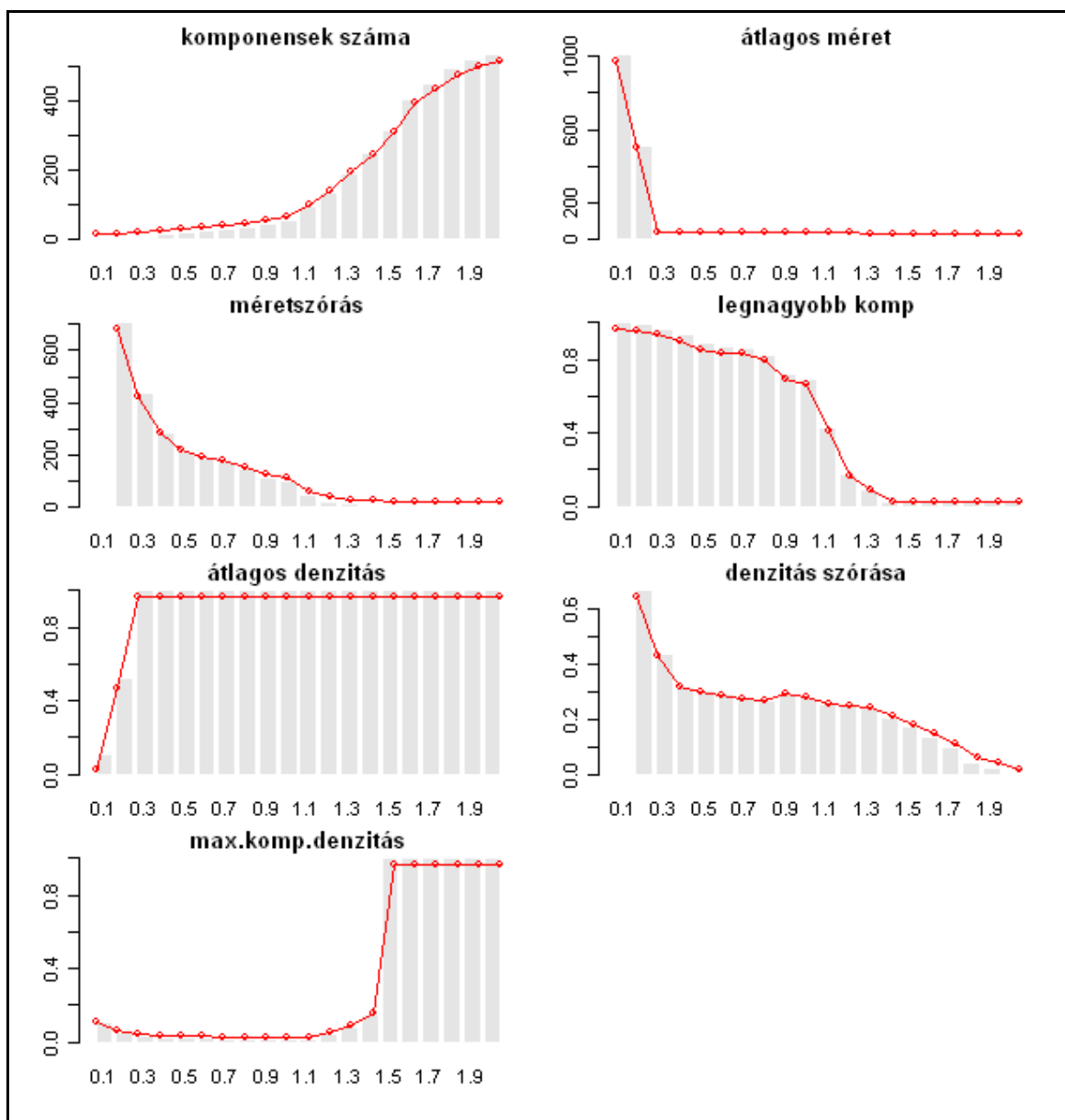
lag ez a szerző-vektorok távolságának meghatározásából állt, amelyet az ún. cosinus-mértékkel definiáltunk).

- A két szerzőhasonlósági mátrixból, még mindig külön kezelve azokat, levezettük a szerzők folyóirat-, illetve kulcsszóalapú hasonlósági hálózatát. Az így keletkezett súlyozott gráfok csomópontjai a vizsgált szerzők, az egyes élek pedig, súlyuk révén, az általuk összekapcsolt szerzők szakterületi hasonlóságának mértékét ábrázolják az adott dimenzió mentén.
- A két hasonlósági mérték, azaz a két gráf kombinációja révén előállítottunk egy „eredő” hálózatot, amely így mindkét tényezőjét kifejezi a szerzők közötti relációnak, ezért robusztusabban viselkedik, mint az eredeti hálók, vagyis kevésbé érzékeny az esetleges egybeesésekre (amelyek nem valódi szakmai kapcsolatokat tükröznek, és torzítják a leírást), mint a kizárólag lelőhely- vagy kulcsszóalapú viszonyrendszer. Az új hasonlósági mértéket az eredeti kettő lineáris kombinációjával határoztuk meg (ahol $S_{komb}(i,j)$ a kombinált, $S_{lelőhely}(i,j)$ a lelőhelyalapú, $S_{kulcsszó}(i,j)$ pedig a kulcsszóalapú hasonlóság mértéke bármely i és j szerző között, α és β pedig az egyes tényezők súlyozására választott paraméterértékek):

$$S_{komb}(i,j) =_{\text{def}} \alpha S_{lelőhely}(i,j) + \beta S_{kulcsszó}(i,j).$$

- A tulajdonképpeni klaszterezés az S_{komb} kombinált hálózatban kirajzolódó koherens szerzőcsoportok detekciójára épült. A csoportdetekció számos lehetséges módszere közül egy gráf redukciós eljárást végeztünk el, amely a „gyenge” kapcsolatok (hasonlóság) kiszűrésével az eredeti gráfból eltávolította azokat az éleket, amelyek súlya egy adott határérték alatt maradt. Az eljárással a hálózatot olyan csoportokra bontjuk fel, amelyek tagjai szorosabban tartoznak össze a vizsgált szakmai dimenziók mentén. Az összetartozás mértékének meghatározása céljából megvizsgáltuk, hogy a határérték megválasztásának függvényében hogyan alakulnak a csoportokra bontott hálózat releváns tulajdonságai (a csoportok – komponensek – száma, méreteloszlása, sűrűsége: 1. ábra).

* A KPA-ból kinyert publikációk nem tartalmazzák a VIII. osztály vonatkozásában releváns két további gyűjtemény, az SZBK- és a KOKI-adatbázis anyagát. A KPA-ból történő legyűjtés a 2009. év januárjában zajlott. Jelen tanulmány revíziójának időpontjában (2009. november) az osztályban nyilvántartott tagok száma 1079. (Forrás: <http://www.mtakpa.hu/akta/stat/oszfok.php>)



1. ábra A kombinált szerzőhasonlósági hálózat tulajdonságainak alakulása a felbontási paraméter (élsúly) függvényében. Az egyes grafikonok a szerzői körök halmazának legfontosabb jellemzőit ábrázolják különböző hasonlósági határértékek mellett

Ennek alapján állítottuk be ezt a paramétert. A döntési szempontok közül első helyen a gráf felbontásából előálló komponensek sűrűsége (denzitása)** szerepelt, tekintve, hogy annak mértéke a csoport koherenciáját jellemzi: a nagyobb denzitású csoportok egységesebbek a vizsgált szempontok szerint. A fentieket összegezve: az irányelv tehát olyan határérték kiválasztása volt, amely mellett viszonylag erős és relatíve koherens kapcsolatrendszer definiálja az egyes szakmai köröket.

Az így kapott szerzőcsoportokat az utolsó lépésben „kiértékeljük”, vagyis az általuk képviselt szakterületet igyekeztünk meghatározni. A szakterület azonosításához az egyes klaszterekhez tartozó folyóiratprofilot használtuk fel: a szakfolyóiratoknak az adott szerzőcsoport mintabeli publikációi által

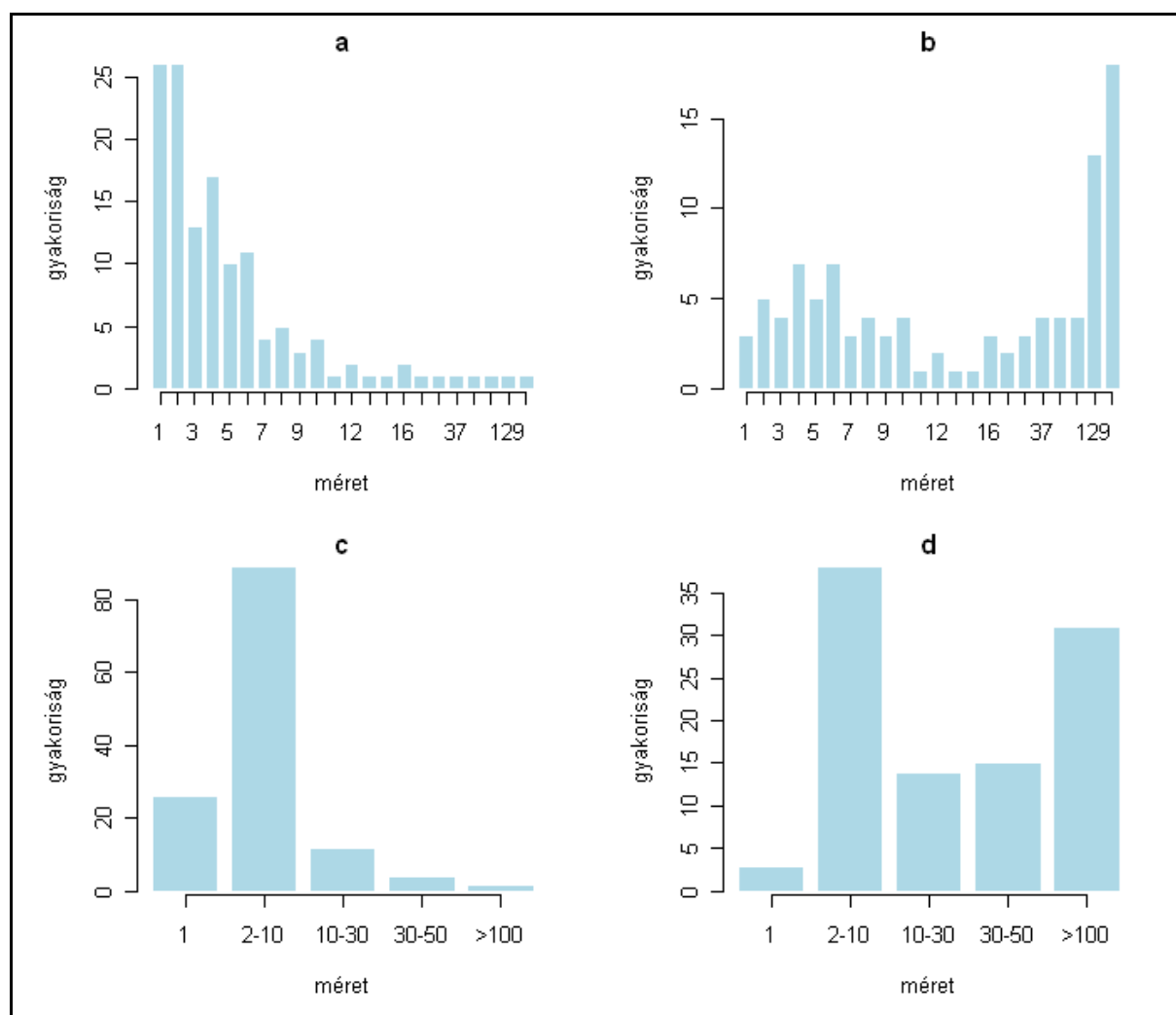
** Egy gráf sűrűségén a csomópontjai közötti tényleges és potenciális kapcsolatok számának arányát értjük.

meghatározott gyakorisági eloszlását. A szakterületet az első tizenöt leggyakrabban előforduló folyóirat részarányai alapján azonosítottuk.

Eredmények

A hasonlósági hálózat viselkedését vizsgálva a legszembevetőbb, hogy a felbontási határérték viszonylag széles tartományában a gráf egyetlen nagy csoportot (komponenst) tartalmaz, amely a vizsgált szerzők többségét tömöríti (1. ábra). Ez a megfigyelés arra utal, hogy a kérdéses osztály szakterületi szempontból meglehetősen homogén.

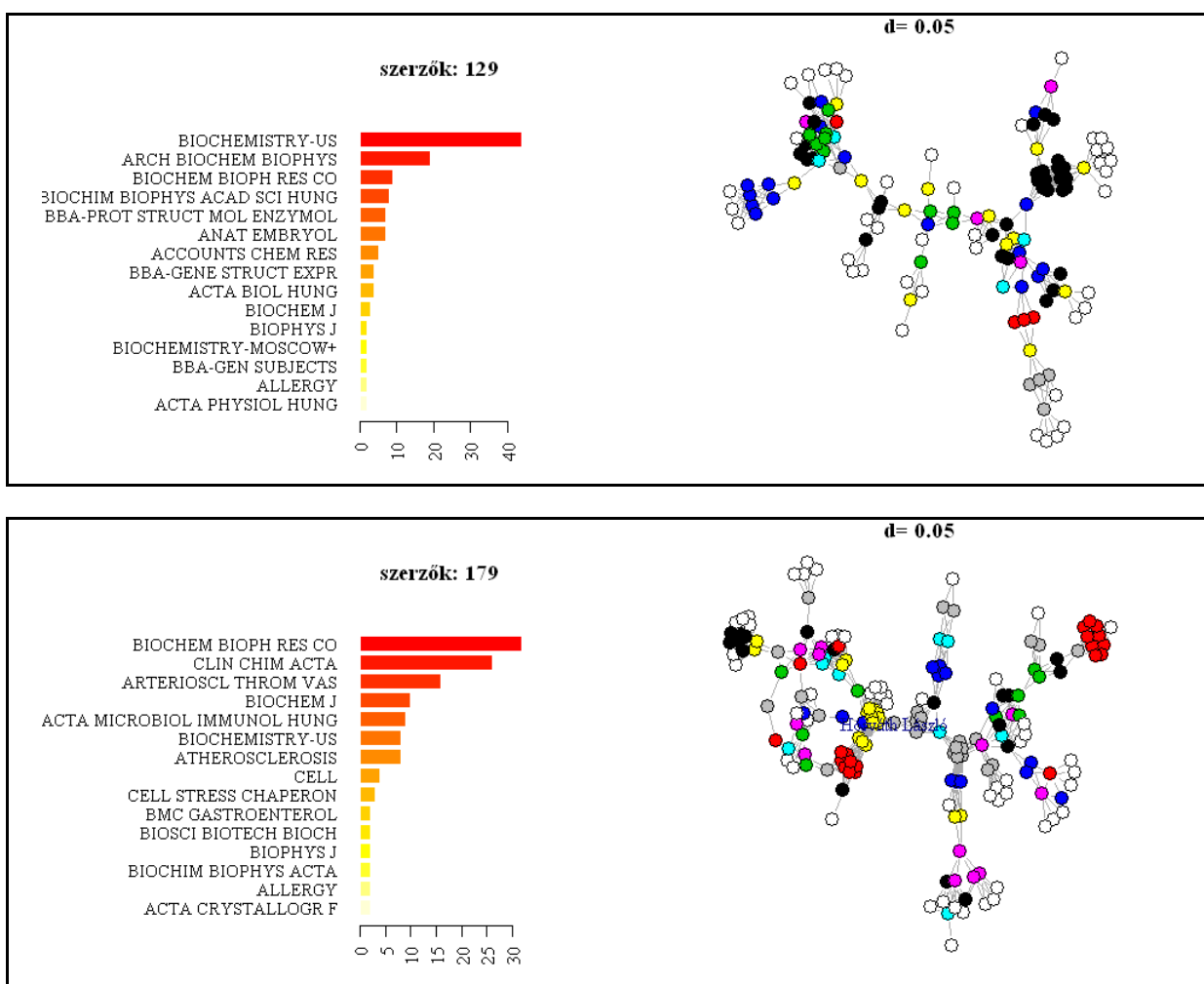
A kérdés, hogy miként jellemezhető ez a homogénnek látszó csoport, és milyen, mennyire koherens csoportok alkotják a „kisebbségeket”. A hasonlóság mértékének alsó határát – a biztonságos, de informatív osztályozás céljából – a jellemzők eloszlása alapján ott vontuk meg, ahol a csoportszám meredek emelkedésnek, a csoportlétszám pedig meredek csökkenésnek indult, miközben a csoportok sűrűsége (koherenciája) ennek a dinamikus tartománynak a többi pontjához képest maximális értéket mutatott ($k=1.2$; a hasonlóságot meghatározó két tényezőt – lelőhely és kulcsszavak – az első kísérletben azonos súllyal vettük figyelembe).



2. ábra A szerzői hálózat választott felbontásának jellemzői
a: a komponensek méreteloszlása,
b: a szerzők eloszlása a különböző méretű komponensek között,
c: a komponensek aggregált méreteloszlása,
d: a szerzők eloszlása az aggregált méretkategóriákban

Az ilyen módon felbontott gráf csoportszerkezetét a 2. ábra szemlélteti. Az a–c. diagram a csoportok méreteloszlását rögzíti: tanúsága szerint a szerzők kevés nagy csoportban (egy 179 és egy 129 tagú), és több kisebb csoportban oszlanak meg. A kisebbek legtöbbször a 2-10-elemű tartományban mozog. Még informatívabb a b–d. grafikon, amely azt mutatja meg, hogy a vizsgált szerzők mekkora hányada alkotja a különböző méretű csoportokat. Mint látható, a legtöbbször a százas létszám feletti, illetve az említett alacsony mérettartományban (2-10) találhatók, vagyis a fent említett klaszterek fedik le a minta legnagyobb részét (70%-át). Fontos megjegyezni, hogy az „egyelemű csoportok”, vagyis a lényegében „sehová sem tartozó” szerzők 3%-ot tesznek ki.

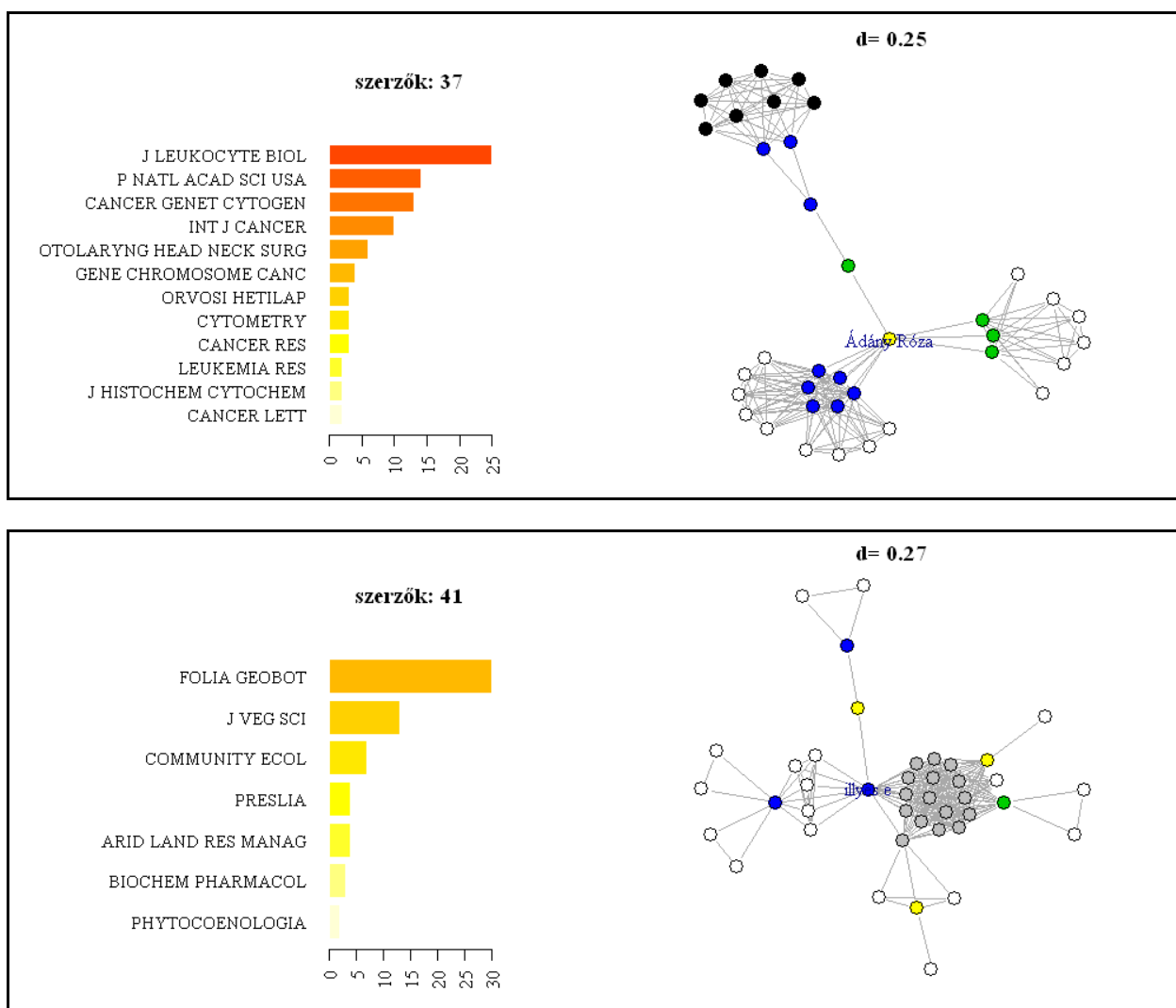
A szakterületi felbontás leírásához célszerű a kapott csoportokat (részhálókat) azok szerkezete – topológiája – és profilja alapján jellemezni. A jellemzéshez azokat a „kellően nagy” csoportokat választottuk ki, amelyek a méreteloszlás felső harmadát képviselik ($n > 6$). Ebben a körben kiugrik a profilja alapján biokémiai, illetve biokémia-fizikai/klinikai kémia fókusszal rendelkező csoport, amely a szerzőknek rendre a 13, illetve 18%-át sorakoztatja fel (3–4. ábra). Mindkét csoport szerkezetére jellemző, hogy globális koherenciája viszonylag alacsony, számos szorosabban összetartozó kisebb szerzőkörből és az ezeket összekapcsoló szerzőkből épül fel, ami úgy értelmezhető, hogy a kellő tematikai egységesség mellett szakmailag sokszínű, interdiszciplináris közösségről van szó.



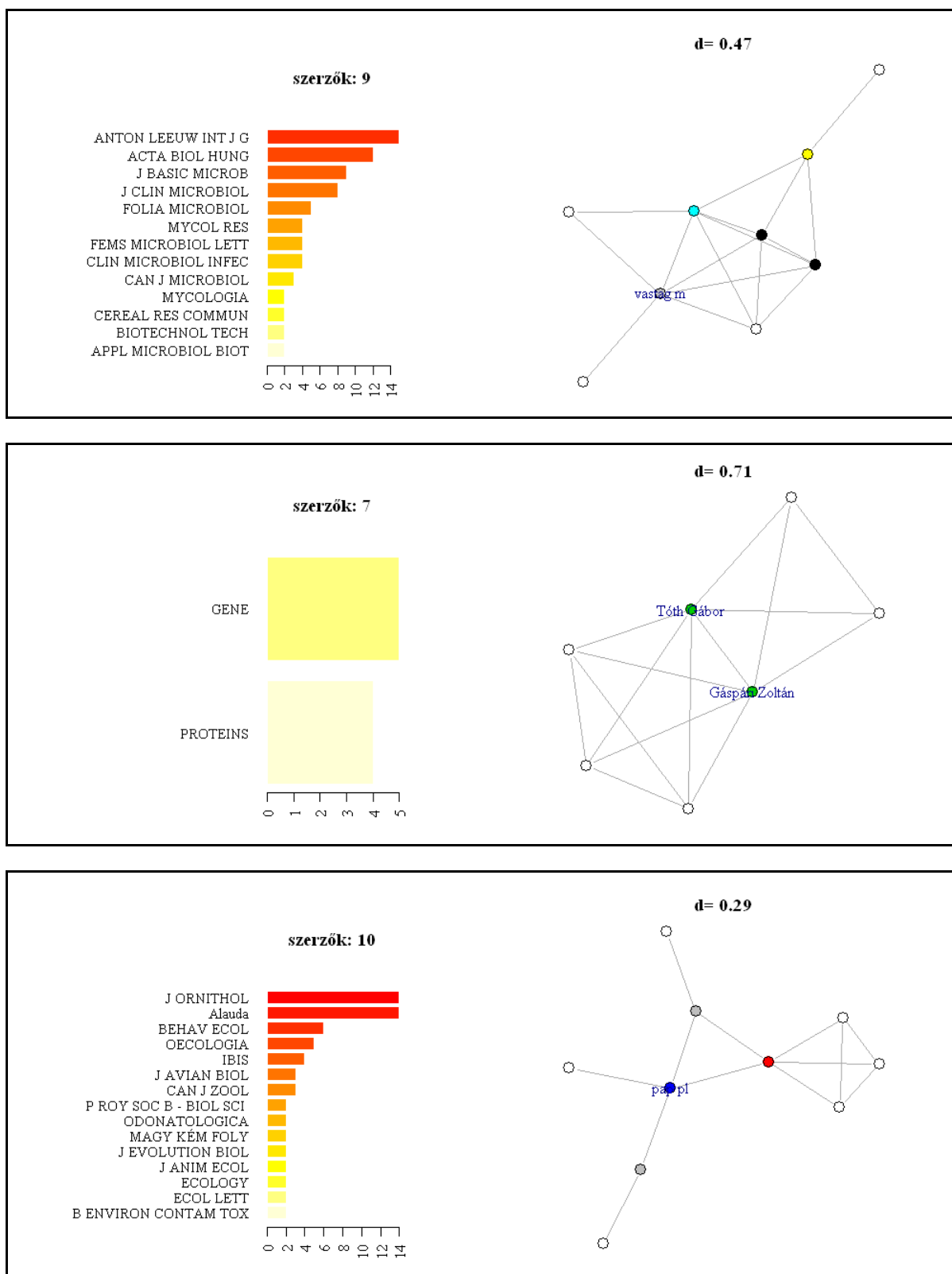
3–4. ábra Biokémia és biofizika

A csoportméret mentén továbbhaladva, a 20-40 fős klaszterekkel jellemzett közepmezőnyben alapvetően kétféle terület bontakozik ki: 1) egyrészt a klinikai orvostudomány és a biomedicina egyes, tematikusan fókuszált csoportjai (sürgősségi ellátás [critical care medicine], rákkutatás, gasztroenterológia), másrészt, 2) az egyed feletti biológia szintén tematizált körei (növénytan-ökológia, etológia, környezettudomány-talajtan). A jellemző csoportokat és felépítésüket az 5–6. ábra illusztrálja. Ezeknek a kisebb csoportoknak a koherenciája erősebb, viszont továbbra is jellemző az a struktúra, amelyben több kisebb alcsoportot az „interdiszciplináris” kapcsolódást megteremtő, összekötő szerepű szerzők fűznek egybe.

A csoportméret alsó tartományában ($n < 20$) hasonló tendenciát figyelhetünk meg, mint a közepes csoportok esetében. Az orvosbiológia és az egyed feletti biológia témacsoportjai mellett kifejezett a mikrobiológia jelenléte (7-9. ábra), amelynek még az alterületei is elkülönülnek (l. virológia). Hasonlóan explicit az élelmiszertudomány, az ökológia–evolúciobiológia és a genetika–genomika–szabályozás megjelenése. Figyelemre méltó, hogy egyes (főként orvosbiológiai) alterületek kicsi, de maximálisan összefüggő és egyöntetű (egyetlen folyóirattal jellemzett) szerzőkörökként válnak ki: ilyen az immunológia és a sejtkutatás.



5–6. ábra Példák közepes csoportokra: rákkutatás (klinikai orvostudomány), növénytan–szünbiológia (egyed feletti biológia)



7–9. ábra Példák kis csoportokra: mikrobiológia, genetika–genomika; állattan–ökológia–evolúcióbiológia

A fentieket összefoglalva, az alkalmazott bibliometriai eszközökkel és mintából a VIII. osztály az alábbi szakterületekre bontható fel:

BIOKÉMIA/KLINIKAI KÉMIA ÉS BIOFIZIKA

ORVOSBIOLÓGIA (KLINIKAI ORVOSTUDOMÁNY ÉS BIOMEDICINA)

- Sürgősségi ellátás
- Rákkutatás
- Gasztroenterológia
- Immunológia
- Neurobiológia és ideggyógyászat

ORGANIZMIKUS ÉS SEJTBOLÓGIA

- Mikrobiológia

- Virologia (mikrobiológia)

- Sejtbiológia

EGYED FELETTI BIOLÓGIA

- Botanika (egyed feletti)
- Viselkedéstudomány
- Ökológia és evolúcióbiológia
- Környezettudomány és talajbiológia

GENOMIKA–GENETIKA

ÉLELMISZERTUDOMÁNY

Az elemzésből kapott és a leírásba bevont csoportok mutatóit és az általuk képviselt folyóiratprofil az 1. táblázat tartalmazza.

1. táblázat

A határérték feletti számosságú klaszterek paraméterei és a hozzájuk tartozó folyóiratok

klasz-ter_ID	méret	denzitás	Folyóiratok
2	179	0.05	ACTA CRYSTALLOGR F, ACTA MICROBIOL IMMUNOL HUNG, ALLERGY, ARTERIOSCL THROM VAS, ATHEROSCLEROSIS, BIOCHEM BIOPH RES CO, BIOCHEM J, BIOCHEMISTRY-US, BIOCHIM BIOPHYS ACTA, BIOPHYS J, BIOSCI BIOTECH BIOCH, BMC GASTROENTEROL, CELL, CELL STRESS CHAPERON, CLIN CHIM ACTA
3	129	0.05	ACCOUNTS CHEM RES, ACTA BIOCHIM BIOPHYS ACAD SCI HUNG, ACTA BIOL HUNG, ACTA PHYSIOL HUNG, ALLERGY, ANAT EMBRYOL, ARCH BIOCHEM BIOPHYS, BBA-GEN SUBJECTS, BBA-GENE STRUCT EXPR, BBA-PROT STRUCT MOL ENZYMOL, BIOCHEM BIOPH RES CO, BIOCHEM J, BIOCHEMISTRY-MOSCOW+, BIOCHEMISTRY-US, BIOPHYS J
8	42	0.15	AM J RESP CRIT CARE, ANN SURG, ANN THORAC SURG, BIOCHEM BIOPH RES CO, BLOOD, BRIT J PHARMACOL, CELL MOL LIFE SCI, CIRCULATION, CRIT CARE MED, EUR J IMMUNOL, EXP DERMATOL, FASEB J, FREE RADICAL BIO MED, GLIA, IMMUNOLOGY
6	41	0.27	ARID LAND RES MANAG, BIOCHEM PHARMACOL, COMMUNITY ECOL, FOLIA GEOBOT, J VEG SCI, PHYTOCOENOLOGIA, PRESLIA
4	37	0.25	CANCER GENET CYTOGEN, CANCER LETT, CANCER RES, CYTOMETRY, GENE CHROMOSOME CANC, INT J CANCER, J HISTOCHEM CYTOCHEM, J LEUKOCYTE BIOL, LEUKEMIA RES, ORVOSI HETILAP, OTOLARYNG HEAD NECK SURG, P NATL ACAD SCI USA
10	33	0.84	J ANIM ECOL, SCIENCE
11	20	0.47	LEUKEMIA LYMPHOMA, ORVOSI HETILAP, SCAND J GASTROENTERO, WORLD J GASTROENTERO

klaszer_ID	méret	denzítás	Folyóiratok
28	16	0.41	CELL STRESS CHAPERON, CLIN VACCINE IMMUNOL, MOL CELL NEUROSCI
45	16	0.7	SCI TOTAL ENVIRON, SOIL BIOL BIOCHEM
32	14	0.38	BIOPHYS J, BMC BIOINFORMATICS, CANCER LETT, CYTOMETRY A, EUR J CANCER, IMMUNOL LETT, MOL CANCER THER, SOIL BIOL BIOCHEM
18	13	0.46	BIOCHIM BIOPHYS ACTA, HISTOCHEM CELL BIOL, P NATL ACAD SCI USA
27	12	0.64	CELL, EMBO J, J GEN VIROL, J VIROL, PLANT CELL, VIROLOGY, VIRUS RES
42	12	0.44	BIOCHEM BIOPH RES CO, BLOOD, J DENT RES, J IMMUNOL
9	11	1	J IMMUNOL
15	10	0.69	CHROMATOGRAPHIA, FOOD CHEM
36	10	0.91	BIOCHEMISTRY-US, P NATL ACAD SCI USA
66	10	0.44	J BIOL CHEM, MOL GEN GENET, NUCLEIC ACIDS RES, P NATL ACAD SCI USA
94	10	0.29	Alauda, B ENVIRON CONTAM TOX, BEHAV ECOL, CAN J ZOOL, ECOL LETT, ECOLOGY, IBIS, J ANIM ECOL, J AVIAN BIOL, J EVOLUTION BIOL, J ORNITHOL, MAGY KÉM FOLY, ODONATOLOGICA, OECOLOGIA, P ROY SOC B - BIOL SCI
1	9	0.47	ACTA BIOL HUNG, ANTON LEEUW INT J G, APPL MICROBIOL BIOT, BIOTECHNOL TECH, CAN J MICROBIOL, CEREAL RES COMMUN, CLIN MICROBIOL INFEC, FEMS MICROBIOL LETT, FOLIA MICROBIOL, J BASIC MICROB, J CLIN MICROBIOL, MYCOL RES, MYCOLOGIA
63	9	0.44	ACTA BIOCHIM BIOPHYS HUNG, ANTIBIOT CHEMOTHER, CYTOMETRY A
73	9	1	CELL
5	8	0.71	APPL MICROBIOL BIOT, CURR GENET
13	8	1	J AM CHEM SOC
46	8	0.71	ARCH BIOCHEM BIOPHYS, J BIOL CHEM
53	8	1	J CEREBR BLOOD F MET
85	8	1	J NEUROVIROL
19	7	0.62	IDEGGYÓGYÁSZATI SZEMLE, J COMP NEUROL
29	7	0.62	APPL SOIL ECOL, COMMUNITY ECOL
82	7	0.71	GENE, PROTEINS
96	7	1	BIOSCI BIOTECH BIOCH

Diszkusszió és következtetések

Az akadémiai osztályok bibliográfiai felbontásához javasolt módszer, illetve hasonlósági metrika értékelésének alapvető mozzanata az olyan módszerekkel való összehasonlítás, amelyek szintén használatosak és hozzáférhetőek a tudományrendszer empirikus feltérképezésében. Jelenlegi céljaink szempontjából ezeknek a módszereknek a többsége a feladathoz felhasznált információ típusa szerint két nagy kategóriába, (1) a hivatkozás-elemzésre, illetve (2) az egyéb, nem bibliometriai információra (pl. szakértői véleményekre) épülő rendszerezés kategóriájába sorolhatók.

A hivatkozási relációkra épülő megközelítés, ha a rendszerezés a szakmai folyóiratok szintjén zajlik, a szaklapok kölcsönös hivatkozásai, és ennek megfelelő hasonlósági viszonyaik alapján alakítja ki a területi klasztereket. A publikációk szintjén ugyancsak alkalmazható, de a hivatkozásokat másképpen felhasználó rendszerezési technika a bibliográfiai csoportosítás (bibliographic coupling), amely a közös hivatkozások aránya alapján rendel egymáshoz tematikusan hasonló közleményeket. Ezek a módszerek azonban értelemszerűen csak olyan minták esetében használhatók eredményesen, ahol a hivatkozások minden esetben maradéktalanul hozzáférhetőek az adatbázisban. A vizsgált adatbázis (KPA) funkciói között ugyanakkor a közlemények hivatkozáslistájának nyilvántartása nem elsődleges, mivel a KPA eredendően nem tudománytérképezési célokat szolgál, a tudomány-metriai aspektust pedig a cikkekre vonatkozó citációk nyilvántartása képviseli. Nem szükségszerű tehát, hogy a hivatkozások minden vizsgált közlemény esetében hozzáférhetőek. Ennek tükrében a fentiekben bemutatott módszer erőssége, hogy csak olyan adatokra támaszkodik, amelyek a KPA-ban minden cikkre vonatkozóan biztosan fellelhetők (cím, leőhely).

A nem bibliometriai információforrások, így a szakértői értékelések alapján készült kategorizációk, bár kétségtelenül autentikusak, bizonyos mértékig szükségképpen személyes tapasztalatokat és látókört, vagyis egyfajta szubjektivitást tükröznek. A szakértői ítélet, a „top–down” megközelítés ideális esetben együttesen alkalmazható a bibliometriai, az adatokból feltárható összefüggésekre építő „bottom–up” jellegű módszerekkel (vö. [3]). Az általunk használt módszer ilyen „bottom–up” típusú, és – bár az eredmények értelmezésében nyilvánvalóan használ egy előzetesen definiált szakterületi felosztást – lényegében kategóriamentes, vagyis

nem támaszkodik *a priori* osztályozásokra. Ebben a tekintetben kiküszöböli a szubjektivitásból eredő torzításokat.

A módszert végül célszerű összevetni egy olyan eljárással, amely természetes módon strukturálja az elemzett mintában előforduló szerzőket: a szerzői kapcsolatháló-elemzés lehetőségével. A társ-szerzői reláció alapján való felosztás a tényleges szakmai együttműködések mentén strukturálja a szerzőket, ebből a szempontból tehát igen megbízható. Hátránya ugyanakkor, hogy kizárólag azokat a szerzőket képes kezelni, akik (legalább egy publikáció erejéig) társszerzői viszonyban vesznek részt. Ez a sajátosság még olyan területeken is bizonytalanná teheti a klaszterezés eredményét, ahol a szerzői együttműködés általános gyakorlat. Az itt bemutatott szerzőhasonlósági metrika ugyanakkor rendelkezik azzal a tulajdonsággal, hogy az egyszemélyes cikkek, illetve magányos szerzőkre is kiterjed, továbbá definíciójából fakadóan tartalmazza a társszerzői információkat is (lévén, hogy a közös publikációk szintjén a cím és a leőhely azonosságából adódóan a két szerző kombinált hasonlósága maximális). Ilyen értelemben az általunk feltárt hálózat tartalmazza a szerzői kapcsolathálót is (mint a hasonlóság egy speciális esetét).

A Biológiai Tudományok Osztályának felbontásáról megállapítható, hogy miközben a klaszterezés az említett bottom-up jelleggel, és kizárólag az osztály részleges, nem reprezentatív publikációs listája alapján készült, a kapott területi struktúra így is összhangban van a tudományrendszer globális feltérképezésének korábbi eredményeivel. A kirajzolódó kategóriák megfelelésbe hozhatók a Glanzel–Schubert javasolta háromszintű, a tudomány-metriai értékelés számára konstruált globális rendszer élettudományi kategóriáival (Glanzel, Schubert, 2003), amely a szakértői és az adatorientált elemzés kombinációjával készült. Látványos továbbá a megfelelés a Rafols–Leydesdorff-féle kísérlet eredményeivel, amely az ISI JCR folyóiratainak hivatkozás-alapú kategorizációját célozta [4]. A szakterületek (subject categories) kapcsolati térképén különösen az orvosi biológiai alterületek rendszere mutat feltűnő hasonlóságot az általunk kapott, az egyes specialitásokat kisebb csoportokban ábrázoló felosztással.

A minta által képviselt csoport belső struktúrája egyaránt jellemző a nagyfokú homogenitás, illetve a szakterületi diverzitás. A kettősség abból adódik, hogy a vizsgált anyag legnagyobb része egy

alterületet képvisel (biokémia–biofizika), de kisebb kategóriákban megjelenik az élettudomány teljes spektruma. Az akadémiai osztálystruktúra ismeretében az összetétel feltűnő vonása, hogy a vizsgált mintában meghatározó a klinikai orvostudomány, amely az *Orvosi Tudományok Osztályával* való erős szakterületi átfedést feltételez. Általánosabban, feltételezhető, hogy a tudományometriai értékelésekhez, összehasonlításokhoz adekvát szakterületi felosztás, szerzőcsoportok meghatározásához a KPA-ban szereplő szerzők olyan csoportosítására van szükség, amely – nagyobb diszciplínákon, például az élettudományon belül – keresztülmetszi az akadémiai osztálystruktúrát. A fentiekben leírt munka következő fázisa ebből adódóan a klaszterezési módszer ennek megfelelő, és immár reprezentatív mintán vagy a teljes adatbázison való alkalmazása: így esetünkben az élettudományi osztályok (biológia, orvos- és agrártudomány) együttes vizsgálata, amelytől így az összehasonlíthatóság kritériumainak jobban megfelelő szerzői körök definícióját remélhetjük.

Irodalom

- [1] TOLNAI Márton: Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak

tükrében. = Magyar Tudomány, 169. köt. 8. sz. 2008. p. 976–988.

- [2] KOLLÁR István – MICHELBERGER Pál: Hozzászólás Tolnai Márton Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak tükrében című cikkéhez. = Magyar Tudomány, 169. köt. 10. sz. 2008. p. 1262–1273.
- [3] GLÄNZEL, W. – SCHUBERT, A.: A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. = Scientometrics, 56. köt. 3. sz. 2003. p. 357–367.
- [4] RAFOLS, I. – LEYDESDORFF, L.: Content-based and Algorithmic Classifications of Journals: Perspectives on the Dynamics of Scientific Communication and Indexer Effects. = Journal of the American Society for Information Science and Technology, megjelenés alatt. URL: <http://users.fmg.uva.nl/leydesdorff/classifications/classifications.pdf>

Beérkezett: 2009. XI. 30-án.



Soós Sándor

az MTA Kutatásszervezési Intézetének tudományos munkatársa.

E-mail: ssoos@colbud.hu

Bevezette a böngészőválasztást a Microsoft

Az amerikai nagyvállalat operációs rendszereit telepítő számítógép-használóknak ezentúl olyan képernyő bukkan fel, amely 12 böngésző közül kínál választást. A felkínált internetes keresők között a jól ismert programok – Internet Explorer, Firefox és az Opera – mellett kevésbé nevesek is szerepelnek, mint amilyen a K-meleon és a Seipnir. A Microsoft európai részlege azt követően kényszerült a szabad böngészőválasztás felkínálására, hogy az *Európai Bizottság* úgy foglalt állást: a szabad verseny megsértését jelenti az a gyakorlat, hogy a vállalat az Internet Explorerrel telepíti az operációs rendszerekre. A cég megbírságolásától eltekintettek, azzal a feltétellel, hogy a Microsoft változtat az alkalmazáson.

A választási lehetőség a Windows XP, a Vista és a Windows 7 felhasználóinak automatikusan ugrik fel. Az Opera, a Google Chrome, a Safari, az Internet Explorer és a Firefox véletlenszerű sorrendben bukkan fel a képernyő egyik – szembeűnő – részén, míg az egyéb böngészők, beleértve a Flock, az Avant és a Maxthon nevű programokat, a monitor másik részén jelennek meg. A választást a keresőkről szóló rövid ismertetővel segítik.

Kitűzőkkel és képeslapokkal kampányol a *Mozilla Alapítvány* a hannoveri *CeBIT* számítástechnikai kiállításon a nyílt forráskódú, szabadon terjeszthető szoftverek pavilonjában. A Firefox böngészőprogram fejlesztője nem titkolja, hogy örül az Európai Bizottság szabad böngészőválasztásról szóló döntésének, azonban aggodalommal figyeli, hogy a felhasználók jelentős része még azt sem tudja, hogy mi a különbség a webes keresőmotor és a böngészőprogram között.

/Sg.hu Hírlevél, 2010. március 3., <http://www.sg.hu/>

(SzP)