

KOSZTYÁN ZSOLT TIBOR–BANÁSZ ZSUZSANNA–CSÁNYI  
VIVIEN VALÉRIA–TELCS ANDRÁS

## Felsőoktatási ligák, parciális rangsorok képzése biklaszterezési eljárásokkal

Napjainkban számtalan felsőoktatási rangsor, illetve különböző szakterületekre, régiókra vonatkozó részrangsor létezik. Az azonban továbbra is nyitott kérdés, hogy mely egyetemet vagy mely országok felsőoktatási rendszerét lehet, és melyeket érdemes összehasonlítani. Tanulmányunk erre a kérdéskörre keres választ. Olyan módszereket javasolunk, amelyek a társadalomtudomány területén még újszerűek, kevésbé alkalmazottak, ugyanakkor segítségükkel lehetőség nyílik az úgynevezett ligák meghatározására. A ligákat értelmezésünkben olyan egyetemek/országok/entitások alkotják, amelyek bizonyos indikátorok alapján összehasonlíthatók. A feladat tehát kettős: egyszerre kell az egyetemeket/országokat és azokat az indikátorokat kiválasztani, amelyek alapján az intézmények vagy az országok felsőoktatási rendszere összehasonlítható.\*

Journal of Economic Literature (JEL) kód: C1, I2.

A legelső egyetemi rangsor 1983-ban jelent meg az U.S. News and World Report gondozásában, amelynek célja az amerikai felsőoktatási intézmények rangsorolása volt. Az első globális egyetemi rangsort csak 2003-ban publikálták, amelyet a Shanghai Jiao Tong University készített, majd ezt követően 2004-ben jelent meg a The Times Higher Education

\* Jelen kutatás az Európai Unió, Magyarország és az Európai Szociális Alap társfinanszírozása által biztosított forrásból az EFOP-3.6.2-16-2017-00017 azonosítójú, Fenntartható, intelligens és befogadó regionális és városi modellek című projekt keretében jött létre.

*Kosztván Zsolt Tibor*, MTA–PE Budapest Rangsor Kutatócsoport, Pannon Egyetem Gazdaságtudományi Kar, Kvantitatív Módszerek Intézeti Tanszék (e-mail: kzst@gtk.uni-pannon.hu).

*Banász Zsuzsanna*, MTA–PE Budapest Rangsor Kutatócsoport, Pannon Egyetem Gazdaságtudományi Kar, Kvantitatív Módszerek Intézeti Tanszék (e-mail: banasz.zsuzsanna@gtk.uni-pannon.hu).

*Csányi Vivien Valéria*, MTA–PE Budapest Rangsor Kutatócsoport, Pannon Egyetem Gazdaságtudományi Kar, Kvantitatív Módszerek Intézeti Tanszék (e-mail: csanyi.vivien@gtk.uni-pannon.hu).

*Telcs András*, MTA–PE Budapest Rangsor Kutatócsoport, Pannon Egyetem Gazdaságtudományi Kar, Kvantitatív Módszerek Intézeti Tanszék, MTA Wigner Fizikai Kutatóintézet, BME Villamosmérnöki és Informatikai Kar, Számítástudományi és Információelméleti Tanszék, MTA Wigner Fizikai Kutatóintézet, Komputációs Tudományok Osztálya, VIAS – Virtual Institute of Advanced Studies (e-mail: telcs.andras@gtk.uni-pannon.hu).

A kézirat első változata 2018. április 16-án érkezett szerkesztőségünkbe.

DOI: <http://dx.doi.org/10.18414/KSZ.2019.9.905>

rangsora. Az azóta eltelt időszakban számos egyetemi rangsor jelent meg, amelyeknek az eredeti fő célja az volt, hogy segítse a hallgatókat a számukra megfelelő felsőoktatási intézmény kiválasztásában (*Lukman és szerzőtársai* [2010], *Shin* [2011]).

Az egyetemi rangsorok a mai napig nagy népszerűségnek örvendenek, annak ellenére, hogy számos kritika éri őket (lásd összefoglalóan *Olcay–Bulu* [2017]). A velük kapcsolatban felmerülő kritikák három nagy témakör köré csoportosíthatók. Az első ilyen nagy témakör a rangsorok intézményekre gyakorolt torzító hatása, hiszen a rangsorokban elért helyezés nagyban függ az intézmények reputációjától (lásd például *Guarino és szerzőtársai* [2005], *Marginson–Van der Wende* [2007]). A második fő probléma a rangsorolás során felhasznált adatokhoz és indikátorokhoz köthető. Egyrészt, a rangsorokban felhasznált indikátorokat és a hozzájuk tartozó súlyokat önkényesen választják meg a rangsorok készítői, másrészt pedig a súlyok és az aggregálási módszer megváltoztatásával egymástól nagyon eltérő rangsorokat kapunk (lásd például *Billaut és szerzőtársai* [2010], *Saisana és szerzőtársai* [2011], *Bengoetxea–Bucla-Casal* [2013]). A harmadik nagy problémakör – ami talán a legvitatottabb pontja a rangsorolásnak –, hogy merőben eltérő felsőoktatási intézményeket, illetve nagyon különböző felsőoktatási rendszereket hasonlítanak össze. Számos tanulmány rámutatott (lásd például *Daraio–Bonaccorsi* [2017]), hogy az egymástól méretben és finanszírozásban eltérő intézményeket nem célszerű összehasonlítani. *Daraio–Bonaccorsi* [2017] részletesen tárgyalja, hogy mikor tekinthető egy felsőoktatási rangsor igazságosnak. Tanulmányunkban egy olyan módszert javasolunk, amelynek végeredményeképpen igazságosabb összehasonlítást tehetünk az egyes országok felsőoktatási rendszereire vonatkozóan.

Az egyetemi, illetve a felsőoktatási rendszereket tekintve „országligákon” – hasonlóan a sportban már régóta használt kifejezéshez – olyan egyetemek/felsőoktatási rendszerek csoportját értjük, amelyek adott szempontok szerint összehasonlíthatók (*Downing* [2013], *Salmi* [2013]). A szerzők egyetértenek *Benneworth* [2010] és *Liu* [2013] felvetésével, amely szerint összehasonlítani csak adott szempontok szerint hasonló intézményeket szabad, ellenkező esetben ezek az összehasonlítások, valamint az így képzett rangsorok meglehetősen torzított képet mutatnak. Ennek kiküszöbölése érdekében keresünk olyan ligákat, ahol az adott országok, illetve intézmények összehasonlíthatók. Tanulmányunkban a *liga* kifejezés egyszerre határozza meg országok, illetve felsőoktatási intézmények egy csoportját, valamint azon indikátorok halmazát, amelyek szerint az adott intézmények összehasonlíthatók. Rangsorokat a ligákon belül képzünk.

A ligák képzésére a társadalomtudományokban eddig kevésbé ismert és alkalmazott módszert, az úgynevezett biklaszterezést (*Cheng–Church* [2000]) alkalmazzuk. Ez a módszer lehetővé teszi, hogy egyszerre válasszunk ki felsőoktatási intézményeket vagy éppen felsőoktatási rendszereket, valamint azokat az indikátorokat, amelyek szerint az összehasonlítás elvégezhető. A módszer egyaránt alkalmas felsőoktatási intézmények ligáinak meghatározására és országok felsőoktatási rendszerének vizsgálatára. Az egyszerűbb bemutatás kedvéért ez utóbbira hozunk példát: az összehasonlítást országok, illetve azok felsőoktatási rendszerei között fogjuk elvégezni.

Tanulmányunk a következőképpen épül fel. Rövid szakirodalmi áttekintést nyújtunk a felsőoktatási rangsorok témakörében. Bemutatjuk a legnépszerűbb

nemzetközi, regionális és nemzeti rangsorokat, továbbá megtárgyaljuk a klaszterezési és biklaszterezési eljárások különbségeit is. Majd ismertetjük az általunk felhasznált adatbázist, és részletes betekintést adunk a biklaszterezési eljárások módszertanába. Ezt követően bemutatjuk és értelmezzük a módszerünk által létrehozott ligákat, majd pedig az egyes ligákba tartozó országok parciális rangsorait. Végül összefoglaljuk eredményeinket, és levonjuk főbb következtetéseinket.

## Szakirodalmi áttekintés

A felsőoktatást érintő rangsorokat különbözőképpen lehet csoportosítani. Az egyik kézenfekvő csoportosítás maga a *területi hatókör*, amely szerint beszélhetünk nemzeti, regionális és nemzetközi rangsorokról. A másik csoportosítási szempont az *aggregálás, összehasonlítás szintje*, azaz hogy felsőoktatási intézményeket vagy országokat – azok felsőoktatási rendszerének egésze szerint – rangsorolunk-e (1. táblázat).

### 1. táblázat

Rangsorok csoportosítása, példákkal

Területi hatókör	Összehasonlítás szintje		
	intézmény	felsőoktatási rendszer	
Nemzetközi	ARWU, THE, QS, CWTS, U-Multirank	U21 (2012–2017) QS (2008, 2016) Lisbon Council (2008)	
Regionális	nemzetközi	QS, THE: Latin-Amerika, Ázsia QS: EECA, BRICS	<i>jelen tanulmány tárgya</i>
	nemzeti	U.S. News: Észak-, Dél-, Középnnyugat-, Nyugat-USA	nem értelmezhető
Nemzeti	Egyesült Királyság: The Complete University Guide in UK Egyesült Államok: Forbes-rangsorok Magyarország: Felvi-, HVG-rangsorok	nem értelmezhető	

*Rövidítések:* ARWU: Academic Ranking of World Universities, THE: Times Higher Education, QS: Quacquarelli Symonds, CWTS: Centre for Science and Technology Studies (Leideni Egyetem), U.S. News: U.S. News & World Report, U21: Universitas21

*Forrás:* saját szerkesztés.

A rangsorok általában az intézményeket veszik számba, azokat hasonlítják össze különböző szempontok szerint, de léteznek az országok felsőoktatását vizsgáló úgynevezett országrangsorok is. Valamennyi összehasonlítás esetén kulcskérdés, hogy kiket, mely országokat vagy mely intézményeket szabad összehasonlítani. Éppen ezért ligák valamennyi esetben képezhetők.

*Nemzetközi, regionális és nemzeti rangsorok*

A teljesen különböző adottsággal rendelkező egyetemek összehasonlíthatósága nem új keletű probléma, ennek egyik következménye az országos (nemzeti) és regionális rangsorok megjelenése. Magyarországon is több intézmény készít intézményi rangsorokat (lásd például a Felvi-, HVG-rangsorokat),<sup>1</sup> de találhatunk rangsorokat az angol (The Complete University Guide in UK)<sup>2</sup> vagy éppen az amerikai felsőoktatási intézményekről is (például a Forbes és a CCAP<sup>3</sup> közös rangsora).<sup>4</sup> A Közgazdasági Szemlében is jelentek már meg olyan tanulmányok, amelyek azt mutatták be, hogy miként lehet Magyarországon a hallgatói jelentkezések alapján preferencia-rangsorokat képezni (*Telcs és szerzőtársai* [2013a], [2013b], *Csató* [2013], [2016]). Jelen kutatásunknak nem ezek az alternatív rangsorképző eljárások a tárgyai, hanem a már meglévő rangsorok által használt indikátorokon képzett ligák és az ezeken belül számolt parciális rangsorok.

A nemzetközi összehasonlíthatósághoz szükség van a nemzetközi rangsorok figyelembevételére is. Ezek közül a legismertebbek: a Sanghaj rangsorként is ismert Academic Ranking of World Universities (ARWU),<sup>5</sup> a Times Higher Education World University Ranking (THE),<sup>6</sup> a Quacquarelli Symonds (QS) által készített World University Ranking,<sup>7</sup> a Leideni Egyetem Centre for Science and Technology Studies (CWTS) rangsora,<sup>8</sup> valamint az U-Multirank<sup>9</sup> (1. táblázat).

A fenti rangsorkészítő szervezetek közül többen is felismerték, hogy nagyon nehéz teljesen eltérő pénzügyi, méretbeli adottsággal rendelkező intézményeket összehasonlítani. Például egy magyar felsőoktatási intézmény kevés eséllyel tud megjelenni a Sanghaj-rangsor élmezőnyében, aminek számos oka van (lásd például *Török* [2006]). Tanulmányunknak nem tárgya ennek részletezése, de a nagyságrendek érzékeltetésére megemlíjtük *Polónyi* [2017] példáját: a Harvard Egyetem költségvetése két és félszer akkora, mint a teljes magyar felsőoktatásé. Éppen ezért mind a THE-rangsorból, mind pedig a QS-rangsorból (*Sowter és szerzőtársai* [2017]) létezik egy-egy földrajzi (például Latin-Amerika országait elemző) vagy éppen gazdasági szempontokat figyelembe vevő, például a BRICS,<sup>10</sup> EECA<sup>11</sup> országait vizsgáló, úgynevezett regionális rangsor (területi hatókör az 1. táblázatban). E nemzetközi regionális rangsorok mellett például az Egyesült Államokban a U.S. News publikál országon belüli, földrajzi (északi, déli, középnnyugati, illetve nyugati) régiókon belüli rangsorokat is.<sup>12</sup> Ez is alátámasztja a parciális rangsorok jelentőségét.

<sup>1</sup> <http://eduline.hu/rangsor>.

<sup>2</sup> <https://www.thecompleteuniversityguide.co.uk>.

<sup>3</sup> *Center for College Affordability and Productivity (CCAP)*.

<sup>4</sup> <https://www.forbes.com/top-colleges>.

<sup>5</sup> <http://www.shanghairanking.com>.

<sup>6</sup> <https://www.timeshighereducation.com/world-university-rankings>.

<sup>7</sup> <https://www.topuniversities.com/university-rankings>.

<sup>8</sup> <http://www.leidenranking.com/>.

<sup>9</sup> <https://www.umultirank.org/>.

<sup>10</sup> BRICS: Brazília, Oroszország, India, Kína, Dél-Afrika (*Brazil, Russia, India, China, South Africa*).

<sup>11</sup> EECA: feltörekvő európai és közép-ázsiai országok (*Emerging Europe and Central Asia*).

<sup>12</sup> <https://www.usnews.com/best-colleges/rankings/regional-universities>.

A régiók (legyen az gazdasági vagy földrajzi) kiválasztása önkényesnek tekinthető. Bár a rangsorok készítői remélik, hogy a regionális rangsorokban lévő intézmények összehasonlíthatósága javul, a régiók önkényes kijelölése e kívánalom megvalósulását megghiúsíthatja. *Jarocka* [2012], *Abankina és szerzőtársai* [2016] és *Ibáñez és szerzőtársai* [2013] már felvetették, hogy gazdasági és regionális rangsorok mellett érdemes lenne az intézmények összehasonlításához klasztereket képezni. A hagyományos klaszterezési eljárások azonban erre csak korlátozottan alkalmasak, hiszen az intézmények közötti hasonlóságot valamennyi (az elemzésbe bevont) indikátor alapján számolják, amelyek egy része nem vagy csak nehezen értelmezhető valamennyi intézményre. Így annak eldöntése, hogy mi alapján hasonlíthatók össze az egyes intézmények, legalább annyira fontos kérdés, mint annak eldöntése, hogy mely intézmények hasonlíthatók össze.

Az intézményi rangsorok mellett ma már az ország egészének felsőoktatási rendszerét is vizsgálják. Természetesen itt nem értelmezhető a nemzeti rangsor (kivételt képezhetne egy szövetségi rendszer államainak összehasonlítása). A nemzetközi országangsorok közül említést érdemel a *Lisbon Council* által 2008-ban készített (17 európai OECD-oroszgot listázó) országangoron (*Ederer és szerzőtársai* [2009]) és a QS által 2008-ban és 2016-ban publikált országangorokon túl (*Hazelkorn* [2015], (QS [2016]) a – vizsgálatunk alapját képező – 50 ország felsőoktatási rendszerét összehasonlító Universitas21 (U21) nevű szervezet országangora, amelyet évente publikálnak 2012 óta.

Tudomásunk szerint országangorokból nem készült semmilyen regionális rangsor, pedig a teljesen eltérő felsőoktatási rendszerek összehasonlítása hasonló kérdéseket vet fel, mint a felsőoktatási intézmények összehasonlítása.

Valóban fontos lehet a hasonló intézmények, illetve hasonló felsőoktatási rendszerek összehasonlítása, de ez a kijelölés nem lehet önkényes. Olyan módszer alkalmazására van szükség, amely képes azokat az intézményeket vagy országokat megtalálni, amelyek bizonyos szempontok szerint összehasonlíthatók. A rangsorok is ezen intézmények/országok esetében képezhetők olyan indikátorok alapján, amelyek indokolják az intézmények/országok összehasonlíthatóságát. Mivel a kialakult ligák nem feltétlenül követik a földrajzi vagy gazdasági értelemben vett regionális besorolásokat, így az általunk javasolt rangsor is inkább parciális vagy részleges rangsornak nevezendő, nem pedig regionális rangsoroknak. A ligák és ezután a parciális rangsorok kialakításánál lényegében egyszerre kell választ kapnunk arra a két kérdésre, hogy 1. mely intézmények/országok, 2. mely szempontok szerint hasonlíthatók össze, ehhez pedig egy, a társadalomtudományban eddig nagyon kevésbé használt módszer, az úgynevezett biklaszterezési eljárás alkalmazása válik szükségessé. Tanulmányunkban három ligát azonosítunk: elit liga (A liga), középmezőny (B liga) és lemaradók (C liga).

### *Klaszterezési és biklaszterezési eljárások összehasonlítása*

A klaszterezési eljárások több évtizedes múlttal rendelkeznek (*Albalate–Minker* [2013], *Csicsman* [1979], *Néda és szerzőtársai* [2008]). A klaszter fogalmának első megjelenése egészen 1739-re vezethető vissza, a klaszterelemzés pedig először 1954-ben jelent meg, azóta pedig valamennyi, statisztikai módszereket is igénylő tudományág alkalmazza.

A módszer használható mind az adatok (a mi esetünkben országok), mind pedig a változók csoportosítására. A biklaszterezési eljárás, amely először 1972-ben jelent meg (Hartigan [1972]), mindezt egy lépésben képes elvégezni. A módszert genetikai adatok elemzésére Cheng–Church [2000] dolgozta ki. A biklasztereket leginkább a bioinformatika alkalmazza, de ma már találkozhatunk ezzel a módszerrel az üzleti tudományokban, például a marketing területén (lásd például Liu és szerzőtársai [2009], valamint Huang [2011]). Raponi és szerzőtársai [2016] az olasz felsőoktatási intézmények csoportosítására használta a biklaszterezési eljárást, azonban nem határozta meg a ligákat, illetve a ligákon belül értelmezhető parciális/részleges rangsorokat.

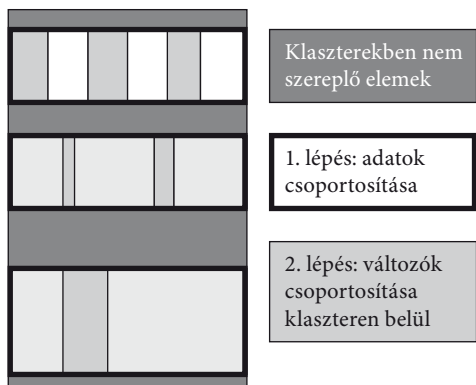
A biklaszterezésnek számos fajtája létezik (Pontes és szerzőtársai [2015]). Az a két eljárás, amelyet a továbbiakban itt bemutatunk, a felsőoktatási rangsorolásban is alkalmazható. Az egyik eljárás eredményeként azonosíthatók a szélsőséges esetek (bizonyos indikátorcsoportok szerint kimagaslóan jól, illetve rosszul teljesítő országok), a másik eljárás pedig éppen a szélsőségektől tekint el azáltal, hogy a legnagyobb homogén (legkisebb szórású) adatcsoportot keresi (ami jelen esetben a középmezőny). Mivel a módszer egyszerre klaszterezi az országokat és az indikátorokat is, ezért a középmezőnyben szereplő országok nem feltétlenül azok, amelyek sem a top, sem a lemaradók között nem szerepelnek, mivel más indikátorok szerint sorolódhatnak az országok az elitbe, a középmezőnybe és a lemaradókhöz.

Az egy és akár több lépésből álló klaszterezési és a biklaszterezési eljárások közötti különbséget az 1. ábra szemlélteti. Látható, hogy a biklaszterezési eljárás során a cél egy homogén részmatrix megtalálása. Ezeket a részmatrixokat nevezzük ligáknak. Az egyes ligák változói és/vagy országai átfedhetik egymást, amely átfedés további érdekes eredményeket szolgáltatathat.

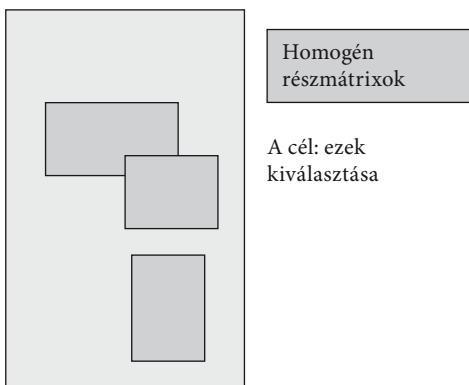
### 1. ábra

Töbllépéses klaszterezési és biklaszterezési eljárások összehasonlítása

#### a) Kétlépéses klaszterezés



#### b) Biklaszterezés



Forrás: saját szerkesztés.

Mind a klaszterezési, mind a biklaszterezési eljárásoknál kulcskérdés a hasonlósági vagy távolságfüggvény megválasztása. A részmatrixok kiválasztása általában

kombinatorikus, NP-nehéz feladat, ezért hasonlóan a klaszterezési eljárásokhoz, közelítő módszereket használ.

A közelítő eljárások közül általában a genetikus algoritmusok alkalmazása a legnépszerűbb (lásd például *Cheng–Church* [2000], *Gusenleitner és szerzőtársai* [2012], *Yang és szerzőtársai* [2003]). Leginkább ezzel a megközelítéssel sikerült nagyobb adatbázisokon is homogén klasztereket találni (lásd például *Yang és szerzőtársai* [2005]). A módszerek egy része megköveteli az adatok binarizálását (lásd például *Gusenleitner és szerzőtársai* [2012]), míg mások csak az adatok normalizálását igénylik (ilyen például *Yang és szerzőtársai* [2005]). A biklaszterezési eljárások közül mi jelen tanulmányban kettővel foglalkozunk részletesebben.

Az első módszer az iBBiG (iterative Binary Biclustering of Gene) eljárás (*Gusenleitner és szerzőtársai* [2012]). A biklaszterezés e fajtájának mozaikszavában az iteratív kifejezés arra utal, hogy a végső biklasztert több lépcsőben, iteratív eljárás segítségével határozza meg. Maga a genetikus algoritmus is, amelyet e módszer alkalmaz, iteratív, ezt a szerzők a módszer nevében is hangsúlyozták. A bináris kifejezés arra utal, hogy a módszer alkalmazása előtt a kiindulási mátrix (amelynek soraiban a vizsgált országok vannak, oszlopaiban pedig a változók) egyes celláinak az (1) egyenlet alapján normalizált értékeit egy előre meghatározott  $\tau$  küszöbérték felett 1-gyé, az ez alattiakat 0-vá konvertálja. A normalizálás során  $b_{ij}$  egy  $\mathbf{B}$  mátrix egy elemét (celláját) jelöli,  $b'_{ij}$  pedig a cella normalizált értékét:

$$b'_{ij} := \frac{b_{ij} - \min_j b_{ij}}{\max_j b_{ij} - \min_j b_{ij}}. \quad (1)$$

A  $\tau$ -érték lehet az adattábla átlaga vagy valamely kvartilise. Mivel az adatok normalizáltak,  $\tau$  értéke 0 és 1 között van. A mozaikszóban lévő gén kifejezés a módszer első felhasználási területére utal, mivel a biklaszterezést a gének vizsgálatával kapcsolatosan publikálták először. A módszer célja, hogy minél nagyobb olyan ( $\mathbf{B}$ ) részmátrixot határozzon meg, amelyen belül az értékek átlaga a  $\tau$  küszöbérték felett van, és az entrópia ( $H_{\mathbf{B}}$ ) alacsony, azaz amelyben hasonló értékek (1-esek) vannak. A felhasználó e két rész cél (1. méret, 2. alacsony entrópia) között egy  $\alpha \in [0, 1]$  paraméterrel döntheti el, hogy melyik rész cél fontosabb számára. A szakirodalom  $\alpha = 0,3$ -et javasol. Ennél alacsonyabb érték esetében nagyobb méretű, míg magasabb értékek esetében kisebb méretű, de homogénebb csoportokat kapunk. A módszer során egy  $\mathbf{B}$  (rész)mátrixra vonatkozóan a (2) egyenlet pontértékét (*score*) maximalizáljuk.

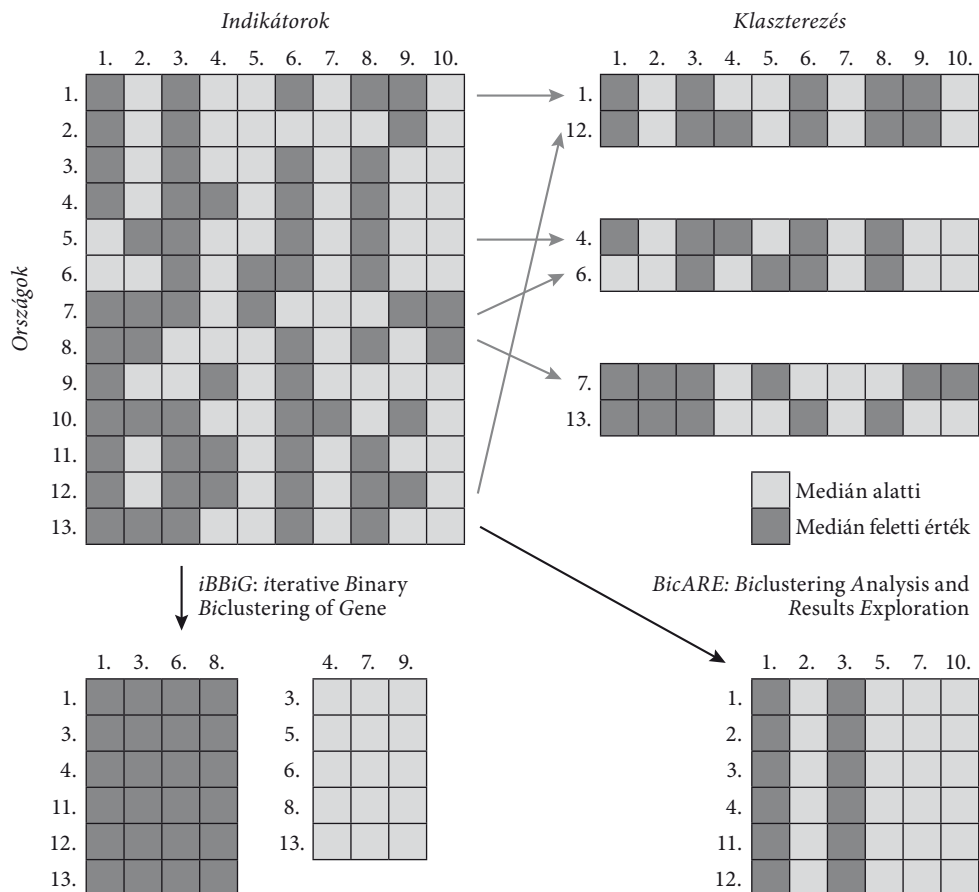
$$\text{score} = f_{\alpha, \tau}(\mathbf{B}) := (1 - H_{\mathbf{B}})^{\alpha} \begin{cases} \sum_i \sum_j [\mathbf{B}]_{i,j}, & \text{ha } M(\mathbf{B}) > \tau \\ 0, & \text{ha } M(\mathbf{B}) \leq \tau \end{cases}, \quad (2)$$

ahol  $\mathbf{B}$  egy (rész)mátrix,  $M(\mathbf{B})$  a (rész)mátrix átlaga,  $H_{\mathbf{B}} = -M(\mathbf{B}) \times \log_2(M(\mathbf{B})) - (1 - M(\mathbf{B})) \times \log_2(1 - M(\mathbf{B}))$  a (rész)mátrix entrópiája,  $\tau \in [0, 1]$  egy előre meghatározott küszöbérték,  $\alpha \in [0, 1]$ .

Azért döntöttünk a fenti módszer mellett, mert egy megfelelő küszöbérték beállításával ez a módszer alkalmas lehet az úgynevezett topligák megtalálására, továbbá a normalizált adatok megfordításával szintén alkalmas a leszakadók azonosítására is (2. ábra).

## 2. ábra

Klaszterezési és biklaszterezési eljárások eredményei



Forrás: saját szerkesztés.

A másik általunk alkalmazott eljárás a BicARE (*Biclustering Analysis and Results Exploration*) (lásd *Gestraud és szerzőtársai* [2014]). Ez a módszer egy továbbfejlesztett változata az eredeti *Cheng–Church* [2000] algoritmusnak, valamint a *Yang és szerzőtársai* [2003] által javasolt FLOC (*FLexible Overlapped biClustering*) eljárásnak. Ezzel a módszerrel lehetőségünk van olyan részmatrixok kiválasztására, amelyek országai között az adott indikátorokat tekintve a korreláció maximális (2. ábra), vagy épp a részmatrixok cellaelemeinek varianciája minimális.

A BicARE módszer alkalmas lehet a középmezőny azonosítására, hiszen a középmezőny országaira jellemző, hogy azok bizonyos indikátorok szerint jól, míg mások szerint rosszul teljesítenek. Bár a középmezőnyben előfordulhatna az is, hogy az oda sorolt országok minden szempontból közepes teljesítményt nyújtanak, azonban jelen vizsgálatunk eredményeiben ide olyan országok sorolódtak, amelyek bizonyos indikátorok szerint jól, míg mások szerint rosszul teljesítenek. A kérdés, hogy vannak-e e tekintetben hasonló országok, országcsoportok.



## A vizsgált adatbázis és az alkalmazott módszerek bemutatása

Vizsgálatunk során az U21 országgrangsorának 2014-es adatait vettük alapul (*Williams és szerzőtársai* [2014]), amely a 2. táblázatban látható. Tettük ezt azért, mert ez az az év, ahol valamennyi részindikátor hozzáférhető volt. A későbbi évekre vonatkozóan már csak a végső rangsor (*Williams és szerzőtársai* [2015]) vagy az ehhez figyelembe vett négy indikátorcsoport aggregált értékei érhetők el (*Williams és szerzőtársai* [2016], [2017]), de ennek további részletei már nem.

### 2. táblázat

U21 rangsorban szereplő országok listája

Országgrangsor				
1–10.	11–20.	21–30.	31–40.	41–50.
1. Egyesült Államok	11. Norvégia	21. Dél-Korea	31. Lengyelország	41. Argentína
2. Svédország	12. Ausztria	22. Tajvan	32. Görögország	42. Thaiföld
3. Kanada	13. Belgium	23. Spanyolország	33. Chile	42. Ukrajna
3. Dánia	14. Németország	24. Portugália	34. Szerbia	44. Horvátország
5. Finnország	15. Hongkong	25. Szlovénia	35. Kína	45. Dél-Afrika
6. Svájc	16. Új-Zéland	26. Csehország	35. Oroszország	46. Mexikó
7. Hollandia	17. Írország	27. Olaszország	37. Szlovákia	47. Törökország
8. Nagy-Britannia	18. Franciaország	28. Malajzia	38. Brazília	48. Indonézia
9. Ausztrália	19. Izrael	29. Magyarország	39. Románia	49. Irán
10. Szingapúr	20. Japán	30. Szaúd-Arábia	40. Bulgária	50. India

Forrás: *Williams és szerzőtársai* [2014].

Az (1) egyenlet alapján normalizált adatok segítségével képeztünk ligákat, majd a liga országait a ligában szereplő indikátorok alapján hasonlítottuk össze, és a ligákon belül képeztünk parciális rangsorokat. Külön vizsgáltuk a biklaszterek átfedéseit is, melyek további részmatrixokat adtak meg, melyek elemzése további érdekes összefüggésekre derített fényt.

### Felhasznált adatok bemutatása

Az U21 országgrangsora 50 ország felsőoktatási rendszerét vizsgálja, amelyek a 2. táblázatban a 2014. évi rangsornak felelnek meg. Az U21 összesen négy csoportba sorolja azt a 24 változót, amelyek szerint az országokat vizsgálja. A változók csoportjai a következők: erőforrások (*R, resources*), környezeti és civilizációs változók (*E, environment*), kapcsolatok (*C, connectivity*) és eredmények, kimenetek (*O, output*). Ezek leírását, valamint az összesített rangsor képzésénél figyelembe vett súlyaikat (*w*) a 3. táblázat tartalmazza.

## 3. táblázat

## Indikátorok és azok súlyainak listája

w (százalék)	Rövidítés	A változók megnevezése
<b>ERŐFORRÁSOK (20 százalék)</b>		
5,0	R1	A felsőoktatásra fordított kormányzati kiadás a GDP százalékában
5,0	R2	A felsőoktatásra fordított összes kiadás a GDP százalékában
5,0	R3	Éves összes kiadás dollárban/hallgató (nappali tagozatos egyenértékben számolva)
2,5	R4	Felsőoktatás K + F kiadásai a GDP százalékában
2,5	R5	Felsőoktatás K + F kiadásai/népesség
<b>KÖRNYEZETI ÉS CIVILIZÁCIÓS VÁLTOZÓK (20 százalék)</b>		
2,0	E1	Női hallgatók aránya
2,0	E2	Női oktatók aránya a felsőoktatásban
2,0	E3	Adatszolgáltatás minősége
14,0	E4	Politikai és szabályozási környezet
<b>KAPCSOLATOK (20 százalék)</b>		
4,0	C1	Nemzetközi hallgatók aránya
4,0	C2	Nemzetközi együttműködésben készült cikkek aránya
2,0	C3	Webometrics átláthatósági ( <i>web transparency</i> ) mutató/népesség
2,0	C4	Webometrics láthatósági ( <i>visibility</i> ) index: az egyetemek honlapjaira mutató külső linkek száma/népesség
4,0	C5	Tudástranszfer az egyetemek és az üzleti vállalkozások között
4,0	C6	Egyetemi és ipari szereplők közösen publikált cikkeinek aránya
<b>EREDMÉNYEK, KIMENETEK (40 százalék)</b>		
13,3	O1	A felsőoktatási intézmények által publikált cikkek száma
3,3	O2	Népességre vetített publikációk száma
3,3	O3	Hivatkozások átlagos száma a Karolinska Intézet normalizált impaktfaktora alapján
3,3	O4	A Sanghaj TOP 500-ban szereplő intézmények pontjaival súlyozott átlag/népesség
3,3	O5	Az ország 3 legjobb egyetemének Sanghaj-rangsorban szereplő pontértéke
3,3	O6	Felsőfokú továbbtanulási ráta
3,3	O7	A 24–64 éves korosztályban a diplomások aránya
3,3	O8	Kutatók száma/népesség
3,3	O9	Diplomás munkanélküliek aránya a 25–64-es korosztályban/ a diplomával nem, csak középfokú végzettséggel rendelkezők munkanélküliségi rátája

Forrás: Williams és szerzőtársai [2014].

Az egyes változókra számított értékeket már az U21 is normalítja oly módon, hogy a legnagyobb értéket tekintik 100-as pontértéknek, majd a további pontértékeket ennek arányában számolják ki. Az egyes változócsoportokra vonatkozóan ( $R, E, C, O$ ) a súlyok felhasználásával kiszámítják az országok pontértékét, majd ezeket szintén úgy arányosítják, hogy a legnagyobb érték 100 legyen. Az összesített pontértéket a változócsoportok szerinti értékek súlyozott összegeként határozzák meg úgy, hogy a legnagyobb pontérték itt is 100 legyen. A pontértékek sorba állításával keletkezik egy sorrend, ahol az egy tizedesjegybeli eltérést már nem különböztetik meg, hanem a két ország rangszámát azonosnak tekintik. Az összehasonlíthatóság érdekében ezt a mechanizmust a ligákon képzett részrangsorok esetében mi is követtük. Bár már az U21 is maximalizálja a változónként elérhető pontértéket, az általunk alkalmazott módszerek megkövetelték a normalizálást, így valamennyi változó esetén az (1) egyenlet szerinti min–max normalizálást követtük.

### *Az alkalmazott módszerek bemutatása*

Tanulmányunkban három típusú ligát keresünk. Valamennyi liga egy részmatrixnak felel meg. Az elit liga megtalálásához olyan részmatrixot keresünk, amely elemeknek átlaga egy adott küszöbérték ( $\tau$ ) felett van. Egyes cellaelemek akár e küszöbérték alatt is elhelyezkedhetnek, de mivel a (2) egyenlet olyan részmatrixokat keres, ahol a cellaértékek entrópiája minimális,  $\alpha$  növelésével homogénebb, de kisebb részmatrixokat kaphatunk. Ha számunkra elegendő, hogy a ligák országai az adott indikátorok alapján a többi országnál jobban teljesítsenek, akkor  $\alpha$  értéke mérsékelhető. Az elit ligát a továbbiakban  $A$  ligának nevezzük. Jelen vizsgálatunkban különböző  $\alpha$ -értékek választása nem eredményezett lényegesen különböző országcsoportokat. Éppen ezért azt az  $\alpha$ -értéket választottuk, ahol a (2) egyenlettel meghatározott pontérték maximális. Ekkor a cél az volt, hogy találjuk meg azt a legnagyobb ligát, amelybe azok az országok tartoznak, amelyeknek teljesítménye az adott indikátorokat tekintve meghalad egy  $\tau$  küszöbértéket. Bár az alacsony  $\alpha$  megválasztásával az entrópia minimalizálása kevésbé volt hangsúlyos, az eredmények azt mutatták, hogy a kiválasztott részmatrixokra vonatkozó varianciák kisebbek, mint a ki nem választott elemek varianciája.

A normalizált adatok ellentettjének (1 *minusz* a normalizált adat) vizsgálatával a leszakadó országok, valamint azon indikátorok, amelyekben ezek az országok egy meghatározott érték alatt teljesítenek, szintén meghatározhatók. Az így kapott ligát a továbbiakban  $C$  ligának nevezzük. Az  $A$  és a  $C$  liga meghatározásához az adattáblát normalizálás után még binarizálni is kell. A binarizálás során 1 értéket kapnak a küszöb ( $\tau$ ) feletti, 0 értéket a küszöb ( $\tau$ ) alatti értékek.

A 3. *ábra* mutatja, hogy hogyan változik a részmatrix mérete a küszöbérték megválasztásának függvényében. Még a rangsor elején szereplő, a biklaszterezés során az  $A$  ligába sorolt országok sem teljesítenek minden indikátorban jól (egy adott küszöbérték felett).

3. ábra

Különböző küszöbértékekre kapott ligák



Forrás: saját szerkesztés.

Ugyanígy elmondható, hogy a  $C$  ligában szereplő országok nem mindegyik indikátorukat tekintve maradnak el a  $\tau$  küszöbként választott 0,5, vagy 0,75-os értéktől. A biklaszterezési eljárások megengedik, hogy az egyes részmatrixok átfedjék egymást, természetesen a  $\tau$  küszöbérték vagy az  $\alpha$ -érték növelésével ez az átfedés megszüntethető, ugyanakkor az átfedő részmatrixok megmutatják, hogy az egyes országok bizonyos indikátorukat tekintve jól, míg másokat tekintve rosszabbul teljesítenek.

A középmezőny vizsgálatára éppen ezért olyan eljárást kerestünk, amely az országok adott indikátorok szerinti korrelációjának maximalizálásával határoz meg részmatrixokat. Ezek az országok tartalmazhatnak küszöb feletti és alatti értékeket is, ugyanakkor az országorrelációk maximalizálása miatt a küszöb feletti és alatti értékek elhelyezkedése hasonló mintát követ, vagyis várhatóan ezek az országok ugyanazon változók szerint fognak jól vagy kevésbé jól teljesíteni. Az így kapott ligát nevezük a továbbiakban  $B$  ligának. A  $B$  liga számításához csak normalizálásra van szükség, az adatok binarizálása nem szükséges.

Összefoglalva, a biklaszterezési módszerek alkalmazásához a következő lépéseket kell követni:

1. LÉPÉS: Adatok változók szerinti normalizálása.

⇒  $B$  ligák meghatározása a BicARE-módszer segítségével.

2. LÉPÉS: adatok binarizálása egy meghatározott  $\tau$  küszöb szerint.

⇒  $A$  ligák meghatározása az iBBiG-módszer segítségével.

3. LÉPÉS: normalizált adatok ellentettjének (1 mínusz a normalizált adat) binarizálása egy meghatározott  $\tau$  küszöb szerint.

⇒  $C$  ligák meghatározása az iBBiG-módszer segítségével.

## Eredmények

Az eredmények ismertetése során először bemutatjuk a vizsgált adattáblát, majd a hiányzó elemek kezelésének leírása után kitérünk arra, hogy miért volt szükség a normalizálásra, különösen az  $A$  és  $C$  ligák meghatározásához. Ezt követően illusztráljuk a kapott ligákat, és részletesen elemezzük a ligák átfedéseiként kapott területeket is. Végül részrangsorokat javasolunk a ligákra, és összehasonlítjuk azokat az eredeti rangsorokkal.

A következőkben bemutatott eredmények és a hozzájuk tartozó  $R$  kódok megtekinthetők a mellékelt HTML-állományban.<sup>13</sup> Elérhetővé teszünk további két kiegészítő fájlt, amelyek egyike a hagyományos modellredukciós módszerekkel (kétlépcsős klaszterezéssel) kapott eredményeket mutatja,<sup>14</sup> a másikban pedig az látható, hogy miként módosulnának a ligák, ha további indikátorokkal bővítjük a vizsgálatot.<sup>15</sup>

<sup>13</sup> [https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/BIC/U21\\_BIC\\_HU.html](https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/BIC/U21_BIC_HU.html).

<sup>14</sup> [https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/TWO-STEP-CLUST/U\\_21\\_BICPCA\\_HU.html](https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/TWO-STEP-CLUST/U_21_BICPCA_HU.html).

<sup>15</sup> [https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/BIC\\_U21\\_AND\\_GCI/U21\\_AND\\_MORE\\_BIC\\_HU.html](https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/BIC_U21_AND_GCI/U21_AND_MORE_BIC_HU.html).

A ligák meghatározása után minden egyes liga esetében stabilitásvizsgálatot végeztünk mind az országokra, mind pedig az indikátorokra vonatkozóan. A Bootstrap-elemzés alapján a bemutatott összes biklaszter stabil. A Bootstrap-vizsgálat eredményei szintén a mellékelt HTML-fájlban olvashatók.<sup>16</sup>

### *Leíró statisztikák*

A 4. táblázat összefoglalja a legfontosabb statisztikai eredményeket.

#### *4. táblázat*

Leíró statisztikák ( $N=50$ , Max = 100,0)

Változók	Hiányzó értékek	Min	Terjedelem	Átlag	Relatív szórás (százalék)
R1	0	22,1	77,9	47,5	34,9
R2	0	25,0	75,0	55,2	30,9
R3	0	3,8	96,2	44,7	54,1
R4	6	2,8	97,2	40,5	57,7
R5	6	0,3	99,7	36,4	82,1
E1	0	79,3	20,7	98,6	4,1
E2	11	35,8	64,2	82,6	16,4
E3	0	68,2	31,8	94,1	7,9
E4	0	53,4	46,6	80,6	13,9
C1	3	0,5	99,5	27,1	96,6
C2	0	22,5	77,5	64,2	30,6
C3	0	4,0	96,0	34,7	73,4
C4	0	2,8	97,2	34,1	72,2
C5	3	27,1	72,9	63,5	31,5
C6	0	0,1	99,9	43,7	58,8
O1	0	0,1	99,9	8,3	196,2
O2	0	0,1	99,9	42,6	74,7
O3	0	23,4	76,6	61,2	32,7
O4	0	0,0	100,0	25,2	109,9
O5	0	0,0	100,0	19,5	95,4
O6	3	23,1	76,9	64,2	28,3
O7	1	6,7	93,3	50,9	45,2
O8	2	1,2	98,8	38,7	65,2
O9	6	33,5	66,5	62,1	26,5

*Forrás:* saját szerkesztés.

<sup>16</sup> [https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/U\\_21\\_HU.html](https://kmt.gtk.uni-pannon.hu/kutatas/u21/HU/U_21_HU.html).

A táblázat második oszlopában látható, hogy viszonylag kevés hiányzó elemmel dolgozhattunk ( $41/(24 \times 50) = 3,42$  százalék). A legtöbb hiányzó elem a női oktatók felsőoktatásbeli aránya ( $E2$ ) változó esetén jelent meg. Az alkalmazott biklaszterezési eljárások képesek kezelni a hiányzó elemeket (*Jiong és szerzőtársai* [2003]), ugyanakkor mivel mind a változócsoportok pontértéke, mind az összesített pontérték rendelkezésre állt, ezekből visszafejthetők voltak a hiányzó adatok (ez csak akkor lehetséges, ha egy változócsoporton belül legfeljebb egy pontérték hiányzik), így esetünkben nem volt szükség a hiányzó adatok pótlására. Bár maga az U21 is 100-as értéket ad a legmagasabb pontértéknek, ezután pedig ezzel arányosítja az összes többi ország e változó szerinti pontértékét, látható, hogy a terjedelmet tekintve jelentős különbségek mutatkoznak. A legkisebb terjedelmű változó, amely szerint a legtöbb ország hasonló értékekkel rendelkezik, a női hallgatók aránya ( $E1$ ), itt a legkisebb a relatív szórás is, mindössze 4,1 százalék. A legnagyobb relatív szórás a publikált cikkek számában mutatkozik meg ( $O1$ : 196,2 százalék), ahol ráadásul az átlagos pontérték (8,3) is extrém alacsony. Az ennyire eltérő terjedelmű változókat biklaszterezési eljárásokkal csak normalizálás után tudjuk összehasonlítani.

#### *Ligák és átfedések ameghatározása*

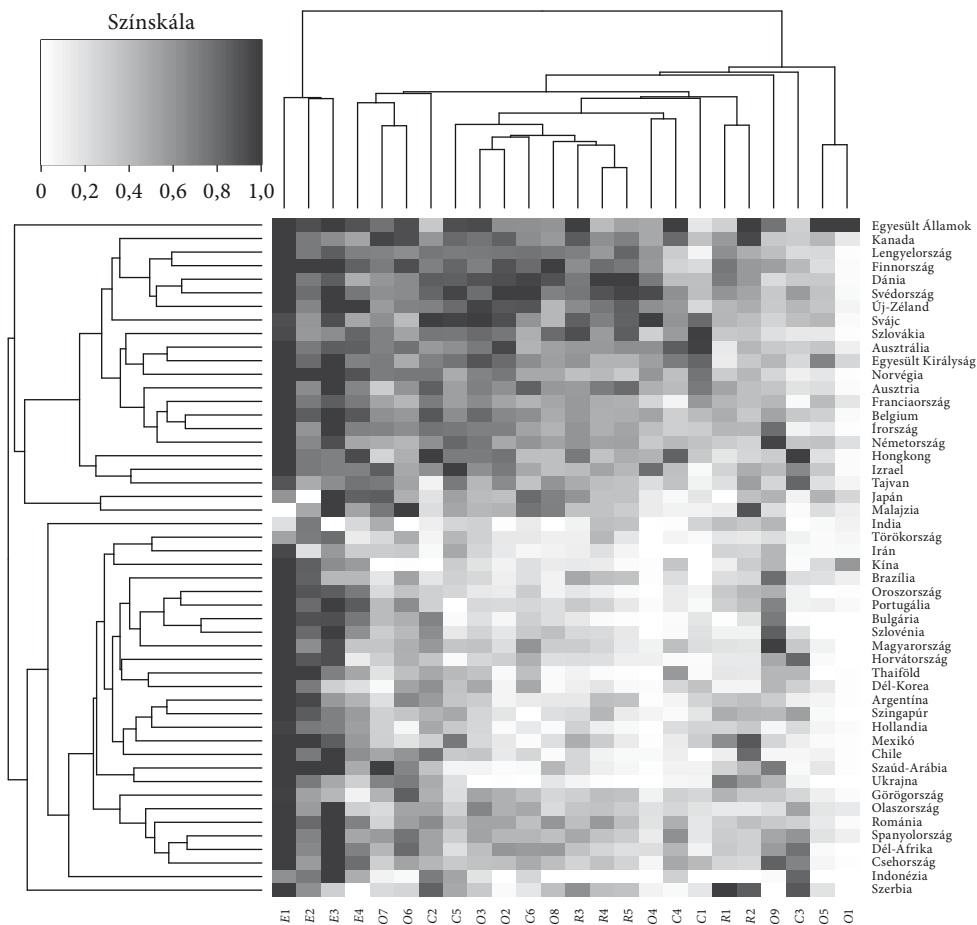
A biklaszterezési eljárások esetén is meg kell határoznunk, hogy hány klasztert keresünk. Ebben segíthet, ha az adatokat úgy rendezzük át, hogy a hasonló értékkel rendelkező cellák egymáshoz minél közelebb kerüljenek (ez az úgynevezett szeriációs eljárás). Maga ez a probléma is már NP-nehéz (*Hahsler és szerzőtársai* [2008]), ugyanakkor az így kapott ábra jó kiindulópontként szolgálhat a klaszterszámok becslése során (4. ábra).

A 4. ábra sötétebb foltjai jelölik a magasabb értékeket, a világosabbak pedig az alacsonyabbakat. Ezek alapján egy  $A$  és egy  $C$  liga azonosítható. Látható, hogy a sötétebb és világosabb mezők nemcsak az országok tekintetében, hanem az indikátorok tekintetében is elkülönülnek. Ez azt jelenti, hogy az  $A$  ligában szereplő országok részben más változók szerint teljesítenek jobban (küszöbszint felett), mint amely változók szerint a  $C$  ligában szereplő országok rosszabbul (küszöbszint alatt).

Bár a normalizált adatok átrendezése után kapott 4. ábra sejteti, hogy legalább egy  $A$  és egy  $C$  liga található a normalizált adatokon, mégis a biklaszterezési eljárással kapjuk csak meg ezeket a részmátrixokat, valamint a részmátrixok meghatározása után kapjuk meg, hogy a sorokra (országokra), valamint oszlopokra (változókra) számolt, a biklaszterbe bevont és onnan kihagyott cellaértékeket összevető  $F$ -statisztika – amelyet a szakirodalom sor-, illetve oszlophatásnak (*row and column effect*) nevez, *Cheng–Church* [2000] – szignifikáns különbséget mutat-e, vagy sem. Az eredmények azt mutatták, amit a 4. ábra alapján még csak sejthettünk, hogy két-két  $A$ , illetve  $C$  liga határozható meg. Ezt támasztja alá az 5. táblázat, amelyben az látható, hogy az  $A$  és a  $C$  ligán belül egy, illetve két ligát keresve: egy liga szignifikáns (a hozzá tartozó  $p$ -értékek mind a sorokra, mind az oszlopokra az 5 százalékos szignifikanciaszint alatt vannak), viszont két liga esetében már csak az oszlopok szignifikánsak, a sorok nem (a hozzájuk tartozó  $p$ -érték nagyobb 0,05-nál).

4. ábra

Rendezett, normált U21-es adattábla



Forrás: saját szerkesztés.

5. táblázat

Pontértékek, oszlop- és sorhatások ( $\alpha = 0,08, \tau = \text{medián}$ )

Liga	Sor-szám	Pontérték	Országok száma	Változók száma	F-próba	
					oszlop-	hatás
					(p-értékek)	
A	1.	287,3794	23	19	0,0000	0,0000
A	2.	78,8670	22	5	0,1866	0,0000
C	1.	535,1661	38	19	0,0000	0,0000
C	2.	52,1089	11	7	0,9350	0,0000

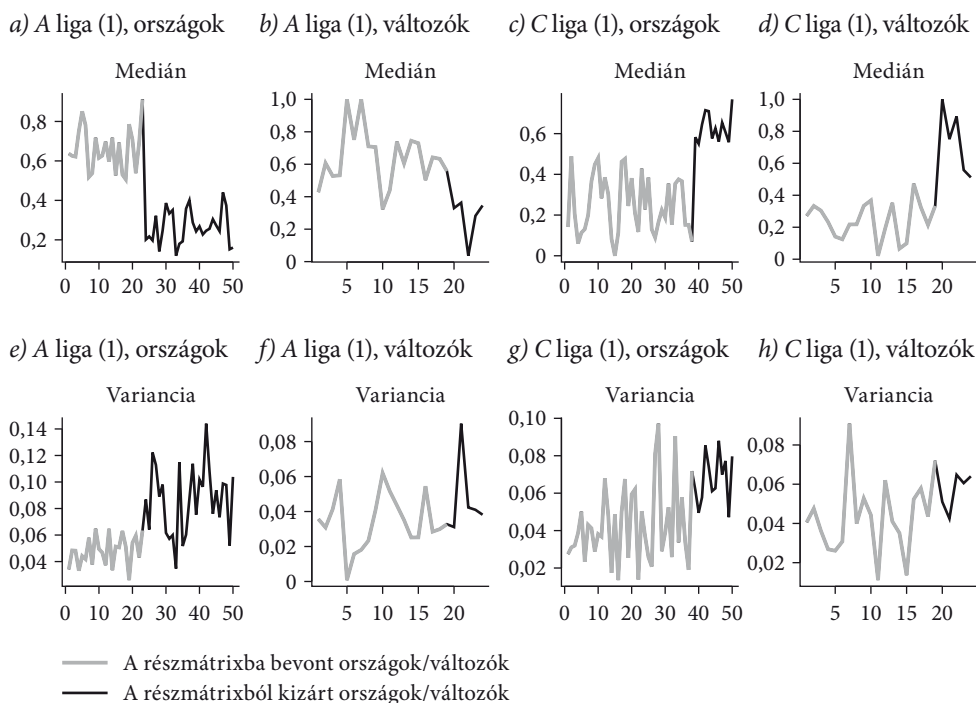
Forrás: saját szerkesztés.



Az egyes ligákban levő, valamint az onnan kimaradó országok közötti különbségek érzékeltetésére célszerű a sorok (az országok adatainak) és az oszlopok (az indikátorok adatainak) mediánját és variációját megrajzoltatni (lásd 5. és 6. ábra). Az A ligában levő országok és indikátorok mediánja magasabb, de variációjája alacsonyabb, és lényegesen kisebb változékonyságot mutatnak, mint a kimaradó országok. A C ligában levő országoknak és indikátoroknak pedig alacsonyabb a mediánjuk és variációjuk. A B ligába sorolt országokról és indikátorokról ugyanez mondható el. Az 5. ábra szemlélteti, hogy a szignifikáns ligákba bevont és onnan kimaradt változók/országok célleértékeinek mediánjai, illetve variációi milyen mértékben különböznek egymástól.

5. ábra

Az A és a C ligákra vonatkozó sor/oszlop mediánok [a)–d)] és variációk [e)–h)]



Forrás: saját szerkesztés.

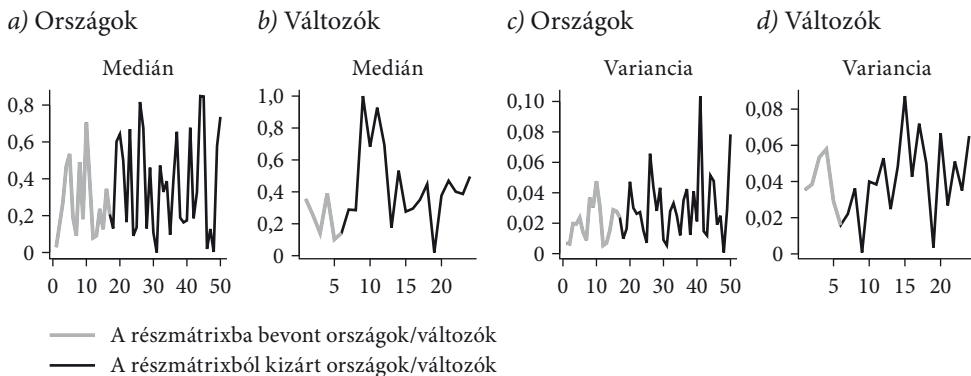
A B ligák kiválasztása esetén is egyetlen szignifikáns biklasztert kaptunk, amely hat változót és 17 országot tartalmazott. Ez a módszer jellegéből adódóan olyan ligát határoz meg, ahol a változók szerinti variancia minimális (6. ábra).

Az egyes ligákba tartozó országok és indikátorok átfedhetnek. A 7. ábrán mind az országokat, mind az indikátorokat jelöltük. Az egyes átfedéseket tartalmazó részmatrixokat is elneveztük a könnyebb hivatkozás érdekében. Az érdekesebbnek ítélt eredményeket külön kiemeljük.

Az 1. táblázat kapcsán említettük, hogy léteznek nemzetközi regionális rangsorok (bár nem az országok felsőoktatási rendszerének egészére, amelyet mi vizsgáltunk, hanem

6. ábra

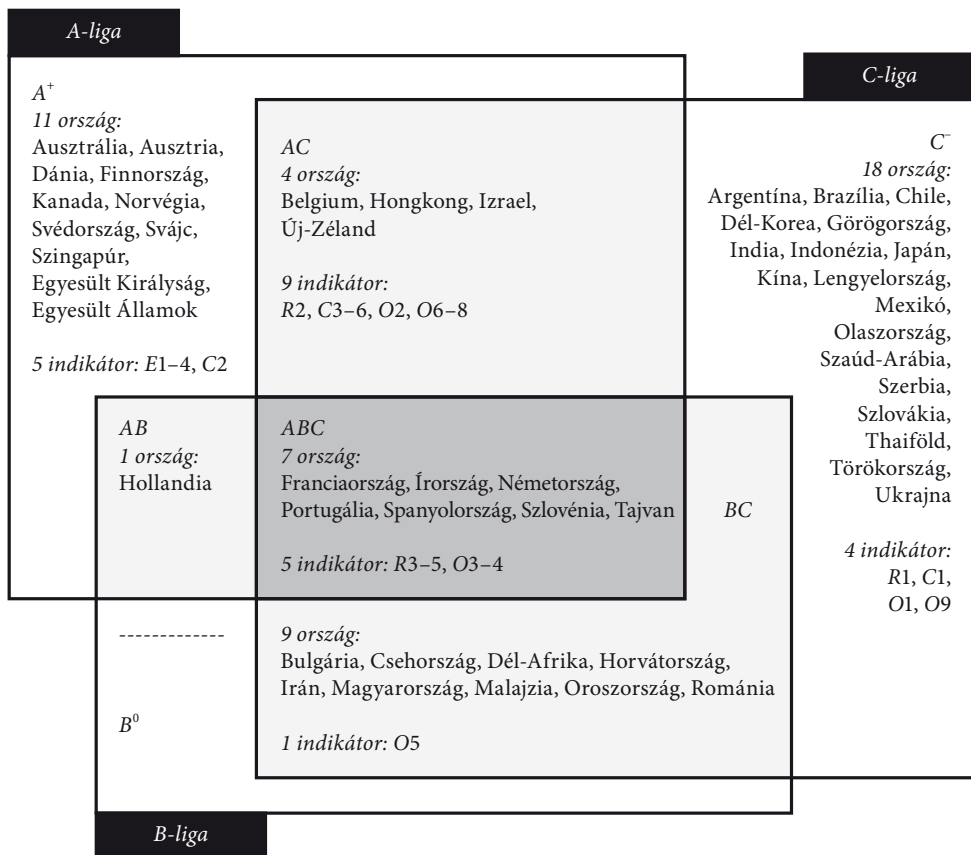
A B ligára vonatkozó sor/oszlop mediánok [a)–b)] és varianciák [c)–d)]



Forrás: saját szerkesztés.

7. ábra

Ligák és a ligák közötti átfedések



Forrás: saját szerkesztés.

a régió egyes egyetemeire). Ezeket a régiókat azonban előre meghatározott gazdasági vagy földrajzi elv alapján képzik (ilyen például a QS-nek a BRICS-országokra vonatkozó rangsora), amely „elvéle elrendeléstől” el kívántunk tekinteni. Módszerünk alapján a BRICS-országok sem teljesen ugyanabban a kategóriában szerepelnek. Bár mindegyik a legkedvezőtlenebb C ligának a tagja, ezen belül csak Brazília, India és Kína sorolható egyértelműen azon országok közé, amelyek a felsőoktatási rendszerük alapján a legrosszabbul teljesítenek ( $C^-$  részmátrix). Viszont a BRICS országok további két tagja – Oroszország, valamint Dél-Afrika – bizonyos indikátoraik alapján már a középmezőnybe is besorolhatók.

A liga: az A liga összesen 23 országot és 19 indikátort tartalmaz. A 23 ország az U21 rangsorában is az élmezőnyben, az első 25 hely valamelyikén helyezkedik el (2. táblázat). A kimaradt változókból nem meglepő a tudományos folyóiratcikkek száma (O1), hiszen ott az átlagos pontérték a legkisebb, a relatív szórás pedig a legnagyobb volt, ami azt mutatja, hogy a magas publikációs teljesítmény csak néhány országra jellemző. Hasonló okok miatt maradt ki az adott ország legjobb egyetemeinek a Sanghaj-rangsor alapján számolt pontértéke (O5) is. Az itt kiválasztott országok esetében is megfigyelhető nagyobb relatív szórás miatt maradt ki az A liga indikátorai közül a diplomás munkanélküliek aránya (O9), a GDP-arányos kormányzati kiadások (R1), valamint a nemzetközi hallgatók aránya (C1). Ez azt jelenti, hogy az A ligában szereplő 23 ország többsége ezekben nem teljesít egységesen jól.

**A<sup>+</sup> részmátrix:** 11 ország csak az A ligába tartozik. Így őket A<sup>+</sup>-szal jelöltük. Nem meglepő módon e 11 ország az U21-es rangsor első 12 helyét foglalja el. Az A<sup>+</sup> részmátrixba 5 indikátor tartozik. Az E1–E4 indikátor (környezeti indikátorok) majdnem minden vizsgált országban magas, ugyanakkor csak ezekre az A<sup>+</sup> országokra jellemző, hogy a nemzetközi együttműködésben készült cikkek aránya (C2) magas. A vizsgált 11 országra vonatkozó U21-es pontérték (C2) átlaga több mint 30 százalékkal magasabb (78,6), mint az összes országra vonatkozó átlag (60,2).

**C liga:** a C liga tartalmazza a legtöbb országot (38) és azokat az indikátorokat (19), amelyek nem tartoznak bele az A<sup>+</sup> részmátrixba. Ez a liga – az országok számát tekintve – nagyobb, mint az A liga, ami azt mutatja, hogy a küszöb alatt (esetünkben medián) teljesítő országból több van. Az A ligába tartozó indikátorok száma megegyezik a C ligába tartozó indikátorok számával (19), és közülük 14 indikátor közös. Az A ligába tartozó országok jól teljesítenek a környezeti indikátorokban (E1–4) és a nemzetközi együttműködésben született publikációk számában (C2). A C ligába tartozó országok azonban általában rosszul teljesítenek a kormányzati kiadások terén (R1), a nemzetközi hallgatók számában (C1), a folyóiratcikkek számában (O1), a Sanghaj-rangsorban elért pozíciójukban (O5) és a munkanélküliségi rátában (O9) is.

**C<sup>-</sup> részmátrix:** a C liga azon részét nevezzük C<sup>-</sup> ligának, amely nincs átfedésben más ligákkal. A C<sup>-</sup> liga 18 országot és 4 indikátort tartalmaz, amelyek közül egy indikátor az erőforrások közé (R1), egy a kapcsolatokhoz (O1) és kettő a kimeneti indikátorok (O1, O9) közé tartozik. Egyetlen környezeti indikátor sem került bele ebbe a ligába, aminek az oka az, hogy a ligában levő 18 ország relatíve magas értékkel rendelkezik ezekből a mutatókból. A 18 ország az U21-es rangsor közepén (20., 21. és 27. helyen) és a rangsor utolsó 20 helyén helyezkedik el. Az A<sup>+</sup> ligával ellentétben a C<sup>-</sup> liga erőforrás-jellegű indikátort is tartalmaz (R1).

**AC és ABC részmatrixok:** az  $A$  és  $C$  halmaz metszete tartalmazza a felsőoktatás tekintetében a legfontosabb indikátorokat. Ide sorolható az öt erőforrás-indikátor közül négy ( $R2-5$ ). Tehát az erőforrások többsége meghatározó mind az  $A$ , mind a  $C$  ligában. Egy ország  $A$  ligába kerüléséhez magas értékkel kell rendelkeznie e négy erőforrás tekintetében, függetlenül attól, hogy a kormányzat a GDP mekkora hányadát költi felsőoktatásra ( $R1$ ). Ha a felsőoktatásra fordított kormányzati kiadás a GDP arányában ( $R1$ ) alacsony, és ezt kevés nemzetközi hallgató ( $C1$ ), valamint magas munkanélküliségi ráta ( $O9$ ) kíséri, akkor ez a kombináció az adott országot rosszabb helyzetbe juttatja, azaz a  $C^-$  részmatrix felé tereli.

Az  $AC$  részhalmazban szereplő országok azért kerültek ide, mert bizonyos indikátorokban jól, míg másokban kevésbé jól teljesítenek. Az  $ABC$  ligába hét ország és öt indikátor tartozik. Ezek az országok és indikátorok, egyszerre vannak jelen mindhárom ligában. Az öt indikátor közül három erőforrás- (= input-) indikátor ( $R3-5$ ), kettő pedig outputindikátor ( $O3-4$ ), így akár az 50 országra vonatkozóan ez az öt változó szerinti input-output elemzés is elvégezhető.

**B liga:** a középmezőnyt reprezentáló  $B$  liga tartalmazza a legkevesebb országot, szám szerint 17-et, és mindössze hat indikátort. Érdekes, hogy a hat indikátor közül itt is három erőforrás- ( $R3-5$ ) és három outputmutató ( $O3-5$ ) található. Mivel az erőforrások inputoknak is tekinthetők, így a  $B$  liga országait egy input-output elemzéssel is meg lehet vizsgálni. A 17 ország – Hollandiát kivéve – az U21-es rangsor közepén és végén helyezkedik el (14–49. helyezés között). Ez az eredmény azt mutatja, hogy az  $A$  liga jobban elkülönül a középmezőnytől, mint a  $C$  liga. Az alkalmazott módszer (BicARE) azokat az országokat és indikátorokat sorolta ebbe a ligába, amelyek nagyon hasonlóak egymáshoz valamilyen tekintetben.

**$B^0$  és AB részmatrixok:** a  $B^0$  ligában egyetlen ország sem található, míg az  $AB$  ligába egy ország került (Hollandia). Ezzel szemben a  $BC$  ligába kilenc ország került, ami szintén jól szemlélteti, hogy mekkora rés tátong a topliga és a másik kettő között.

**BC részmatrix:** a  $B$  és  $C$  liga közötti átfedés kilenc országot és egy indikátort tartalmaz, az ország három legjobb egyetemének Sanghaj-rangsorban szereplő pontértékét ( $O5$ ), amely outputjellegű mutató. Ha egy ország jól teljesít ebben a mutatóban, akkor jobb ligába kerülhet. Ha egy adott ország jobban tud teljesíteni az  $O5$  mutató tekintetében, akkor a  $C$  ligából feljebb kerülhet a  $B$  ligába. Idetartozik Magyarország is, amely a három legjobb egyetemének Sanghaj-rangsorban szereplő pontértékében ( $O5$ ) meglehetősen szerény teljesítményt mutat.

### *Parciális rangsorok meghatározása*

Az egyes ligákhoz tartozó – az eredeti 24-nél kevesebb – indikátorból az U21 által alkalmazott módon állítottuk elő az egyes ligába tartozó országok (parciális) rangsorát. Ehhez az eredeti U21-súlyokat használtuk, és követtük az U21-nek azt a módszerét is, amely szerint kategóriánként ( $E$ ,  $R$ ,  $C$ ,  $O$ ) a legnagyobb értéket 100-nak tekintettük, majd a többi értéket ehhez arányosítottuk, és így számoltuk ki a kategóriák pontszámainak súlyozott összegéből a végső pontszámot, amely a rangsort adja.

A 6. táblázatban a ligán belül számolt rangsorokat vetítettük vissza az U21 rangsoraira, és azt vizsgáltuk, hogy az eredeti rangsorhoz képest a parciális rangsorok mekkora eltérést mutatnak.

Bár a 6. táblázatban látható, hogy az U21 rangsorához képest átrendeződés tapasztalható a ligákon belül, a három ligán belüli rangsor és az U21 rangsora szoros korrelációt mutat (7. táblázat).

6. táblázat

Parciális rangsorok meghatározása

	Rangsor U21	A ligán belüli rangsor		B ligán belüli rangsor			C ligán belüli rangsor			
		U21	liga	különbség	U21	liga	különbség	U21	liga	különbség
Egyesült Államok	1.	1.	3.	-2						
Svédország	2.	2.	1.	1						
Dánia	3.	3.	2.	1						
Kanada	3.	3.	6.	-3						
Finnország	5.	5.	4.	1						
Svájc	6.	6.	7.	-1						
Hollandia	7.	7.	5.	2	1.	1.	0			
Egyesült Királyság	8.	8.	11.	-3						
Ausztrália	9.	9.	9.	0						
Szingapúr	10.	10.	10.	0						
Norvégia	11.	11.	8.	3						
Ausztria	12.	12.	14.	-2						
Belgium	13.	13.	13.	0				1.	2.	-1
Németország	14.	14.	18.	-4	2	2.	0	2.	1.	1
Hongkong	15.	15.	12.	3				3.	8.	-5
Új-Zéland	16.	16.	16.	0				4.	5.	-1
Írország	17.	17.	17.	0	3.	4.	-1	5.	3.	2
Franciaország	18.	18.	19.	-1	4.	3.	1	6.	6.	0
Izrael	19.	19.	15.	4				7.	4.	3
Japán	20.							8.	7.	1
Dél-Korea	21.							9.	9.	0
Tajvan	22.	20.	20.	0	5.	7.	-2	10.	10.	0
Spanyolország	23.	21.	23.	-2	6.	6.	0	11.	11.	0
Portugália	24.	22.	21.	1	7.	5.	2	12.	13.	-1
Szlovénia	25.	23.	22.	1	8.	8.	0	13.	12.	1
Csehország	26.				9.	10.	-1	14.	14.	0
Olaszország	27.							15.	15.	0

## A 6. táblázat folytatása

	Rangsor	A ligán belüli rangsor			B ligán belüli rangsor			C ligán belüli rangsor		
	U21	U21	liga	különbség	U21	liga	különbség	U21	liga	különbség
Malajzia	28.				10.	12.	-2	16.	20.	-4
Magyarország	29.				11.	9.	2	17.	17.	0
Szaúd-Arábia	30.							18.	18.	0
Lengyelország	31.							19.	22.	-3
Görögország	32.							20.	16.	4
Chile	33.							21.	26.	-5
Szerbia	34.							22.	24.	-2
Kína	35.							23.	19.	4
Oroszország	35.				12.	14.	-2	23.	21.	2
Szlovákia	37.							25.	26.	-1
Brazília	38.							26.	23.	3
Románia	39.				13.	16.	-3	27.	31.	-4
Bulgária	40.				14.	17.	-3	28.	34.	-6
Argentína	41.							29.	30.	-1
Thaiföld	42.							30.	33.	-3
Ukrajna	42.							30.	28.	2
Horvátország	44.				15.	13.	2	32.	25.	7
Dél-Afrika	45.				16.	11.	5	33.	29.	4
Mexikó	46.							34.	36.	-2
Törökország	47.							35.	32.	3
Indonézia	48.							36.	38.	-2
Irán	49.				17.	15.	2	37.	35.	2
India	50.							38.	37.	1

*Megjegyzés:* a döntetlen helyek dőlt betűsek, a világosszürke hátrébb sorolást, a sötétszürke előrébb sorolást jelöl.

*Forrás:* saját szerkesztés.

## 7. táblázat

Parciális rangsorok korrelációja

	A liga			B liga			C liga		
	korrelációs együttható	$p$ -érték	$N$	korrelációs együttható	$p$ -érték	$N$	korrelációs együttható	$p$ -érték	$N$
Kendall-féle $\tau_B$	0,824	0,000	23	0,583	0,001	17	0,855	0,000	38
Spearman-féle $\rho$	0,956	0,000	23	0,785	0,000	17	0,966	0,000	38

*Forrás:* saját szerkesztés.

## Következtetések

A 6. táblázat azt mutatja, hogy a javasolt ligák közül az A liga valóban a fejlettebb oktatási rendszerrel rendelkező országokat tartalmazza, míg a C liga az U21-rangsor közepén és végén elhelyezkedőket. Ugyanakkor a javasolt módszer azt is megerősíti – amit a 4. ábrán is látni –, hogy az A liga országai nem minden területen teljesítenek a medián felett. Jó példa erre Norvégia, amely tagja az A ligának, sőt része az A<sup>+</sup> részmatrixnak is, ugyanakkor a nemzetközi hallgatók fogadásában a 17,8-es mediánhoz képest 7,4-es pontértékével jelentősen alulteljesít. Hasonlóképpen, Nagy-Britannia kormányzati kiadásai (32-es pontérték) is elmaradnak a mediántól (ami 44,8-es pontérték). Ugyanez igaz fordítva a C ligára is. Szaúd-Arábia és Indonézia a C liga tagjai, ugyanakkor a nemzetközi együttműködéssel írt cikkek számában (C2), amely nem része a C liga változóinak (87,5, valamint 78,2-es pontértékkel) a medián (67,8-es pontérték) felett teljesítenek.

A B liga ezzel szemben jóval színesebb országcsoporthoz mutat, ugyanakkor inkább az U21-es lista közepéről és végéről tartalmaz országokat. A B és C liga kialakulásában alapvető (a B liga kialakulásában kizárólagos) az erőforrások szerepe (illetve annak hiánya). Két outputtényező is fontos szerephez jut a B és C ligában: a publikációs teljesítmény, illetve a nemzetközi rangsorokban mérhető teljesítmény.

Az egyik legérdekesebb eredmény, hogy a felsőoktatásra fordított kormányzati kiadások GDP-n belüli arányának (R1) nem kell magasabbnak lennie ahhoz, hogy egy ország az élvonalba tartozzon [a topligába kerülést leginkább a nemzetközi társszerzős publikációk magas aránya (C2) biztosíthatja], viszont ezen indikátor alacsony volta szükséges ahhoz, hogy a legrosszabbak közt szerepeljen. Bár az R1 mutató nem szerepel az A liga indikátorai között, az R2 viszont igen, ezért a kormányzati kiadások arányának (R1) csökkentése – ezt ellentételező források hiányában –, további (C1, O1, O9) indikátorok alacsony szintje mellett, alacsonyabb ligába juttathat egy országot.

## Összefoglalás, további tervek

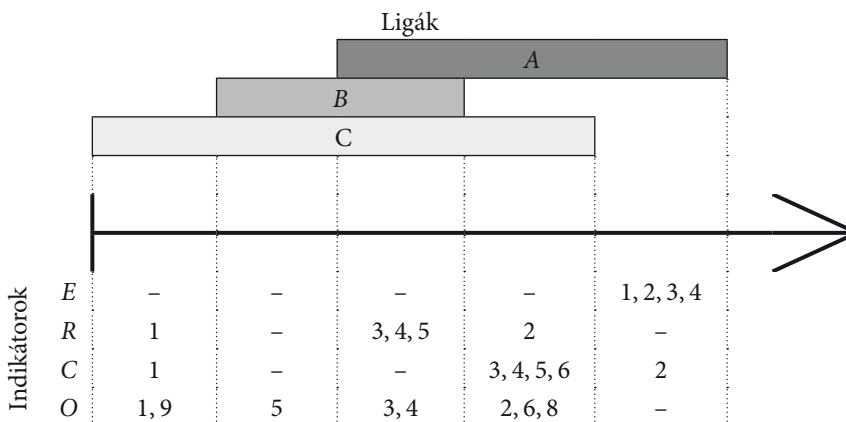
A bemutatott módszer egy fontos elv megvalósításához vihet közelebb, nevezetesen, hogy a hasonlót a hasonlóval hasonlítsuk össze. Az eddigieknél sokkal méltányosabb összehasonlításokat kaphatunk, ha az egy ligába tartozó országok csoportján képzünk rangsorokat. Továbbá érdemes az azonos ligában levő országokon hatékonyságelemzést is végezni.

A biklaszterezési eljárás jól meghatározott célfüggvény esetén segíthet a ligák meghatározásában. A ligák képzése során megengedtük az átfedés lehetőségét, hiszen így további érdekes információkat kaphattunk. A 8. ábra összefoglalja, hogy a négy változócsoporthoz közül melyek érintettek az egyes ligák kialakulásában. Ez az eredmény egyfajta fejlesztési irányt is kijelölhet.

Azok az országok, illetve indikátorok kerültek az A ligába, amelyek medián felettek, a C ligába azok, amelyek medián alattiak, a B ligába pedig azok, amelyeknek

## 8. ábra

Ligák és fejlődési lehetőségek



Forrás: saját szerkesztés.

alacsony a szórása. A 8. ábra az indikátorokat mutatja az egyes ligákban. A 7. ábra szerint Magyarország a B és a C ligában helyezkedik el, azaz legfeljebb a középmezőnybe sorolható, vagyis a B ligabeli indikátorokban homogén a többi ide sorolt országgal, és a C ligabeli indikátorokban medián alatti. Magyarország úgy kerülhet be az A ligába, hogy először az R3–5, illetve O3–4 mutatókat fejleszti, vagyis az egy hallgatóra jutó felsőoktatásra fordított (nem feltétlenül kormányzati) összeget kellene növelni (R3), valamint a felsőoktatás K + F kiadásait a GDP arányában (R4), illetve lakosságszámra vetítve (R5), továbbá a hivatkozások számát (O3), és előrébb kellene kerülnie a Sanghaj-rangsorban (O4). Fontos megállapításnak tartjuk, hogy a 6. táblázat alapján az A liga jobban elkülönül a középmezőnytől (a B ligától), mint a C liga.

A bemutatott módszer kellően általános ahhoz, hogy a természettudományok mellett a társadalomtudományok területén is sikerrel legyen alkalmazható. Tanulmányunkban a biklaszterezési eljárást ligák és részrangsorok meghatározására alkalmaztuk, de e módszer alkalmazása minden olyan esetben érdekes lehet, amikor egyidejűleg kell indikátorokat (változókat) és eseteket kiválasztani. Részrangsorok meghatározására ez a módszer nemcsak ország, hanem intézményi szinten is alkalmas, ha rendelkezésre állnak az indikátorok, részindikátorok értékei.

## Hivatkozások

- ABANKINA, I.–ALESKEROV, F.–BELOUSOVA, V.–GOKHBERG, L.–KISELGOFF, S.–PETRUSHCHENKO, V.–SHVYDUN, S.–ZINKOVSKY, K. [2016]: From equality to diversity: Classifying Russian universities in a performance oriented system. *Technological Forecasting and Social Change*, Vol. 103. Supplement C, 228–239. o. <https://doi.org/10.1016/j.techfore.2015.10.007>.
- ALBALATE, A.–MINKER, W. [2013]: *State of the Art in Clustering and Semi-Supervised Techniques*. Wiley–Blackwell, 2. fejezet, 15–89. o. <https://doi.org/10.1002/9781118557693.ch2>.



- BENGOETXEA, E.–BUELA-CASAL, G. [2013]: The new multidimensional and userdriven higher education ranking concept of the European Union. *International Journal of Clinical and Health Psychology*, Vol. 13. No. 1. 67–73. o. [https://doi.org/10.1016/S1697-2600\(13\)70009-7](https://doi.org/10.1016/S1697-2600(13)70009-7).
- BENNEWORTH, P. S. [2010]: *A University Benchmarking Handbook. Benchmarking in European Higher Education*. ESMU, Brüsszel.
- BILLAUT, J.-C.–BOUYSSOU, D.–VINCKE, P. [2010]: Should you believe in the Shanghai ranking? *Scientometrics*, Vol. 84. No. 1. 237–263. o. <https://doi.org/10.1007/s11192-009-0115-x>.
- CHENG, Y.–CHURCH, G. M. [2000]: Biclustering of expression data. Megjelent: *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, 93–103. o.
- CSATÓ LÁSZLÓ [2013]: Rangsorolás páros összehasonlításokkal: Kiegészítések a felvételizői preferencia-sorrendek módszertanához. *Közgazdasági Szemle*, 60. évf. 12. sz. 1333–1353. o.
- CSATÓ LÁSZLÓ [2016]: Felsőoktatási rangsorok jelentkezői preferenciák alapján. *Közgazdasági Szemle*, 63. évf. 1. sz. 27–61. o.
- CSICSZMAN JÓZSEF [1979]: A klaszterelemzés módszerei és alkalmazási lehetőségei a statisztikában. *Statisztikai Szemle*, Vol. 57. No. 2. 137–145. o.
- DARAI, C.–BONACCORSI, A. [2017]: Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*, Vol. 68. No. 2. 508–529. o. <http://dx.doi.org/10.1002/asi.23679>.
- DOWNING, K. [2013]: What's the use of rankings? Rankings and Accountability in Higher Education: Uses and Misuses. UNESCO, 11. fejezet, 197–208. o.
- EDERER, P.–SCHULLER, P.–WILLMS, S. [2009]: University systems ranking: citizens and society in the age of the knowledge. *Educational Studies*, No. 3. 169–202. o.
- GESTRAUD, P.–BRITO, I.–BARILLOT, E. [2014]: Bicare: Biclustering analysis and results exploration. R package, <http://www.bioconductor.org/packages/release/bioc/vignettes/BicARE/inst/doc/BicARE.pdf>.
- GUARINO, C.–RIDGEWAY, G.–CHUN, M.–BUDDIN, R. [2005]: Latent Variable Analysis: A New Approach to University Ranking. *Higher Education in Europe*, Vol. 30. No. 2. 147–165. o. <https://doi.org/10.1080/03797720500260033>.
- GUSENLEITNER, D.–HOWE, E. A.–BENTINK, S.–QUACKENBUSH, J.–CULHANE, A. C. [2012]: iBBiG: iterative binary bi-clustering of gene sets. *Bioinformatics*, Vol. 28. No. 19. 2484–2492. o. <https://doi.org/10.1093/bioinformatics/bts438>.
- HAHSLER, M.–HORNİK, K.–BUCHTA, C. [2008]: Getting things in order: an introduction to the R package seriation. *Journal of Statistical Software*, Vol. 25. No. 3. 1–34. o.
- HARTIGAN, J. A. [1972]: Direct clustering of a data matrix. *Journal of the American Statistical Association*, Vol. 67. No. 337. 123–129. o. <https://doi.org/10.1080/01621459.1972.10481214>.
- HAZELKORN, E. [2015]: *Rankings and the reshaping of higher education: The battle for world-class excellence*. Palgrave Macmillan UK, London.
- HUANG, Q.-H. [2011]: Discovery of time-inconsecutive co-movement patterns of foreign currencies using an evolutionary biclustering method. *Applied Mathematics and Computation*, Vol. 218. No. 8. 4353–4364. o. <https://doi.org/10.1016/j.amc.2011.10.011>.
- IBÁÑEZ, A.–LARRAÑAGA, P.–BIELZA, C. [2013]: Cluster methods for assessing research performance: exploring Spanish computer science. *Scientometrics*, Vol. 97. No. 3. 571–600. o. <https://doi.org/10.1007/s11192-013-0985-9>.
- JAROCKA, M. [2012]: University ranking systems. From league table to homogeneous groups of universities. *World Academy of Science, Engineering and Technology*, Vol. 6. No. 6. 1377–1382. o.

- JIONG, Y.–WANG, H.–WANG, W.–YU, P. [2003]: Enhanced biclustering on expression data. Third IEEE Symposium on Bioinformatics and Bioengineering, 2003. Proceedings. 321–327. o. <https://doi.org/10.1109/BIBE.2003.1188969>.
- LIU, N. C. [2013]: The Academic Ranking of World Universities and its future direction. *Rankings and Accountability in Higher Education: Uses and Misuses*. UNESCO, Párizs, 1. fejezet, 23–39. o.
- LIU, S.–CHEN, Y.–YANG, M.–DING, R. [2009]: Bicluster algorithm and used in market analysis. Megjelent: 2009 Second International Workshop on Knowledge Discovery and Data Mining. 504–507. o. <https://doi.org/10.1109/WKDD.2009.224>.
- LUKMAN, R.–KRAJNC, D.–GLAVIČ, P. [2010]: University ranking using research, educational and environmental indicators. *Journal of Cleaner Production*, Vol. 18. No. 7. 619–628. o. <https://doi.org/10.1016/j.jclepro.2009.09.015>.
- MARGINSON, S.–VAN DER WENDE, M. [2007]: To rank or to be ranked: The impact of global rankings in higher education. *Journal of Studies in International Education*, Vol. 11. No. 34. 306–329. o. <https://doi.org/10.1177/1028315307303544>.
- NÉDA ZOLTÁN–RAVASZ MÁRIA–FLORIAN, RĂZVAN–LIBÁL ANDRÁS–GYÖRGYI GÉZA [2008]: Klaszterezés és fázisátalakulás frusztrált hálózatokban. *Műszaki Szemle*, 11. évf. 42. sz. 3–8. o.
- OLCAY, G. A.–BULU, M. [2017]: Is measuring the knowledge creation of universities possible? A review of university rankings. *Technological Forecasting and Social Change*, Vol. 123. Supplement C, 153–160. o. <https://doi.org/10.1016/j.techfore.2016.03.029>.
- POLÓNYI ISTVÁN [2017]: Oktatáspolitikai kísérletek és kudarcok. Nemzetközi színvonalú oktatási rendszer létrehozása. Megjelent: *Jakab, András–Urbán László* (szerk.): Hegymenet. Társadalmi és politikai kihívások Magyarországon. Osiris Kiadó, Budapest, 379–400. o.
- PONTES, B.–GIRÁLDEZ, R.–AGUILAR-RUIZ, J. S. [2015]: Biclustering on expression data: A review. *Journal of Biomedical Informatics*, Vol. 57. Supplement C, 163–180. o. <https://doi.org/10.1016/j.jbi.2015.06.028>.
- QS [2016]: Higher Education System Strength Rankings, 2016. A ranking of national higher education system. <https://www.topuniversities.com/system-strength-rankings/2016>.
- RAPONI, V.–MARTELLA, F.–MARUOTTI, A. [2016]: A biclustering approach to university performances: an italian case study. *Journal of Applied Statistics*, Vol. 43. No. 1. 31–45. o. <https://doi.org/10.1080/02664763.2015.1009005>
- SAISANA, M.–D’HOMBRES, B.–SALTELLI, A. [2011]: Rickety numbers: Volatility of university rankings and policy implications. *Research Policy*, Vol. 40. No. 1. 165–177. o. <https://doi.org/10.1016/j.respol.2010.09.003>.
- SALMI, J. [2013]: If ranking is the disease, is benchmarking the cure? *Rankings and Accountability in Higher Education: Uses and Misuses*. UNESCO, Párizs, 13. fejezet, 235–255. o.
- SHIN, J. C. [2011]: Organizational effectiveness and university rankings. Megjelent: *Shin, J. C.–Toutkoushian, R. K.–Teichler, U.* (szerk.): *University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education*. Springer Netherlands, Dordrecht, 19–34. o. [https://doi.org/10.1007/978-94-007-1116-7\\_2](https://doi.org/10.1007/978-94-007-1116-7_2).
- SOWTER, B.–HIJAZI, S.–REGGIO, D. [2017]: Ranking World Universities. *IGI Global A Decade of Refinement, and the Road Ahead*. 1–24. o. <https://doi.org/10.4018/978-1-5225-0819-9.ch001>.
- TELCS ADRÁS–KOSZTYÁN ZSOLT TIBOR–TÖRÖK ÁDÁM [2013a]: Hallgatói preferencia-sorrendek készítése az egyetemi jelentkezések alapján. *Közgazdasági Szemle*, 60. évf. 3. sz. 290–317. o.
- TELCS ADRÁS–KOSZTYÁN ZSOLT–TÖRÖK ÁDÁM [2013b]: Reflexiók Csató László vitairatára. *Közgazdasági Szemle*, 60. évf. 12. sz. 1354–1356. o.

- TÖRÖK ÁDÁM [2006]: Az európai felsőoktatás versenyképessége és a lisszaboni célkitűzések. Mennyire hihetünk a nemzetközi egyetemi rangsoroknak? *Közgazdasági Szemle*, 53. évf. 4. sz. 310–329. o.
- WILLIAMS, R.–DE RASSENFOSSE, G.–JENSEN, P.–MARGINSON, S. [2014]: U21 ranking of national higher education systems, 2014. Melbourne Institute of Applied Economic and Social Research University of Melbourne, [http://mycc.my/document/files/PDF%20Dokumen/University%20Ranking/U21/U21\\_2014.pdf](http://mycc.my/document/files/PDF%20Dokumen/University%20Ranking/U21/U21_2014.pdf).
- WILLIAMS, R.–LEAHY, A.–DE RASSENFOSSE, G.–JENSEN, P. [2015]: U21 ranking of national higher education systems, 2015. Melbourne Institute of Applied Economic and Social Research University of Melbourne, <http://mycc.my/document/files/PDF%20Dokumen/University%20Ranking/U21/153U21Rankings2015.pdf>.
- WILLIAMS, R.–LEAHY, A.–DE RASSENFOSSE, G.–JENSEN, P. [2016]: U21 ranking of national higher education systems, 2016. Melbourne Institute of Applied Economic and Social Research University of Melbourne, <https://universitas21.com/sites/default/files/2018-03/2016%20full%20report.pdf>.
- WILLIAMS, R.–LEAHY, A.–JENSEN, P. [2017]: U21 ranking of national higher education systems, 2017. Melbourne Institute of Applied Economic and Social Research University of Melbourne, <https://universitas21.com/sites/default/files/2018-03/2017%20full%20report.pdf>.
- YANG, J.–WANG, H.–WANG, W.–YU, P. S. [2003]: Enhanced biclustering on expression data. Megjelent: Third IEEE Symposium on Bioinformatics and Bioengineering, 2003. Proceedings. 321–327. o. <https://doi.org/10.1109/BIBE.2003.1188969>.
- YANG, J.–WANG, H.–WANG, W.–YU, P. S. [2005]: An improved biclustering method for analyzing gene expression profiles. *International Journal on Artificial Intelligence Tools*, Vol. 14. No. 5. 771–789. o. <https://doi.org/10.1142/S0218213005002387>.