

TELCS ANDRÁS–KOSZTYÁN ZSOLT TIBOR–TÖRÖK ÁDÁM

Hallgatói preferencia-sorrendek készítése az egyetemi jelentkezések alapján

Intézményi rangsorokat sokféleképpen alkotnak. Egyszerre több szempont alapján értékelik a felsőoktatási intézményeket – például a minősített oktatók száma vagy az egy oktatóra jutó hallgatók száma alapján. Ez a rangsorolás számos kérdést felvet: egyrészt az egyes szempontok figyelembevétele és relevanciája, másrészt a rangsorolási eljárások adekvát volta és interpretációja tekintetében. Magát az eredményeket torzíthatja a vizsgálat készítőjének piaci érdeke, a vizsgált intézmények érdekei, valamint az irányításukért felelős kormányzati politika várható reakciója, nem beszélve a rendelkezésre álló adatok minőségéről, mennyiségéről és torzítottságának fokáról. Végül, de nem utolsósorban: számos esetben a rangsorokat, illetve a készítésükhöz alkalmazott szempontokat kormányzati döntésekhez is felhasználják. Ezért Magyarországon is megkezdődött az intézmények minősítési szempontok szerinti „feljavítása”, azaz a mérés visszahat magára a mérés tárgyára is, ami módszertani fordulatot sürget a felsőoktatási rangsorok készítésében. Tanulmányunkban egy olyan megközelítést próbálunk megalapozni, amelynek segítségével elkerülhetők az említett veszélyek, illetve új elemekkel gazdagítható a rangsorkészítés módszertana.*

Journal of Economic Literature (JEL) kód: C44, I23.

A felsőoktatási intézmények rangsorolásának nemzetközi tapasztalatai alapján Magyarországon 2002-ben kezdődött meg rangsorok készítése és publikálása. Mihályi Péter már számos indikátor alapján vizsgálta a diplomák értékét és az azokat kibocsátó intézmények népszerűségét (*Mihályi [2002]*).

* Ez a tanulmány a TÁMOP-4.2.2/B-10/1-2010-0025. számú a Tudományos képzés műhelyeinek támogatása a Pannon Egyetemen elnevezésű projekt keretében készült. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásában valósul meg. A szerzők köszönetüket fejezik ki a Educatio Társadalmi Szolgáltató Nonprofit Kft.-nek az adatok rendelkezésre bocsátásáért. Hálásak vagyunk *Abonyi Jánosnak* hasznos tanácsaiért, kérdéseierért, valamint *Farkas Andrásnak*, akitől rendkívül értékes útbaigazítást, tanácsokat kaptunk a rangsorolás elméleti kérdéseiben és annak irodalmára vonatkozóan.

Telcs András, Pannon Egyetem Kvantitatív Módszerek Tanszék.

Kosztján Zsolt Tibor, Pannon Egyetem Kvantitatív Módszerek Tanszék.

Török Ádám, Pannon Egyetem Közgazdaságtan Tanszék, BME Közgazdaságtan Tanszék, MTA–PE Regionális Innovációs és Fejlesztési Hálózati Kutatócsoport.

A Felsőoktatási Műhely tanulmányai 2002 óta jelennek meg (Fábri [2008]). A műhely nagy adatbázison széles körű, sokrétű kutatásokat végez a felsőoktatás területén (Fábri–Torda [2005]).

Az első lépéseket hamarosan követte a HVG, majd más kiadók hasonló összeállításai. (Kiváló áttekintést olvashat a 2001–2009 időszakban kialakult rangsorokról és azok hazai szakirodalmáról az olvasó Horváth–Kiss [2009] tanulmányában.) Napjainkban a Felvi honlapjáról egy interaktív, igen rugalmasan személyre szabott, a jelentkező szempontjaira épülő rangsorkészítő módszer is rendelkezésre áll. Ilyen például az egy oktatóra jutó hallgatók száma, a publikációs tevékenység mérőszámai, a végzés és elhelyezkedés között eltelt idő, illetve az első kereset. Nagyon sok és érdekes mutatót dolgoztak ki eddig is, ami jól mutatja, hogy a felsőoktatás mint szociális tevékenység, illetve szolgáltatás rendkívül összetett, jó néhány szálon kapcsolódik – a magánszférától kezdve a makrogazdaságig – számos területhez. Ennek megfelelően az indikátorok kiválasztása, súlyozásuk meghatározása távolról sem egyértelmű és egyszerű feladat. (A módszerek, indikátorok kijelöléséről lásd Katona–Balogh [2010].) Az intézményeket lehet vizsgálni *1. az inputból kiindulva* a hallgatói toborzás oldaláról, a bevételek összessége és összetétele alapján, a szolgáltatásaik mennyiségi, illetve minőségi mutatói alapján, gazdasági, hatékonysági szempontok szerint, *2. az output alapján* a végzett hallgatók képzettsége, felkészültsége, munkaerő-piaci pozíciója.

A rangsorkészítők a felsőoktatási piacon számba vették mind a nemzeti, mind a nemzetközi versenyt, ez utóbbit egyrészt az Európai Unióhoz csatlakozás, másrészt a lisszaboni célkitűzések sikere szempontjából (Török [2006]–[2009d], Török–Kovács [2011]). A gazdasági megközelítés mellett megvan a létjogosultsága a regionális (Csuka–Szívós [2010a], [2010b]), szociológiai és marketingszemléletű figyelembevételnek is, mint például a hallgatói elégedettség, a felvettek szociális összetétele, a családok, jelentkezők körében az intézmények presztízse. Hasonlóan fontos mutató, hogy milyen presztízst tulajdonítanak a vállalatvezetők az oktatási intézményeknek, illetve milyen végzettséggel vesznek fel szívesebben pályakezdőt vagy tapasztaltabb munkatársat.

Tamándl [2011] kísérletet tett a mutatószámok hierarchikus szerkezetbe rendezésére és szisztematikus felépítésére, s eredményei jól mutatják, hogy az indikátorokat és a kialakított rangsort egyenként és összességükben a társadalomtudományi megközelítés módja, közgazdasági, szociológiai, illetve menedzsmentszemléletek határozzák meg. Közhely, hogy nincs tökéletes rangsor, ezért is korrekt a több-dimenziós rangsorépítés és még inkább a Felvi testre szabott megközelítése. Ugyanakkor jól érzékelhető az igény egységes, könnyen értelmezhető és a közfelfogással, előzetes igényekkel többé-kevésbé összhangban lévő rangsorokra. A megvalósítás nehézsége abban áll, hogy a célközönség is igen heterogén. Elsősorban idetartoznak a felvételiző diákok és családjuk, de az intézmények számára is fontos, hogy a piac miképpen ítéli meg őket, továbbá a felsőoktatás irányítói is figyelemmel kísérik a mutatószámokat és rangsorokat, akárcsak a munkaerőt kereső vállalatok. Ennek megfelelően a rangsorokkal kapcsolatos igények is sokfélék. Ebben a közegben kívánnak a rangsorokat közzétévők egyrészt szakmailag

megalapozott, jól publikálható, a piacon mind nagyobb elfogadottságot nyerő rangsorokat kialakítani.

Az elmúlt tíz évben összegyűlt magyar és a nagyobb múltra visszatekintő nemzetközi felsőoktatási rangsorkészítési tapasztalat, gyakorlat fényében nem kecsesget túlzott sikerrel egy újabb rangsor felállítása. Tanulmányunkban preferencia-sorrend felállítására teszünk kísérletet, ami új irány lehet, szakítva az elődök által kicsiszolt módszertan hagyományaihoz képest. Nem alkalmazunk input-output adatokat, minőségi, mennyiségi mutatókat, presztízsindexeket, hanem kizárólag a jelentkező diákok jelentkezési szándékaira alapozzuk a kialakítandó rangsort. Ezzel egyrészt megszabadulunk a társadalomtudományi alapvetés általi determinációtól, másrészt a mutatók súlyozásának önkényes megválasztásától és az azzal kapcsolatos kivédhetetlen kérdőjelektől és kritikáktól.

A hallgatók egy intézmény kiválasztásakor a „lábukkal” szavaznak. Egy hallgató, aki különböző intézményekbe jelentkezik, valamilyen szempont alapján (például képzés színvonala, lakóhely közelsége stb.) „sorrendbe rakja” azokat. Kutatásunkban arra kerestük a választ, hogyan lehet a hallgatói jelentkezések alapján egy preferencia-sorrendet felállítani szakok, karok és intézmények között kizárólag e hallgatói jelentkezésekre alapozva.

2001-től 2011-ig rendelkezésünkre állt a Felvi-honlap (<http://felvi.hu>) által nyilvántartott, hallgatói jelentkezéseket tartalmazó teljes adatbázis. Egy-egy évre vonatkozó több mint százezer jelentkezés több mint százezer részleges rangsort jelentett, amelyekből a végső sorrend kialakításához több módszer is rendelkezésre áll. Munkánk során kiválasztottuk a feladatnak leginkább megfelelő, valójában kivitelezhető és megfelelő értelmezést lehetővé tevő eljárásokat.

Írásunk egy hosszabbra tervezett cikksorozat első eleme, amelynek keretében a jelentkezések sokoldalú elemzésére törekszünk. A kutatást itt egy torzításmentes preferencia-sorrend kialakításával alapozzuk meg. Bemutatjuk, hogyan vizsgálhatók a különböző szempontú jelentkezési preferenciák, hogyan kell az egyedi adatokat tárolni és aggregálni, majd példákon mutatjuk be a kapott felsőoktatási preferencia-sorrendeket.

A módszerek bemutatása

A módszerek bemutatása során először az általunk felhasznált forrásadatokkal foglalkozunk, majd azzal, hogyan szűrhetjük az adatokat lényeges információvesztés nélkül, és hogyan lehet ezeket az információkat veszteségmentesen egy gráfrepresentáció segítségével tárolni, itt ismertetjük a sorrendeket megadó eljárásokat. Végül a robusztusságvizsgálatra kifejlesztett eszközöket mutatjuk be. Ez a fejezet részben azonos a *Telcs és szerzőtársai* [2012] cikk módszertani fejezetével, de úgy véljük, az olvasót segíti, ha a preferencia-sorrendek területén alkalmazott egyesítési módszereket itt is kellő részletességgel bemutatjuk.

Olyan, kizárólag a hallgatói jelentkezéseket figyelembe vevő módszereket kerestünk, amelyek segítségével úgy lehet teljes preferencia-sorrendet létrehozni, hogy az a leginkább összhangban legyen az egyes jelentkezők preferenciájával.

A forrásadatok szerkezete

2001–2011-ig rendelkezésünkre állt a <http://felvi.hu> által nyilvántartott hallgatói jelentkezéseket tartalmazó teljes adatbázis. Ez az adatbázis tíz oszlopot tartalmazott:

1. ÉV: 2001–2011
2. ELJÁRÁS: normál vagy keresztfélév
3. EGYEDI AZONOSÍTÓ
4. JELENTKEZÉSI HELY: az adott hallgató az intézménybe hányadikként jelentkezett
5. INTÉZMÉNY
6. KAR
7. SZAK
8. A KÉPZÉS FORMÁJA: alap, mester vagy osztatlan
9. A KÉPZÉS MÓDJA: nappali vagy levelező
10. KÉPZÉS FINANSZÍROZÁSA: állami vagy költségtérítéses

Minden egyes év több mint 400 ezer rekordot és 100 ezer jelentkezést tartalmazott. Részletesebben mi a 2011. évi jelentkezési adatokat tekintettük, de elemzési módszereink a korábbi évek adatait is fel tudják dolgozni. E tíz adat valamennyi rekord esetén hiánytalanul szerepelt. Ugyanis e nélkül nem érvényes a jelentkezés.

Szűrések, objektumok definiálása

Módszerünk tárgyalása során az első felmerülő kérdés, hogy milyen adatokat tekintünk, és mi legyen egy objektum. Az adatok szűrését a következőképpen végeztük:

1. definiáltuk azon szakok/karok/intézmények körét, amelyek adatait a későbbi elemzések során vizsgálni szerettük volna (például közgazdasági terület képzései);
2. a hallgatói jelentkezések során csak azokat a jelentkezéseket tekintettük, amelyek a fent definiált területen szereplő képzésekre vonatkoztak.

A következő lépésben el kell döntenünk, hogy mit tekintünk egy objektumnak, amelyek között szeretnénk sorrendeket felállítani. Kézenfekvőnek tűnik az intézményeket választani objektumoknak, hiszen intézmények között szoktuk a sorrendeket keresni. Azt azonban nem szabad elfelejtenünk, hogy a hallgató alapvetően nem intézményekre, hanem azon belül szakokra fog jelentkezni. Elképzelhető az is, hogy egy hallgató több szakra jelentkezik egy intézményen belül. Éppen ezért az információvesztés elkerülése érdekében egy intézményen belül egy szakot tekintettünk egy objektumnak. A következő alfejezetben bemutatjuk, hogy a gráfobjektumok összevonásával nemcsak szakok között, hanem karok, intézmények között is lehet sorrendet értelmezni. Ehhez azonban egy megfelelő gráfrepresentációt kell alkalmazni (hasonlóan a *Chebotarev–Shamis* [1999] és *Slikker–Borm–van den Brink* [2012] tanulmányokhoz), mely a jelentkezések adatait redundanciamentesen, de információvesztés nélkül tudja tárolni.

A sorrendek kialakításánál néhány általános elvet követünk:

1. nincs különbség a preferenciák erőssége között;
2. közvetett preferenciák is számítanak (az első helyen megjelölt intézmény a harmadiknál is jobb);
3. megjelölt szakok preferáltak az összes kihagyotthoz képest;
4. a nem megjelölt szakok kevésbé preferáltak, mint bármelyik megjelölt;
5. a nem megjelölt szakok egyenrangúak, közöttük semmilyen megkülönböztetést nem teszünk.

Gráfrepresentáció

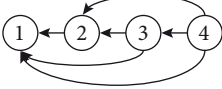
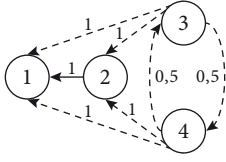
Tegyük fel, hogy m darab objektumunk van (például m darab szak között szeretnénk sorrendet felállítani). Legyen \mathbf{A} n darab hallgatói jelentkezést mint jelentkezési vektorokat tartalmazó halmaz. $\mathbf{A} = \{\mathbf{a}^1, \dots, \mathbf{a}^n\}$, $\mathbf{a}^i = [a^i_1, \dots, a^i_m]^T$, $i = 1, 2, \dots, n$. A gráf csúcsai jelölik az objektumokat (amelyek lehetnek szakok, karok vagy intézmények), az élek pedig a preferenciákat. Ha egy i -edik objektumból fut j -edik objektumba egy irányított él, akkor ez azt jelenti, hogy i -edik objektumot preferáltuk j -edikkel szemben.

Ha a hallgatói jelentkezéseket szeretnénk gráfok segítségével reprezentálni, akkor két szabályt kell szem előtt tartanunk. Egyrészt jelölnünk kell a közvetett preferenciákat is a gráfban (például ha egy hallgató jelentkezési sorrendje: $\mathbf{a}^1 = [1, 2, 3, 4]^T$, ez azt jelenti, hogy a hármas objektumnál nemcsak a kettést, hanem az egyest is előrébb sorolta). Másrészt azokat az objektumokat (szakokat), amelyekre nem adtak be jelentkezési lapot, az utolsó helyen megjelölt szaknál is hátrább sorolták. Minden egyes preferenciaválasztásnál egy „szavazatot” adunk annak az intézménynek, amelyet előrébb jelölt meg a hallgató egy másik intézménnyel szemben. Így a gráfon ezen élek súlya egy lesz, függetlenül attól, hogy azért húztunk be egy élt két objektum között, mert mindkettő szerepelt a hallgató jelentkezési sorrendjében, vagy azért, mert a két objektumból az egyik nem szerepelt a hallgató által megjelölt szakok között.

Azon szakok között nem tudunk preferencia-sorrendet felállítani, amelyek nem szerepeltek egy hallgatói jelentkezési lapon. Itt nem tudjuk egyik szakot sem egy-egy szavazattal „jutalmazni” a másikhoz képest, így a szavazatot 50-50 százalékban megosztjuk a két szak között úgy, hogy mindkét objektumból egy-egy él mutat a másik felé 0,5-ös súllyal. (Tegyük fel, hogy $m = 4$; $\mathbf{a}^1 = \{1, 2, 3, 4\}$ $\mathbf{a}^2 = \{1, 2\}$). Ekkor a gráfban a hármas és a négyes objektumból kimenő éleket szaggatottal, a gráfot reprezentáló úgynevezett szomszédsági vagy adjacenciamátrixban a cellákat dőlt betűvel jelöltük (1. táblázat). Itt a cellaértékek a gráf csúcsait összekötő éleknek a súlyát jelenítik meg, amely jelen esetben vagy 1, vagy 0,5 lehet.

Az általunk alkalmazott gráfrepresentációban éleket húztunk a nem preferált intézményektől a preferált (például a felvételi lapon szereplő) intézmények felé. Másrészt azok a nem preferált intézmények, amelyeket a hallgató nem rangsorolt, ott a nem preferált csúcspontok között 0,5 súlyú élek szerepelnek.

1. táblázat
Jelentkezési sorrendek gráfrepresentációja

Gráfrepresentáció egyéni preferenciagráf	Szomszédsági mátrix egyéni preferenciamátrix, $\mathbf{M}^{(1)}$	Gráfrepresentáció egyéni preferenciamátrix	Szomszédsági táblázat egyéni preferenciamátrix, $\mathbf{M}^{(2)}$																																																		
$\mathbf{a}^1 = [1, 2, 3, 4]^T, m = 4$  Élsúlyok: 1	<table border="1"> <tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th></tr> <tr><th>1</th><td></td><td>-</td><td>1</td><td>1</td></tr> <tr><th>2</th><td></td><td></td><td>-</td><td>1</td></tr> <tr><th>3</th><td></td><td></td><td></td><td>-</td></tr> <tr><th>4</th><td></td><td></td><td></td><td></td></tr> </table>		1	2	3	4	1		-	1	1	2			-	1	3				-	4					$\mathbf{a}^2 = [1, 2]^T, m = 4$ 	<table border="1"> <tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th></tr> <tr><th>1</th><td></td><td>-</td><td>1</td><td>1</td></tr> <tr><th>2</th><td></td><td></td><td>-</td><td>1</td></tr> <tr><th>3</th><td></td><td></td><td></td><td>-</td></tr> <tr><th>4</th><td></td><td></td><td></td><td></td></tr> </table>		1	2	3	4	1		-	1	1	2			-	1	3				-	4				
	1	2	3	4																																																	
1		-	1	1																																																	
2			-	1																																																	
3				-																																																	
4																																																					
	1	2	3	4																																																	
1		-	1	1																																																	
2			-	1																																																	
3				-																																																	
4																																																					
Oszlopösszegek:	Σ 0 1 2 3	Oszlopösszegek:	Σ 0 1 2,5 2,5																																																		

HIBAFÜGGVÉNY, INHOMOGENITÁSI INDEX • A hibafüggvény segítségével meg tudjuk határozni, hogy az egyéni preferenciamátrixok összegeként adódó aggregált preferenciamátrix (Chebotarev–Shamis [1999]) alapján hány olyan hallgatói jelentkezés található, amely egy adott preferencia-sorrenddel ellentétes jelentkezési sorrendet tükröz.

Legyen \mathbf{M} a jelentkezési sorrendekből képzett gráfok szomszédsági mátrixainak összege: az úgynevezett aggregált szomszédsági mátrix ($\mathbf{M} \in \mathbb{N}^{m \times m}$), ahol $m_{ij} \in \mathbf{M}$, $i \neq j$ elem azt mutatja, hogy hányszor preferálták az i intézményt a j -vel szemben.

Legyen $\mathbf{b} = [b_1, b_2, \dots, b_m]^T$ egy tetszőleges sorrend (\mathbf{b} vektor $\{1, 2, \dots, m\}$ számok permutációi). Legyen továbbá $\mathbf{M}_{\mathbf{b}}$ az \mathbf{M} mátrix átrendezettje. Ekkor a hibafüggvény a következőképpen számítható:

$$h(\mathbf{M}, \mathbf{b}) = \sum_{i=1}^m \sum_{j=i+1}^m m_{b_j b_i}, \quad \forall m_{b_j b_i} \in \mathbf{M}_{\mathbf{b}}. \tag{1}$$

A hibafüggvény tehát megadja, hogy egy \mathbf{b} sorrendet tartalmazó vektort tekintve hányan preferálták ellenkező sorrendben az egyes objektumokat (például szakokat). Egy \mathbf{b} preferencia-sorrendre vonatkozó hibafüggvény tehát egy $\mathbf{M}_{\mathbf{b}}$ mátrix alsó háromszögében lévő elemek összessége lesz. Maximális értéke: $n[m(m - 1)/2]$, ahol n a hallgatók, m pedig az intézmények számát jelöli. Minél több hallgató jelentkezik egy adott évben, annál nagyobb lehet a hibafüggvény értéke, így ha ezzel a maximális $n[m(m - 1)/2]$ értékkel elosztjuk a hibafüggvény értékét, akkor egyszázalékos értéket kapunk, amely értéket már összehasonlíthatjuk más intézményi és hallgatói létszámmal végzett vizsgálatokkal is.

Az $I_{\mathbf{M}_{\mathbf{b}}}^n = h(\mathbf{M}, \mathbf{b})/n[m(m - 1)/2] \in [0, 1]$ értéket \mathbf{b} sorrend inhomogenitási indexének nevezzük. Értéke akkor lenne 1, ha valamennyi hallgató \mathbf{b} sorrenddel ellentétes módon preferálta volna az intézményeket. \mathbf{b} értéke nulla, ha mindenki \mathbf{b} sorrenddel megegyező módon preferálja valamennyi intézményt.

Tegyük fel, hogy \mathbf{b}^* sorrend esetén $h(\mathbf{M}, \mathbf{b}^*)$ értéke minimális. Ekkor $I = I_{\mathbf{M}_{\mathbf{b}^*}}^n = h(\mathbf{M}, \mathbf{b}^*)/n[m(m - 1)/2]$ értéke is minimális. Az alacsony I érték arra utal, hogy a \mathbf{b}^* -gal ellentétes preferenciák száma is alacsony. A preferencia-sorrendek konzisz-

tens módon ugyanabba az irányba mutatnak. A magas I érték arra utal, hogy bár találtunk \mathbf{b}^* sorrendet, amelyre a hibafüggvény értéke minimális, mégis sok hallgató ezzel ellentétes sorrendben jelöli meg az intézményeket. Érdekes kérdés lehet olyan klasztereket képezni, ahol I értéke a klaszterben minimális. Ugyanis ekkor homogén klasztereket kapnánk. Későbbi kutatásainkban kitérünk arra, hogy I értéke lehet-e egy régió vonzáskörzetében lakó hallgatók jelentkezései alapján számolt sorrendre nézve jellemző az adott régióra. Továbbá vajon az egyes régiókon belül lakó hallgatók jelentkezései alapján lehet-e régióspecifikus homogén sorrendeket meghatározni.

OSZLOPÖSSZEG-, SORÖSSZEGMÓDSZER • Felhasználva, hogy egy (aggregált) szomszédsági mátrixban az oszlopok/sorok összege megadja a csúcsból kimenő élek számát – amely egyben azt is jelenti, hogy az adott objektumot hányszor sorolták hátrébb, más objektumokkal szemben –, egy gyors heurisztikus módszer konstruálható (hasonlóan *Chebotarev–Shamis* [1999] publikációkban leírtakhoz). A lépések a következők.

1. Készítsük el az \mathbf{M} aggregált szomszédsági mátrixot!

2. Számítsuk ki az összegeket, majd az elemeket rendezzük nagyság szerint növekvő sorrendbe. Az így megadott \mathbf{b} sorrend lesz a javasolt sorrend!

A 2. táblázat mutatja a módszer lépéseit. Tegyük fel, hogy adott az \mathbf{M} aggregált szomszédsági vagy aggregált szomszédsági mátrix (a 2. táblázat bal oldala).

2. táblázat

Az oszlopösszeg- (vagy sorösszeg-) módszer lépései

	Aggregált szomszédsági mátrix (\mathbf{M})						Aggregált szomszédsági mátrix (\mathbf{b} szerinti) átrendezettje (\mathbf{M}_b)				
	1	2	3	4	$h(\mathbf{M}, [1, 2, 3, 4]^T) = 17$		1	2	3	4	$h(\mathbf{M}, [1, 3, 4, 2]^T) = 15$
1	–	5	5	3		1	–	5	3	5	
2	3	–	1	5		2	2	–	2	3	
3	2	3	–	2		3	1	3	–	5	
4	1	5	3	–		4	3	1	5	–	
Σ	6	13	9	10	$\mathbf{b} = [1, 3, 4, 2]^T$	Σ	6	9	10	13	

A módszer egyszerűsége mellett meg kell jegyezni, hogy ha az oszlopösszegek egyenlők, akkor nem állítható fel egyértelmű sorrend. Ha hiányos preferencia-sorrendekből képeztünk aggregált szomszédsági mátrixot, akkor az oszlopösszegek alapján felállított sorrend szerint átrendezett mátrix nem ad feltétlenül minimális hibafüggvényt.

A RANGSZÁMÖSSZEG-MÓDSZER GRÁFREPREZENTÁCIÓJA • Az úgynevezett rangszámösszeg-módszer lényege, hogy a végső preferencia-sorrend kialakításához össze kell adni az egyedi sorrendek ($\mathbf{a}^i, i = 1, 2, \dots, n$) rangszámait. Az ebből kiszámolt rangok összegét sorrendbe állítva, megkapjuk a végső preferencia-sorrendet ($\mathbf{b} \in \mathbb{N}^m$). Ha ezek a preferencia-sorrendek teljeseek, akkor a végső sorrend hallgatók preferen-

cia-sorrendjeitől vett négyzetes különbségeinek összege minimális, amelyet a (2) képlettel határozhatunk meg.

$$\sum_{i=1}^n \sum_{j=1}^m (a_j^i - b_j)^2, \text{ ahol } a_j^i \in \mathbf{a}^i, b_j \in \mathbf{b}. \quad (2)$$

Ha a hallgatói preferencia-sorrendek nem teljesek, akkor egyrészt itt is feltételezhetjük, hogy azokba az intézményekbe, ahová nem adott be a hallgató jelentkezési lapot, azt az utolsó helyen bejelölt szaknál (illetve intézménynél) is hátrább sorolta. Másrészt, mivel a nem megjelölt intézmények között nem tudunk preferencia-sorrendet felállítani, így ezeket a fennmaradó rangszámokat egyenletesen osztjuk el közöttük. Ezzel a hallgatói preferencia-sorrendeket kiegészítjük a ki nem osztott rangszámok átlagával, és az i -edik hallgatóra kapunk egy $\mathbf{s}^i \in \mathfrak{R}^m$ vektort, itt $\mathbf{s}^i = [s_1^i, s_2^i, \dots, s_m^i]^T$, ahol

$$s_j^i = \begin{cases} k, & \text{ha } a_k^i = j \in \mathbf{a}^i, \\ \frac{1}{m - m_i} \sum_{k=m_i+1}^m k, & \text{különben} \end{cases} = \begin{cases} k, & \text{ha } a_k^i = j \in \mathbf{a}^i, \\ \frac{m_i + 1 + m}{2}, & \text{különben} \end{cases}. \quad (3)$$

Ezt a vektort nevezzük a továbbiakban az i -edik hallgató preferenciavektorának. (Legyen például $m = 4$, az i -edik hallgató preferencia-sorrendje $\mathbf{a}^i = [1, 3]^T$ – elsőként az egyes, másodikként a hármas szakra jelentkezett –, $m_i = 2$, ekkor az i -edik hallgató rangszámait tartalmazó vektor a következőképpen írható fel: $\mathbf{s}^i = [1, 3, 5, 2, 3, 5]^T$.)

A rangszámösszeg-módszert a már bemutatott gráfrepresentáció segítségével is kiszámíthatjuk.

Legyen $\mathbf{s} = \sum_{i=1}^n \mathbf{s}^i$ a rangszámok összegét tartalmazó vektor, míg $\bar{\mathbf{s}} = \mathbf{s}/n$ az átlagos rangérték. Legyen továbbá $\tilde{\mathbf{s}}^i = \mathbf{s}^i - \mathbf{1}$ egy úgynevezett korrigált preferenciavektor, ahol $\mathbf{1}$ egy m dimenziós 1 értékeket tartalmazó vektor. A korrigált rangszámösszeg: $\tilde{\mathbf{s}} = \sum_{i=1}^n \tilde{\mathbf{s}}^i = \sum_{i=1}^n \mathbf{s}^i - \mathbf{1}n$. A korrigált átlagos rangszámérték pedig: $\bar{\tilde{\mathbf{s}}} = \tilde{\mathbf{s}}/n = \mathbf{s}/n - \mathbf{1}$.

Az oszlopösszegmódszer során bemutatott gráfrepresentációhoz tartozó egyéni preferenciamátrixok oszlopösszegei a korrigált preferenciavektorokat adják.

$$\tilde{\mathbf{s}}^i = \mathbf{s}^i - \mathbf{1} = [s_1^i - 1, s_2^i - 1, \dots, s_j^i - 1, \dots, s_m^i - 1]^T = [\tilde{s}_1^i, \tilde{s}_2^i, \dots, \tilde{s}_j^i, \dots, \tilde{s}_m^i]^T,$$

$$\text{ahol } \tilde{s}_j^i = \begin{cases} k - 1, & \text{ha } a_k^i = j \in \mathbf{a}^i, \\ \frac{m_i - 1 + m}{2}, & \text{különben} \end{cases}. \quad (4)$$

Az \tilde{s}_j^i i -edik ($i = 1, 2, \dots, n$) hallgató j -edik szakra vonatkozó oszlopösszegértéke pontosan $k - 1$, ha a j -edik szak szerepel a hallgatói jelentkezési lapon k -adikként, hiszen ekkor pontosan $k - 1$ szakot sorolt előrébb a j -edikhez képest. Ha a j -edik szak nem szerepelt a hallgató jelentkezési lapján, akkor m_i darab 1 súlyú élt húztunk be a hallgatói jelentkezési lapon megjelölt szakok felé, valamint 0,5 súllyal azon szakok felé, amelyeket a hallgató nem szerepeltetett a jelentkezési lapján. Ha önmagába nem húztunk élt, akkor ennek száma $m - m_i - 1$. Ekkor az oszlopösszeg

értéke: $\tilde{s}_j^i = 1m_i + 0,5(m - m_i - 1) = (m_i - 1 - m_i)/2$, ($j \notin \mathbf{a}^i$). Változás csak a nem preferált szakokra vonatkozóan van.

Az egyéni preferenciamátrixok összege megadja az aggregált preferenciamátrixot, az aggregált preferenciamátrix oszlopösszegei pedig megadják a korrigált rangszám-összegek értékét. Így az oszlopösszegek meghatározásával mind az átlagos rangszámérték, mind pedig a rangszámok összege kiszámítható.

Látható, hogy az oszlopösszeg- és a rangösszegmódszer ekvivalens eredményre vezet, az előző fejezetben bemutatott gráfrepresentáció alkalmazásával. Ekkor az így kialakított rangsortól az egyéni preferencia-sorrendek euklideszi távolságainak négyzetes összege minimális.

Így a gráfrepresentáció alapján számított aggregált preferenciamátrix oszlopösszegei egyben rangösszeget is adnak.

A PÁROS ÖSSZEHAJONLÍTÁS MÓDSZERE HIÁNYOS RANGSOROKRA • A páros összehasonlítási módszerek alkalmazása során kihasználjuk, hogy a hallgatói jelentkezési sorrendek egy részben rendezett egyéni preferencia-sorrendet adnak meg. Azokat a szakokat, amelyeket a hallgató megjelölt, jelentkezési sorrend alapján sorba rendezte. Azon szakok esetén, amelyekre a hallgató nem adott be jelentkezési lapot, ott csak annyit feltételezhetünk, hogy ezeket a szakokat a legutolsó helyen megjelölt szakhoz képest is hátrább sorolta. Arról azonban nincs tudomásunk, hogy a nem megjelölt intézmények között milyen preferencia-sorrend állítható fel. Az ilyen problémát nem teljesen kitöltött páros összehasonlítási problémának (*incomplete pairwise problem*) nevezzük (lásd *Farkas–Lancaste–Rózsa* [2003]). A probléma megoldására számos tanulmány született (*Alonso* [2005], *Tanino* [1984], *Fedrizzi–Giove* [2007], *Bozóki–Fülöp–Rónyai* [2010]). A páros összehasonlítási módszerek nagy előnye, hogy végeredményül nemcsak egy preferencia-sorrendet kapunk, hanem arra is következtetéseket tudunk levonni, hogy két szak között a hallgatók jelentkezéseit figyelembe véve, mekkora különbség mutatkozik egymáshoz képest. Eredményül tehát sorrendi skála helyett arányskálát kapunk.

Mi Michele Fedrizzi és Silvio Giove módszerét alkalmaztuk (*Fedrizzi–Giove* [2007]). Itt a bemeneti páros összehasonlítási mátrixot (*pairwise comparison matrix*, *PCM*) az aggregált preferenciamátrixból (\mathbf{M}) képeztük, úgy, hogy $\forall pcm_{i,j} \in \mathbf{PCM}$, $\forall m_{i,j} \in \mathbf{M}$ esetén

$$pcm_{i,j} = \begin{cases} \frac{m_{i,j}}{m_{i,j} + m_{j,i}}, & \text{ha } m_{i,j} + m_{j,i} > 0 \\ 0,5 & \text{különben} \end{cases} \quad (5)$$

Meg kell jegyezni, hogy a páros összehasonlítás során a mátrix főátlójának elemeit nem vesszük figyelembe a számítások során.

A *PCM* egy olyan mátrix lesz, ahol a főátlóra szimmetrikus elemek összege 1. Cel-laértékei azt jelentik, hogy a hallgatói jelentkezési sorrendeket figyelembe véve, hány százalékban preferálták az i -edik szakot a j -edikkel szemben, másképpen fogalmazva: hány százalék annak a valószínűsége, hogy az i -edik szakot előbbre sorolták a

j -edikkel szemben. A módszer lépései vázlatosan a következők: a **PCM** mátrixból képezünk egy standardizált z értékekből álló **Z** mátrixot, amelyek cellaértékei z_{ij} a **PCM** mátrix i -edik sorának összege mínusz a **PCM** mátrix j -edik oszlopának összege ($i, j = 1, \dots, m, i \neq j$). A standardizált z értékek tehát a sorok és oszlopok várható értékeinek különbségei (a várható értékek jelentik a skálaértékeket). Könnyen belátható, hogy a **Z** mátrix oszlopösszegeinek/sorösszegeinek összege 0 lesz. A z értékek oszlopösszegeit átlagolva, jó becslést kapunk a szakok skálán elfoglalt pozícióira. Jelöljük ezeket z_1, z_2, \dots, z_m -mel! Ezeket az értékeket nevezzük preferenciaértékeknek. Ha nagyság szerint sorrendbe rakjuk ezeket az értékeket, akkor megkapjuk a módszer által szolgáltatott \mathbf{b}_z preferencia-sorrendet. A sorrenden kívül azonban a z_1, z_2, \dots, z_m értékekből arra is következtethetünk, hogy mekkora különbség van két szak preferenciaértéke között.

Mivel az előző fejezetben javasolt gráf-reprezentációban a nem preferált intézményektől is húztunk éleket, valamint a nem preferált intézményeket is összekötöttük 0,5-ös élekkel, így nem fordul elő, hogy az ebből számolt aggregált preferenciamátrix különböző számú összehasonlításokat tartalmaz. Így az így számolt relatív értékek sem lesznek torzítottak.

A következőképpen ellenőrizhetjük, hogy a kialakított skálaértékek milyen mértékben tükrözik a szakok közötti preferenciakülönbségeket. A skálán mért távolságok mátrixát jelöljük **D**-vel, ahol d_{ij} cella értéke a következőképpen számítandó: $d_{ij} = z_i - z_j, i \neq j$. Ezután határozzuk meg **P** mátrixot úgy, hogy annak elemei p_{ij} a d_{ij} helyen vett standard normális eloszlás értékei legyenek. Az így kapott értékeket a skálaértékekből becsült valószínűségi értékeknek nevezzük. Egy χ^2 -próbalával pedig azt vizsgálhatjuk, hogy a **PCM** és **P** mátrix értékei mennyire térnek el egymástól/mennyire illeszkednek egymáshoz. Az eltérést egy úgynevezett stresszmutatóval jellemezhetjük, amely a következőképpen számítható:

$$\text{stressz} := \sqrt{\frac{\sum_{i=1}^m \sum_{j=1, i \neq j}^m (pcm_{i,j} - p_{i,j})^2}{\sum_{i=1}^m \sum_{j=1, i \neq j}^m pcm_{i,j}^2}}. \quad (6)$$

A kis érték jó illeszkedést mutat, ami azt jelenti, hogy a skálaértékekből kis hibával vissza tudjuk állítani az eredeti **PCM** mátrixot. Ez a lépés egyfajta ellenőrzése is lehet a módszer által szolgáltatott eredményeknek. A módszer során kapott z_1, z_2, \dots, z_m értékek egy skálatranszformáció után minősítő skála kialakítására is alkalmasak lehetnek.

PAGERANK-MÓDSZER • A PageRank a Google internetes keresőmotor legfontosabb eleme. Larry Page és Sergey Brin (a Google alapítói) fejlesztették ki 1998-ban a Stanford Egyetemen (*Page-Brin* [2009]).

A Google arra a feltételezésre épít, hogy a honlapok készítői általában azokra az oldalakra linkelnek a saját lapjukról, amelyeket jónak tartanak, vagyis minden hiperlink felfogható egy-egy szavazatként a céloldalra. Minél több szavazatot kap egy oldal, annál fontosabb, de azt is figyelembe kell venni, hogy a szavazatot leadó oldal mennyire fontos. (Ez egy rekurzív definíció: az a fontos oldal, amire fontos

oldalak mutatnak.) A PageRank a fontosság számszerűsítése, amelyet $PR(E)$ -vel jelölünk. Az alapötlet szerint kezdetben minden oldalnak egy egység szavazata van, amelyet egyenlően szétoszt azok között az oldalak között, amelyekre hivatkozik, és a más oldalaktól kapott szavazatokat is ugyanígy továbbosztja.

Ez a módszer is használható hiányos sorrendek kezelésére (*Kóczy-Nichifor* [2012]). Itt az oldalak helyett intézményeket tekintünk, a linkek helyett pedig hallgatói jelentkezéseket. Eredményként minden csúc (intézmény) kap egy súlyszámot (szavazatot), amely alapján itt is képezhető egy sorrend. Az eredményeinket e módszer által meghatározott sorrendekkel vetettük össze.

GENETIKUS ALGORITMUSOK ALKALMAZÁSA • A genetikus algoritmusok megalkotása során abból indulunk ki, hogy meg kell határoznunk m szak esetén egy $\{1, 2, \dots, m\}$ számok lehetséges \mathbf{b} vektorban leírt permutációját. Bármely \mathbf{b} vektorra meg tudjuk határozni $h(\mathbf{M}, \mathbf{b})$ hibafüggvényt, hiszen ez nem jelent mást, mint hogy \mathbf{M} mátrixot kell \mathbf{b} szerint átrendezni, és ezután az alsó háromszögben szereplő értékeket kell összeadnunk. Ekkor \mathbf{b} vektorban szereplő tagok legyenek egy m csúcsot tartalmazó gráf csúcspontjai. Ha a csúcsokat \mathbf{b} vektor szerint járjuk be, akkor egy Hamilton-utat kapunk, hiszen \mathbf{b} vektor az $\{1, 2, \dots, m\}$ számok egy lehetséges permutációja, így \mathbf{b} vektorban minden szám, vagyis minden csúcspont csak egyszer szerepel. Legyen egy ilyen Hamilton-út költsége a $h(\mathbf{M}, \mathbf{b})$ hibafüggvény értéke. Ekkor a feladat az, hogy meg kell találni azt a \mathbf{b} vektort (más szavakkal: azt a Hamilton-utat), amelyre $h(\mathbf{M}, \mathbf{b})$ értéke minimális, vagyis egy utazóügynök-problémához jutottunk.

Genetikus algoritmusokat széles körben alkalmaznak utazóügynök-problémák megoldására. Bár itt nem az élék költsége adja a Hamilton-út költségét, ebben az esetben is meghatározható valamennyi Hamilton-útra a hibafüggvény értéke, és ezzel a Hamilton-út költsége is. A genetikus algoritmusok alkalmazása során nem használjuk ki az élék költségének ismeretét, csak a Hamilton-út költségének ismerete szükséges.

Az algoritmus kiinduló lépése egy véletlen populáció generálása. Minden egyes permutáció egy egyed lesz a populációban. A jósági függvény (*fitness function*) a hibafüggvény értéke egy adott permutációra. A szelekció során az egyedhez tartozó kisebb hibafüggvényérték nagyobb esélyt biztosít a következő generációba való kiválasztásra. A mutáció során két vektorelem véletlenszerűen helyet cserél egymással (például $m = 10$ -re $[1\ 2\ 3\ 4 \leftrightarrow 10\ 5\ 6\ 7\ 8\ 9]^T \rightarrow [1\ 2\ 3\ 10\ 4\ 5\ 6\ 7\ 8\ 9]^T$), míg a rekombináció két szekvencia olyan közös részét jelenti, amely maga is permutációja $\{1, 2, \dots, m\}$ -nek (például $m = 10$ -re $[1\ 2\ 3\ 4\ 6\ 10\ 5\ 7\ 8\ 9] + [3\ 2\ 1\ 6\ 4\ 10\ 9\ 8\ 7\ 5] \rightarrow [1\ 2\ 3\ 6\ 4\ 10\ 9\ 8\ 7\ 5]$). A rekombináció, mutáció és szelekció operátorok beállítása során Heinrich Braun cikkét követtük (*Braun* [1991]).

CSEMÓPONTOK AGGREGÁLÁSA • A hallgatók szakokat jelölnek meg a jelentkezési lapokon, de nemcsak szakok, hanem karok vagy intézmények közötti sorrendekre is kíváncsiak lehetünk.

Rendeljünk minden egyes szakhoz egy természetes számot, amely lehet egy intézmény/kar/régió stb. kódja! Azokat a csomópontokat tekintjük a továbbiakban új cso-

mópontnak, amelyekhez ugyanazt az intézményi/kari/régiós kódot rendeltük. Az új preferenciagráfból azokat az éleket elhagyjuk, amelyek ugyanazon intézményhez/karhoz stb. tartozó szakok között szerepelnek. A többi élt megtartjuk.

Példaképpen tegyük fel, hogy $m = 4$ a vizsgált szakok száma. Az i -edik hallgató jelentkezési sorrendje $\mathbf{a}^i = [1, 2]^T$. Tegyük fel továbbá, hogy 1-gyel és 2-vel kódolt szakok az I-es, míg a 3-mal és 4-gyel kódolt szakok a II-es intézményhez tartoznak. Ekkor az egyéni preferenciagráf csúcspontjai a 3. táblázatban látható módon aggregálhatók.

3. táblázat

Csomópontok (szakok) aggregálása

Szakok közötti preferenciák	Szakok aggregálása	Intézmények közötti preferenciák																																																											
	<p>Preferenciagráf</p>																																																												
<p>Preferenciamátrix</p> <table border="1" style="margin: auto;"> <tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td></tr> <tr><td>1</td><td>-</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td></td><td>-</td><td>1</td><td>1</td></tr> <tr><td>3</td><td></td><td></td><td>-</td><td>0,5</td></tr> <tr><td>4</td><td></td><td></td><td>0,5</td><td>-</td></tr> </table>		1	2	3	4	1	-	1	1	1	2		-	1	1	3			-	0,5	4			0,5	-	<p>Preferenciamátrix</p> <table border="1" style="margin: auto;"> <tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td></tr> <tr><td>1</td><td>-</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>2</td><td></td><td>-</td><td>1</td><td>1</td></tr> <tr><td>3</td><td></td><td></td><td>-</td><td>0,5</td></tr> <tr><td>4</td><td></td><td></td><td>0,5</td><td>-</td></tr> </table>		1	2	3	4	1	-	1	1	1	2		-	1	1	3			-	0,5	4			0,5	-	<table border="1" style="margin: auto;"> <tr><td></td><td>I</td><td>II</td></tr> <tr><td>I</td><td>-</td><td>4</td></tr> <tr><td>II</td><td></td><td>-</td></tr> </table>		I	II	I	-	4	II		-
	1	2	3	4																																																									
1	-	1	1	1																																																									
2		-	1	1																																																									
3			-	0,5																																																									
4			0,5	-																																																									
	1	2	3	4																																																									
1	-	1	1	1																																																									
2		-	1	1																																																									
3			-	0,5																																																									
4			0,5	-																																																									
	I	II																																																											
I	-	4																																																											
II		-																																																											

A 3. táblázatbeli példán egyes és kettes szakot hármas és négyes szaknál is jobban előbbre sorolta a hallgató, hiszen sem a hármas, sem a négyes szakra nem adott be jelentkezési lapot. E példán látható, hogy ha két szak ugyanahhoz az intézményhez tartozik, akkor azok az élek elvesznek, amelyek a szakok közötti preferenciákat jelölik.

Mind az egyéni, mind az aggregált preferenciamátrixból képezhető a fenti módon vett egyéni, illetve intézményi preferenciamátrix. Könnyen látható, hogy a szakokra vonatkozó egyéni preferenciamátrixokból képzett intézményi preferenciamátrixok aggregálása és a szakokra vonatkozó aggregált preferenciamátrixokból képzett aggregált intézményi preferenciamátrix ugyanazt az eredményt adja. Vagyis a csomópontok és a mátrixok aggregálása felcserélhető.

Felmerülhet a kérdés, hogy miért van szükség ilyen módszerekre, ha pusztán a jelentkezők száma alapján is sorrendbe tudjuk rakni az intézményeket. Példaként tekintsünk két hallgatót és három szakot. Az első hallgató jelentkezési sorrendje legyen A, B , a másodiké C, B . Ekkor a B szakra jelentkeztek a legtöbben, pedig nem tűnik úgy, mintha ez lenne a leginkább preferált.

Egyszerű sorrend a jelentkezők száma alapján *versus* preferencia-sorrend

Miért nem megfelelő a jelentkezők száma alapján rangsorolni az intézményeket (illetve karokat, szakokat)? Ez a kérdés – amelyet kollégánk, *Abonyi János* tett fel számunkra – nemcsak teljesen jogos, hanem rávilágít a nemzetközi felsőoktatási rangsorok egyik alapvető módszertani problémájára is (*Török [2006]*). Mégpedig, hogy az egymással versengő intézmények mérete a felsőoktatási teljesítmény-összehasonlítások számos elemét befolyásolja. A teljesítménymutatók egy részét ugyanis nem fajlagos mutatóként kezeljük, ezért pedig értelemszerű, hogy a nagyobb intézménynek nemcsak több publikációra, szabadalomra van lehetősége/kapacitása, mint a kisebbnek, hanem például – méretgazdaságossági okokból – kedvezőbb oktató/kiszolgáló személyzet arány (illetve a PhD-fokozatot szerettek magasabb aránya) mellett is tud működni, mint a kisebbek. A világon ezért több helyen – Kínában, Szingapúrban, Mexikóban, Oroszországban és máshol – megfigyelték, hogy az egyetemek koncentrációs folyamata számottevően javította az adott ország nemzetközi felsőoktatási rangsorpozícióit.

Szűkebb témánkra visszatérve: az egyszerű, formális válasz az lenne, hogy a ke-retszámok, az intézmény mérete erősen befolyásolja a jelentkezők számát, és nem feltétlenül tükrözi a jelentkezők preferenciáit. Természetesen a jelentkezők száma alapján rendkívül egyszerű sorrendet felállítani, a sorrend tartalma is közérthető, de interpretálhatósága már kérdéses, hiszen a jelentkezők számában számos olyan tényező játszhat szerepet, ami nem a hallgató motivációit, illetve a választott hely érdemi pozícióját tükrözi a preferencia-sorrendben, azaz a jelentkezések száma valamivel erősebben függ az intézmény méretétől, mint a preferencia-sorrend.

Az erősorrend és a preferencia-sorrend sok esetben akár meg is egyezhet. Ugyanakkor, ha adott két diák és három szak. Az első diák sorrendje *A, B*, és a másodiké *C, B*. Ekkor a *B* szakra jelentkeztek a legtöbben, pedig mindkét diák mást részesít előnyben.

A két módszer más jelenség visszatükrözésére alkalmas, más kérdésre ad választ. Ezért tehát nem is lehet közöttük értéksorrendet felállítani. Mint azt a bevezetőben is leszögeztük, célunk preferenciákat vizsgálni, és mint illusztrációs példánk mellett a későbbi, a jelentkezési adatokból kapott eredmények is mutatják, a két sorrend eltérhet, sőt el is tér egymástól. Ezért érdemes a bonyolultabb, de árnyaltabb preferencia-sorrendet létrehozó módszerek valamelyikét alkalmazni.

Robusztusságvizsgálat

A robusztusságvizsgálat során arra voltunk kíváncsiak, hogy szimulált preferencia-sorrendeket tekintve az egyes módszerek milyen eredményekre vezetnek. Vizsgáltuk azt az esetet is, amikor egy hallgató valamennyi intézménybe beadja a jelentkezését, tehát teljesen tekinthető a preferencialista, másrészt azzal az esettel is foglalkoztunk, amikor a hallgatók csak néhány intézményt jelölnek meg. Ezen belül is kétfajta módszerrel foglalkoztunk. Az első módszer során feltételeztünk egy elméleti sorrendet, amely a legtöbbször szerepel a szimulált adatbázisban, illetve olyan preferencia-

sorrendeket is megvizsgáltunk, amelyek egy euklideszi értelemben vett d távolságra vannak ettől az elméleti sorrendtől. Minél nagyobb volt az elméleti sorrend és a véletlenül meghatározott sorrend távolsága, annál kisebb mértékben szerepelt a szimulált adatbázisban. Ezzel a vizsgálattal arra voltunk kíváncsiak, hogy az egyes módszerek visszaadják-e az elméleti sorrendet. Az elméleti sorrendtől eltérő preferencia-sorrendek zajként is felfoghatók. A robusztusság itt pedig a zajérzékenységet jellemzi.

A másik vizsgálat ehhez hasonló. Itt feltételeztük, hogy a hallgató az intézmény hasznossága alapján választ intézményt. Ezeket a hasznosságokat bizonyos szempontok, tényezők alapján határozza meg. Azt az intézményt fogja megjelölni, amelyik hasznossága a legnagyobb. A jelentkezését pedig e hasznosságok alapján rakja sorrendbe. A preferencia-sorrendek kialakításánál itt is feltételezünk egy olyan e hibatagot, amely befolyásolhatja, módosíthatja a preferencia-sorrendeket. Itt is arra vagyunk kíváncsiak, hogy e zajtényező mennyiben befolyásolja az elméleti sorrend alapján felállított, hasznosságok szerint rendezett sorrend megtalálását.

A robusztusságvizsgálat során feltételeztünk egy $\mathbf{b} \in \mathbb{N}^m$ sorrendet. Arra voltunk kíváncsiak, ha egy olyan adatbázist generálunk, amelyben \mathbf{b} perturbációi, valamint adott gyakorisággal más sorrendek is szerepelnek, akkor ebből visszakapjuk-e becsült sorrendként a \mathbf{b} vektort.

Két módszerrel vizsgáltuk a bemutatott módszerek robusztusságát. Mindkét módszerhez felhasználtuk a preferenciavektorok euklideszi távolságát.

Legyen m a vizsgált szakok/intézmények, n pedig a hallgatói jelentkezések száma; az \mathbf{A} a hallgatói jelentkezéseket tartalmazó vektorokból álló halmaz, ahol $\mathbf{a}^i \in \mathbf{A}$ esetén $\mathbf{s}^i \in \mathbb{R}^m$ az i -edik hallgató preferenciavektora ($i = 1, 2, \dots, n$). Ha $\mathbf{b} \in \mathbb{N}^m$ minden szak/intézmény csak egyszer szerepel, akkor \mathbf{b} vektorhoz tartozó \mathbf{s}^b preferenciavektor egyben a \mathbf{b} vektort adja ($\mathbf{s}^b = \mathbf{b}$).

Ekkor az i -edik hallgató preferenciavektorának euklideszi távolsága \mathbf{b} vektortól a következőképpen adható meg:

$$d_{i,\mathbf{b}} = \sqrt{\sum_{j=1}^m (s_j^i - b_j)^2} = \|\mathbf{s}^i - \mathbf{b}\|_2, \text{ ahol } \mathbf{s}^i = [s_1^i, \dots, s_m^i]^T, \mathbf{b} = [b_1, \dots, b_m]^T, i = 1, \dots, n. \quad (7)$$

A (7) formula segítségével meg lehet határozni tetszőleges $m_i \leq m$ hosszúságú jelentkezési sorrendből meghatározott egyéni preferenciavektornak a \mathbf{b} vektorhoz képesti euklideszi távolságát.

A (7) távolságfüggvény felhasználásával, n hallgatói jelentkezést generálhatunk, ahol például a hallgatói jelentkezéseket úgy szimulálhatjuk, hogy leggyakrabban \mathbf{b} vektor forduljon elő \mathbf{A} halmazban, minden más preferenciavektor egy $f(x)$ függvényit követve egyre kisebb mértékben fordul elő.

Másik javasolt módszerünkben a hallgatói választást szimuláljuk egy multinomiális logit modellel. Tegyük fel, hogy a hallgató t szempont alapján választ egy adott szakot. Feltesszük, hogy az i -edik hallgató ($i = 1, 2, \dots, n$) ezek alapján rakja sorrendbe az m szakra adott jelentkezését, vagyis $\mathbf{a}^i = [a_1^i, a_2^i, \dots, a_m^i]^T$ esetén $U_{a_1^i} < U_{a_2^i} < \dots < U_{a_m^i}$, ahol $U_{a_j^i}$ az i -edik hallgató j -edik helyen megjelölt a_j^i -edik szak hasznosságát jelöli t különböző szempont figyelembevételével. Az egyszerűség kedvéért először feltesz-

szük, hogy a hallgató valamennyi szakot sorrendbe rakja. A szakokhoz rendelt hasznosság a következőképpen jellemezhető:

$U_{a_j} = V_{a_j} + \varepsilon_{a_j} = \beta_{1,a_j}x_{1,a_j} + \beta_{2,a_j}x_{2,a_j} + \dots + \beta_{t,a_j}x_{t,a_j} + \varepsilon_{a_j}$, ahol ε_{a_j} logisztikus eloszlású véletlen változó. E modell szerint az i -edik hallgató az j -edik helyen megjelölt a_j^i -edik szak U_{a_j} hasznossága x_{1,a_j} , x_{2,a_j} , ..., x_{t,a_j} szempontokat β_{1,a_j} , β_{2,a_j} , ..., β_{t,a_j} súllyal veszi figyelembe döntésénél. Egy végső $\mathbf{b} = [b_1, \dots, b_m]^T$ sorrend valószínűsége a hasznosságok alapján ebben a modellben a (8) szerint számítható:

$$P(U_{b_1} < U_{b_2} < \dots < U_{b_m}) = \prod_{j=1}^m \frac{\exp(\mathbf{v}_{b_j})}{\sum_{l=1}^m \exp(\mathbf{v}_{b_l})}, \text{ ahol} \quad (8)$$

$$\mathbf{V}_{b_i} = [V_{a_1^i}, V_{a_2^i}, \dots, V_{a_m^i}]^T, \quad i = 1, 2, \dots, m.$$

A módszer segítségével, a súlyok megválasztásával generált hallgatói preferenciá-sorrendekből meghatározzuk a legvalószínűbb végső sorrendet, valamint a korábban bemutatott módszerekkel is kiszámítjuk a becült sorrendeket, majd ezeket összehasonlítjuk.

Eredmények

Hosszas töprengés után választottuk ki azokat a szakokat, amelyeken konkrét jelentkezési adatok alapján módszereinket bemutatjuk. Olyan szakokat szerettünk volna választani, amelyek elég sok intézményben megtalálhatók, másrészt elég nagy létszámmal működnek, azaz a rendelkezésre álló adatok mennyisége megnyugtatóan nagy. Ennek alapján természetesen esett a választás néhány népszerű szakra a gazdálkodás és az informatika területén. Ezt a kutatói választást kizárólag a jelzett szakok kiemelkedő országos népszerűsége motiválta, azaz a minél szélesebb körű összehasonlíthatóság. Intézményi vagy egyéb érdekek tehát nem játszottak szerepet benne. Az eredményeket is kizárólag a preferenciavizsgálat tekintetében ismertetjük, azaz ebben a tanulmányban elsősorban módszertani jellegű következtetésekre törekedtünk. Három nappali tagozatos, államilag finanszírozott szakot vizsgáltunk a 2011. évi jelentkezések alapján: gazdálkodás és menedzsment, gazdasági informatikus és műszaki informatikus alapképzések. Elemeztünk egy mesterszakot (levelező mesterszakot) is. Egyes táblázatokban egyes intézmények neve és az azokra jellemző adatok helyett XX, illetve nnn került közlésre, elkerülendő a közlésből fakadó esetleges vitákat, jogi következményeket.¹ Az adatok forrása: Felvi-honlap (<http://felvi.hu>).

¹ A táblázatokban szereplő rövidítések a következő intézményeket takarják (zárójelben a korábbi rövidítések): ÁVF: Általános Vállalkozási Főiskola; BCE: Budapesti Corvinus Egyetem (BKÁE); BGF: Budapesti Gazdasági Főiskola; BKF: Budapesti Kommunikációs és Üzleti Főiskola; BME: Budapesti Műszaki és Gazdaságtudományi Egyetem; DE: Debreceni Egyetem; DF: Dunaújvárosi Főiskola; EJF: Eötvös József Főiskola; EKF: Eszterházy Károly Főiskola; ELTE: Eötvös Loránd Tudományegyetem; GDF: Gábor Dénes Főiskola; HJF: Harsányi János Főiskola; IBS: IBS Nemzetközi Üzleti Főiskola; KF: Kecskeméti Főiskola; KJF: Kodolányi János Főiskola; KRF: Károly Róbert Főiskola; ME: Miskolci Egyetem; MÜTF: Modern Üzleti Tudományok Főiskola; NYF: Nyíregyházi Főiskola; NYME: Nyugat-

A módszerek által meghatározott rangsorok összehasonlítása a 2011. évi jelentkezési adatok alapján

A gazdasági informatika nappali tagozatos államilag finanszírozott alapszakra 2011-ben jelentkezett hallgatók preferencia-sorrendjét vizsgálva, valamennyi módszer ugyanarra az eredményre vezetett (4. táblázat). Bár a 12. helyre sorolt Nyugat-magyarországi Egyetemet 115-en, míg az első helyen szereplő Szegedi Tudományegyetemet 506-an jelölték meg, ez a különbség a preferenciákban nem mutatkozott meg ilyen élesen. Mind a $PR(E)$, mind a páros összehasonlításra alapuló Z értékek, mind pedig az átlagos rangértékek közel esnek egymáshoz. Az inhomogenitási index értéke pedig 47,08 százalék. A stresszteszt alacsony értéke azt mutatja, hogy a páros összehasonlítás eredményeképpen kapott Z értékekből nagy pontossággal visszaállítható az eredeti preferenciamátrix.

4. táblázat

Gazdasági informatika nappali tagozatos államilag finanszírozott alapszakra 2011-ben jelentkezett hallgatók preferencia-sorrendje különböző módszerek szerint

Helyezés	Intézmény	Rangösszeg (átlagos rangérték)	Páros összehasonlítás		PageRank $PR(E)$	Jelentkezés	
			Z	$\Phi^{-1}(Z)$		összes	első helyen
1.	SZTE	6,0901	0,0860	0,5343	0,3065	506	467
2.	BCE	6,1234	0,0790	0,5315	0,3021	477	453
3.	DE	6,1386	0,0758	0,5302	0,3057	483	435
4.	BGF	6,2256	0,0588	0,5234	0,2762	442	286
5.	XX	6,3827	0,0247	0,5099	0,2776	362	307
6.	PTE	6,4638	0,0077	0,5031	0,2778	323	270
7.	XX	6,6927	-0,0402	0,4840	0,2699	207	168
8.	ME	6,7193	-0,0458	0,4817	0,2791	196	153
9.	KRF	6,7238	-0,0467	0,4814	0,2901	194	154
10.	DF	6,7536	-0,0529	0,4789	0,2940	181	136
11.	ZSKF	6,8170	-0,0663	0,4736	0,3081	150	77
12.	NYME	6,8835	-0,0802	0,4680	0,2731	115	81

Hibaérték $h(\mathbf{M}, \mathbf{b})$: 92 819.

Inhomogenitási index (I): 47,08 százalék.

Stresszteszt: 0,0043.

A jelentkezések alapján felállított sorrend alapján a Debreceni Egyetem megelőzné a Budapesti Corvinus Egyetemet, ugyanakkor valamennyi olyan módszer a Budapesti

magyarországi Egyetem; OE: Óbudai Egyetem (BMF); PE: Pannon Egyetem (VE); PPKE: Pázmány Péter Katolikus Egyetem; PTE: Pécsi Tudományegyetem; SE: Semmelweis Egyetem; SZE: Széchenyi István Egyetem; SZIE: Szent István Egyetem; SZTE: Szegedi Tudományegyetem; TPF: Tomori Pál Főiskola; TSF: Tessedik Sámuel Főiskola; ZSKF: Zsigmond Király Főiskola.

6. táblázat
Gazdálkodás és menedzsment nappali tagozatos államilag finanszírozott alapszakra 2011-ben jelentkeztetett hallgatók preferencia-sorrendje

Helyezés	Genetikus algoritmus		Rangösszeg		Páros összehasonlítás			PageRank	
	intézmény	összes jelentkező	intézmény	átlagos rangérték	intézmény	Z	$\Phi^{-1}(Z)$	intézmény	P(R)
1.	BGF	3292	BGF	11,7555	BGF	0,3037	0,619322	BGF	0,2523
2.	BCE	1956	BCE	13,4404	BCE	0,1494	0,559381	BCE	0,2265
3.	BME	1440	BME	13,9031	BME	0,0862	0,534346	BME	0,2118
4.	SZIE	1359	SZIE	14,0637	SZIE	0,0737	0,529375	SZIE	0,2100
5.	ÁVF	1139	ÁVF	14,2798	ÁVF	0,0485	0,519341	ÁVF	0,2050
6.	PTE	1048	DE	14,3616	PTE	0,0385	0,515355	PTE	0,2033
7.	DE	958	PTE	14,3832	DE	0,0281	0,511209	DE	0,1996
8.	BKF	906	BKF	14,4302	BKF	0,0222	0,508856	BKF	0,1987
9.	OE	767	OE	14,5935	OE	0,0045	0,501795	OE	0,1954
10.	SZE	734	ZSKF	14,6143	SZE	0,0006	0,500239	SZE	0,1948
11.	ZSKF	717	SZE	14,6363	ZSKF	-0,0034	0,498644	SZTE	0,1933
12.	SZTE	639	SZTE	14,7673	SZTE	-0,0100	0,496011	ZSKF	0,1933
13.	ME	532	ME	14,9058	ME	-0,0237	0,490546	ME	0,1910
14.	MÜTF	508	XX	14,9564	MÜTF	-0,0270	0,489230	MÜTF	0,1907
15.	XX	469	MÜTF	14,9598	XX	-0,0310	0,487635	PE	0,1894
16.	NYME	406	NYME	15,0654	NYME	-0,0389	0,484485	NYME	0,1885
17.	NYF	358	NYF	15,0942	NYF	-0,0449	0,482094	NYF	0,1870
18.	KJF	292	KJF	15,1850	KJF	-0,0532	0,478786	KJF	0,1857

Corvinus Egyetemet rangsorolja előrébb, amely a preferencia-sorrend alapján állítja fel a helyezéseket. Még szembevetőbb a jelentkezések és a preferencia-sorrendek eltérése abban a példában, ahol a műszaki informatika szakra való jelentkezéseket vizsgáltuk (5. táblázat).

2011-ben műszaki informatika alapszakra az Óbudai Egyetemre 1905, míg a Budapesti Műszaki és Gazdaságtudományi Egyetemre 1776 jelentkezés érkezett. A PageRank módszerrel kivül azonban minden más módszer a Budapesti Műszaki és Gazdaságtudományi Egyetemet helyezi az első helyre ezen a listán. A hallgatók általában előrébb preferálták a műszaki egyetemet, mint az Óbudai Egyetemet. Ez látszik az elsőhelyes jelentkezéseken is. Ugyanakkor pusztán az elsőhelyes jelentkezések nem adnak teljes képet a hallgatók preferenciáiról, hiszen ekkor csak az első helyre való jelentkezést vesszük figyelembe. A 7. táblázat egy, különösen sok intézményben folyó (gazdálkodási és menedzsment) képzést mutat. A táblázatból kitűnik, hogy amennyiben csak az elsőhelyes jelentkezéseket vennénk figyelembe, akkor merőben más sorrendet kapnánk, mint preferencia-sorrendet meghatározó módszerekkel. Az utóbbi módszerek eredményei ugyanakkor alig térnek el egymástól.

A legkisebb hibaértéket a genetikus algoritmusok alkalmazásával kapjuk. Ettől csak kismértékben tér el a páros összehasonlítás módszere. A páros összehasonlítás módszerével nem pusztán sorrendet kapunk, hanem az egyes intézmények egymáshoz képesti „relatív távolságát” is. Tehát nemcsak azt tudjuk meg, hogy egy intézményt előrébb soroltak egy másik intézményhez képest vagy sem, hanem azt is, hogy mekkora a preferenciatávolság. Az átlagos rangérték ugyancsak fontos információ szolgál. A 7. táblázatban a mesterszakra jelentkezéseket elemezzük.

7. táblázat

Mesterszakra 2011-ben jelentkezett hallgatók preferencia-sorrendje

Helyezés	Intézmény	Rangösszeg átlagos rangérték	Páros összehasonlítás		PageRank PR(E)	Jelentkezés	
			Z	$\Phi^{-1}(Z)$		összesen	első helyen
1.	XX	3,1593	0,3925	0,6527	0,5150	nnn	nnn
2.	BME	3,8488	0,0822	0,5328	0,3909	102	94
3.	XX	3,9140	0,0647	0,5258	0,3876	nnn	nnn
4.	DE	4,2640	-0,0567	0,4774	0,3529	54	40
5.	PTE	4,3977	-0,1153	0,4541	0,3329	35	29
6.	BCE	4,5791	-0,1759	0,4302	0,3162	14	9
7.	SZTE	4,6093	-0,1914	0,4241	0,3107	9	7

Hibaérték $h(\mathbf{M}, \mathbf{b})$: 3810, inhomogenitás (I): 42,19 százalék, stresszteszt: 0,0038.

A mesterszakra jelentkezések vizsgálatában valamennyi módszer ugyanazt a sorrendet adja. Mindegyik egy kicsivel többet árul el a „miértéről”. Amíg a genetikus algoritmus „csak” a preferencia-sorrendeket kapjuk meg, addig a PageRank már egy arányskálán mért pontszám (úgynevezett *score* érték). A páros összehasonlítás eredményeként pedig az intézmények közötti távolságokat is láthatjuk. Amíg 2–7.

helyezett intézmények átlagos rangértékei, Z és $PR(E)$ értékei alig különböznek egymástól. A első helyezett intézmény $score$ értékei: átlagos rangértékek, $PR(E)$, Z értékei „kiugranak” ebből a mezőnyből.

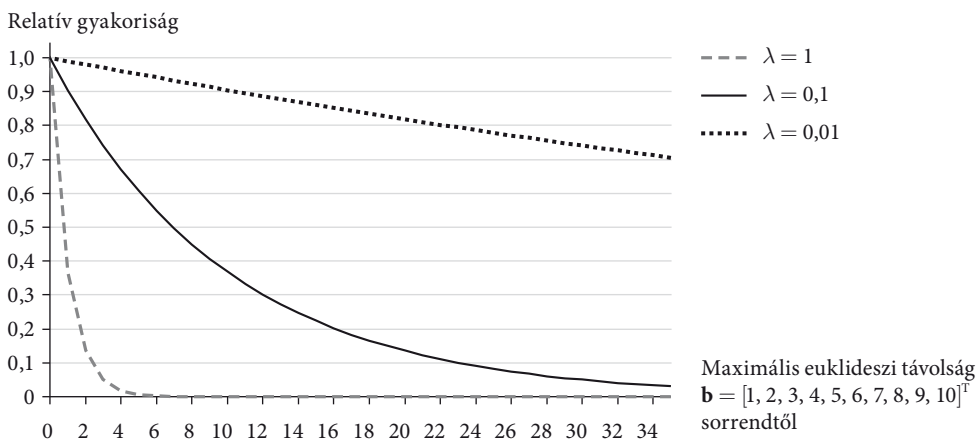
A robusztusságvizsgálat eredményei

A robusztusságvizsgálat során az előzőkben tárgyalt módszerek szerint generáltunk sorrendeket. Itt nem volt célunk, hogy a generált sorrendek a jelentkezési sorrendek minden szempontját visszaadják. A jelentkezések néhány jellemzőjét azonban a sorrendek generálásánál is figyelembe vettük. Az egyik ilyen jellemző az egy hallgató által elküldött jelentkezések száma. 2012-ben 161 731-en jelentkeztek valamilyen felsőoktatási intézménybe. A hallgatók közel kétharmada maximum három helyre jelentkezett (66,2 százalék). Ennek oka az lehet, hogy a jelentkezési lapon maximum három intézményt lehetett megjelölni, minden további jelentkezésért külön jelentkezési lapon kellett fizetni. Több mint 10 intézményt a jelentkezők kevesebb mint 1 százaléka jelölt meg az 51 intézménybe, amelyet megjelölhettek, így a vizsgálat során mi is maximum 10 hosszúságú jelentkezési sorrendeket tekintettünk.

Az első vizsgálat során feltételeztük, hogy egy adott $\mathbf{b} = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]^T$ a valódi sorrend. Feltételeztük, hogy olyan hallgatói sorrendek, melyek euklideszi távolságát a (7) képlet szerint számoltuk, $\lambda e^{-\lambda x}$ valószínűséggel jelenhetnek meg a hallgatói jelentkezések között. Itt λ csillapítási tényező, míg x a \mathbf{b} és a generált i -edik hallgató preferencia-sorrendje közötti euklideszi távolság. Csillapítási tényezőnek $\lambda_1 = 1$; $\lambda_2 = 0,1$; $\lambda_3 = 0,01$ értékeket választottunk. Az 1. ábrából és a 8. táblázatból látható, hogy minél nagyobb a csillapítási tényező értéke, annál kisebb mértékben jelenik meg a mintában az elméleti sorrendtől különböző preferencia-sorrend. A csillapítási tényezővel tehát a „zaj” mértékét lehet befolyásolni. Ha kicsi a csillapítási tényező értéke, akkor elméleti sorrendtől nagyon különböző sorrendek is megjelennek a mintában.

1. ábra

Hallgatói jelentkezések generálása



8. táblázat

Hallgatói jelentkezések generálása

(a módszerek és különböző csillapítások esetén szolgáltatott eredményeik)

Helyezés	Átlagos rang-érték	Z	$\Phi^{-1}(Z)$	PR(E)	Helyezés	Átlagos rang-érték	Z	$\Phi^{-1}(Z)$	PR(E)
$\lambda = 1$					$\lambda = 0,1$				
1.	1,0420	1,5280	0,9367	0,9593	1.	2,4027	1,0714	0,8580	0,5274
2.	1,9683	1,1255	0,8698	0,2780	2.	2,5293	1,0582	0,8550	0,5136
3.	2,9914	0,7320	0,7679	0,0489	3.	3,0530	0,8622	0,8057	0,4412
4.	3,9983	0,3765	0,6467	0,0086	4.	3,9689	0,5401	0,7054	0,3460
5.	5,0002	0,0595	0,5237	0,0016	5.	5,4094	-0,0142	0,4943	0,2303
6.	6,0028	-0,4560	0,3242	0,0004	6.	6,3460	-0,3093	0,3785	0,1827
7.	7,0125	-0,8058	0,2102	0,0001	7.	7,0188	-0,5091	0,3054	0,1506
8.	8,0418	-1,0965	0,1364	0,0000	8.	7,3132	-0,5594	0,2879	0,1475
9.	8,9428	-1,4631	0,0717	0,0000	9.	8,3234	-0,9567	0,1694	0,0894
10.	9,9580	-2,0231	0,0215	0,0000	10.	8,6353	-1,1833	0,1183	0,0699
<i>I</i> (százalék): 0,25, stresszteszt: 0,0055.					<i>I</i> (százalék): 16,11, stresszteszt: 0,0135.				
$\lambda = 0,01$									
1.	4,1500	0,3668	0,6431	0,3971					
2.	4,1947	0,3526	0,6378	0,3945					
3.	4,4722	0,2956	0,6162	0,3717					
4.	5,0917	0,1334	0,5531	0,3264					
5.	5,2027	0,0818	0,5326	0,3262					
6.	5,3167	0,0478	0,5190	0,3238					
7.	5,7594	-0,0656	0,4739	0,2888					
8.	6,0257	-0,1331	0,4470	0,2735					
9.	6,7814	-0,3396	0,3671	0,2282					
10.	8,0055	-0,7397	0,2298	0,1432					
<i>I</i> (százalék): 45,4, stresszteszt: 0,0234.									

Bár a csillapítási tényező csökkentésével mind az inhomogenitás, mind a stresszteszt értéke növekszik, valamennyi módszer visszaadta a $\mathbf{b} = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]^T$ sorrendet.

A következő vizsgálatban logit modell segítségével generáltunk hasznossági értékeket. Az előzőkben nem volt célunk, hogy az együtthatók értéke a valóságot tükrözze, hiszen ekkor csak a módszerünk robusztusságát szeretnénk volna vizsgálni. Mivel azonban kutatásunk folytatásában éppen arra keressük a választ, hogy milyen tényezők befolyásolják a jelentkezéseket, nem szeretnénk volna teljesen elrugaszkodni a valóságtól sem. Feltételeztük, hogy a hasznosságot három tényező

befolyásolja leginkább: a jelentkező lakóhelye és az intézmény távolsága (*distance*), a tanulás költségei (*compulsory fees*), valamint az intézmény elismertsége (*faculty awards*). Feltételeztük, hogy a távolság és a tanulás költségei negatívan, míg az intézmény elismertsége pozitívan hat a hasznossági értékekre. (E három változó értékét *a priori* információ hiányában egységnyinek választottuk, míg a β együtthatók értékeit a fenti megszorításokkal véletlenszerűnek.) Ezek alapján a hasznossági értékek alakulását a 9. táblázat mutatja.

9. táblázat

Hallgatói jelentkezések generálása

A szimulált együtthatók és hasznossági értékek

Sorrend		Az intézmény távolsága (x_1)	A tanulás költségei (x_2)	Az intézmény elismertsége (x_3)	V
	x	1	1	1	
1.	β_1	-0,05	-0,25	0,18	-0,12
2.	β_2	-0,06	-0,28	0,15	-0,20
3.	β_3	-0,06	-0,17	0,02	-0,22
4.	β_4	-0,16	-0,21	0,10	-0,26
5.	β_5	-0,17	-0,13	0,00	-0,30
6.	β_6	-0,58	-0,27	0,03	-0,82
7.	β_7	-0,55	-0,40	0,04	-0,91
8.	β_8	-0,71	-0,40	0,14	-0,97
9.	β_9	-0,81	-0,29	0,09	-1,00
10.	β_{10}	-1,00	-0,32	0,16	-1,16

A magyar felsőoktatásba 161 731-en jelentkeztek a vizsgált évben. Így vizsgálatunkban mi is ennyi „jelentkező” hasznossági értékeit generáltunk oly módon, hogy $U_{ij} = V_i + \varepsilon_{ij}$ ($i = 1, \dots, 10; j = 1, \dots, 161\ 731$) képletben az ε_{ij} $m = 0, b = 1$ paraméterű logisztikus eloszlást kövessen. Az U_{ij} hasznossági értékeket sorrendbe rakva arra voltunk kíváncsiak, hogy az egyes módszerekkel visszacapjuk-e V_i ($i = 1, \dots, 10$) hasznossági értékeken alapuló $\mathbf{b} = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]^T$ sorrendet (10. táblázat).

A 10. táblázatban feltételeztük, hogy 161 731 hallgató valamennyi intézménybe beadta a jelentkezését. Ha azt is figyelembe vesszük, hogy a valóságban egy hallgató hány helyre adta be a jelentkezését, akkor a hallgató preferencialistájából elhagyjuk az utolsó m darab helyen szereplő intézményt. A megjelölt intézmények elhagyásánál igyekeztünk a 2011. évi jelentkezési adatoknak megfelelően szimulálni a hallgatók intézményekbe való jelentkezését. 2011-ben legalább két intézményt 128 286 hallgató, legalább hármat 107 070 hallgató jelölt meg (lásd a 11. táblázat első 2. oszlopát).

A szimuláció során ezeket az értékeket vettük alapul, így a logit modell által szolgáltatott 10 hosszúságú sorrendeket úgy csonkoltuk, hogy 10 hosszúságú sorrend 2215; minimum 9 hosszúságú 2995; minimum 8 hosszúságú 5132 stb. preferencia-

10. táblázat

A módszerek által szolgáltatott eredmények

Sorrend	Átlagos rangérték	Z	$\Phi^{-1}(Z)$	PR(E)
1.	3,2543	0,6681	0,7480	0,4517
2.	5,0303	0,5406	0,7056	0,4400
3.	5,1982	0,1622	0,5644	0,3405
4.	4,9706	0,1438	0,5572	0,3348
5.	3,5575	0,1234	0,5491	0,3254
6.	8,4251	0,1032	0,5411	0,3060
7.	6,3743	0,0803	0,5320	0,2937
8.	5,3050	-0,2337	0,4076	0,2285
9.	5,1297	-0,7008	0,2417	0,1449
10.	7,7845	-0,8869	0,1876	0,1238

I (százalék): 30,22, stresszteszt: 0,0147.

11. táblázat

A módszerek által szolgáltatott eredmények

Jelentkezések		Eredeti sorrend	Rangösszeg	Páros összehasonlítás	Genetikus algoritmusok	PR
sorrendje	száma					
1.	161 731	1.	1	1.	1.	1.
2.	128 286	2.	2	2.	2.	2.
3.	107 070	3.	3	3.	3.	4.
4.	42 738	4.	4	4.	4.	3.
5.	26 008	5.	5	5.	5.	6.
6.	16 906	6.	7	6.	6.	5.
7.	7 365	7.	6	7.	7.	8.
8.	5 132	8.	9	8.	8.	7.
9.	2 995	9.	8	10.	9.	9.
10.	2 215	10.	10	9.	10.	10.

I (százalék): 33,85, stresszteszt: 0,0459.

sorrend maradjon a szimulált preferencia-sorrendek között. Ez azt jelentette, hogy 2995 – 2215 = 780 esetben az utolsó helyen preferált, 5132 – 2995 = 2137 esetben az utolsó két helyen preferált intézményt elhagytuk a preferencia-sorrendekből, majd a korábban tárgyalt módszereket e csonkolt preferencia-sorrendeket tartalmazó adathalmazon futtattuk.

Az eredményekből kitűnik, hogy a PageRank eredmény adta legkevésbé vissza az eredeti sorrendet. A genetikus algoritmus alkalmazásával visszakaptuk az eredeti sorrendeket, igaz, ez a módszer a leglassúbb heurisztikus módszer, és csak sorrendet

kapunk, míg a többi módszer által szolgáltatott eredmény a sorrendi információk kívül egy pontszámot (*score* értéket) is megad. A páros összehasonlításra alapuló módszer segítségével arra is következtethetünk, hogy a végső sorrendként kapott helyezések milyen távolságra vannak egymástól.

Következtetések és további tervek

Az alkalmazott gráfrepresentáció segítségével információvesztés nélkül tároltuk a hallgatói jelentkezéseket. Az általunk javasolt módszerek alkalmasak arra, hogy a hiányos preferencia-sorrendekként tekinthető hallgatói jelentkezésekből intézményi vagy kari sorrendet képezzünk. Az optimális sorrend megtalálása kombinatorikus probléma, de genetikus algoritmusok használatával tetszőlegesen megközelíthető. Ha egy átlagos rangértéket szeretnénk meghatározni, akkor a rangösszegmódszert, ha pedig az intézmények közötti relatív preferenciatávolságokat is, akkor a páros összehasonlítás módszerét javasoljuk.

További terveink között szerepel, hogy a hallgatói jelentkezésekből kialakított intézményi preferencia-sorrendek kialakulását megértsük. Kíváncsiak vagyunk arra, hogy mi vezet egy hallgatót arra, hogy megjelöljön egy adott intézményt. A célunk azonban nem egy összesített, több szempontot is figyelembe vevő többdimenziós rangsor kialakítása, hanem a hallgatói jelentkezési sorrendek kialakulásának megértése.

Ennek illusztrálására lássuk a következő példát, amely egy összetett rangsor alapján határozza meg a legjobbnak vélt intézményeket. Az egyik legelismertebb intézményi rangsor a HVG című hetilapban megjelent, a <http://eduline.hu/rang-sor> honlapon is elérhető intézményi rangsor. Itt két fő területen értékelik az intézményeket: 1. hallgatói kiválóság (elsőhelyes jelentkezők száma, felvettek pontát-laga, nyelvvizsgálóval felvettek száma, középiskolai versenyen helyezettek száma); illetve 2. oktatói kiválóság alapján (minősített oktatók száma, egy minősített oktatóra jutó hallgatók száma stb.). A 12. táblázat az első 12 magyar intézményt mutatja oktatói kiválóság alapján rangsorolva, s tartalmazza az általunk javasolt genetikus algoritmusmal számolt hallgatói jelentkezések alapján kialakított intézményi preferencialistát is.

Látható, hogy az első három helyen ugyanazok az intézmények szerepelnek, pedig az oktatói kiválóság alapján (részben szubjektív módon) számolt rangsor és a hallgatói jelentkezések alapján kialakított objektív intézményi preferencialista módszertanában teljes mértékben különbözik egymástól. Az oktatói kiválóság alapján számolt rangsor objektív tényezőket tartalmaz ugyan, a végső sorrend kialakításánál azonban ezeknek a tényezőknek a súlyát szakértők szubjektív módon határozták meg. A hallgatói rangsor viszont a hallgatói jelentkezéseken alapul, így teljes mértékben objektívnek tekinthető.

Természetesen nem vitatjuk, hogy egy intézmény megítélésében szerepet játszanak más tényezők is, mint a hallgatói jelentkezések, de szeretnénk e tényezőket kü-

12. táblázat

Oktatói kiválóság alapján kialakított rangsor (2011) és az intézményi preferencialista (2001–2011) eredményeinek összehasonlítása

Sorrend	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	Oktatói kiválóság (2011)
1.	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE	ELTE
2.	SZTE	DE	BGF	SZTE	SZTE	SZTE	SZTE	SZTE	DE	DE	DE	SZTE
3.	PTE	SZIE	SZTE	DE	PTE	PTE	DE	DE	SZTE	SZTE	SZTE	DE
3.	DE	SZTE	PTE	PTE	DE	DE	PTE	PTE	PTE	PTE	PTE	NYME
5.	BGF	PTE	DE	BGF	BCE	BCE	BCE	BCE	BCE	BCE	BCE	SE
6.	SZIE	BGF	SZIE	BKÁE	BGF	BGF	BGF	BME	BME	BME	BGF	PTE
7.	NYF	NYF	BKÁE	NYF	BME	BME	BME	BGF	BGF	SZIE	SZIE	XX
8.	KJF	BKÁE	BME	SZIE	NYF	NYF	ME	NYME	SZIE	BGF	BME	BCE
9.	BME	BME	NYF	SZE	PPKE	ME	BMF	ME	ME	NYME	NYME	ME
10.	ME	ME	ME	BME	VE	SZIE	NYF	SZIE	NYME	ME	ME	XX
11.	EKF	KJF	BMF	ME	SE	PPKE	SZIE	XX	BMF	SE	SE	BME
12.	TSF	TSF	EKF	BMF	EKF	EKF	SE	BMF	PPKE	BMF	SZE	NYF

lönválasztani az objektív jellemzőktől. Az egyezőségre úgy tekintünk, mint annak a további vizsgálatnak az igazolására, hogy a hallgatók vajon az oktatói kiválóság szempontjából előrébb sorolt intézményt választják-e. Fontos-e a hallgatóknak, hogy milyen rangos intézményben tanulnak, és milyen más szempontok (távolság, kereseti lehetőségek stb.) dominálnak? E kutatás során most a magyarított oldalt határoztuk meg. Más kutatásokkal szemben azonban mi a következő lépésben nem kérdőíves megkérdezéssel szeretnénk megtalálni az okokat, hanem objektív mutatókkal szeretnénk magyarázni a hallgatói jelentkezések kialakulását.

E tanulmány egy hosszabb kutatás első szakaszáról számolt be. Cikkünkben bemutattuk, hogyan lehet a hallgatói jelentkezéseket redundanciamentesen tárolni. Kifejtettük, hogyan lehet szakra való jelentkezésekből szakai, kari, illetve intézményi szintű preferencia-sorrendet kialakítani. Bemutattuk, hogy a különböző módszerek (például páros összehasonlítás módszere, rangösszeg módszere) felhasználásával további információt is kaphatunk a preferált intézmények preferenciatávolságáról. Az eredményül kapott sorrend, illetve a kapott pontszám egy későbbi kutatás eredményeként a hallgatói jelentkezéseket befolyásoló tényezőket feltáró kutatás magyarított változója lehet. A javasolt módszerek robusztusak, s véleményünk szerint alkalmazási lehetőségei szélesebbek is, mint az itt említett hallgatói jelentkezések aggregálása. Minden olyan esetben ugyanis, amikor hiányos rangsorokból kell végső sorrendet kialakítani, az itt bemutatott módszerek gyors és korrekt megoldást ígérnek.

Hivatkozások

- ABONYI JÁNOS [2012]: Szóbeli közlés.
- ALONSO, S. CHICLANA, F.–HERRERA, F. –HERRERA-VIEDMA, E.–ALCALA-FDEZ, J.–PORCEL, C. [2005]: A consistency based procedure to estimate missing pair-wise preference values, tech. rep. Department of Computer Science and Artificial Intelligence. University of Granada.
- BRAUN, H. [1991]: On solving travelling salesman problems by genetic algorithms. Lecture Notes in Computer Science, Vol. 496. 129–133. o. DOI: 10.1007/BFb0029743.
- BOZÓKI SÁNDOR–FÜLÖP JÁNOS–RÓNYAI LAJOS [2010]. On optimal completion of incomplete pairwise comparison matrices. Mathematical and Computer Modelling. Vol. 52. No. 1–2. 318–333. o.
- CHEBOTAREV P. Y.–SHAMIS E. [1999]: Preference fusion when the number of alternatives exceeds two: indirect scoring procedures. Journal of the Franklin Institute, Vol. 336. No. 2. 205–226. o.
- CSUKA GYÖNGYI–SZÍVÓS MIHÁLY [2010a]: Rangsorok és tükörképek – ahogy a hallgatójelöltek választanak. *Competitio*, 9. évf. 2. sz. 78–91. o.
- CSUKA GYÖNGYI–SZÍVÓS MIHÁLY [2010b]: Vonzáskörzet, döntési szempontok, intézményválasztás, a hallgatójelöltek új egyetem- és főiskola-választási szokásai. Megjelent: *Csuka Gyöngyi–Kovács Bernadett–Szívós Mihály* (szerk.): Regionális gazdasági koncentráció, felsőoktatási térszerkezet, innováció. MTA–PE–BME Regionális Innovációs és Fejlesztési Hálózati Kutatócsoport, Pécs.
- FARKAS ANDRÁS–LANCASTER PÉTER–RÓZSA PÁL [2003]: Consistency adjustments for pairwise comparison matrices. *Numerical Linear Algebra with Applications*, Vol. 10. No. 8. 689–700. o. DOI: 10.1002/nla.318.
- FÁBRI GYÖRGY [2008]: Magyar felsőoktatási rangsorok – 10 év tükrében. Hozzászólás Török Ádám cikkéhez. *Közgazdasági Szemle*, 55. évf. 12. sz. 1116–1119. o.
- FÁBRI GYÖRGY–TORDA JÚLIA (szerk.) [2005]: Felvi-rangsor. Egyetemek, főiskolák mérlegén 2006. *Educatio Társadalmi Szolgáltató Kht.–Országos Felsőoktatási Információs Központ*, Budapest.
- FEDRIZZI, M.–GIOVE, S. [2007]: Incomplete pairwise comparison and consistency optimization. *European Journal of Operational Research*, Vol. 183. No. 1. 303–313. o.
- HORVÁTH DÁNIEL–KISS LÁSZLÓ [2009]: Rangsorok a közbeszédben. A felsőoktatási rangsorok megjelenése a hazai médiában és a felsőoktatási intézmények kommunikációjában. *Felsőoktatási Műhely*, IV., 45–57. o. http://www.felvi.hu/pub_bin/dload/FeMu/2009_04/oldal45_58_horvath_kiss.pdf.
- KÓCZI Á. LÁSZLÓ–NICHIFOR, A. [2012]: The intellectual influence of economic journals: quality versus quantity. *Economic Theory*, megjelenés alatt, DOI 10.1007/s00199-012-0708-0.
- KATONA TAMÁS–BALOGH MIKLÓS [2010]: A felsőfokú tanintézetek összehasonlító rangsorát meghatározó indikátorrendszer kidolgozása. *Statisztikai Szemle*, 4. sz. 417–432. o.
- MIHÁLYI PÉTER [2002]: Mit érnek a közgazdász diplomák? *Figyelő*, 37. sz. 46–54. o.
- PAGE, L.–BRIN, S. [2009]: Google Press Center: Fun Facts. www.google.com. <http://web.archive.org/web/20090424093934/>. <http://www.google.com/press/funfacts.html>.
- SLIKKER M.–BORM P.–BRINK R. [2012]: Internal slackening scoring methods. *Theory and Decision*, Vol. 72. No. 4. 445–462. o.
- TAMÁNDL LÁSZLÓ [2011]: A felsőoktatási intézmények versenyképességi tényezői, különös tekintettel a diplomás pályakövetésre. Doktori értekezés, Széchenyi István Egyetem Regionális és Gazdaságtudományi Doktori Iskola, Győr.

- TANINO T. [1984]: Fuzzy preference orderings in group decision making. *Fuzzy Sets and Systems*, 12. 117–131. o.
- TELCS ANDRÁS–KOSZTYÁN ZSOLT TIBOR–TÖRÖK ÁDÁM [2012]: Unbiased one dimensional University Ranking – application based preference ordering. Közlésre benyújtva.
- TORBEN, D.–CHRISTOPHER, M. [2006]: How do students choose a university? An Analysis of Applications to Universities in Ontario. Canada, *Research in Higher Education*, Vol. 47. No. 7. 781–800. o. DOI: 10.1007/s11162-006-9015-6.
- TÖRÖK ÁDÁM [2006]: Az európai felsőoktatás versenyképessége és a lisszaboni célkitűzések. Mennyire hihetünk a nemzetközi rangsoroknak? *Közgazdasági Szemle*, 53. évf. 4. sz. 310–329. o.
- TÖRÖK ÁDÁM [2007]: Tükör által – homályosan? *Közgazdász szemmel a nemzetközi egyetemi rangsorokról*. Felsőoktatási Műhely, őszi szám, 91–97. o.
- TÖRÖK ÁDÁM [2008a]: Felsőoktatási rangsorok. Lemaradásunk torzképei. Figyelő, szeptember 11–17. 31–32. o.
- TÖRÖK ÁDÁM [2008b]: A mezőny és tükörképei. Megjegyzések a magyar felsőoktatási rangsorok használatáról és korlátairól. *Közgazdasági Szemle*, 55. évf. 10. sz. 874–890. o.
- TÖRÖK ÁDÁM [2009a]: Közös érdeklődés – eltérő nézőpont. Válasz Fábri Györgynek. *Közgazdasági Szemle*, 56. évf., 1. sz. 93–95. o.
- TÖRÖK ÁDÁM [2009b]: On the economics of the university ranking lists: intuitive remarks on intuitive comparisons. Megjelent: *Varga Attila* (szerk.): *Universities, Knowledge Transfer and Regional Development*. Edward Elgar.
- TÖRÖK ÁDÁM [2009c]: Teljesítménymérés és rangsorolás a magyar felsőoktatásban. Megjelent: *Hubos Ildikó–Török Imre* (szerk.): *Intézményi menedzsment a felsőoktatásban. Szemelvények kiemelt témakörökben*. Műegyetemi Kiadó.
- TÖRÖK ÁDÁM [2009d]: Verseny a felsőoktatásban – így mértek ti. Megjelent: *Muraközy László* (szerk.): *A jelen a jövő múltja. Járatlan utak – járt úttalanságok*. Akadémiai Kiadó, Budapest, 241–294. o.
- TÖRÖK ÁDÁM–KOVÁCS BERNADETT [2011]: A nemzetközi felsőoktatási verseny mérési problémáiról. Megjelent: *Magyar felsőoktatás 2010. Konferencia dokumentumok*, Budapesti Corvinus Egyetem. NFKK Füzetek, 6. Budapest, 15–16. o.