

Magyar nyelvjárási adattárak vizsgálata interaktív dialektometriai térképekkel

Vargha Fruzsina Sára – Vékás Domokos

A dialektometriai projekt megvalósítását
2007. november 1. és 2008. augusztus 31. között
a Bihalbocs támogatása tette lehetővé.

A számítógépes dialektológiai projektek jellemzően kiegészítik egymást, egymásra is épülnek. Éppen ezért már előzetesen is köszönetet szeretnénk mondani mindazoknak, akik korábbi, szintén Bihalbocs technológiákat használó projektekben részt vettek, és így közvetve segítették a mostani dialektometriai vállalkozást.

Elhangzott a Magyar Nyelvtudományi Társaság felolvasóülésén, 2009. március 24-én

A dialektometria helye és súlya

- a számítógépes dialektológia viszonylag kis területe (Vékás 2007)
- prioritás a számítógépes dialektológiában: beszéltnyelvi (orális) korpuszok építése
- dialektometriára jutó erőforrások és erőfeszítések: mintegy 10%

Miről lesz szó?

- a dialektometria kialakulása és mai fogalma
- kutatópontok nyelvi rokonítása különböző kvantitatív módszerekkel
- a nyelvföldrajzi perspektíva változása: honnan nézzük a nyelvterületet?
- nyelvjárások közötti “genetikai” kapcsolatok vizsgálatának kérdése
- területek elhatárolásának kérdése

Hagyományos nyelvföldrajzi módszerek

- feladat: nyelvjárások klasszifikációja
- probléma: nyelvjárási határok – kontinuum
- hagyományos módszerek:
 - izoglosszák (izoglosszanyalábok = határok)
 - vonallal nem mindig választhatók el területek
 - szubjektív, önkényes (fokozatos változásnál)
 - izoglosszák ritkán esnek egybe
 - fonémarendszerek
 - különböző nyelvjárások, azonos fonémarendszer

A dialektometria

- reakció hagyományosabb nyelvföldrajzi módszerek vélt elégtelenségére
- Magyar dialektológia (Kiss J.)
 - kvantitatív módszerek használata a kritérium
 - nem térképközpontú
 - nyelvjárásiasság/köznyelviesség mértéke
- Séguy (szomszédos k.pontok összevetése)
- Mátrix alapú interaktív térképek
 - Goebel (munkatérképek manuális csoportosítással)
 - Nerbonne, Heeringa (Levenshtein algoritmus)

A térképközpontú magyar dialektometria előjelei

- “... térképen vagy más formában ábrázolható elvont jelenségekhez juthatunk, ahol a konkrét adatokhoz való kapcsolódás sokszorosán áttételes.” Vékás (1999)
- “Az új nyelvatlaszok adatainak dialektometriai, azaz nyelvjárás mérési elemzése közelebb visz bennünket a dialektológia régi-új problémájának a megnyugtatóbb megoldásához, a nyelvjárások tipizálásához is.” Kiss (2001)

Eldöntendő: saját fejlesztések szükségessége

- A fejlesztések roppant erőforrásokat igényelnek, és fenntartásuk nehezen biztosítható (Goebel esete)
- Van kész módszer!
 - RuG/L04 (<http://www.let.rug.nl/~kleiweg/L04/>)
- A saját fejlesztések korlátlan lehetőségeket biztosítanak

Kutatópontok összevetése

a Somogy–zalai nyelvatlaszból, fő- és alkategóriák szerint,
Király Lajos csoportosítása alapján

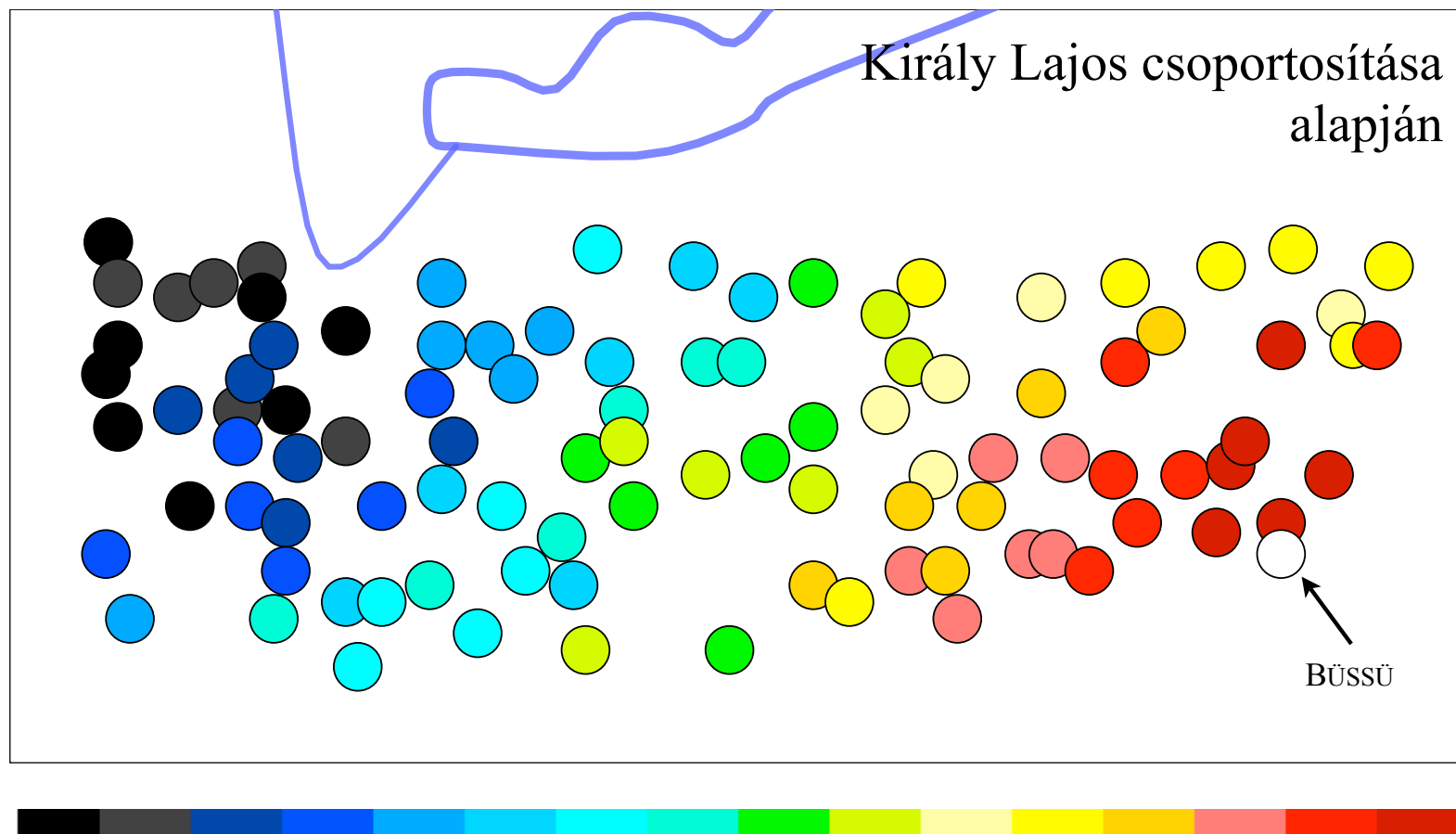
KUTATÓP.	ADAT (85. TÉRKÉPLAP)	KAT. (85.)	BÜSSÜHÖZ KÉPEST (85. T.)	BÜSSÜHÖZ KÉPEST (ÖSSZES T.)
Büssü	marasztajja	1.1.	100%	100,0%
Ráksi	marasztàjja	1.1.	100%	86,1%
S.vámos	m̄ar̄asztùja	1.2.	50%	68,7%
Gelsesziget	maraszt̄alla	2.1.	0%	43,2%

Mátrix kialakítása (S–ZA.)

Király Lajos csoportosítása alapján

%	Büssü	Ráksi	S.vámos	Gelsesziget
Büssü	100,0			
Ráksi	86,1	100,0		
S.vámos	68,7	68,3	100,0	
Gelsesziget	43,2	43,6	47,4	100,0

S–ZA. dialektometriai térképek 1.



Levenshtein algoritmus

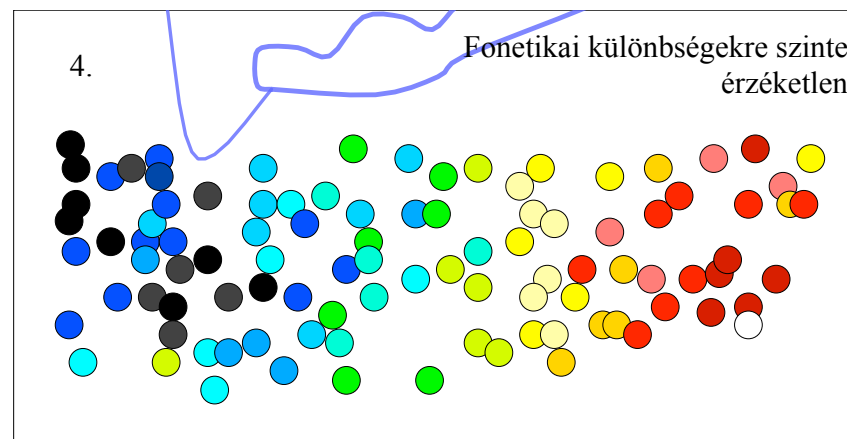
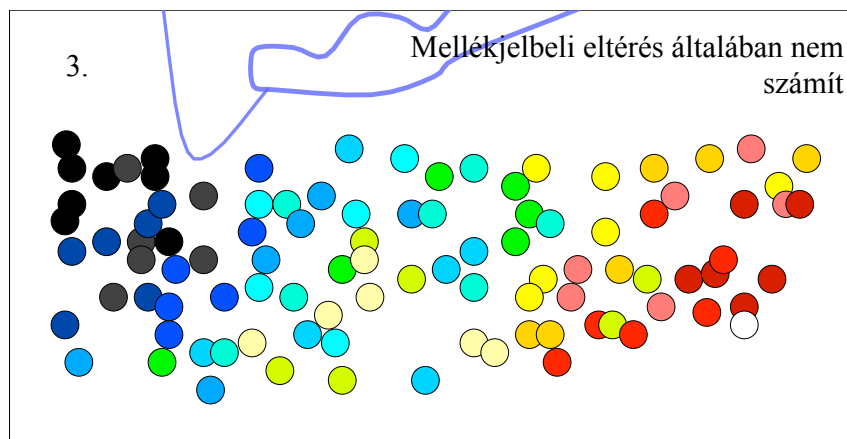
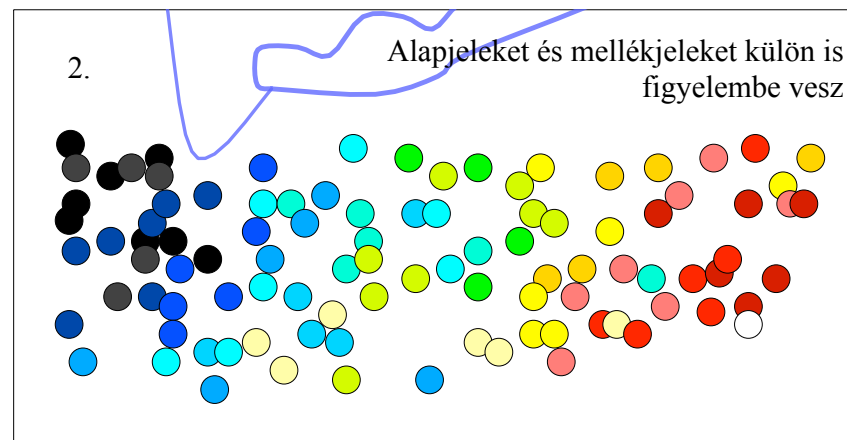
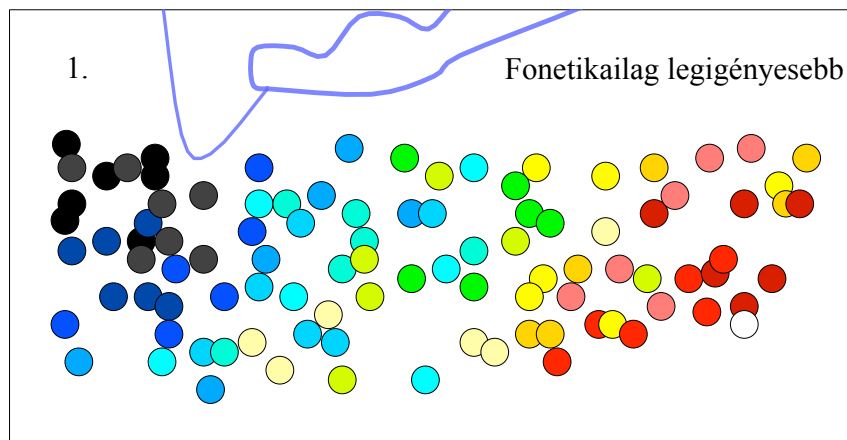
- távolság mérése két betűlanc között
 - távolság mérése két nyelvváltozat között adatpárok összehasonlításával
- a betűlancok közötti transzformációhoz szükséges beillesztések, törlések, helyettesítések legkevésebb költséges (legkevesebb lépésből álló) kombinációja
- első dialektológus alkalmazó: Kessler (1995)
- tökéletesítők: Nerbonne és Heeringa

Adatok összevetése Levenshtein algoritmussal

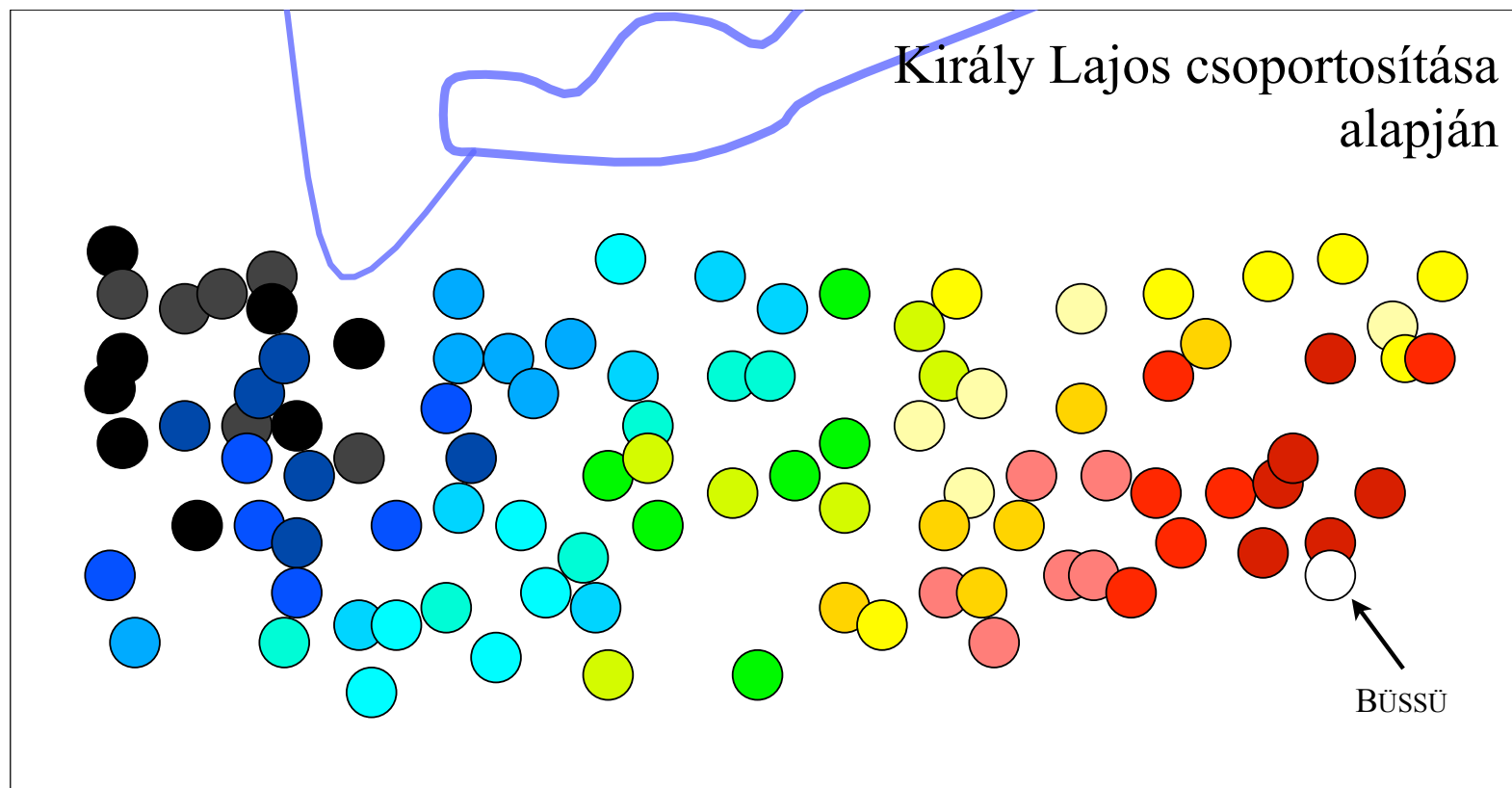
vakondak	ES	dj0	dj2	dj4
vakondak	1000	1000	1000	1000
vākondāk	600	625	750	1000
vākānd	250	500	500	750
vakondi [^] ēk	584	875	875	1000
vākhānD	375	250	375	625
patkány	125	125	125	375
höncsök	250	250	250	500

S–ZA. dialektometriai térképek 2.

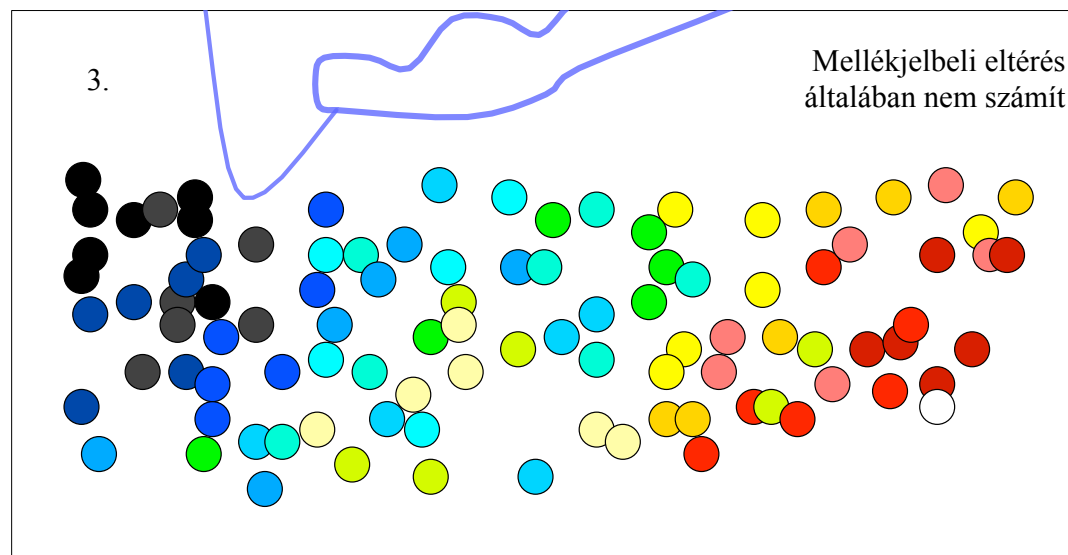
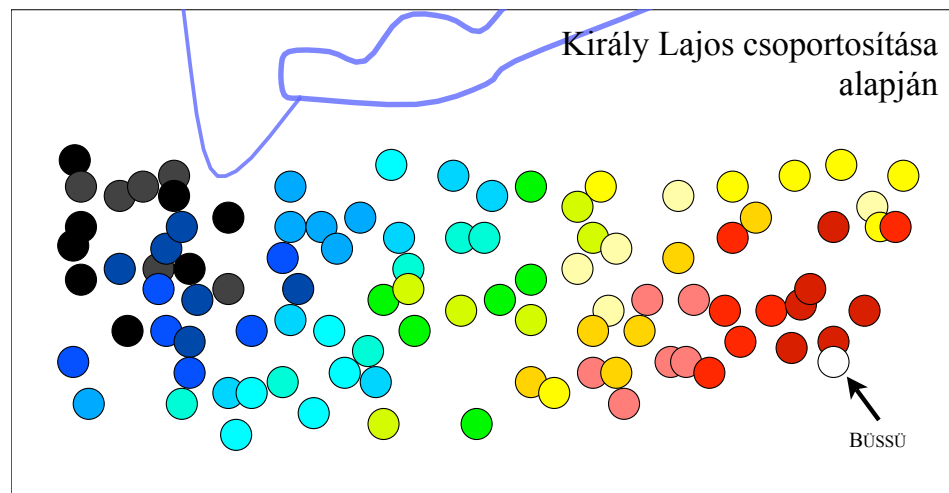
viszonyítási pont: Büssü, különböző mátrixok



S–ZA. dialektometriai térképek 3.



S–ZA. dialektometriai térképek 3.



Számítási módszerek összevetése

Viszonyítási pont: Büssü

‰	KL	dj0	ES	dj2	dj4
Büssü	1000	1000	1000	1000	1000
Ráksi	861	870	870	878	958
S.vámos	687	755	772	774	926
Gelsesziget	432	608	617	673	871

KL: Király Lajos csoportosítása alapján

dj0: két hang csak akkor azonos, ha a mellékjelek szintjén is az

ES: az alapjelek egyezését figyelembe veszi, de a mellékjelek is számítanak

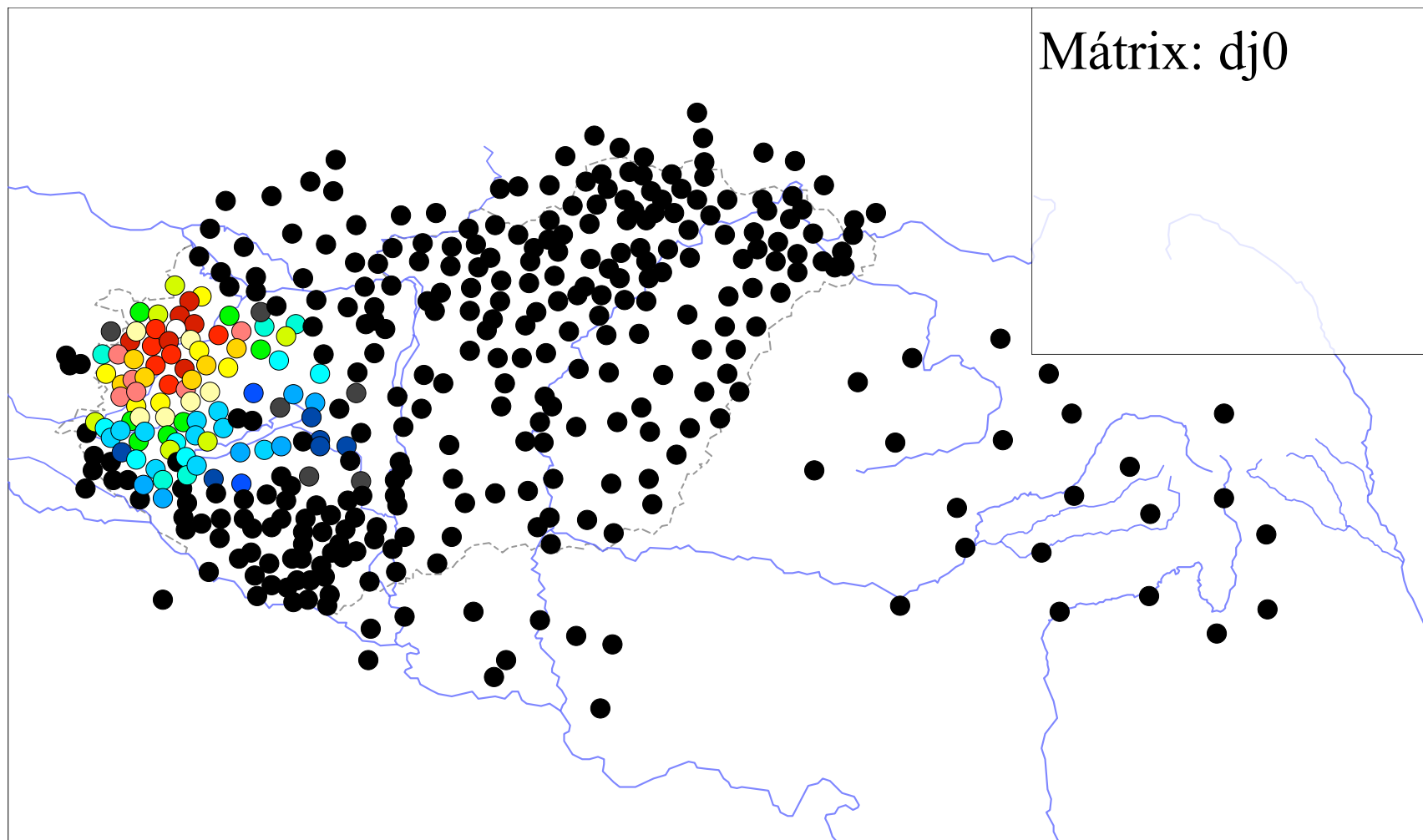
dj2: a mellékjelek nem számítanak, kivéve pl. nyílt *e* esetén

dj4: magánhangzók közti különbségek nem számítanak, több rokon mássalhangzót is
összemosunk (pl. *ty* = *gy*)

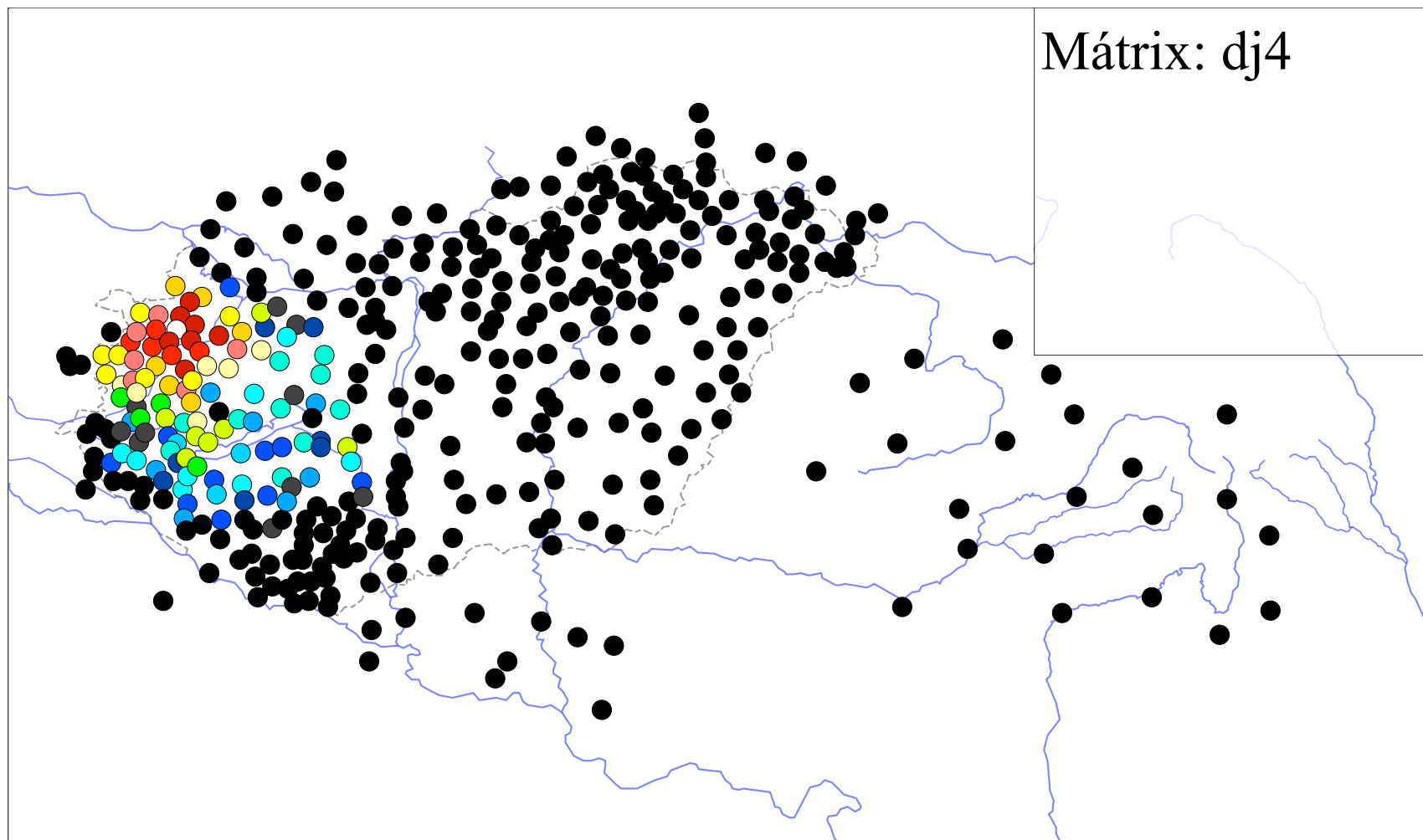
MNyA. dialektometria

- 395 kutatópont
- mátrix mérete: 156.025 darab számadat
- 476 felhasznált térképlap
 - gyakorlati (nem nyelvészeti) szempontú válogatás
 - mindegyik térkép sűrű kutatóponthálózatú
 - az újkeletűnek minősített adatok nélkül

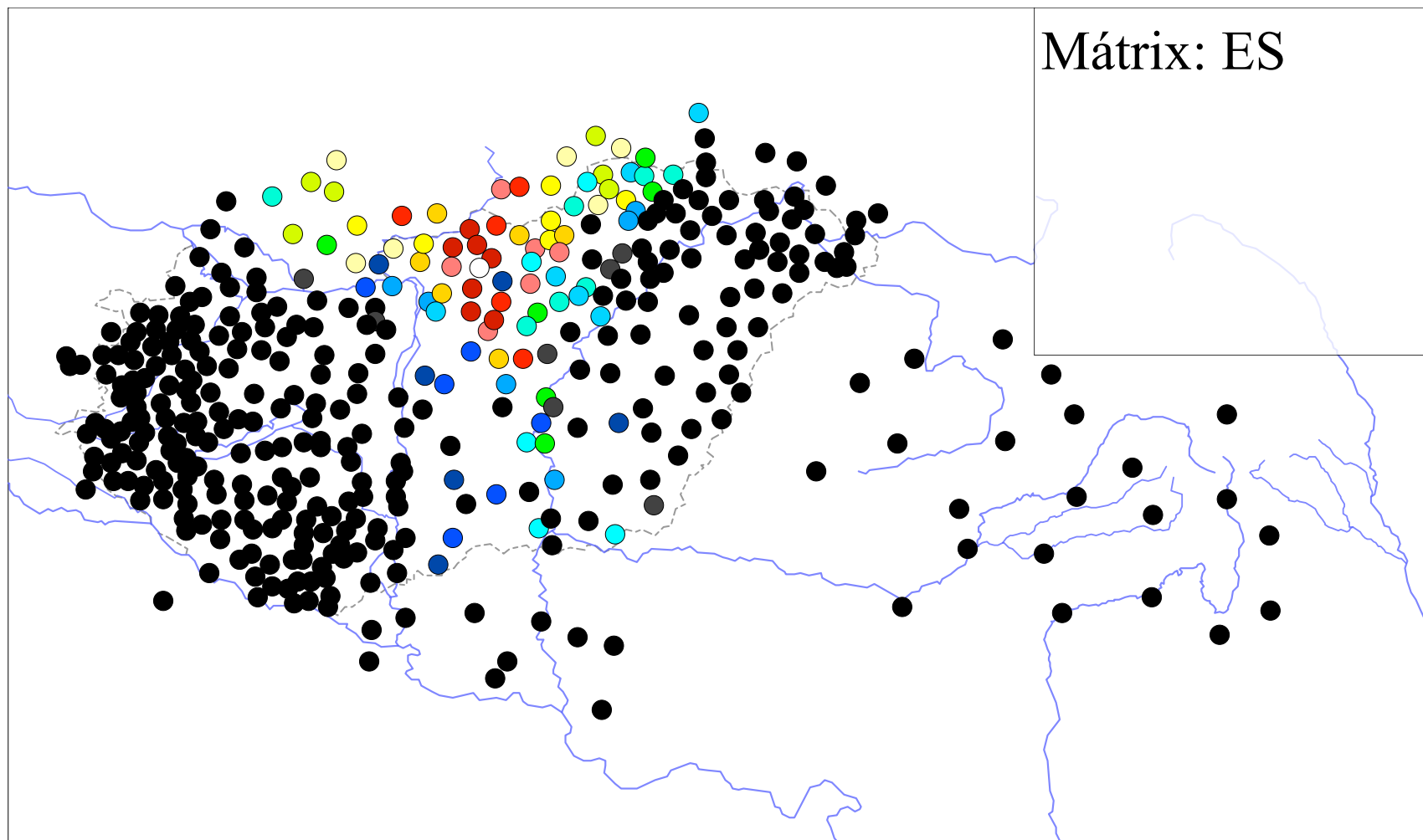
Nyelföldrajzi kapcsolatok kimutatása különböző mátrixok használatával: Mihályi



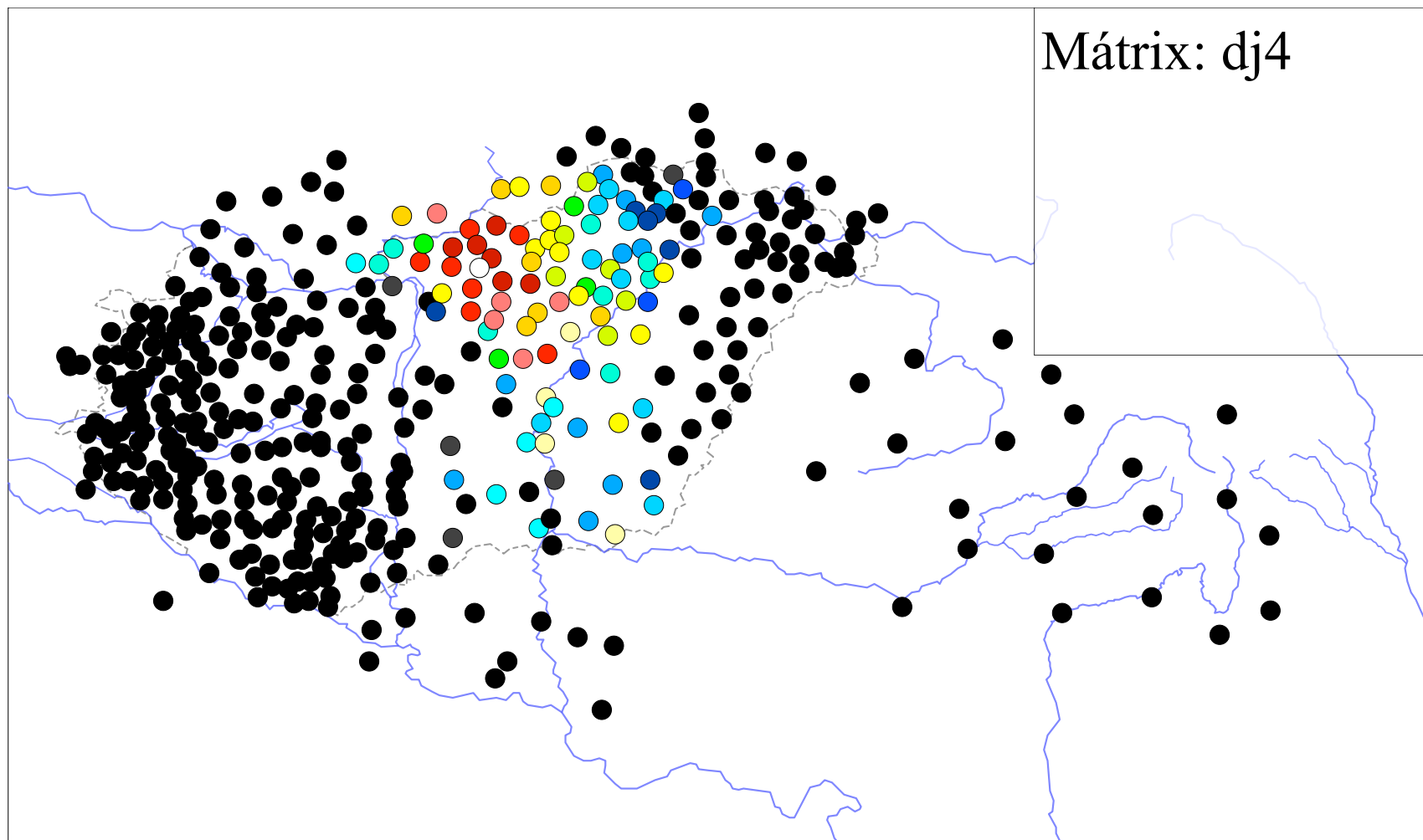
Nyelföldrajzi kapcsolatok kimutatása különböző mátrixok használatával: Mihályi



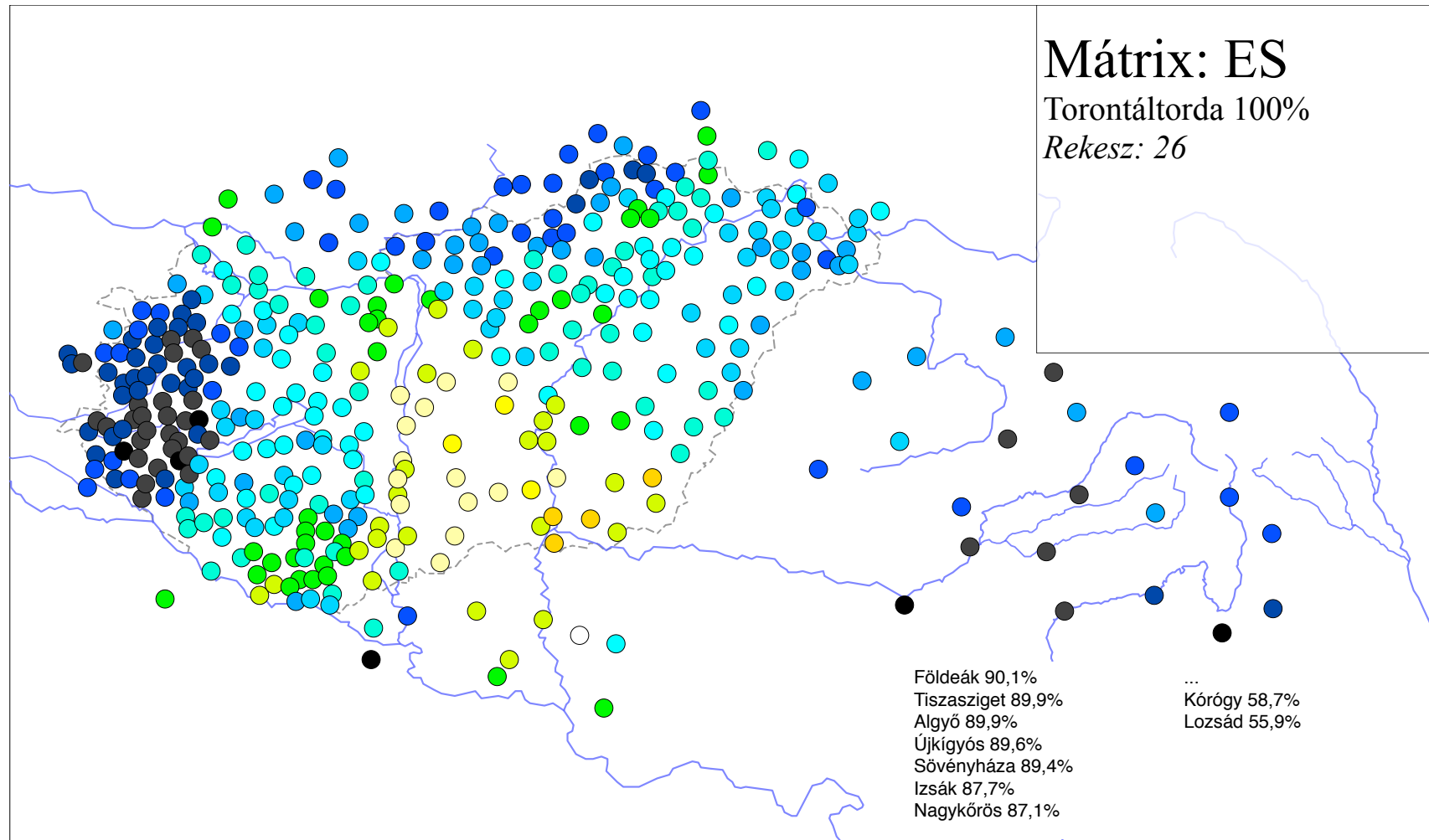
Nyelföldrajzi kapcsolatok kimutatása különböző mátrixok használatával: Ecseg



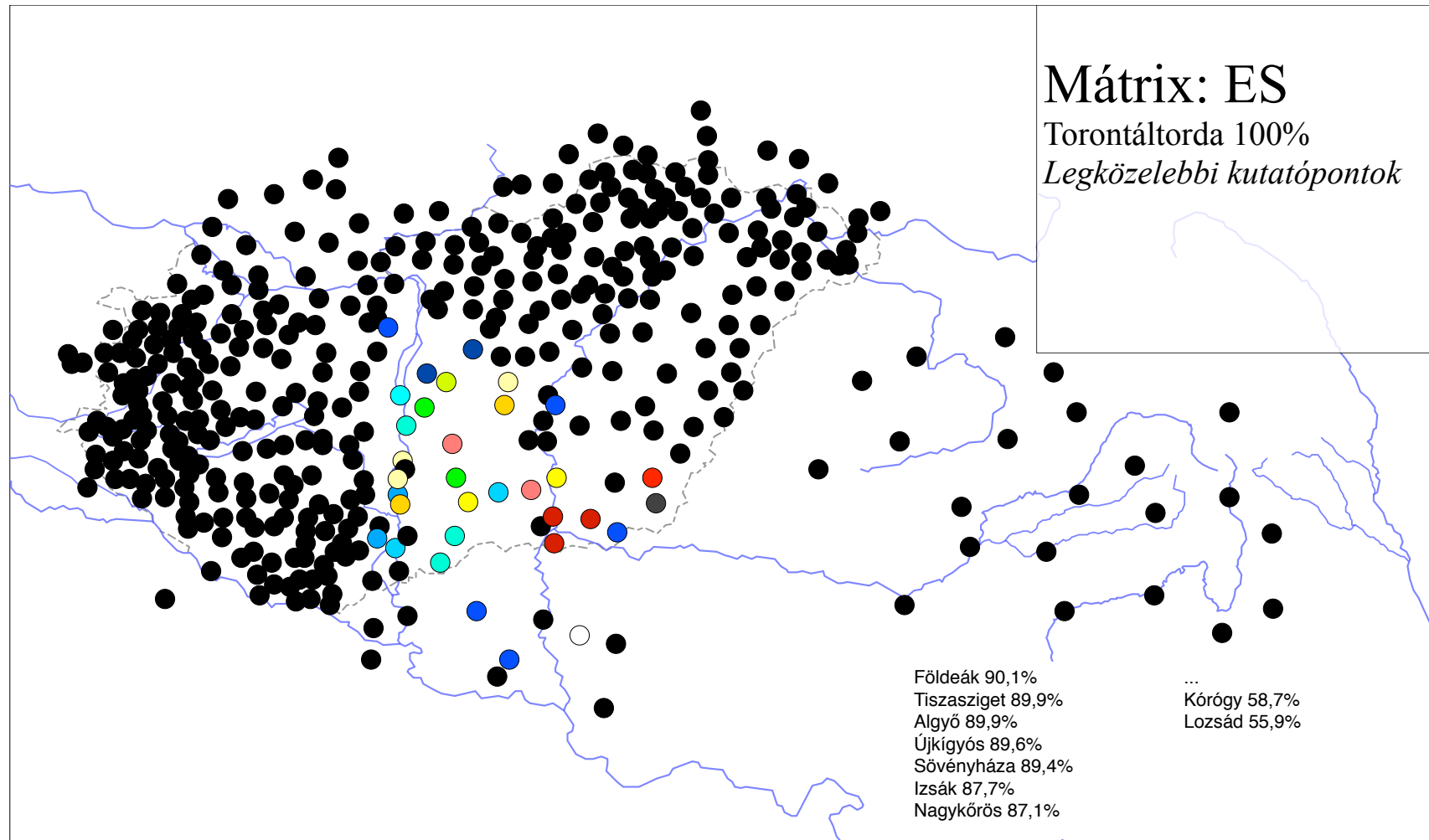
Nyelföldrajzi kapcsolatok kimutatása különböző mátrixok használatával: Ecseg



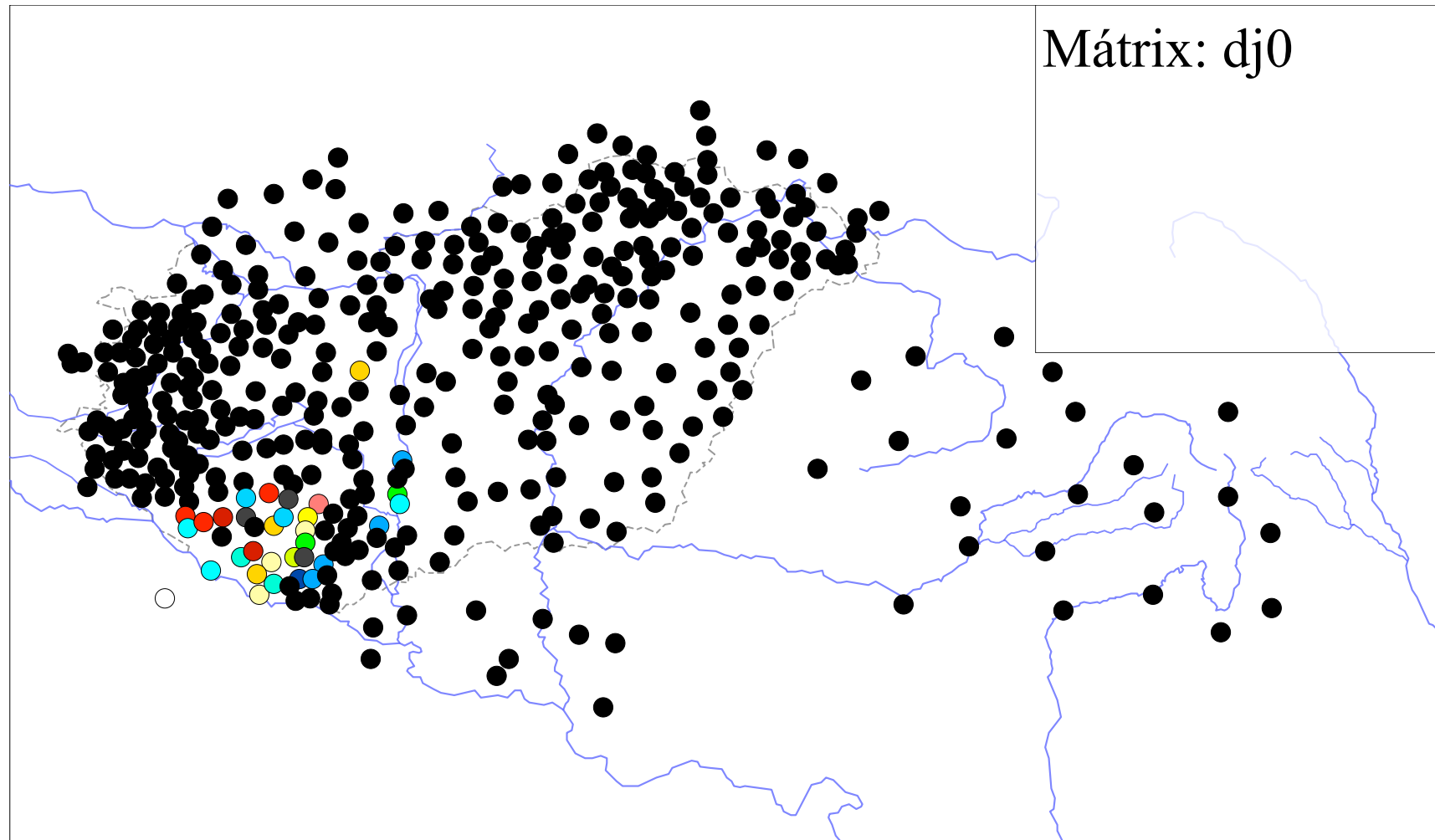
Torontáltorda



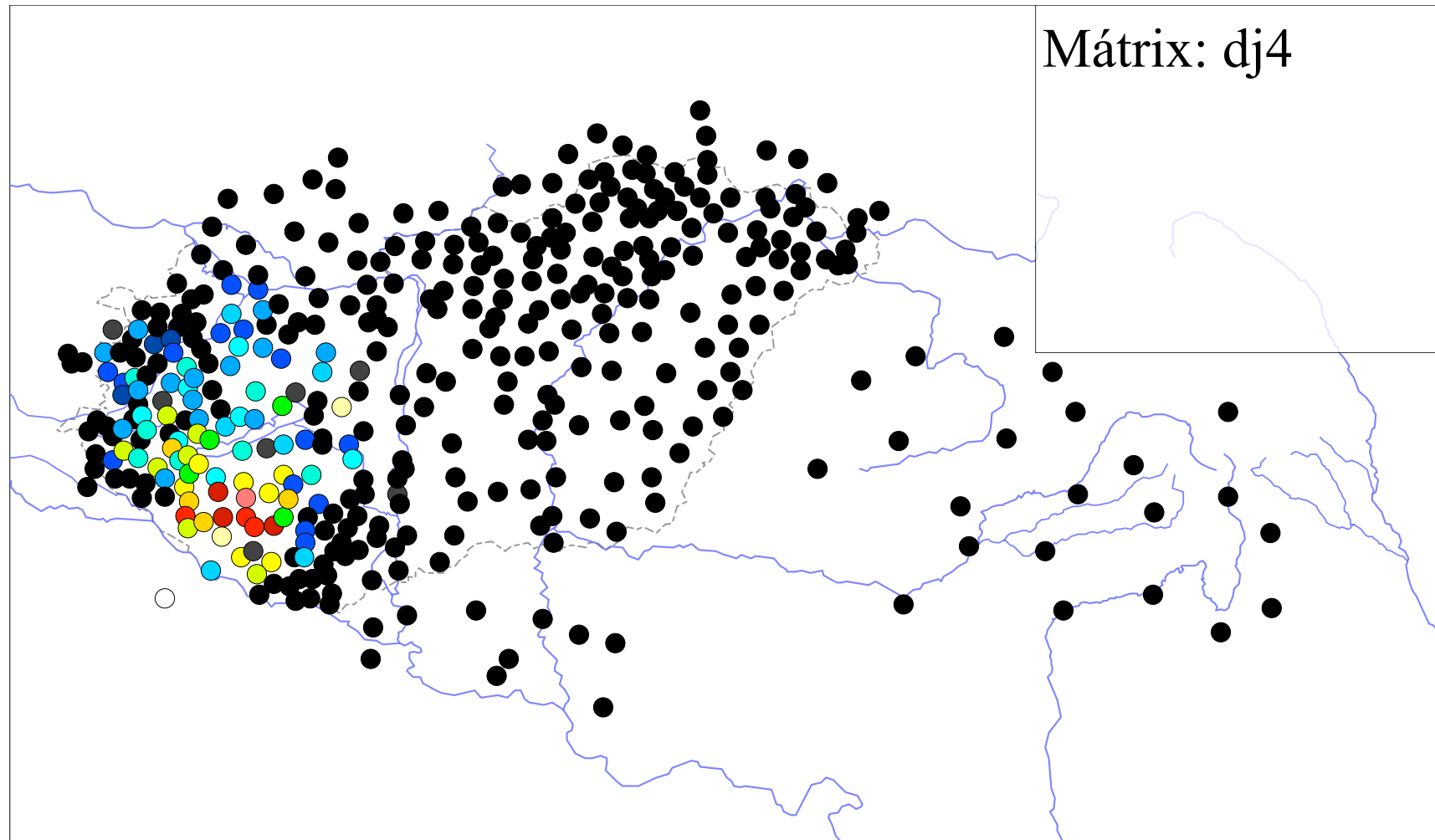
Torontáltorda



Velika Pisanica

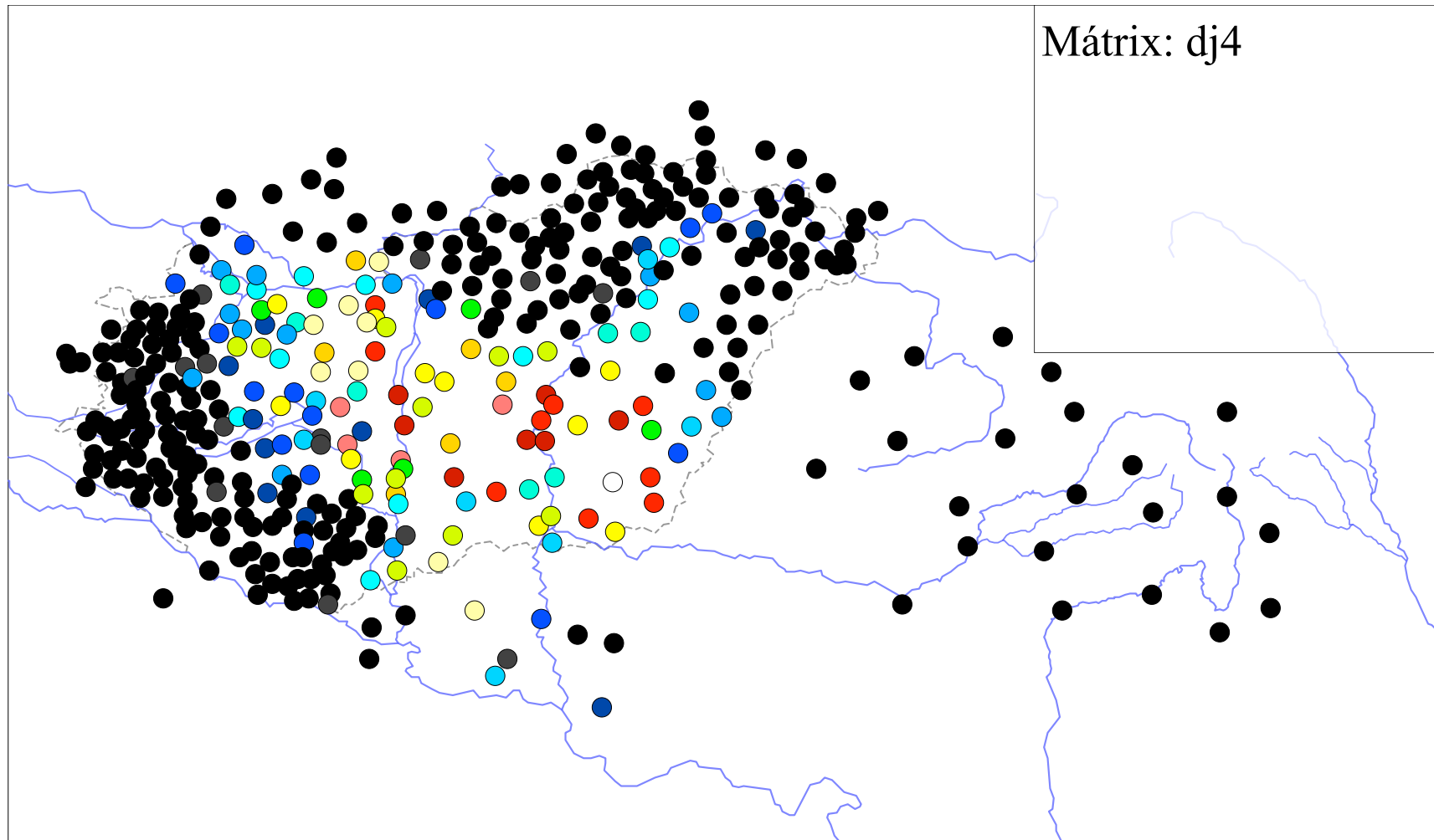


Velika Pisanica



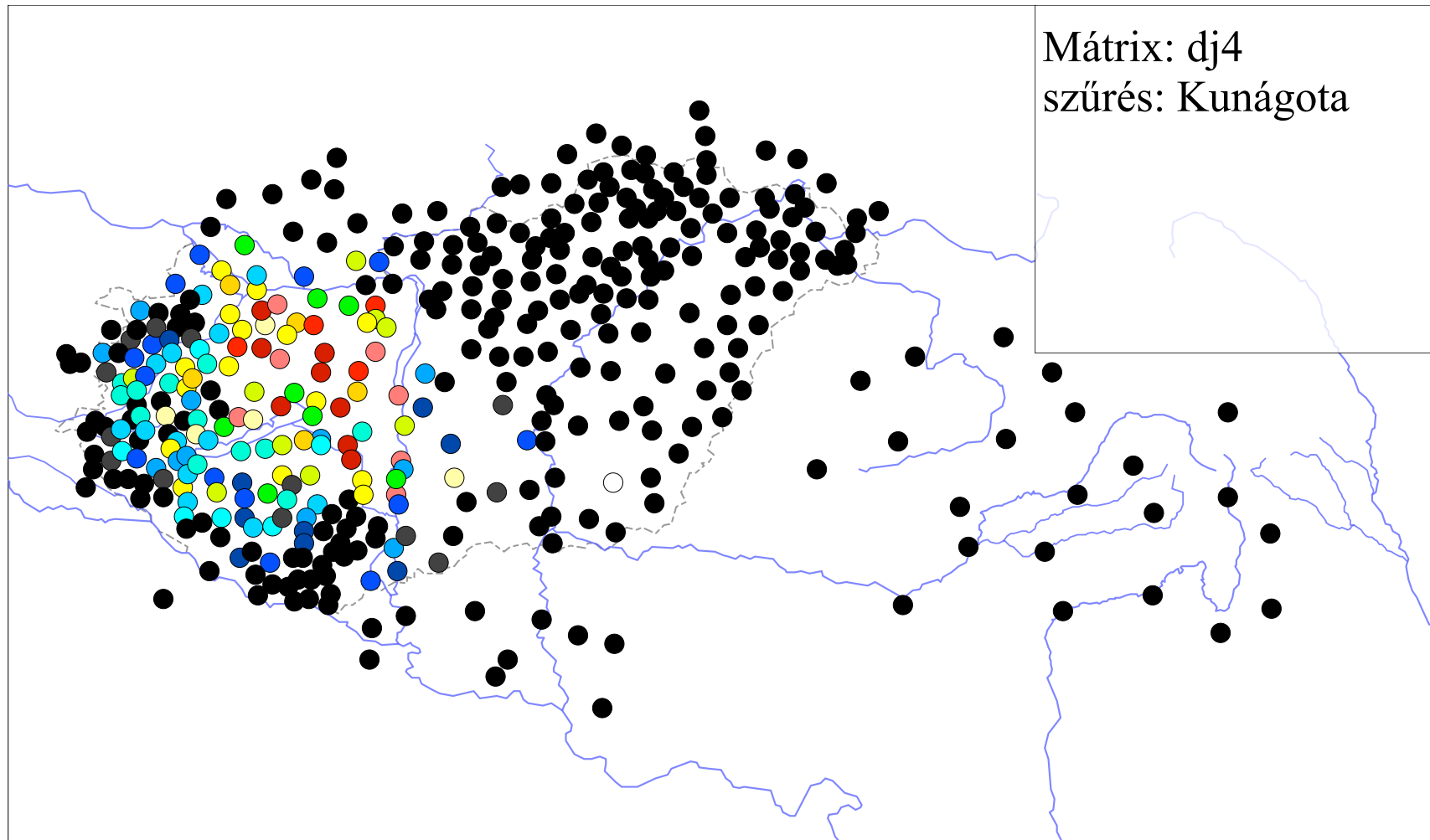
Orosháza (Szentetornya):

fonetikailag érzéketlen mátrix

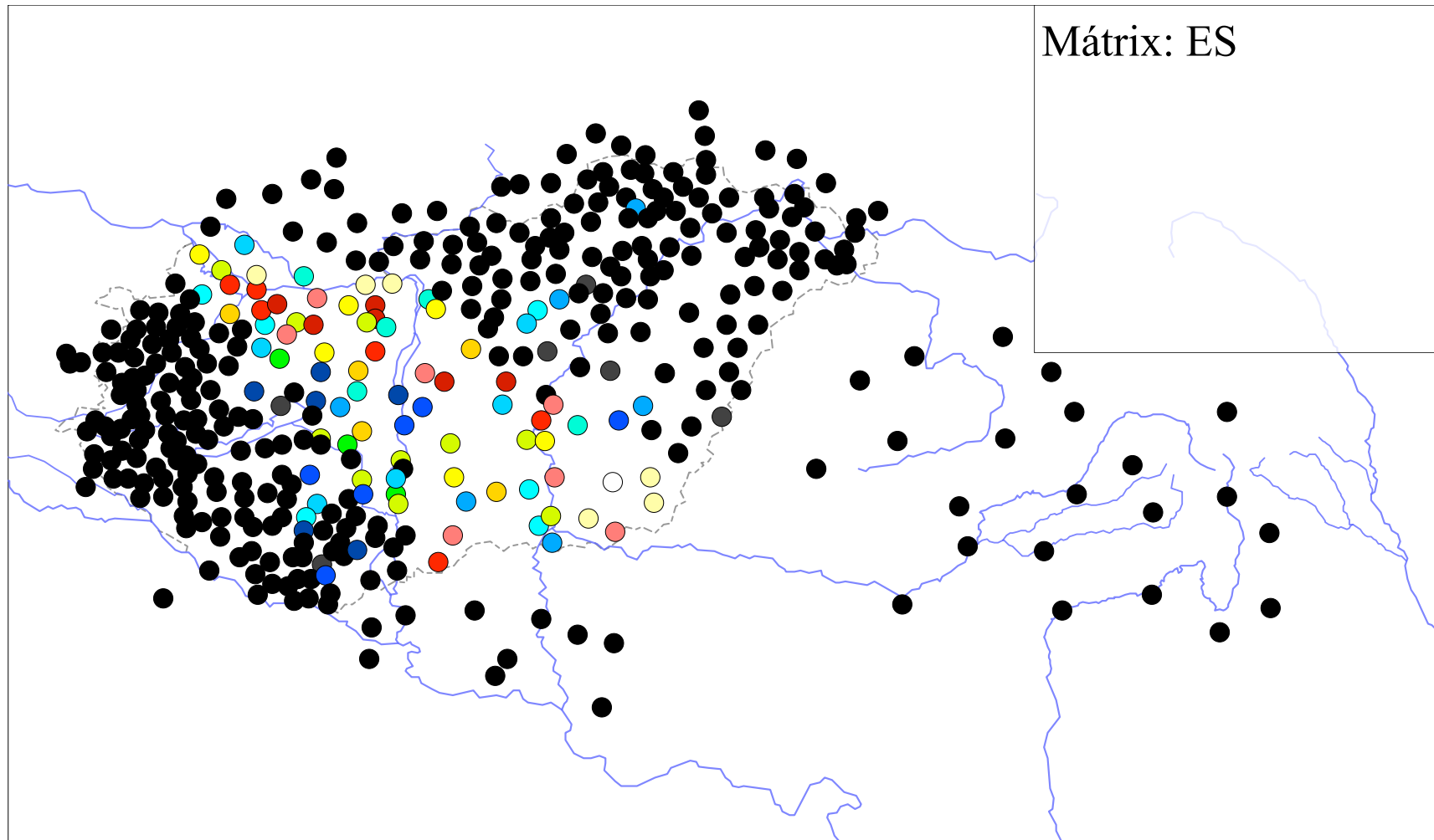


Orosháza (Szentetornya):

fonetikailag érzéketlen mátrix

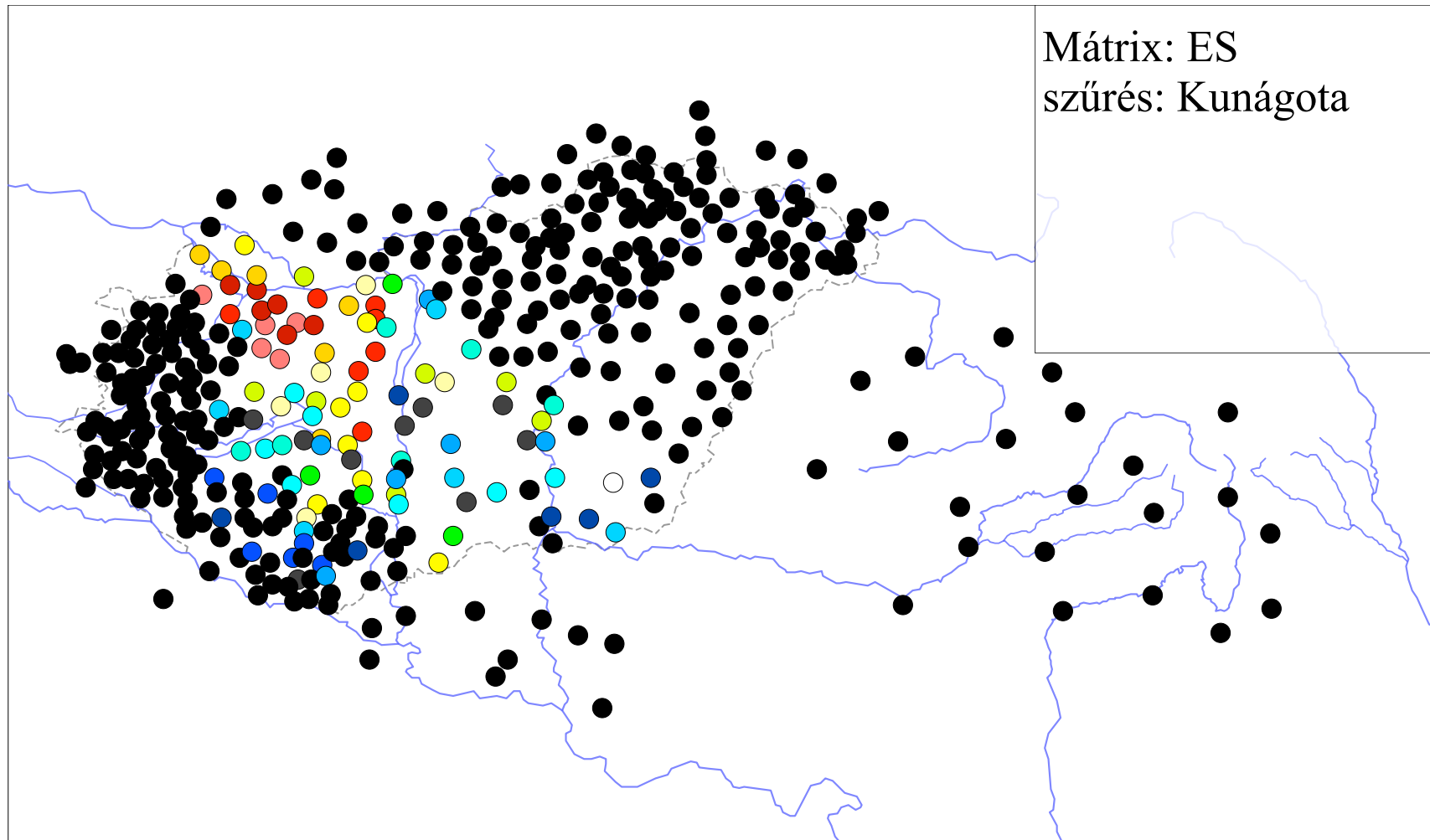


Orosháza (Szentetornya): fonetikailag érzékeny mátrix

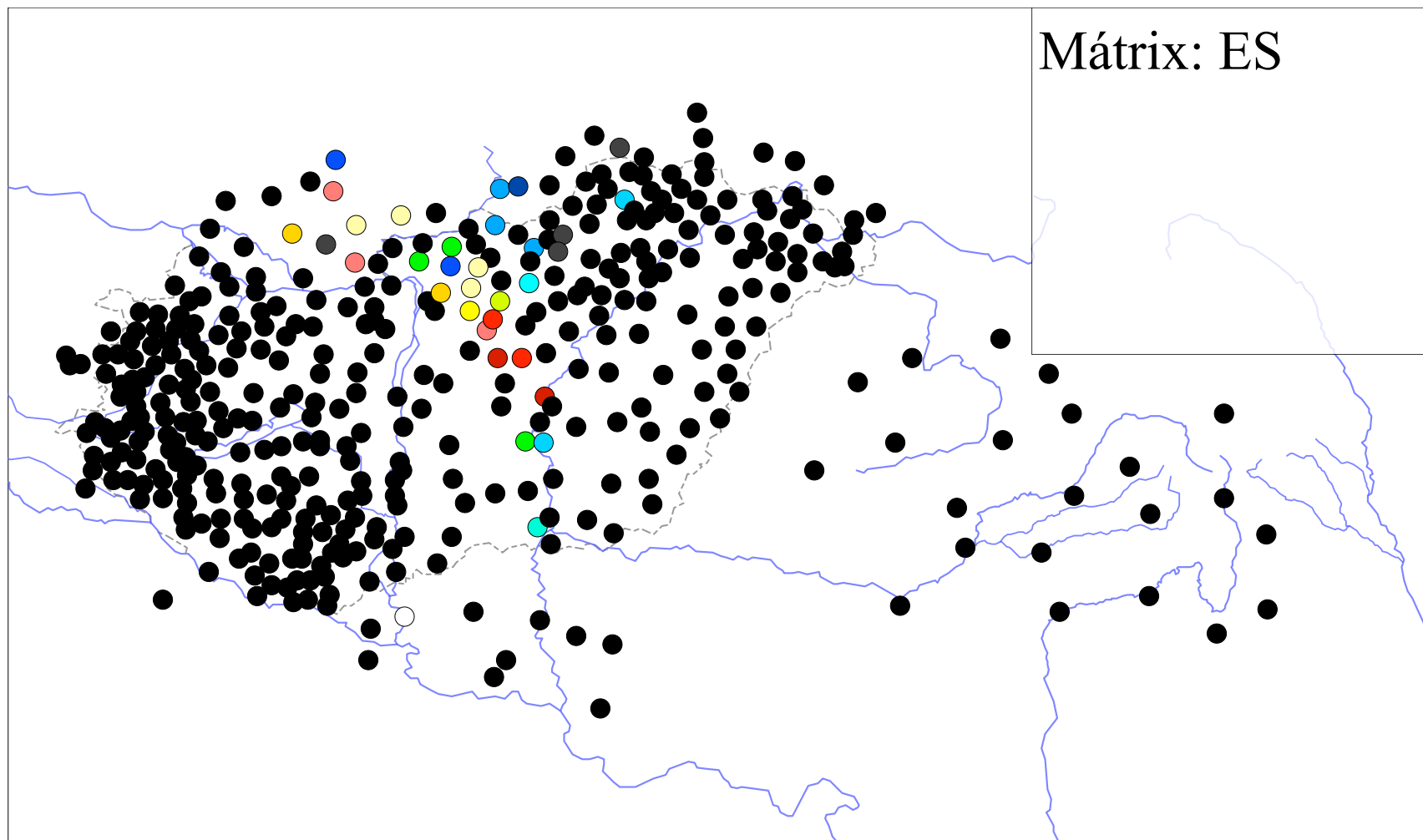


Orosháza (Szentetornya):

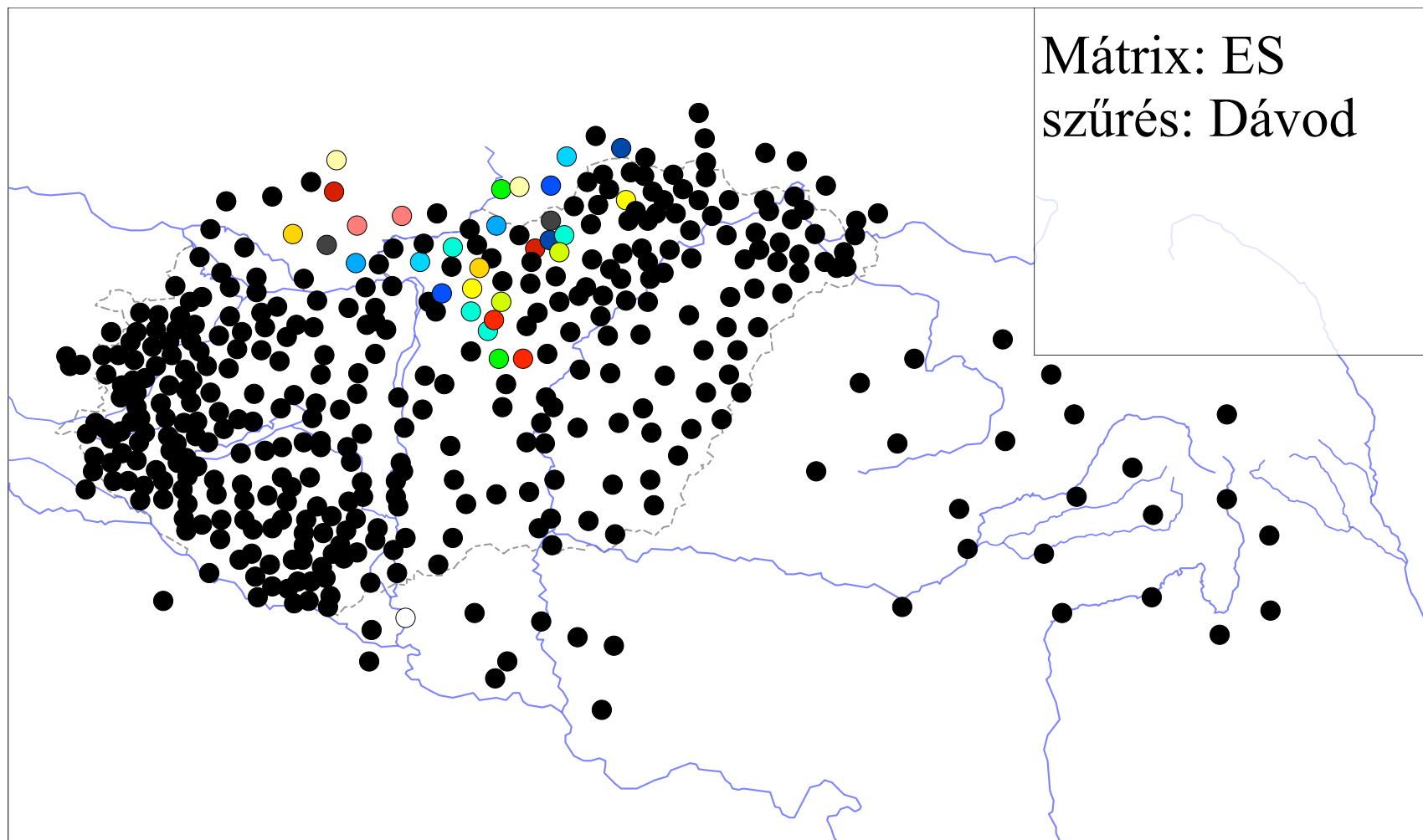
fonetikailag érzékeny mátrix



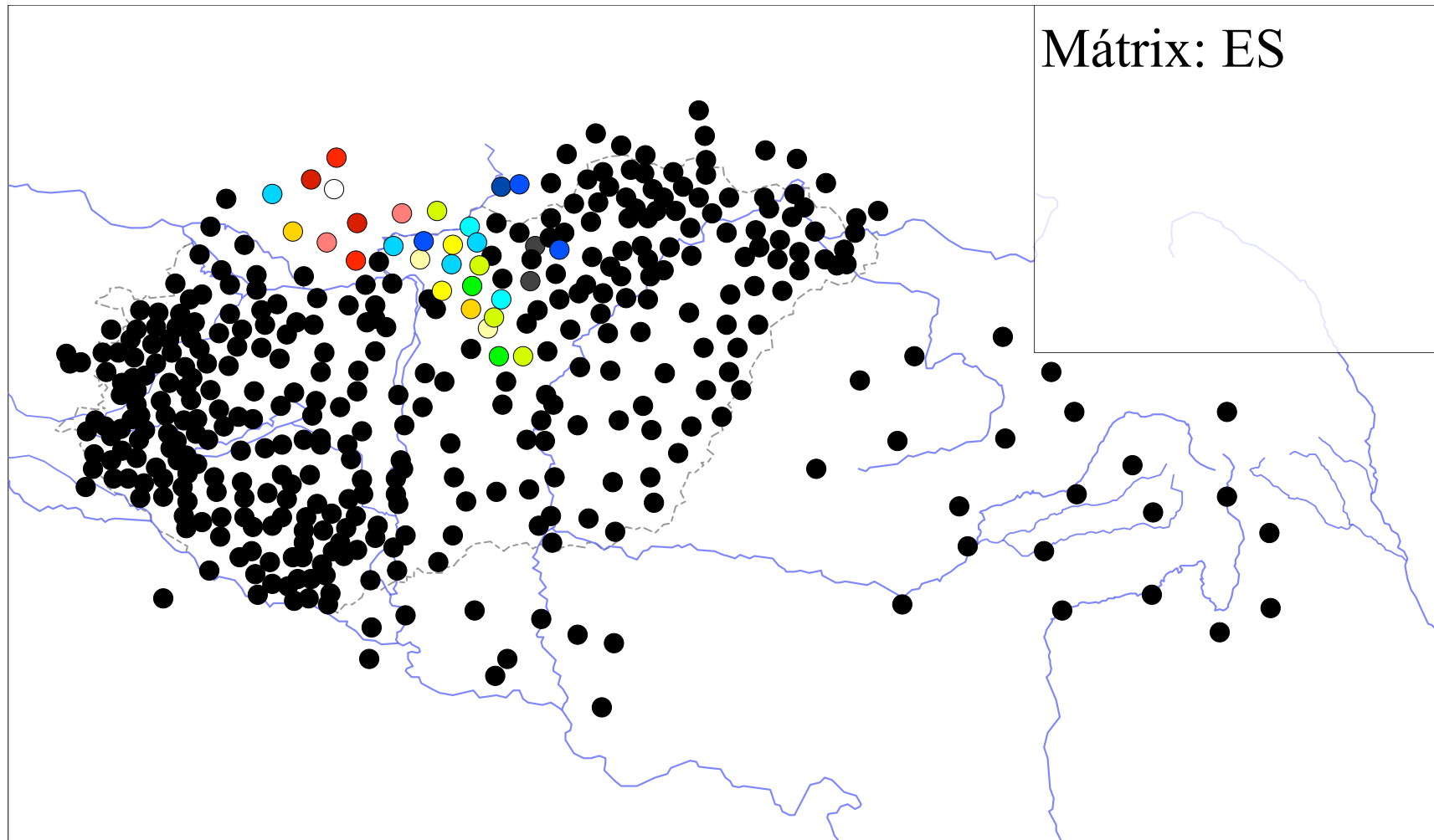
Kupuszina (Bácskertes)



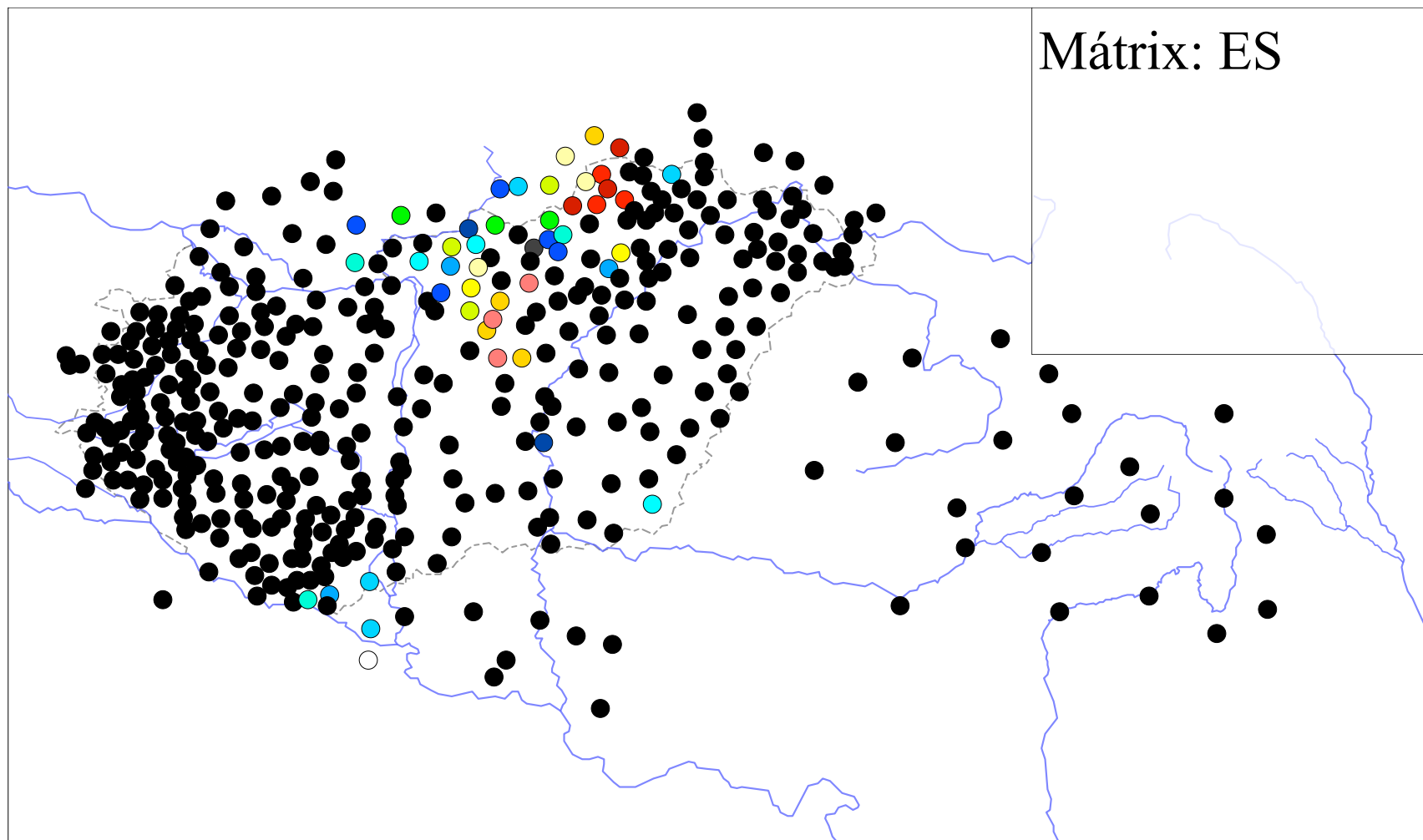
Kupuszina (Bácskertes)



Nagyhind

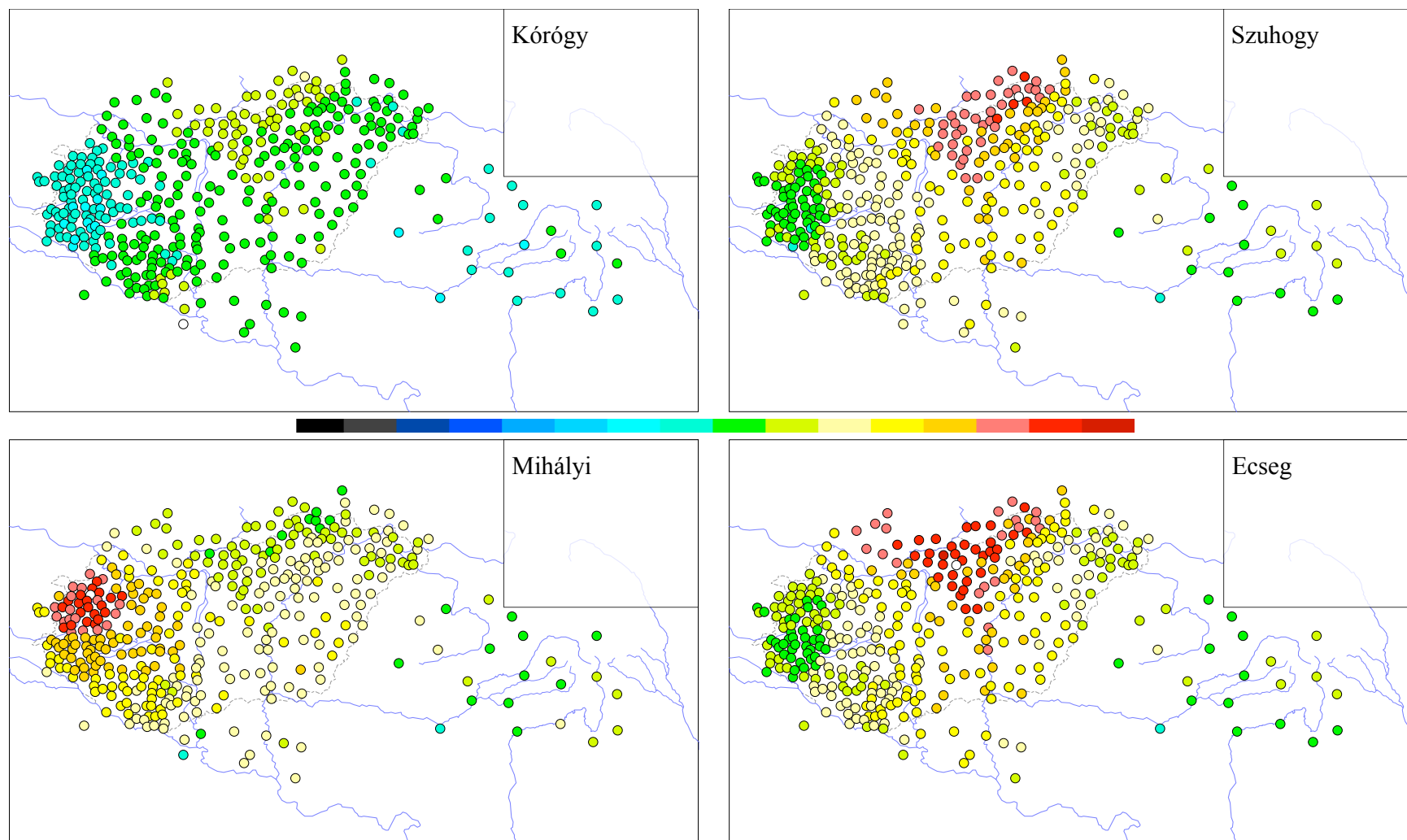


Kórógy 1.

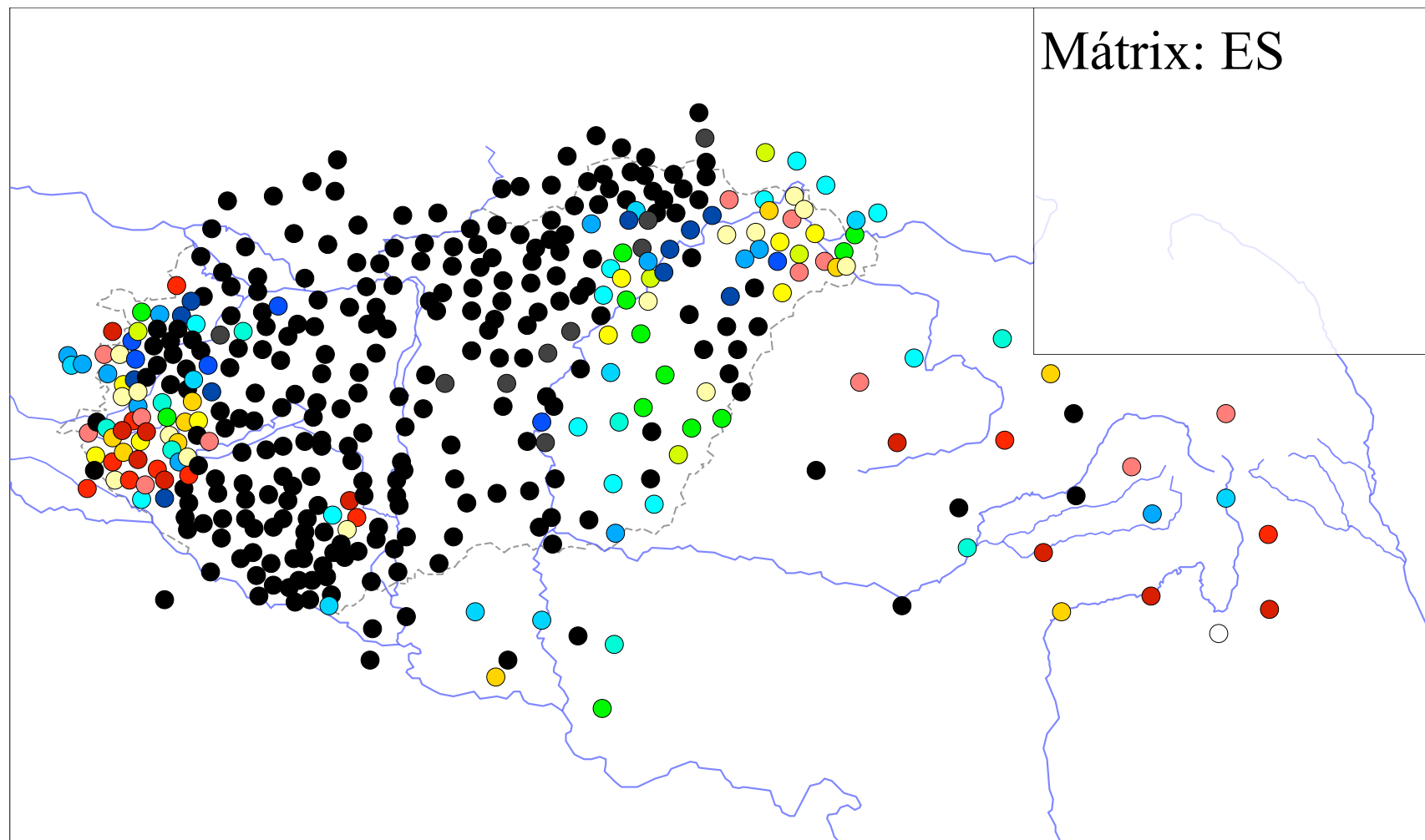


Települések kapcsolatai állandó arányokkal

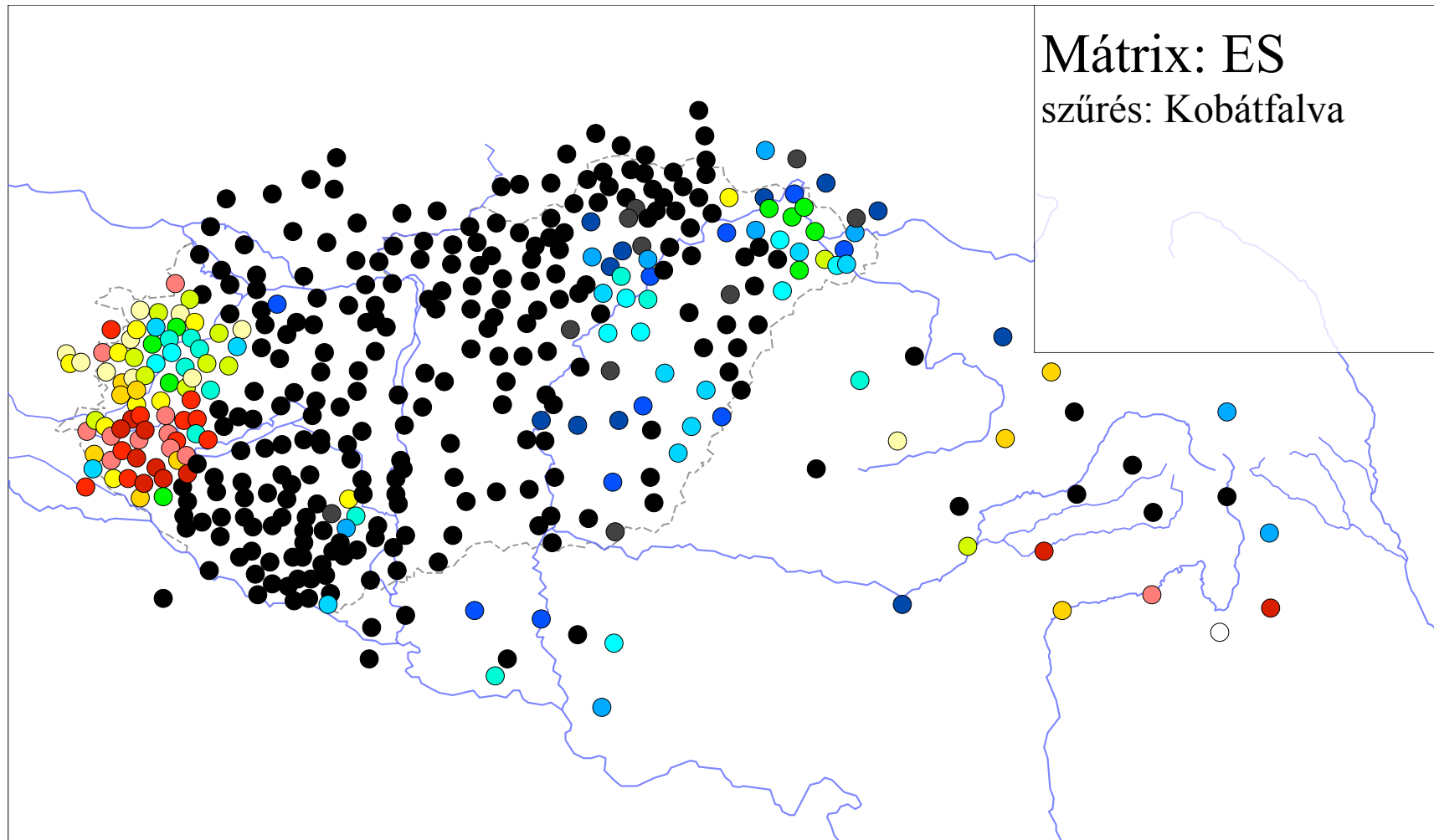
Mátrix: ES; rekesz: 57



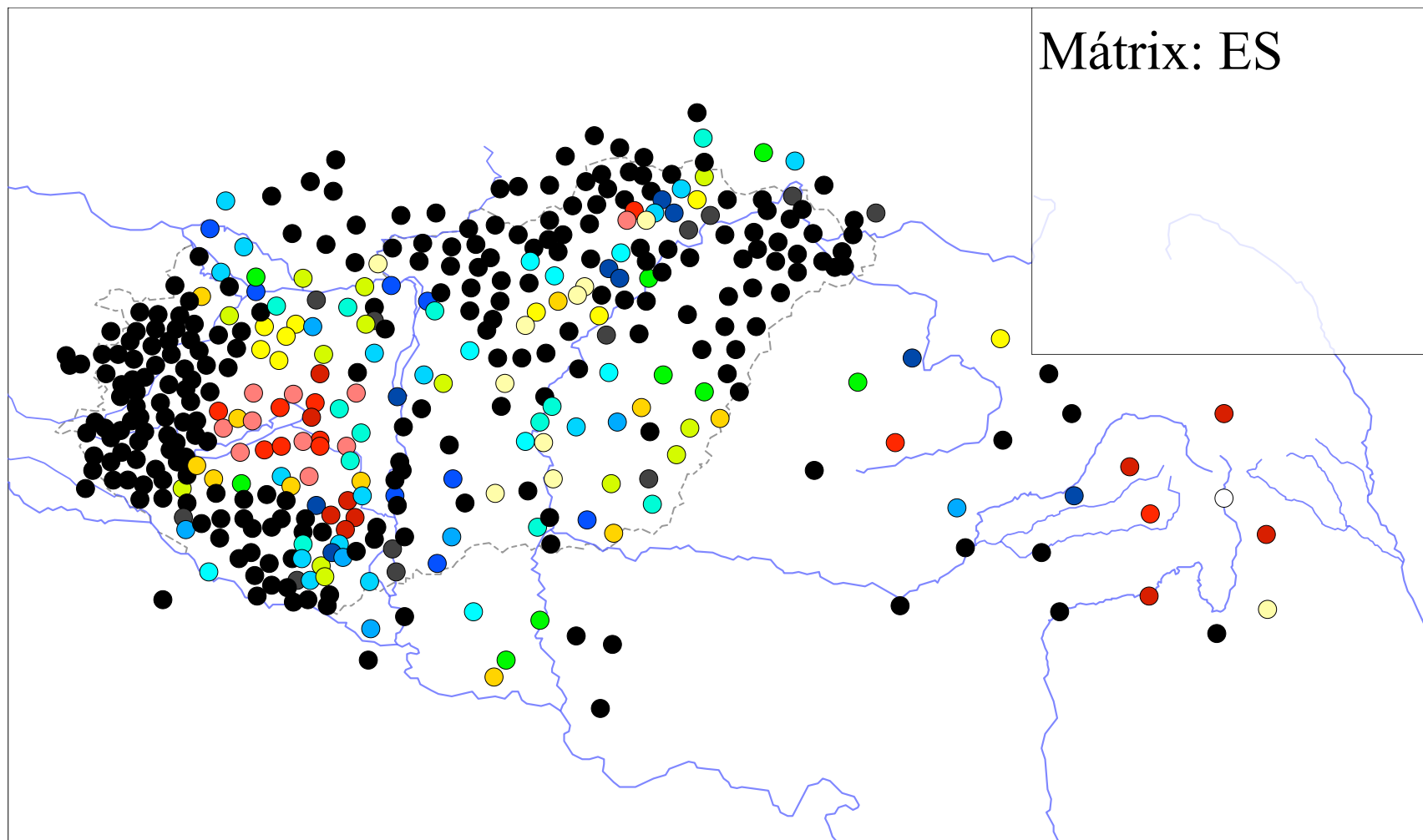
Csernátfalu



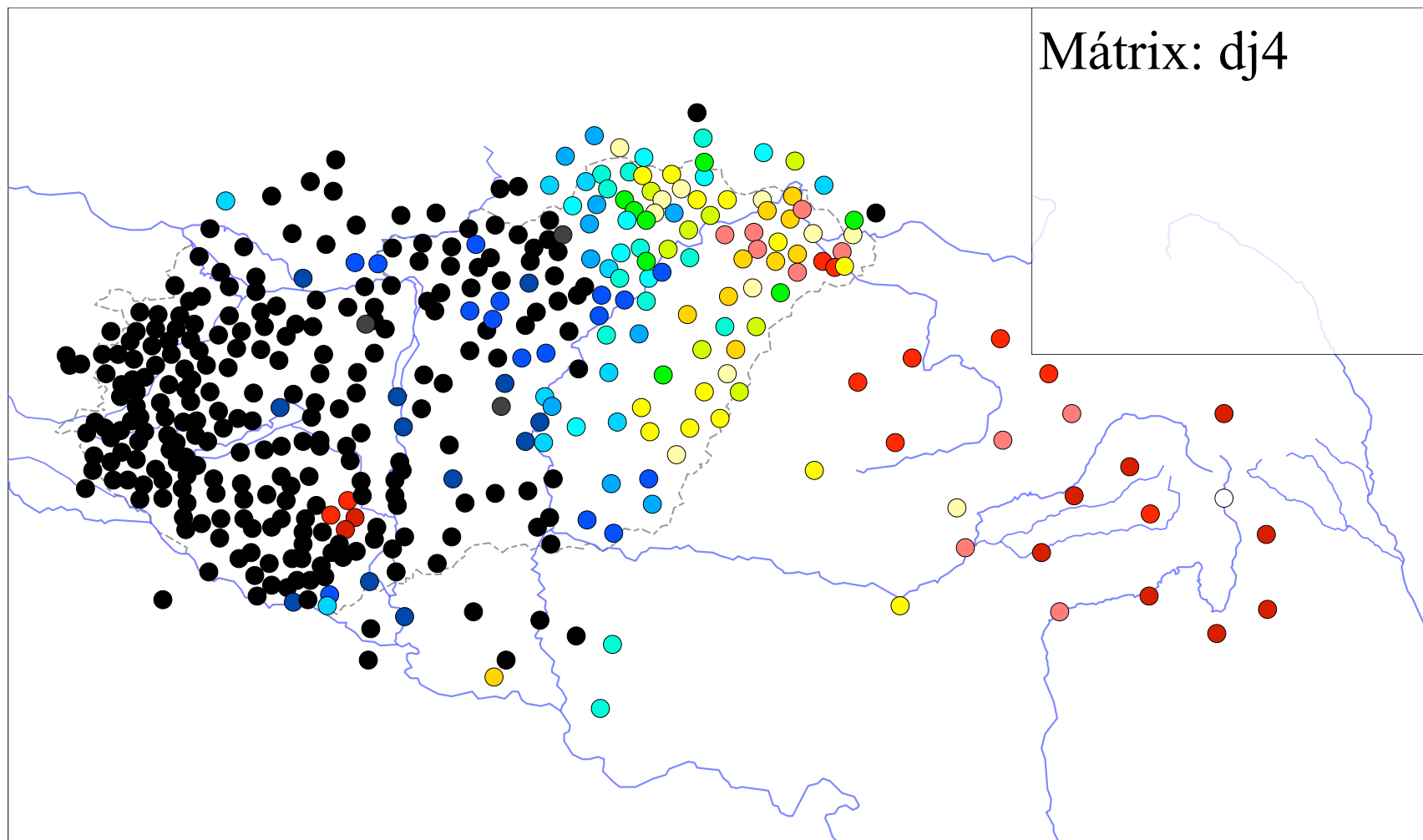
Csernátfalva



Csíkrákos

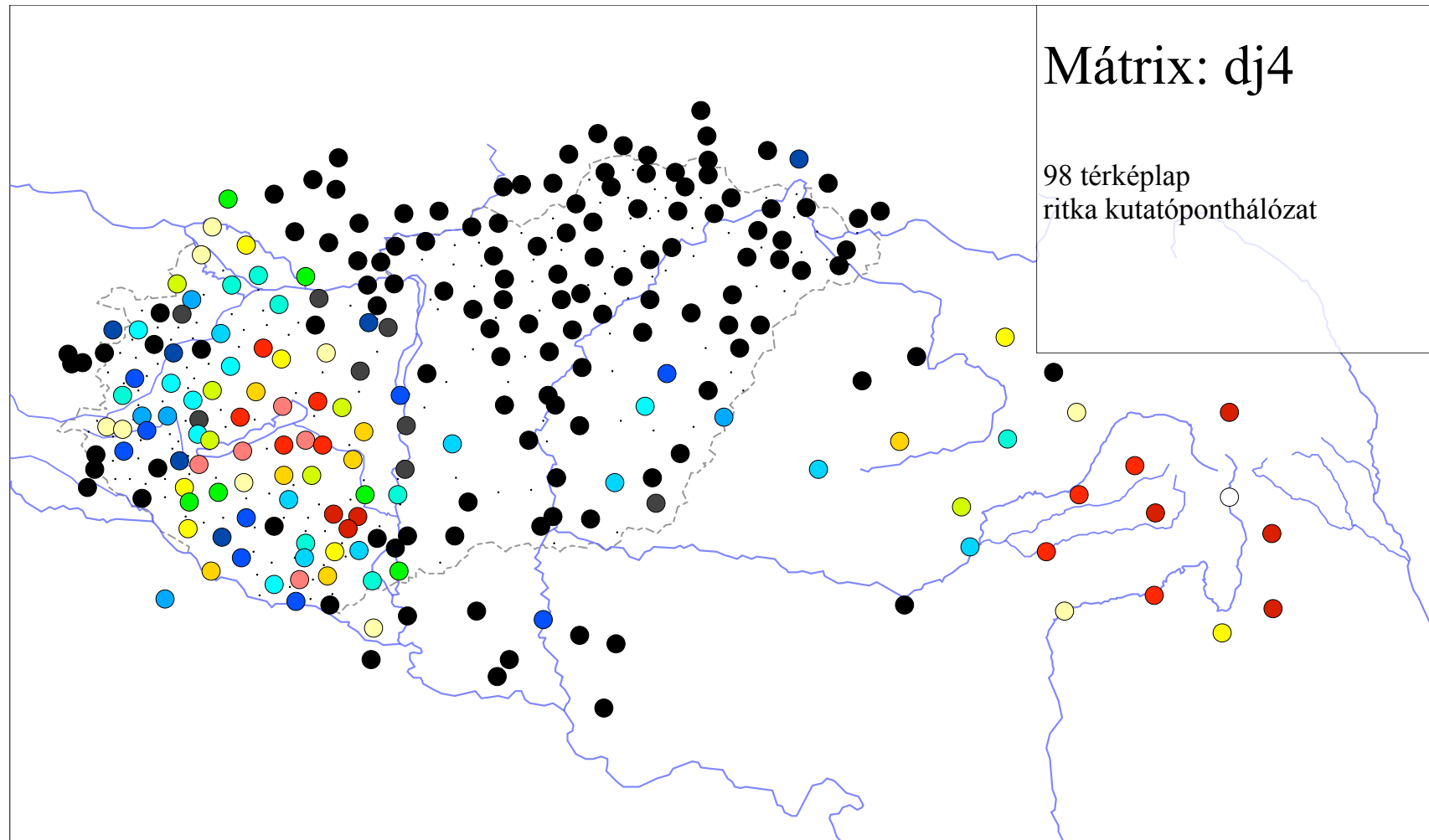


Csíkrákos



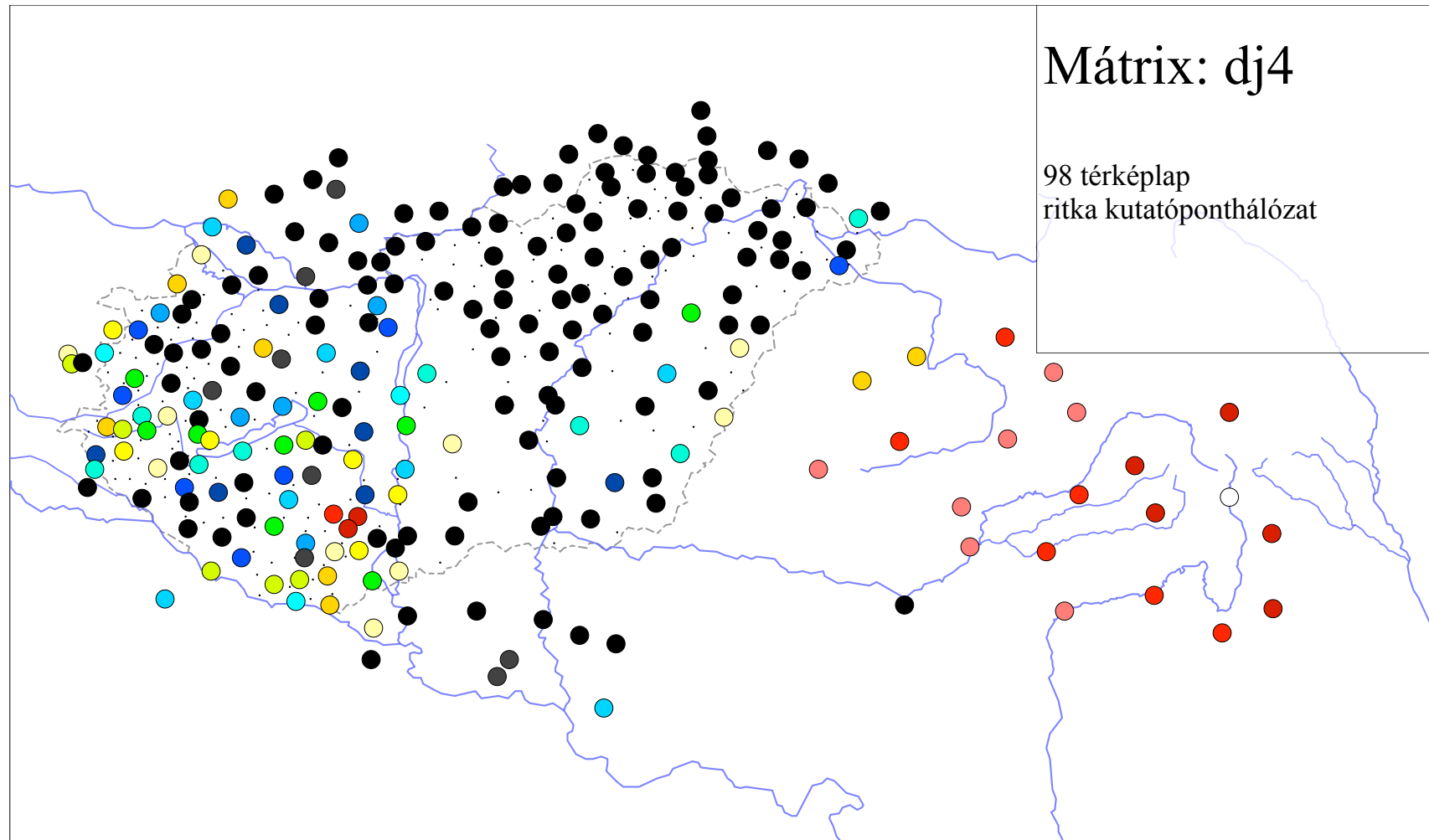
Benkő L. válogatása (1966)

Csíkrákos

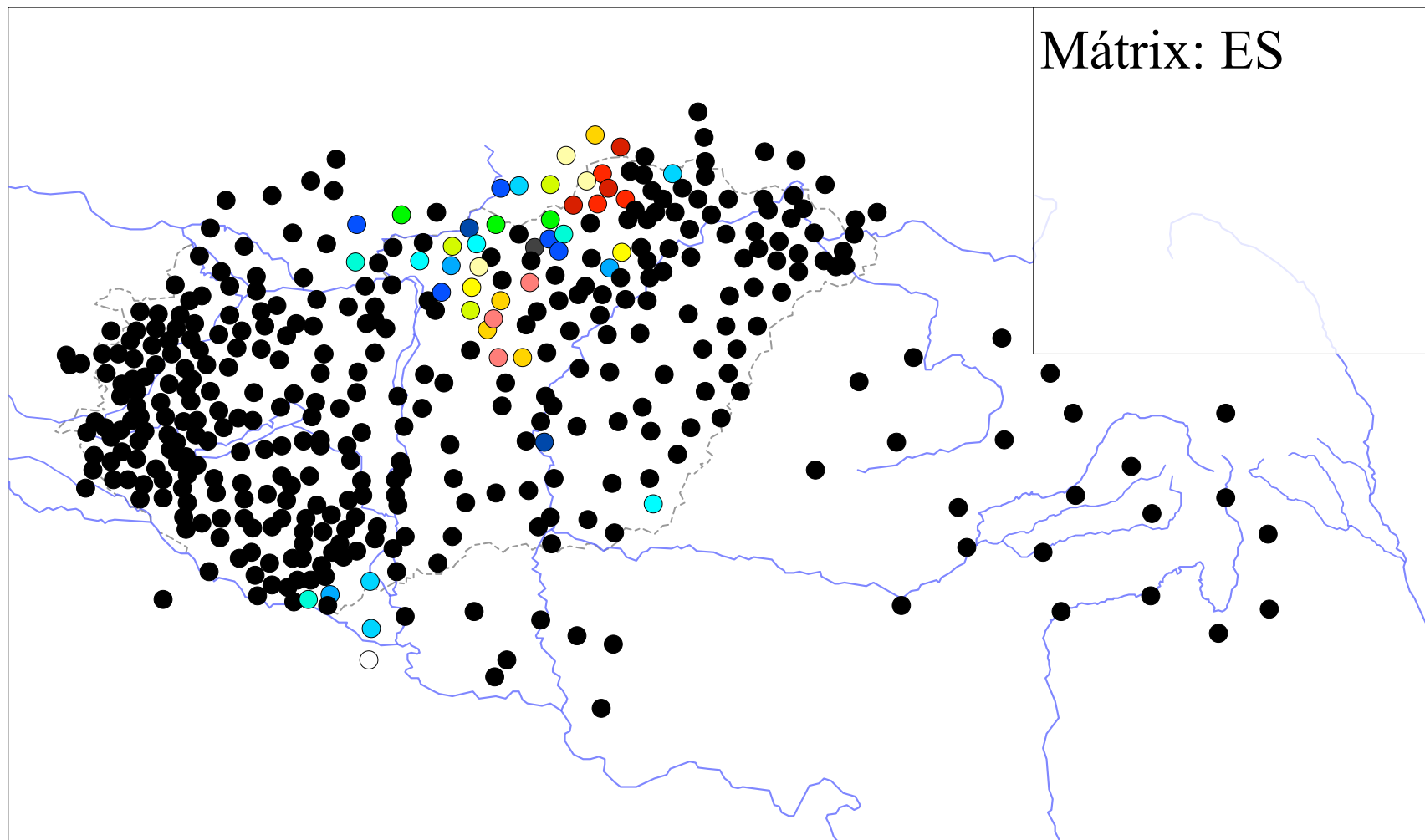


Benkő L. válogatása (1966)

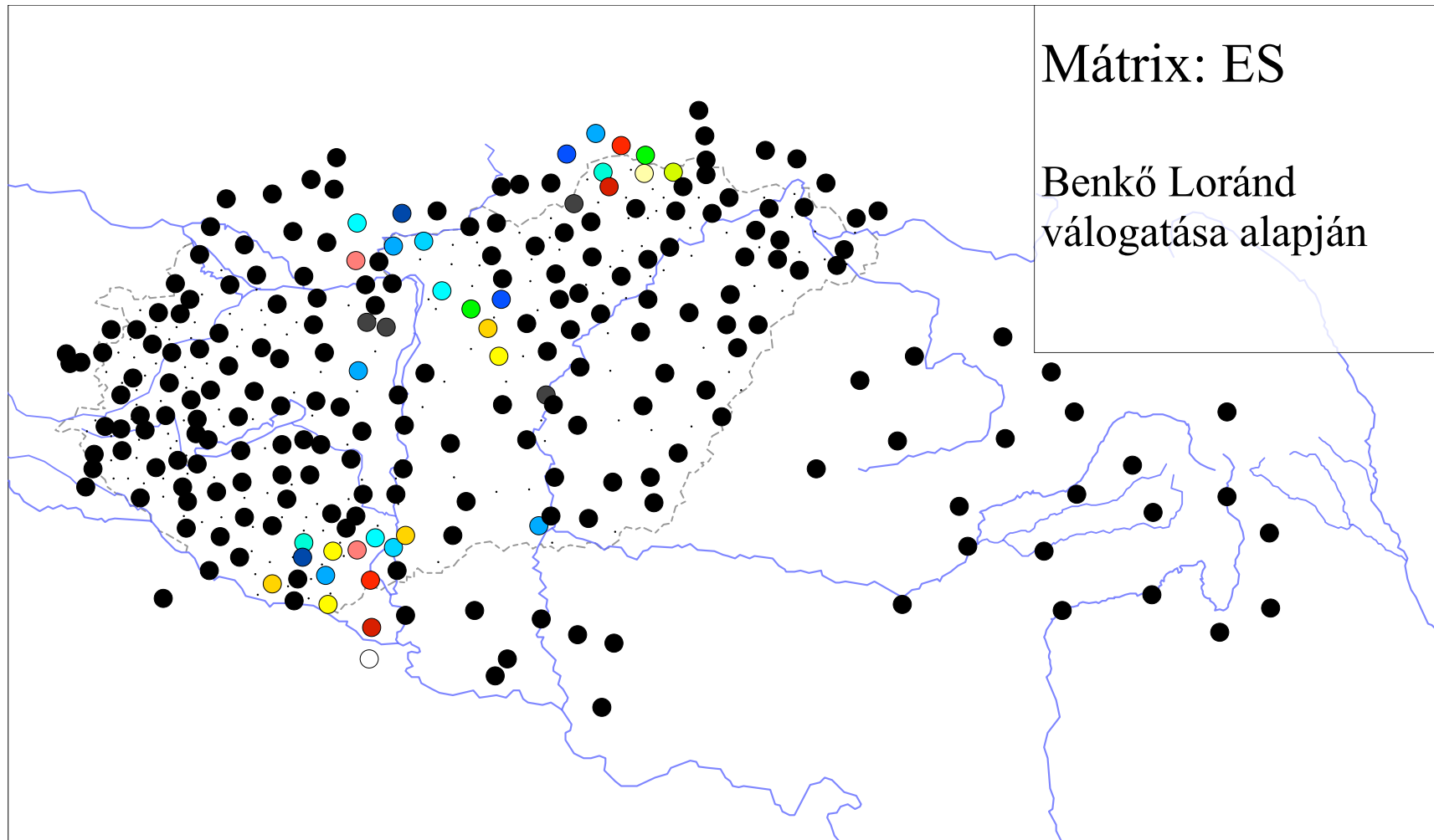
Csíkrákos



Kórógy 2.



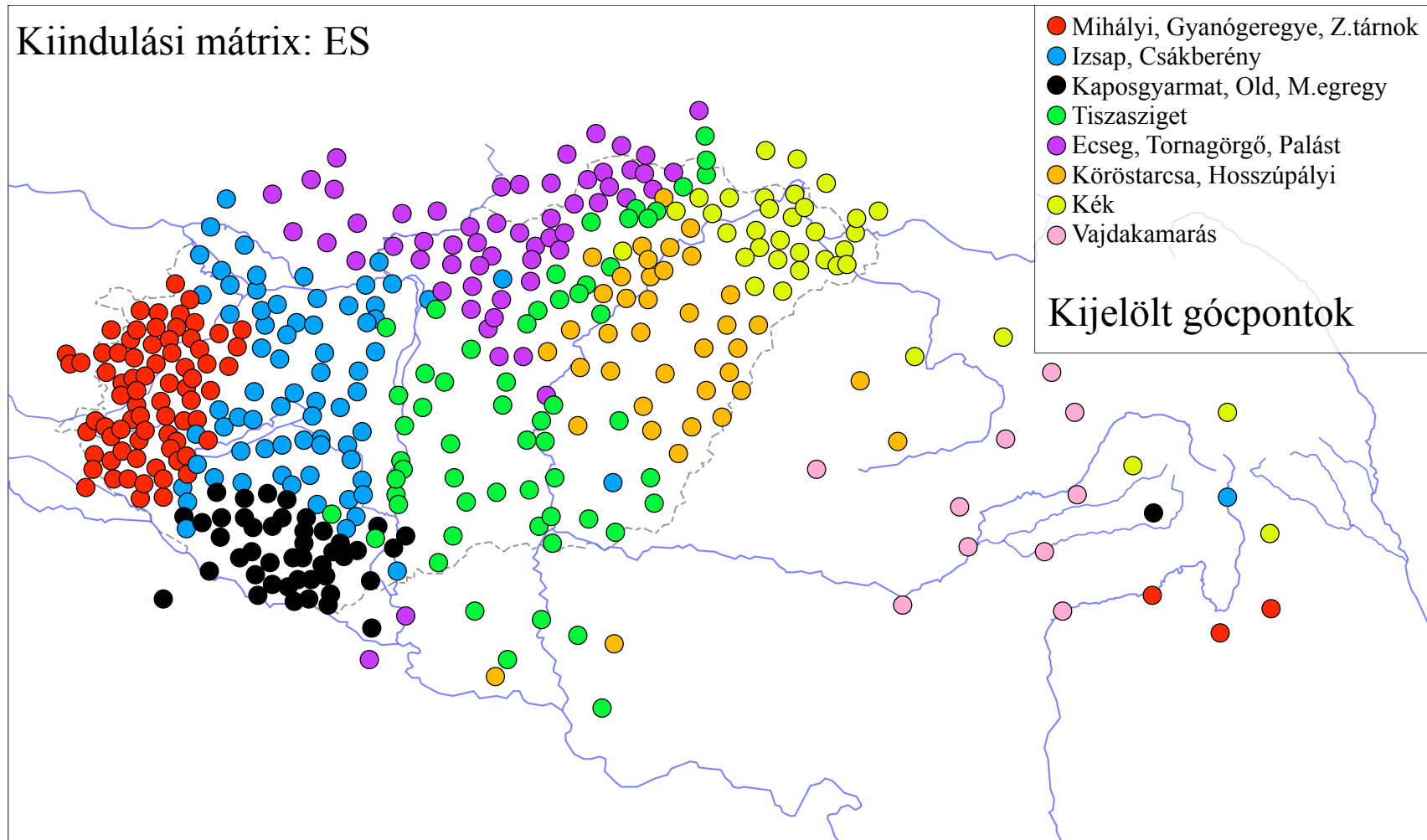
Kórógy 2.



Kutatópontok csoportosítása, területek meghatározása

- automatizált csoportosítás hasonlóság alapján, nyelvész közreműködése nélkül
 - klaszteranalízis
 - többdimenziós skálázás
- félig automatizált csoportosítás (a gócpontokat nyelvész jelöli ki)

Nyelvjárási régiók :-)



Miből lesz térkép?

- A dialektometriai térképek feltétele: informatizált nyelvjárási adattár létrehozása
 - MNyA.: Balogh Lajos, Bodó Csanád, Bükkösi Hajnal, Vékás Domokos (és mások)
 - S–ZA.: Király Lajos, Vargha Fruzsza, Cs. Nagy Lajos (és mások)

Kitekintés

- adattárak dialektometriai célú integrálása
- dialektometriai térkép akusztikai adatokból
- a kiadatlan MNyA.-térképlapok informatizálása

**KÖSZÖNJÜK A
FIGYELMET!**

www.bihalbocs.hu