

Földtudományok Doktori Iskola

A közelmúltbeli európai tiltakozások többnyelvű  
közösségi média adatainak tér- és időbeli, illetve  
szemantikai vizsgálata gépi tanulással  
– Twitter adatok alapján –

A Doktori értekezés tézisei

Kovács Tamás

Témevezetők:

Dr. Németh Ádám

Prof. Dr. Siposné Kecskeméthy Klára

Pécsi Tudományegyetem  
Természettudományi Kar  
Pécs, 2022

## 1. BEVEZETÉS

Különböző helyszíneken időről időre megjelennek tiltakozások és társadalmi megmozdulások. Sok tényezőtől függő, sokféle jellemzővel bíró események. Sok esetben sikeresen növekednek, másokat is bevonnak, így egyre nagyobb komplexitású hálózatot építve. A tüntetések kiindulópontját, okát, a történeti és szociológiai kutatásokban is trigger eseménynek nevezik. Az ilyen jellemzőkre példák az olyan események, amelyek általában nagyobb, egymástól függő csoportokat érintenek. Elég csak a francia forradalomra gondolni, amely egy élelmiszerhiányból indulva felforgatta a legapróbb francia falvak életét, később pedig Európa mindennapjai is megváltoztak miatta.

Az igazságtalanság vagy egyszerűen a jó és a rossz kiegyensúlyozott helyzetének felborulása gyakran arra készíti egy csoportot (vagy akár az egész társadalmat), hogy a politikát és a hatalmat érintő változást mozdítsanak elő, például tiltakozások formájában. Az elmúlt évtizedekben az internet alapú kommunikációs technológiák megjelenése és rohamos térhódítása átalakította mindennapjainkat, a szerveződési rendszerek is átalakultak. Az ún. kapcsolati cselekvésen (connective action) keresztül sokan ha nem is keresik a változáskeresés alapját, az ok és az igazolás egyre könnyebben talál rájuk. Az ember és szándékai mellett napjainkban az információ szerepe, terjedése, ill. annak egyre akadálymentesebb elérése egyre nagyobb jelentőséget kap. A közösségi média platformok, mint például a Twitter a kollektív cselekvés új közvetítőjének számítanak, ahol a civil mozgalmak különféle formái egyesülnek a nyilvános posztok körül, gyakran közös hashtaget használva, ezzel is erősítve az információáramlás és közvetve egy-egy mozgalmat.

## 1. PROBLÉMAFELVETÉS

Egy, a közösségi média bejegyzéseire és tevékenységeire támaszkodó, adat vezérelt analitikus megközelítésnek számos erőssége lehet a mozgalmak elemzésében – különös tekintettel a nagy időbeli felbontásra és arra, hogy a felhasználók gyorsan reagálnak bizonyos hírekre és információkra, miközben azzal sincsenek tisztában, hogy éppen egy elemzés tárgyai. A Twitter-adatok egyedülálló és hasznos információforrásként szolgálnak a civil mozgalmak elemzéséhez, mivel az elemzés fontos mintázatokat tárhat fel tér-időbeli és hangulati vonatkozásokban, ami szintén segíthet megérteni a tiltakozás tér- és időbeli eszkalációját, motivátorait. A közösségi média vizsgálata olyan események esetében, mint a 2020-as fehéroroszárszági elnökválasztás után kibontakozó tiltakozó mozgalom, vagy éppen a 2018-as szlovákiai újságírógyilkosság fontos eszköznek tűnhet az emberek reakcióinak nyomon követésére és megértésére, ellentétben más módszerrel és információforrásokkal szemben.

A jelenlegi doktori kutatásban kidolgozott módszertani munkafolyamat az idősorok klaszterezését ötvözi a georeferált közösségimédia-adatokon végzett szemantikus témamodellezéssel és hangulatelemzéssel, amely multimodális betekintést nyújt a közvélemény adott politikai eseményre adott reakcióiba. A javasolt megközelítés magában foglalja a többnyelvű korpuszfordítást, valamint a

hely- és hangulelemzést, gépi tanulási és témamodelllezési módszerekkel annak érdekében, hogy a kollektív cselekvés rejtett érdekeit és motivátorait feltárhassuk. Ezáltal az értekezésben felvázolt megközelítés határozott előnyt élvez azokkal a korábbi vizsgálatokkal szemben, amelyek elsősorban vagy a hashtag-aktivitásra (a térbeli dimenziók figyelmen kívül hagyásával), vagy éppen ellenkezőleg, csak helyspecifikus hashtagok feldolgozására irányultak. Ezzel szemben a szinte teljesen automatizálható gépi tanulási algoritmusok és technikák alkalmazásával az elemzés a bemeneti adatok sokkal szélesebb körét fedi le, mint a meglévő tanulmányok, ahol a kutatók kizárólag manuálisan értékelik a bejegyzéseket. Összességében ezzel a vegyes megközelítésű módszerrel a jelenlegi munka túlmutat a társadalmi mozgalmakkal kapcsolatos kortárs kutatások azon korlátain, amelyek főként egy nyelvre és egy korlátozott térbeli területre összpontosítanak.

A disszertációban elemzett közösségimédia-adatokat a Twitter Streaming Application Programming Interface (API), az USA-beli hálózati és mikroblog szolgáltatás segítségével szereztük be. Az első adatkészlet kezdő dátuma a Ján Kuciak szlovák újságíró meggyilkolásának első hivatalos bejelentéséhez, míg az utolsó nap Robert Fico miniszterelnök lemondásának legkorábbi nyilatkozatához (2018. február 26. és március 15.) igazodik. A második adatkészlet kezdő dátuma a 2020-as fehérorosz elnökválasztás napjához, az utolsó nap pedig Alekszandr Lukasenko fehérorosz elnök hivatalos beiktatásához (2020. augusztus 9. és szeptember 23.) igazodik. Mindkét adatkészlet a tweetek tartalmából és további attribútumokból áll, mint például a felhasználónév, a felhasználó helye és a tweet közzétételének időbélyege.

## **2. CÉLKITŰZÉSEK**

A disszertáció egy átfogó módszertani megközelítést javasol a jelenlegi módszerek korábban említett korlátjainak leküzdése érdekében, valamint a tiltakozások összetettségének elemzésére. Ez a munka a Twitter-adatok térbeli, időbeli és érzelmi markereinek hasonlóságait és összefüggéseit vizsgálja egy új, adatalapú kombinált megközelítés kidolgozásával, amely két különböző kelet európai tiltakozó mozgalmat vizsgál esettanulmányokon keresztül. Elsőként a szlovák újságíró, Ján Kuciak és menyasszonya, Martina Kusnirova 2018-as merényletének európai hatását, valamint a 2020-as fehérorosz elnökválasztás hatását, amelyet a korabeli kutatás ez idáig nem elemzett.

A tiltakozó mozgalmak, tüntetések dinamikus és kihívásokkal teli természetének multispektrális értelmezése hatékony analitikai módszert igényel, amely elősegíti ezen események dinamikájának átfogóbb megértését. Ennek érdekében ez a munka két szempontból is túlmutat a jelenlegi kutatás megközelítésén. Először is bemutatja, hogy a georeferált közösségimédia-adatok hogyan használhatók fel politikai események elemzésére, különös tekintettel a kisebb térbeli és társadalmi léptékre, valamint nem kizárólag angol nyelvre koncentrálna. Másodszor, módszertani szempontból egy új algoritmikus munkafolyamatot javasol, amely ötvözi az idősoros klaszterezést szemantikus témamodelllezéssel és érzelmek elemzésével a georeferált közösségimédia-adatokon.

Ezen túlmenően ez a kutatás erőfeszítéseket tesz az elmúlt évtized releváns kutatásainak áttekintésére.

Az eredeti perspektíva bemutatásával és a meglévő elemzések bemutatott korlátjainak figyelembevételével jelen disszertációban a következő kutatási kérdésekre kívánunk választ adni:

1. Időbeli és térbeli vonatkozások:

1. Hogyan változott a Kuciák meggyilkolásával kapcsolatos Twitter-tevékenység Európa-szerte? (RQ1a)
2. Hogyan változott a fehéroroszországi elnökválasztáshoz kapcsolódó tweetelési tevékenység Európa-szerte? (RQ1b)
3. Lehetséges azonosítani bizonyos események és incidensek hatását, mint például a médiajelentések, külpolitikai döntések vagy a nyomozás megállapításait a tweetelési tevékenységen alapján? (RQ1c)

2. Tartalmi szempontok:

1. Hogyan változik a tweetek hangulata az idő múlásával, és hogyan kapcsolódik konkrét eseményekhez és hírekhez? (RQ2a)
2. Milyen témák körül koncentrálnak az üzenetek? (RQ2b)

3. Profilok:

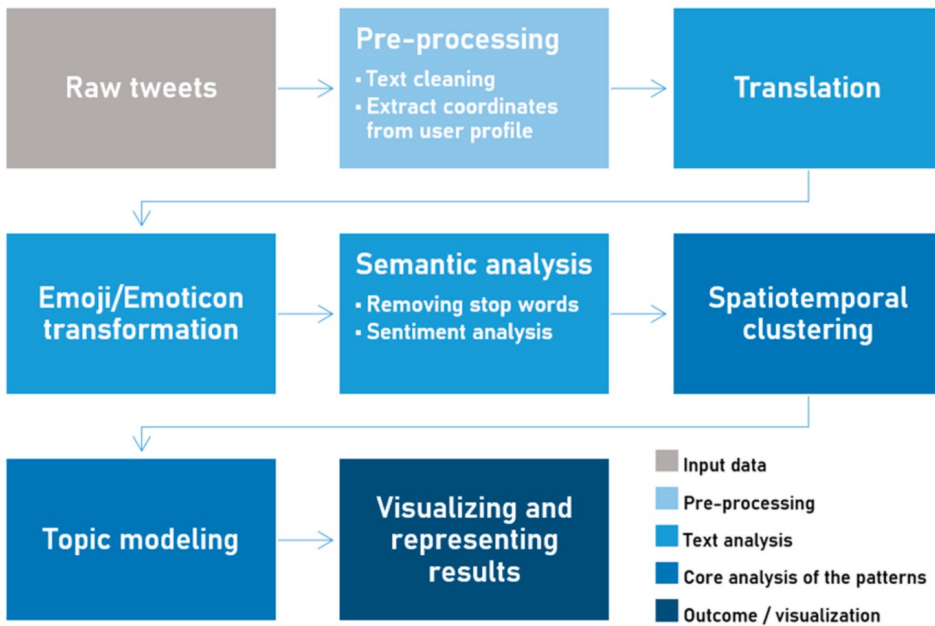
1. Hogyan jellemezhetjük az egyes országokat a tweetelési tevékenység időbeli vonatkozásai és a tweetek hangulata alapján? (RQ3a)
2. Az országok kategorizálása is tükrözi-e az azonosított témák eltéréseit vagy a hangulati értékek időbeli változását? (RQ3b)

Az összetett kérdéssor elegendő adatot szolgáltat ahhoz, hogy klasszikus kapcsolati keretrendszerben is értelmezzük a vizsgált eseteket.

### **3. MÓDSZERTAN**

Az ebben a munkában elemzett Twitter-adatokat a Twitter Streaming Application Programming Interface szolgáltatta. Az adat-előfeldolgozási megközelítés (amely a "nyers" tweeteken hajtottunk végre) egy alapos szövegtisztítási munkafolyamatból és a rendelkezésre álló értelmes helyinformációk koordinátákká történő átalakításából áll, a térelemzési lépésben történő további hasznosítás érdekében. Az előfeldolgozás első részében a módszertan egy elsődleges szűrés alkalmaz az adatkészleten, annak érdekében, hogy kizárja azokat a rövid tweet üzeneteket, amelyeknek nem rendelkeznek értelmezhető szemantikai jelentőséggel. Továbbá az ún. stop szavak, illetve a ritka, valamint a túl gyakori szavakat is eltávolításra kerülnek az adatkészlet normalizálása és a tweetekből származó redundáns zaj csökkentése érdekében. Ezt követően történik a felhasználók elérhető helyinformációinak felhasználásával az üzenetek térbeli lokalizálása (más néven geoparsing), amely a geokoordinátával nem rendelkező üzeneteket érinti. Ez a lépés alapvető jelentőségű annak érdekében, hogy az adatkészlet mérete növelhető legyen

a soron következő a gépi tanulás alapú fordításhoz valamint a térbeli és időbeli elemzéshez.



1. *Ábra*: Munkafolyamat fontosabb lépései

### 3.1. Előzetes szövegtisztítás

A szövegtisztítás célja a későbbi fordítási folyamat hatékonyságának növelése. Ennek érdekében eltávolításra kerülnek a rövid (egy szót tartalmazó) tweetek, illetve azok a bejegyzések, amelyek csak hashtageket vagy URL-eket tartalmaznak, hiszen nehezen értelmezhető vagy semmilyen szemantikai értékkel nem rendelkeznek. Ezt követik az ún. válaszjelölők (@user\_name) eltávolítása, mivel ez az elemzésben szükségtelen zajnak minősülnek; ezzel szemben a hashtagek megmaradnak, de a (#) jelük eltávolításra kerül. E lépés mögött az a megfontolás áll, hogy a felhasználók hajlamosak a hashtageket a szintaxis szerves részeként használni. Végző tisztítási lépésként eltávolításra kerülnek az „újsor” karakterek, valamint a további szóközők az üzenetekből.

### 3.2. Tweetek lokalizálása a koordináták és felhasználói profiladatok segítségével

Általában véve a koordinátákat rendelkező tweetek az összes üzenetnek csak egy kis szeletét (1-2%) teszik ki. Ennek a korlátnak a leküzdésére a vázolt analitikai megközelítés megkísérli lokalizálni azokat az üzeneteket is, amelyeknek nem rendelkeztek geokoordinátával.

A térbeli hivatkozással rendelkező tweetek mennyiségének növelése

érdekében a Twitter-felhasználók profiljából kinyerhetőek olyan (hely)információk, amelyek adott esetben proxyként szolgálhatnak ahhoz, hogy a felhasználó aktivitási helyét azonosítsuk (pl. munkahely). A user\_location adatokkal kapcsolatos információk megszerzéséhez először az egyes user\_location adatok gyakoriság szerint rangsorolásra kerülnek, majd a Python beépített mapping funkcióját és az általam fejlesztett fordító szótárat használjuk a helyelemek átalakítására (egységesítésére) és csoportosítására (pl. Germania; Németország; Deutschland... → Germany). A user\_location transzformáció utáni következő lépés az OpenStreetMap Geolocator funkciójának alkalmazása a Geopy-n Python csomagon keresztül, amely szélességi és hosszúsági koordinátákat rendel a nevekhez, amelyeket a későbbi lekérdezési alkalmazásokhoz használhatunk

### **3.2.1.Fiók helye és felhasználói leírása**

A felhasználó helye (user location) és a felhasználói leírás (user description) olyan mezők, amelyek földrajzi adatokat tartalmazhatnak néha akár elvont formában is. Ezek a mezők a felhasználó által szerkeszthetőek, így az információ hitelessége is a felhasználón múlik. Általában ezeket az értékeket nem gyakran módosítják, és nem feltétlenül írják le a tweet pontos helyét, de városi szinten a felhasználó lakóhelyét vagy munkahelyét jelölhetik. A legtöbb felhasználó rendeltetésszerűen használja ezt a mezőt; egyesek azonban anyanyelvükön töltik ki, míg mások kreatív módokon osztják meg helyüket, akár koordinátákkal, akár fiktív helyekkel, akár hangulatjelekkel kommunikálják azt, így zajt keltve a szabványos földrajzi helymeghatározási megközelítésekhez. Ezek a korlátok kiterjedt előfeldolgozási lépésekkel jelentősen enyhíthetők.

### **3.2.2.A különböző koordinátaformátumok átalakítása DMS-re**

Egyes felhasználók pontos földrajzi koordinátákkal jelölik lakhelyüket, ezt azonban különböző formátumokkal (DMS, MinDec vagy DegDec) teszik a user\_location mezőben. A disszertáció módszertana egyszerű regex mintakeresést használ ezen értékek kinyerésére, majd DMS-re történő konvertálására, amely használhatóvá teszi ezeket az adatokat a későbbi a földrajzi helymeghatározásnál.

### **3.2.3.Emoji szöveggé konvertálása**

A kinyerhető felhasználói helymeghatározás egy másik típusa az emoji zászlókon alapul. A user\_location korlátozott beviteli karaktermezőjének minimalizálása érdekében egyesek emoji-zászlóként, vagyis apró országpiktogramként adják meg saját országukat. A módszertan ezeket a piktogramokat a a Demoji Python könyvtár segítségével szöveggé alakítja, így ezek az adatok is felhasználhatóvá válnak a földrajzi helymeghatározásnál.

### **3.2.4.Konzisztens helyesírás biztosítása a felhasználó által szerkeszthető mezőkben**

Noha a vizsgált adatkészlet többnyelvű elemeket tartalmaz, ennek ellenére a javasolt módszertan helyesírás-ellenőrzést javasol a felhasználói hely és leírás mezőkben,

amely kijavíthatja a hibás gépelést, ezzel is növelve a későbbi földrajzi helymeghatározás hatékonyságát.

### **3.2.5. Geoparsing és geokódolás**

A módszertan geoparsing lépése POS tagging eljárással kivonja a helyneveket a szövegből, majd a GeoNames adatbázis segítségével lokalizálja őket. A megközelítés gép tanuláson és a Mordecai keretrendszerén alapul, amely egy geoparsing rendszer, amely a Spacy POS-taggerét, valamint Elasticsearch-et használ a földrajzi információkat lekérdezésénél.

### **3.2.6. Fordítás Google API és BERT transzformátor segítségével**

A tweetek többsége (68%) az angoltól eltérő nyelven íródott, ezért vagy le kell fordítani vagy „transzformálni” kell őket, mielőtt a következő elemzési lépésekben felhasználásra kerülnének. A kisebb szlovák adatkészlethez a TextBlob szövegfeldolgozó könyvtárat használtam a Google Fordító API-val, míg a nagyobb fehérórosz adatkészletnél a BERT előtanított „xlm-r-100langs-bert-base-nli-stsb-mean-tokens” többnyelvű transformer modellt alkalmaztam a különböző nyelvek egységesítésére.

### **3.2.7. Emoji/Emoticon szöveggéalakítás**

Általánosságban a szöveg-előfeldolgozási megközelítések eltávolítanak minden hangulatjelet a szövegből. Az ilyen megközelítések az a fő problémája, hogy a felhasználók ezeket a kis képeket és karaktereket a közösségi média lingua franca-jaként használják érzéseik és ötleteik kifejezésére. Eltávolításuk a szemantikai elemzés torzulásához vezethet. A módszertan ezért ezeket szöveges formátumra konvertálja, hogy megőrizze a szemantikai információt a további hangulatelemzés során.

## **3.3. Szemantikai elemzés**

A módszertani szemantikai szövegelemzési folyamata két szakaszra oszlik: először az algoritmusban szereplő stopszavak listáját bővítjük az adatkészletünk jellemzői alapján (outlierek), majd eltávolítjuk ezeket a szavakat a szövegből, másodszor pedig egy szótár alapú hangulatelemzés alkalmazunk, amely osztályozza az egyes tweetekben található szubjektív érzelmeket.

### **3.3.1. Stop-szavak eltávolítása**

A szakirodalom a segédigéket, kötőszavakat és az írott szöveg egyéb, jelentős szemantikai jelentéssel nem bíró részeit „stopszavaknak” tekinti. Ezeknek a szavaknak a listáját a Natural Language Toolkit (NLTK) előre meghatározza, a disszertációban használt algoritmus erre épít. Mindazonáltal a stopszó szótárat további, a szerkesztetlen szövegre, illetve az elemzett korpuszra jellemző szavakkal egészítettem ki, beleértve a speciális keresztneveket is. Ezeket a szavakat is eltávolítjuk az adatkészletből a három vagy kevesebb karakterből álló szavakkal együtt, mivel ezeknek szintén korlátozott a szemantikai jelentősége.

### **3.3.2. Hangulatelemzés**

A hangulatpontoszámok segítségével megállapítható, hogy egy adott tweet szövege mennyire pozitív vagy negatív. Ez az azonosítás a pozitív és negatív kifejezések mennyisége közötti különbség kiszámításával történik pozitív és negatív szavakat tartalmazó szókincs segítségével, automatizált módon. Erre a célra a VADER (Valence Aware Dictionary and sEntiment Reasoner) lexikont választottuk, amely egy szabályalapú hangulatelemző eszközt, amely lexikonalapú módszerrel és heurisztikával értékeli a bemeneti adatokat. Ez a módszer a közösségi médiában megjelenő érzelmekre szabványosított, és a legújabb szakirodalom tükrében magasabb osztályozási pontossággal rendelkezik, mint a többi módszer.

### **3.3.3. Tér- és időbeli adatfeldolgozás és klaszterezés**

A tiltakozások közösségimédia-aktivitáson alapuló eszkalációjának megismerése érdekében, a tweetek segítségével tér-időbeli elemzést is végeztem. A lekérdezés eredményeiben szereplő tweetek Európán kívülről is érkeztek; mindenesetre az elemzés során csak az európai országokat vettem figyelembe, köztük Oroszországot és Törökországot. Az adatok ezt követően országos szinten összesítésre kerültek, mivel a nyelv és az ország politikai jellemzői erősebben befolyásolhatják a tweetelési viselkedést, mint a városi szintű egyéb jellemzők vagy más finomabb térbeli léptékek.

Ezt követi a térbeli klaszterezést, annak érdekében, hogy megtaláljuk a hasonló tweetelési trendekkel rendelkező országokat, amelyek valószínűleg azt is jelezhetik, hogy mikor került sor a tiltakozásra, vagy más befolyásoló paraméterek jelenlétére, mint például a média vagy a politika. Erre a célra az ArcGIS Pro 2.8-ban az idősoros klaszterezést használtam, ahol az idősorok adatai három kritérium alapján csoportosíthatók: hasonló értékek az időben, amelyek hajlamosak egyszerre növekedni és csökkenni, vagy hasonló ismétlődő mintázatok. A hasonló mintázatú országok azonosításával feltárhatjuk a befolyásoló paramétereket és azt, hogy ezek a paraméterek hogyan változtak időben. Sőt, az eredmény tömörebb és informatívabb megjelenítését és értelmezését nyújtja, mint egyenként 39 ország statisztikai értékei. Ezen túlmenően ez a megközelítés képes elkerülni a különböző népességmérettekkel kapcsolatos problémákat, és nincs szükség népességen alapuló normalizálásra a klaszterezéshez.

### **3.3.4. Témamodellzés látens Dirichlet-allokációs (LDA) módszerrel**

A témamodellzés első lépése a tokenizálás, a szövegek kisebb egységekre bontása. Megközelítésemben a Gensim könyvtár egyszerű előfeldolgozó funkcióját használom, amely iteratív módon konvertálja a tokeneket Unicode karakterláncokká, eltávolítja az ékezetes jeleket, és kisbetűssé teszi a karakterláncot. Ezután a mondatfolyamból kiszűröm a leggyakoribb bi- és trigramokat (két- és háromszavas kifejezéseket). A megfelelő szűrési küszöb beállításához azonban először manuálisan vizsgálom meg ezeket a többszavas kifejezéseket a scikit-learn CountVectorizer segítségével, amely a szöveget tokenszámok mátrixává alakítja. Ezután a Gensim-



küszöböt úgy állítom be, hogy figyelmen kívül hagyja azokat a bi- és trigramokat a tér-időbeli elemzésben azonosított egyes klasztereknél, amelyek széles körben ismert információkat hordoznak, mint például a gyilkosság ténye (a szlovákiai adatállomány esetében) vagy az áldozat foglalkozása.

### **Lemmatizálás és vektorra alakítás**

A lépés során előbb egy adott szó ragozási és származékos formáit közös alapalakra redukáljuk, hasonlóan a töképző megközelítéshez. Lemmatizálás esetén azonban először meg kell határozni egy szó szófaját, például (POS tag) szimbólumokat, számokat vagy igéket, és a normalizálási szabályok az egyes szófajokra eltérőek lesznek, így lexikailag kifinomultabb. Ez a módszer magában foglalja az egyes szavak ragozott alakjainak csoportosítását is, amelyeket a szó lemmája vagy szótári alakja azonosít, így azok egyetlen elemként elemezhetők (jó, jobb, legjobb → jó), ezzel is növelve a téma-szó asszociációk jelentőségét. A lemmatizáláshoz a spaCy Lemmatizer-t használom, amely szabályalapú lemmatizálást biztosít azzal a beállítással, hogy csak a tulajdonneveket, igéket és főneveket engedélyezzünk az LDA korpuszunkhoz, mert kutatásunk olyan témákra koncentrálna, amelyek elsősorban arra a kérdésre keresnek választ, hogy ki, mit csinált és hol.

A téma modellezéséhez LDA-t használtam, Gensim könyvtárral a Python földrajzi helymeghatározott tweetek alapján. Megjegyzendő, hogy a mai napig nem létezik általánosan elfogadott a priori paramétermodellezési megközelítés az LDA számára. A legmegfelelőbb paraméterek, az alfa, béta és az adathalmazból kinyerhető témakörök számának megtalálása érdekében hiperparaméter-optimalizálást alkalmaztam, amely a legjobb beállítást keresi egy validációs korpuszban (75%). A teljesítmény-összehasonlításhoz a témakoherencia mérőszámot ( $C_v$ ) használtam, amely a legerősebb korrelációt mutatja az emberi értékelésekkel. A  $C_v$  értéke egy közvetett megerősítő mértéket kombinál, amely normalizált pontonkénti kölcsönös információt (NPMI. Pointwise mutual information), koszinusz-hasonlóságot, egy logikai csúszó ablakot valamint a felső szavak egykészletes szegmentálását használja. Ezt az optimalizálási megközelítést alkalmaztam a tér és időbeli elemzésben azonosított összes klaszterre. Végül a tweeteket a legnagyobb valószínűséget produkáló téma szerint osztályoztam; így létrehozva minden témakörhöz a 10 leggyakoribb kulcsszót.

### **3.3.5. Alternatív témamodellezési lehetőség a Bertopic segítségével**

Bár a Latent Dirichlet allokáció (LDA) napjaink egyik legelterjedtebb témamodellezési technikája, kipróbáltam egy másik módszert is a Transformers Bidirectional Encoder Representations (BERT) segítségével, amelyet a második nagyobb fehér orosz adatkészlet terjedelmi kényszerített ki. A témamodellezéshez a Bertopic modellt az előre betanított "xlm-r-100langs-bert-base-nli-stsb-mean-tokens" nyelvű beágyazási modellel használtam. A beágyazási modell segítségével kinyerhető az összes token kontextusba helyezett szóreprezentációja, amit ezt

követően átad a BERTopicnak, amely az Uniform Manifold Approximation and Projection (UMAP) algoritmust használja a beágyazás dimenziójának csökkentésére, megőrzve a beágyazások lokális és globális szerkezetét. Következő lépésként a sűrűség alapú algoritmus (HDBSCAN) klaszterezi a tweeteket, lehetővé téve a kiugró értékek azonosítását. Noha a módszer kétségtelen előnye, hogy automatikusan határozza meg a témák számát, azok jobbra sematikusak, így nem segítik egy komplex értelmezés kialakítását.

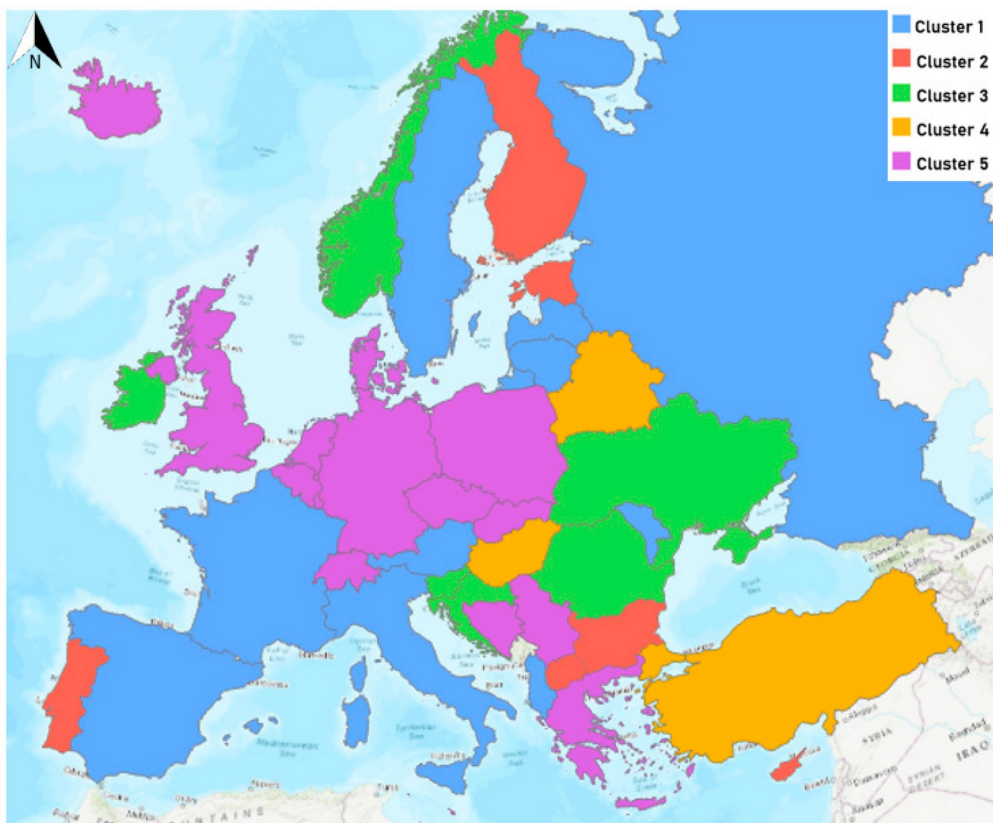
#### **4. EREDMÉNYEK**

4.1. Bár a Twitter Developer Platform szerint a tweetek körülbelül 1-2%-a rendelkezhet geocímkevel, míg a tweetek ~30-40%-a tartalmazhat további helyi kapcsolatos profilinformációt. A bemutatott vegyes geolokációs model módszerünk képes volt az eredeti adatkészlet körülbelül 60%-át lokalizálni.

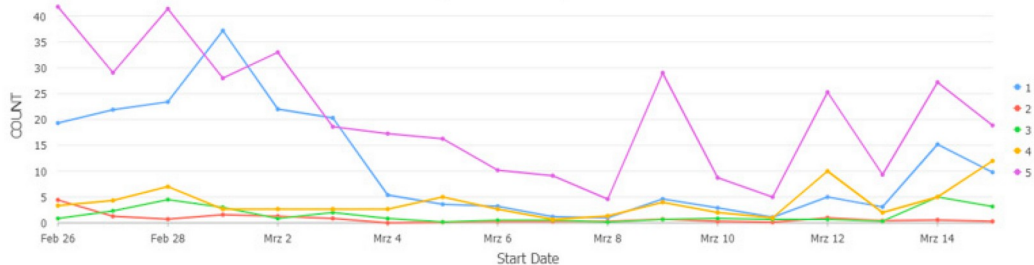
4.2. A kutatás esettanulmányokon keresztül bizonyította, hogy a Twitter-adatok egyedülálló és hasznos információforrásként szolgálnak a civil mozgalmak elemzéséhez, hiszen az elemzés fontos mintázatokat tárhat fel tér-időbeli és hangulati vonatkozásokban, ami a térben és időben történő tiltakozás eskalációjának megértésében is segíthet. A közösségi média vizsgálata olyan események esetében, mint egy újságíró meggyilkolása vagy a fehérorszországi tüntetések, minden más módszertől és információforrástól eltérően fontos eszköznek tűnik az emberek azonnali reakcióinak nyomon követésére és megértésére. Egy alapos adat-előfeldolgozást követően, beleértve a térinformációk kinyerését a felhasználók profiljából és a nem angol tweetek fordítását, az európai országokat az elemzési időszak tweetelési tevékenységének időbeli mintázatai alapján csoportosítottam, és megvizsgáltam, hogyan változtak a tweetek és a témák idővel ezekben a klaszterekben. Ez a megközelítés kimutatta, hogy a tweetelési tevékenység nemcsak konkrét eseményekre rezonál, hanem bizonyos esetekben egy adott ország politikai narratívájára is, amely befolyásolja a közbeszéd lefolyását.

4.3. A Twitternek, mint adatforrásnak számos előnye van a kollektív cselekvések elemzésében, különösen a nagy időbeli felbontásban, köszönhetően az emberek azonnali reagálásának az adott eseményekre, hírekre, amit a többrétegű elemzéssel illusztrálni kívántam. A doktori kutatásban kidolgozott módszertani munkafolyamat az idősorok klaszterezését ötvözi a georeferált közösségimédia-adatokon végzett szemantikus témamodellezéssel és hangulatelemzéssel, amely multimodális betekintést nyújt a közvélemény politikai eseményre adott reakcióiba. A javasolt megközelítés magában foglalja a többnyelvű korpuszfordítást/átalakítást, valamint a hely- és hangulatkivonást, gépi tanulási témamodellezési módszerekkel a kollektív cselekvés rejtett érdekeinek és motivátorainak feltárására. Ezáltal a megközelítés határozott előnyt élvez azokkal a korábbi vizsgálatokkal szemben, amelyek elsősorban vagy a hashtag-aktivizmusra (a térbeli dimenziók figyelmen kívül hagyásával), vagy éppen ellenkezőleg, csak helyspecifikus hashtagok felhasználására irányultak. Míg a szinte teljesen automatizálható gépi tanulási

algoritmusok és technikák alkalmazásával az elemzés a bemeneti adatok sokkal szélesebb körét fedi le, mint a meglévő tanulmányok, ahol a kutatók kizárólag manuálisan értékelik a bejegyzéseket. Az elemzés révén reflektálni lehetett arra, hogy mi is zajlik a kutatásban érintett országokban. Ezen túlmenően maga a munkafolyamat, amely magában foglalja a teljes előfeldolgozást és az azt követő elemzési lépéseket az országok klaszterezésére, valamint az adatokban szereplő témák és érzésmintázatok vizsgálatára, átvihetőnek tekinthető, és hasonló esetek elemzésére használható. Ezen túlmenően ennek az általános, feltáró jellegű elemzésnek az eredményei erős alapot jelenthetnek specifikusabb elemzésekhez, például országspecifikus vizsgálathoz vagy egy konkrét témára vagy napra való összpontosításhoz. Bár a jelenlegi módszert a Twitter adatokra támaszkodva alkalmaztuk, miután egy adott esemény már megtörtént, érdemes megjegyezni, hogy mind az elemzés tárgya, mind a régiója felcserélhető más témákkal, valamint más közösségi média adatforrásokkal, amelyek rendelkeznek a helyszínrre vonatkozó kötelező attribútumok.



Average Time Series per Cluster



## 2. *Ábra*: Kuciak adatkészlet klaszterezés

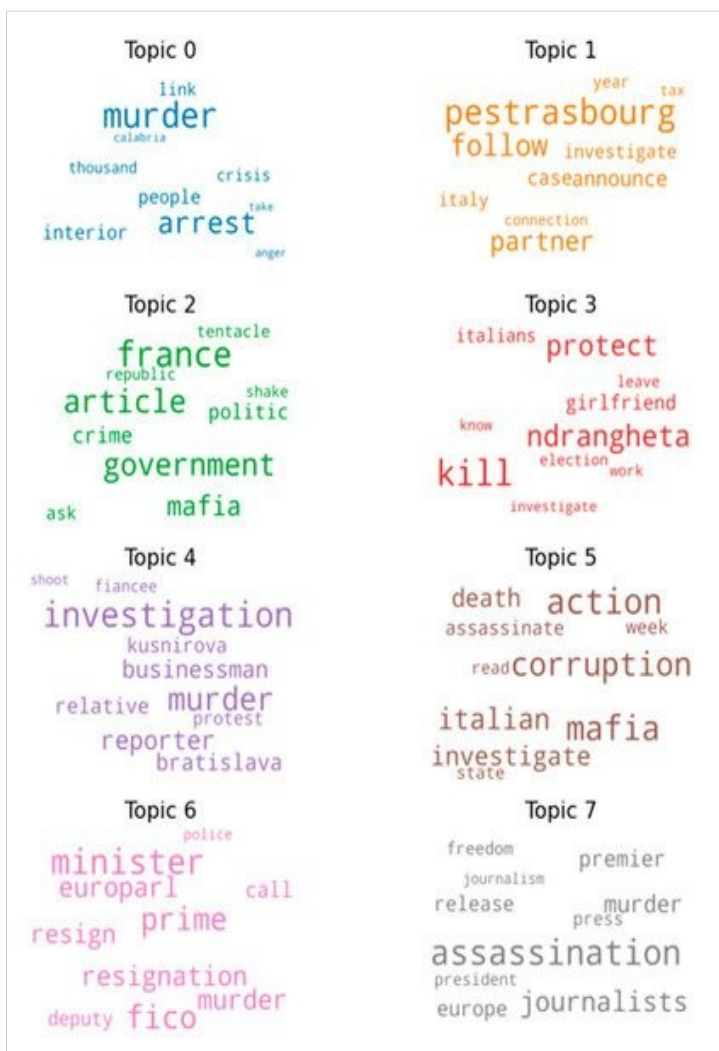
4.4. A Jan Kuciak meggyilkolása utáni Twitter-adatok elemzése során öt csoportot különítettünk el, amelyek közül két nagyobb tartalmazta a tweetek több mint 90%-át, és adta meg a legtöbb részletet az aktivitási csúcsokról. Az első klaszter kevésbé volt aktív az elemzési időszak második felében, ahol a temetés vagy a miniszterelnök lemondása történt. Ide tartoznának olyan országok, mint Olaszország, Franciaország vagy Ausztria. A második nagy klaszterbe többek között Szlovákia, Németország és az Egyesült Királyság tartozik, amelyek jelentősebb

csúcst értek el mind a vizsgált időszak elején, amikor kiderült a gyilkosság és a szlovák politikusok első kapcsolata az olasz maffiával, mind pedig a második szakaszban (temetés, miniszterelnök lemondása). A fennmaradó három klaszter csak kisebb számú tweetet tartalmazott, ezért nem végeztünk részletes elemzést, hogy elkerüljük a reprezentativitással kapcsolatos problémákat.

Topic #	Topic Contribution %	Keywords	Representative Text
0	0.771081	murder, arrest, interior, people, link, crisis, thousand, calabria, anger, take	in paris a citizens rally takes place on atpm in front of the slovak embassy
1	0.690083	pestrasbourg, follow, partner, announce, case, investigate, italy, year, connection, tax	the slaughter of colleagues continues now its the turn of jan kuciak in slovakia the coordination for the safety of journalists set up in italy is a model in the european states
2	0.770364	france, article, government, mafia, crime, politic, ask, tentacle, republic, shake	the last article of jan kuciak the tentacles of the mafia in
3	0.741201	kill, ndrangheta, protect, girlfriend, italians, election, leave, work, know, investigate	ndrangheta friends of the ndrangheta jan kuciak investigated the infiltration of ndrine in slovakia
4	0.810029	investigation, murder, reporter, businessman, bratislava, relative, kusnirova, protest, fiancee, shoot	we condemn the murder of investigative journalist jan kuciak and his fiancee martina kusnirova
5	0.709867	action, mafia, corruption, italian, investigate, death, assassinate, week, state, read	absolutely sick slovakia is truly going back to its post communists era rip to jan kuciak
6	0.739461	minister, fico, prime, resignation, europarl, murder, resign, call, deputy, police	slovak prime minister robert fico resigns after murder of laguardia journalist
7	0.791084	assassination, journalists, murder, premier, europe, release, press, freedom,	president tiber gaspar confirms the murder of jan kuciak might have been

		president, journalism	related to journalistic work the government offers mil euro as a reward for information about the murder radio slovakia
--	--	-----------------------	---

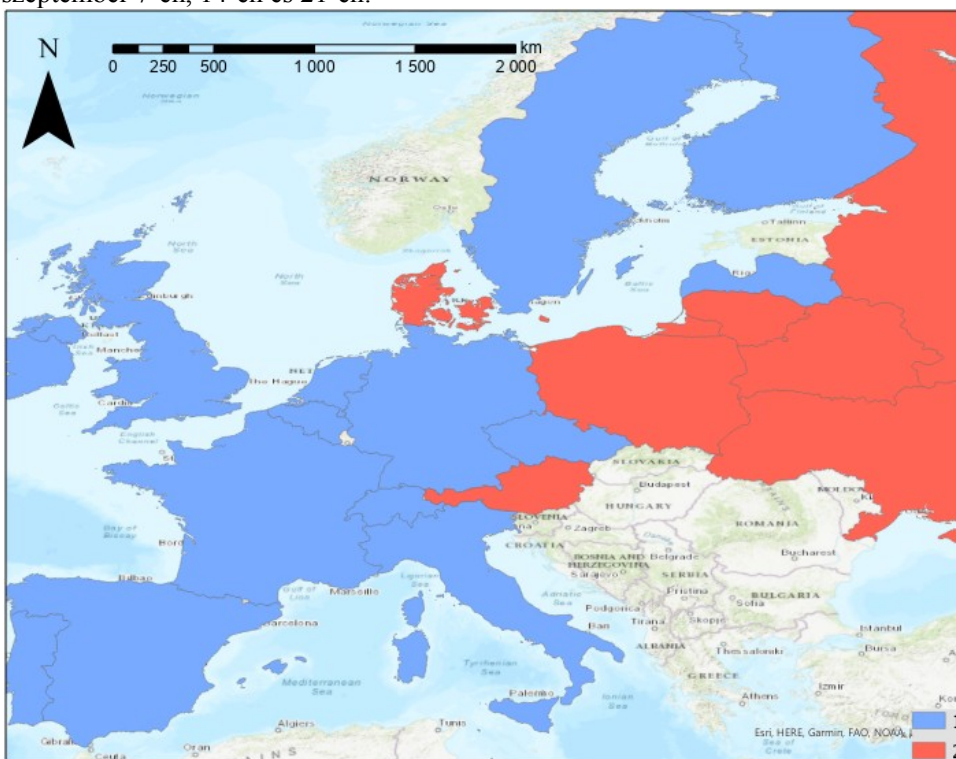
3. *Ábra*: Kuciak adatkészlet kulcsszavak, reprezentatív szöveg



4. *Ábra*: Kuciak 1. klaszter téma-alapú címkefelhője

4.5. A disszertációban használt témamodellező algoritmus képes volt azonosítani a kulcsfontosságú témákat mindkét adatkészlet esetében, mint például az olasz maffiával való kapcsolat, más újságírók meggyilkolása vagy a szlovákiai miniszterelnök lemondása vagy Oroszország és Fehéroroszország kapcsolata, illetve azokat a vitákat, amelyek blokkolták az Fehéroroszország ellenes EU szakciók sikerességét.

4.6. A fehéroroszországi választások utáni tiltakozások esetében a felhasználók aktivitási mintáit tekintve két fő országtípust különböztethetünk meg Európában. Az 1. és 2. klaszter közötti jelentős trendkülönbség elsősorban az első részben mutatható ki augusztus 24-ig. Az 1. klaszter a politikai döntések és események rendkívül ésszerű csoportját képviseli. Ezzel szemben a 2. klaszter országaiban a tweetelési aktivitás alacsony aktivitási szintet és csökkenő tendenciát mutat a teljes elemzési periódusban, három kisebb csúcsponttal augusztus 24-én, 31-én és szeptember 7-én, 14-én és 21-én.



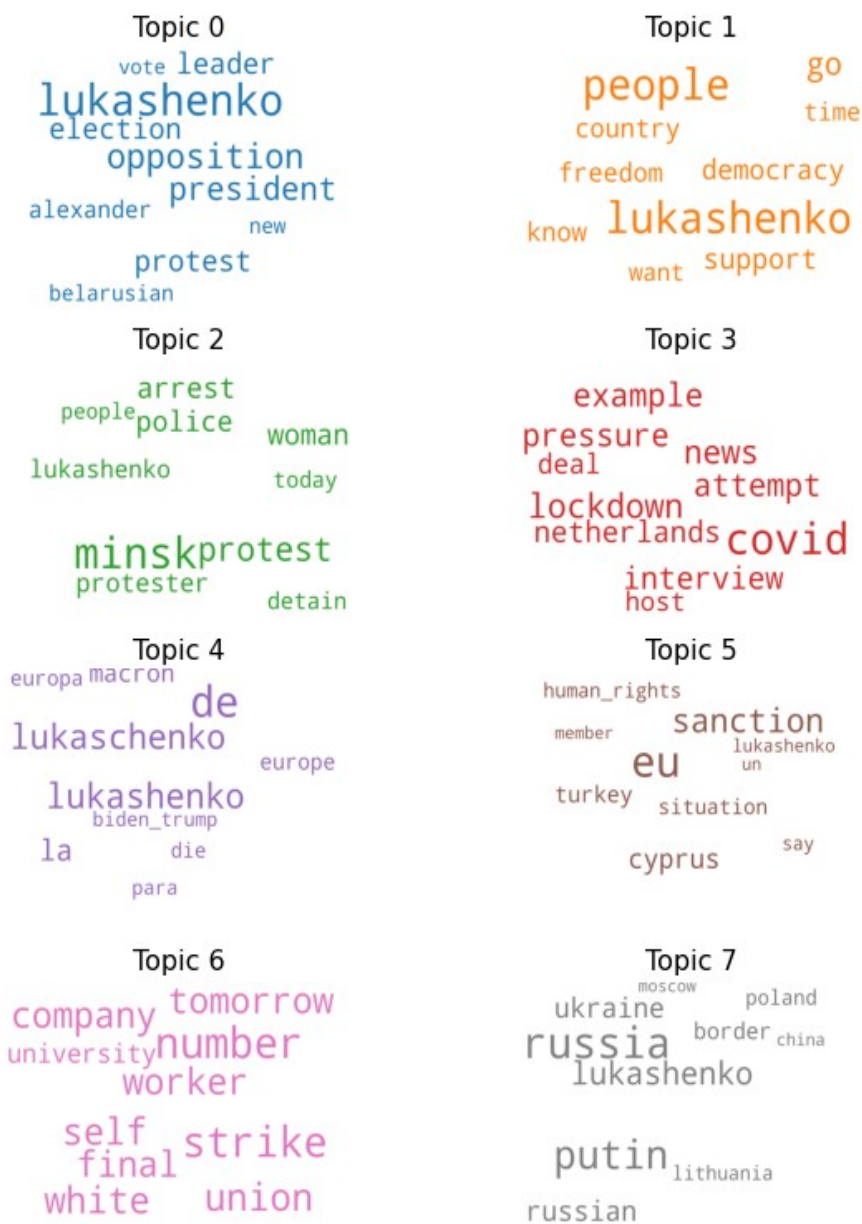
5. Ábra: Belarus adatkészlet klaszterezés

4.7. Az 1. klaszter felhasználói az időbeli átlagos hangulati értékeket figyelembe véve több olyan eseményt is pozitívabban értékelték szeptemberben, különösen

szeptember 11-én, 15-én, 19-én és 21-én, amelyek elsősorban a politikai helyzet fejlődési szakaszai voltak. Az Európai Unió például elítélte az erőszakot, és felszólította a fehérorosz hatóságokat, hogy engedjék szabadon az összes korábban fogva tartott tüntetőt.

A legnegatívabb értékeket a legnagyobb fehéroroszországi tiltakozás és azok utóhatásai jellemezték, mint például a szolidaritás menet megszervezése (augusztus 13.), valamint az, hogy három magas rangú ellenzéki szervező eltűnt egy nappal azután, hogy tízezer tüntető vonult át Minszken. Ezt alátámasztotta a tweetek napi és óránkénti megoszlásának vizualizálása is, amely erős koncentrációt mutat a vasárnapi és hétfői, vagyis a tüntetések idején. A szignifikáns különbség az 1. és 2. klaszter között az, hogy a hangulati értékek élesebben változtak. A 2. klaszter felhasználóinak érzelmeit erősen befolyásolták a tiltakozások, ugyanakkor pozitívan értékelték, amikor a fehérorosz gyári munkások általános sztrájkot kezdtek. Továbbá, míg az 1. klaszter felhasználói pozitív jelnek tekintették azt, hogy az Európai Parlament elismerte az ellenzék szervezetét (Koordinációs Tanács) a fehérorosz nép ideiglenes képviselőjének. A 2. klaszter felhasználói ezzel szemben a negatív fehéroroszországi eseményekre összpontosítottak, mint például arra az önfeláldozási kísérletre, amikor egy tüntető felgyújtotta magát vagy amikor 390 nőt vettek őrizetbe a fehérorosz erők szeptember 19-én. Összességében elmondható, hogy a 2. klaszter országainak lakói a mindennapi életükre és a fehéroroszországi belpolitikai eseményekre, míg az 1. klaszter országainak lakói a politikai helyzet folyamatos fejlődésére (pl. Fehéroroszország elleni szankciók; a Koordinációs Tanács nemzetközi elismerése) koncentráltak.





6. *Ábra*: Belarus 1. klaszter téma-alapú cimkefelhője

4.8. A Belarus adatkészlet esetében a témák érintik a klaszter csúcsaihoz kapcsolódó eseményeket, híreket. A legfontosabb témákat azok az üzenetek jelentették, amelyekben más országok vezetői nyíltan elítélték a fehéroroszországi eseményeket. Ilyen például Franciaország elnöke, Emanuel Macron, aki azt mondta, hogy Lukashenkónak „mennie kell”, vagy Biden amerikai elnök nyíltan elmarasztalta a korábbi elnököt, Donalt Trumpot, amiért nem volt hajlandó szót emelni a fehérorosz elnökkel szemben. Ezeket a témákat Lukashenkio orosz elnökhöz fűződő szívélyes kapcsolata, valamint a fehéroroszországi eseményekre és szomszédjaira gyakorolt hatásai követik. A harmadik legtöbbet vitatott téma a Fehéroroszország elleni uniós szankciók, valamint a helyi sztrájk gazdasági kapcsolatokra gyakorolt hatása volt, mivel az EU 2020-ban Fehéroroszország második vezető gazdasági partnerének számított. Ezt a témát követi egy másik, amely a demokrácia szükségességére fókuszál, főleg mivel Lukasenka Európa „utolsó diktátoraként” emlegetik. A klaszter az utolsó témái a tiltakozásokra koncentráltak, különösen a nők tiltakozó felvonulásaira.

### Topic 0

cyprus putin  
russian country  
turkey sanction eu  
lukashenko  
russia  
ukraine

### Topic 2

detain  
minsk police  
today  
woman protester protest  
belarus protest people  
lukashenko

### Topic 4

energy union green new deal coal  
energy economic  
european gas  
oil community start let  
pan\_european  
create great  
future of europe covid  
interdependence humanity first

### Topic 6

journalist  
woman  
school revolution  
use belarusian flower  
flowers come weapon  
airport  
remove

### Topic 1

belarusian  
president  
lithuania  
eu  
opposition member lukashenko  
election call  
human rights

### Topic 3

artist exhibition stop dictator  
prayer for belarus lukashenko  
art  
help belarus twiby  
lukashenko  
prayer for belarus stop dictator

### Topic 5

director child  
radio employee  
worker strike  
nexta company plant  
tractor

### Topic 7

support  
freedom  
world want help  
know live  
go people thank

7. *Ábra*: Belarus 2. klaszter téma-alapú címkefelhője

Az események összetett és dinamikus jellegének többspektrális értelmezése hatékony elemzési módot adott a tiltakozás dinamikájának alaposabb megértéséhez. Következésképpen munkánk két különböző módon is felülmúlja a korábbi teljesítményt:

1. A dolgozat megmutatta, hogy a georeferált közösségimédia-adatok felhasználhatók politikai események elemzésére, még kisebb földrajzi és társadalmi léptékben, az angoltól eltérő nyelveken is.

2. A jelen tanulmány egy újszerű algoritmikus megközelítést javasol, amely integrálja az idősoros klaszterezést a szemantikai témamodellezéssel és a georeferált közösségimédia-adatokon a érzelmek elemzésével a társadalmi jelenségek elemzésében.

Az eredeti perspektíva bemutatásával és a meglévő elemzések bemutatott korlátainak figyelembevételével ez a kutatás erőfeszítést tett az elmúlt évtized releváns kutatásainak mélyreható megértésére. A kutatás konkrét eredményeit a dolgozat Eredmények című fejezete ismertette ötven oldalon keresztül. Szeretnénk azonban ismét hangsúlyozni, hogy mind a szakirodalmi áttekintés, mind a módszertani összeállítás egyedülálló hozzájárulást jelent a kutatáshoz. Az alábbiakban röviden, tételesen összefoglalom a legfontosabb eredményeket:

#### • **Jobb geoparshing pontosság**

A Twitter Fejlesztői Platform szerint a tweetek hozzávetőleg 1-2%-a földrajzi címkével van ellátva, míg a tweetek ~30-40%-a tartalmaz bizonyos profilhely-információkat. Fehéroroszországi adatkészletünkben 655 423 sor volt a user\_location mezőben és 10 353 hely a user\_bio mezőben. A leírt geoparsher módszer 346 162 helyzint talált a user\_location mezőben és 10 353 helyet a user\_bio mezőben. Összességében a bemutatott vegyes modell geolokációs megközelítés 356 515 földrajzi hely azonosítására volt képes, ami a fehérorosz adatállomány 54,3%-a. Ugyanennek a módszernek az alkalmazása az #Allforjan adatkészleten 8069 tweet koordinátáit adta vissza, ami az eredeti adatkészlet 61,2%-a. Kapanova és Stoykova (2019) egy összehasonlítható vizsgálatot végeztek, és körülbelül napi 29 földrajzi hely szerinti tweetet tudtak elemezni, míg a módszertanunk átlagosan 100-at biztosított minden napra Szlovákia esetében. Ez átlagosan 15–20%-os javulást jelent ahhoz képest, amit a Twitter Developer Platformról tudunk, és többszörös növekedést jelent a hasonló tanulmányukhoz képest.

#### • **Az adatok jobb szemantikai megértése**

A közösségi médiával kapcsolatos korábbi szociológiai tanulmányok elsősorban a hashtag-tevékenységre összpontosítottak, figyelmen kívül

hagyva az írott nyelvet, amely az emberi kommunikáció döntő összetevője. Ez a disszertáció a szövegre összpontosít, beleértve a többnyelvű korpuszfordítást és érzéskivonást, gépi tanulási témamodellezési technikákat alkalmazva a kollektív cselekvés mögöttes érdekeinek és motivációinak feltárására. Következésképpen technikánk határozott előnnyel rendelkezik a korábbi tanulmányokhoz képest, amelyek gyakran figyelmen kívül hagyták ezeket a tényezőket.

• **A térbeli tényező jobb megértése a kortárs tiltakozó mozgalmak dinamikájában**

A csoportdinamika térbeli aspektusainak megértése érdekében a közösségi média szociológiai vizsgálata elsősorban a hashtag-aktivitáson múlott, ami korlátozott megértést biztosít. A közösségi média aktivitáson alapuló demonstrációk eszkalációjának megértéséhez a disszertáció módszertana tér-időbeli elemzést végzett geolokált tweetek segítségével. Az adatok nemzeti szinten összesítésre kerültek, mivel előfordulhat, hogy egy ország nyelvi és politikai tényezői hatással lehetnek az értékelt mozgalomra. Az ezt követő klaszterezési algoritmus összehasonlítható tweet-mintákkal rendelkező nemzeteket azonosít, amelyek arra is utalhatnak, hogy mikor történtek demonstrációk, vagy egyéb befolyásoló jellemzők létezését. A hasonló mintázatú országok kiválasztásával módszerünk rávilágított a hozzájáruló jellemzőkre és ezek időbeli alakulására. Sőt, egytől egyig tömörebb és szemléletesebb megjelenítést és magyarázatot adott az eredményről, mint az európai országokra vonatkozó statisztikai értékek.

• **Az általános térbeli jellemzők fejlett vizualizációval feltárták az #Allforjan mozgalom támogató országait**

Az értekezés többretegű térbeli elemzésének térbeli eredményei közvetlen összefüggést mutattak az európai országokkal. Bár az abszolút aktivitási statisztika azt mutatta, hogy Szlovákia, Németország, Olaszország és Franciaország a legaktívabb nemzetek, módszerünk kiküszöbölte a népességszám hatásából adódó torzítást. Az egyes országok népességének normalizálása és a jobb megjelenítés két kulcsfontosságú tényezőt tárt fel: az aktív országokat és a kiugró értékeket. Belgium, a Cseh Köztársaság, Németország, Franciaország, Magyarország, Lengyelország, Szlovákia és Svájc volt az a nyolc ország, ahol az elemzésünkben értékelt 18 nap során következetes tweet-aktivitás volt tapasztalható. Csehország, Magyarország és Lengyelország három szomszédos ország, közös múlttal. A vizualizáció három kiugró értéket mutatott: Málta, Olaszország és Szlovákia. Néhány hónappal korábban Kuciak meggyilkolásával egy máltai újságíró, Daphne

Caruana Galiziát is meggyilkolták. Ez felkeltette Málta kíváncsiságát. Az olasz 'Ndrangheta maffia és a szlovák hivatalnokok közötti kapcsolat miatt az olasz álláspontját széles körben vitatták az egész időszakban. Svájcban volt annak a folyóiratnak a kiadója, amely Kuciakot alkalmazta. Az módszertan megtalálta a három legjelentősebb témacsoportot, amelyek hozzájárultak a tiltakozó hálózat létrejöttéhez.

- **Az általános időbeli jellemzők elemzése feltárta az #Allforjan mozgalom döntő napjait**

Az egyik legfigyelemreméltóbb kiugrás február 28-án következett be, amikor kiderült Kuciak hiányos tanulmánya az olasz maffia és a szlovák politikusok kapcsolatairól, míg a tiltakozás fő napján (március 9.) és a szlovák belügyminiszter lemondásával újabb kisebb csúcsokkal járt, ahogy a további lemondások is (március 12. és 14.).

- **Az országok hasonlóságon alapuló klaszterezése feltárta az #Allforjan mozgalom fő jellemzőit és dinamikáját**

Az algoritmus öt csoportot állított fel: az 1. klaszter (kék) magas aktivitást mutatott (március 1.) az első napokban, és egy második, kisebb csúcsot, amikor PM Fico lemondott (március 14.). Olaszországban van a legtöbb tweet ebben a csoportban, főként a maffiának a szlovákiai gyilkosságokban és korrupcióban való részvétele miatt. A 2. klaszterben (piros) volt a legkevesebb Kuciakkal kapcsolatos tweet. A 3. klaszter (zöld) szerény aktivitási szinttel rendelkezik, mint a 2. klaszter, két kisebb csúccsal február 28-án és március 14-én. A 4. klaszter csak három országot foglal magában, és nincs csúcs, amikor a gyilkosság kiderült. Március 12-e és 15-e a miniszterelnök lemondása és a gyilkosság vagy az újságírói munka egyéb közvetett hatásai miatt van csúcspont ezekben az országokban. A különböző országok eredményei azt mutatják, hogy a gyilkosságról és az azt követő eseményekről szóló tweeteknek politikai narratívája lehet. A tendencia nyilvánvaló, de a tweetek abszolút mennyisége alacsony (hasonlóan a 3. klaszterhez), ami arra utal, hogy ezekben az országokban nem a Twitter a legnépszerűbb közösségi média eszköz. Azok, akik ezt használják, kevésbé reprezentatívak a szélesebb lakosság körében, mint más nemzetekben.

- **Az országokonkénti hangulatelemzés és annak eloszlása feltárta az #Allforjan mozgalomra érkezett reakciókat, és betekintést nyújtott a közös identitás formálódásába**

Figyelembe véve a kormány maffiához fűződő kapcsolatait, a legtöbben elégedettek voltak a politikai következményekkel, például a miniszterelnök lemondásával. A kezdeti (a hangulatosztályok

alkalmazása előtt) kapott érzelmi összetett pontszámot felhasználva azt tapasztaltuk, hogy Szlovákia és Németország azon nemzetek közé tartozik, amelyekben az elemzési időszakban legalább egy tweet szerepel naponta, ami statisztikailag emelkedő tendencia. (Németország 95%-os, Szlovákia 99%-os volt)

- **A legaktívabb klaszterek (1. és 5.) témamodelljezése feltárta az #Allforjan mozgalom fő mozgatórugóit és más politikai tényezők súlyát.**

Az 1. klaszter magában foglalja a miniszterelnök távozását, az olasz maffiát és az EU szerepét. A legtöbb tweet elítéli a gyilkosságot, ezt követi a sajtószabadság és -biztonság miatti aggodalmak. Kuciak cikke, amely halála után, február 28-án jelent meg angolul és szlovákul, később más nyelveken is megjelent (pl. franciául). A cikk összefüggéseket mutatott be a szlovák kormány és a szervezett bűnözés között. Az 5. klaszter témamodelljezési eredményei 7 témát mutatnak be. A 2. klaszter témái nem csak a szlovák miniszterelnökről, hanem a kormány szerepéről, a tiltakozásokról, a maffiáról, Kuciak menyasszonyáról és Orbán Viktorról is szóba kerültek, bár Magyarország nem volt sem ebben a klaszterben, sem a legaktívabb országok között. Több tweet egyértelmű összefüggést állított Kuciak és Daphne Caruana Galizia, egy korábban meggyilkolt máltai oknyomozó újságíró halála között. A túlerepresentáltság azt mutatja, hogy Caruana Galizia esete megerősítette a Kuciak mozgalmat az interneten. Azok a tweetek, amelyek Galizia halálát Kuciak menyasszonyának ügyével hozták összefüggésbe, a nők elleni erőszak egyértelmű elítélésének tekinthetők, amely szintén segíthette a mozgalmat. Martina Kusnirovát név szerint és menyasszonyként, lányként, barátnőként is emlegetik, ami arra utal, hogy a klaszter felhasználói különbséget tettek a két halál között, az ártatlan halála és a munkával összefüggő haláleset között. A témamodelljezés során a magyar miniszterelnökről és a magyar származású amerikai milliárdosról, Soros Györgyről szóló tweeteket is azonosítottak. Ennek két oka lehet. Először is, hat hónappal Kuciak meggyilkolása előtt a magyar kormány országos óriásplakát-kampányt indított Sorostól, ami bűnbakként kezelte, másrészt Sorost Fico szlovák miniszterelnök is nyilatkozatban említtette hasonló képet festve róla, mint ahogyan az Magyarországon történt. A klaszterezés egy csoportba helyezte Magyarországot, Fehéroroszországot és Törökországot, ami az országokban zajló hasonló tendenciákra utalhat.

- **Az általános térbeli jellemzők fejlett vizualizációval feltárták a #Belarusprotest mozgalom támogató országait**

A kutatás többretegű térbeli elemzése összefüggést mutatott ki a tiltakozás és más európai országok között. A populáció normalizálása és a vizualizáció javítása aktív országokat és kiugró értékeket tárt fel. 44 napos elemzésünk során 33 ország folyamatosan tweetel. Ezek közül a nyugati szomszédok Lettország, Litvánia és Észtország voltak a legfontosabbak. Ezek a nemzetek először Fehéroroszországot szankcionálták. A litván parlament augusztus 18-án gazdasági szankciókat rendelt el Alekszandr Lukasenkóval és 30 fehérorosz tisztviselővel szemben. Ezek az országok beépítették az Európai Unió ítéleteit is. Az elemzések egy kiugró jelenséget tártak fel, nevezetesen Ausztriát, ahol a narratíva egy esetleges energetikai válsághoz kötődött.

- **Az általános időbeli jellemzők feltárták a # Belarusproteszt meghatározó napjait**

Az egyik legjelentősebb kiugrás akkor következett be, amikor Konstantin Shishmakov (29), a Vaukavysk Bagration Hadtörténeti Múzeum igazgatója 2020. augusztus 15-én és 16-án eltűnt. A választási bizottság helyetteseként nem volt hajlandó átvenni és aláírni a hamisított dokumentumokat. 17 órakor felhívta a feleségét, és azt mondta, hogy hazajön, de nem érkezett meg. Augusztus 18-án egy grodnói folyóban találták meg Shishmakov holttestét. Augusztus 15-től kisebb csúcsok jelennek meg. Valószínűleg ezek a legnagyobb tiltakozó napok a "szolidaritási lánc" felvonulások miatt, amelyek a választásokat követő fehéroroszországi fellépés és az azt követő rendőri erőszak miatt tiltakoztak, ami több halálesethez és letartóztatáshoz vezetett. Josep Borrell, az EU főképviseleje augusztus 14-én Brüsszelben kijelentette, hogy az EU eljárást indít a választásokat meghamisító és erőszakot alkalmazó fehérorosz tisztviselők ellen. Augusztus 16-án és 23-án, Fehéroroszország legnagyobb tiltakozó napjain, Minszkben „Az egység menetét” tartották 200 000 résztvevővel és minden nagyobb regionális központtal. Homelban 6000-en, Hrodnában 4000-en, Bresztben, Vitebszkben és Mogilevben pedig 3000-en tiltakoztak, ami megmagyarázhatja, miért említették több ízben az eseményt Európa-szerte.

- **Az országok hasonlóságon alapuló klaszterezése feltárta a #Belarusprotest mozgalom jellemzőit és dinamikáját**

A klaszterezés legfontosabb eredménye a klaszterek száma, amely kettő, ami jelentős térbeli megosztottságot, korlátozottságot és elszigeteltséget mutat. Az első klaszterbe jobbra az európai nemzetek kerültek, magas aktivitással, amely a elsősorban az európai politikai válaszreakciókhoz kötődött. Az első csúcs (augusztus 18.) után csökkenő mintát mutat,



kisebb csúcsokkal augusztus 20-án, 24-én, 28-án és 31-én, valamint szeptember 7-én, 14-én és 21-én. Szeptember 7. és 9. között ráadásul fokozódik a feszültség. Hétfőn egyes csúcsok vannak, ami vasárnapi tiltakozásra utal. Az 1. klaszter felhasználói a tiltakozó vezetőkre összpontosítottak, akiket elraboltak és az ukrán határhoz hurcoltak. Az 1. és 2. klaszter trendjei augusztus 24-ig jelentősen eltérnek egymástól: Az Európai klaszter a politikai gazdasági hatásokra összpontosított, míg a Belarus klaszter a tüntetésekre és azok eredményére.

- **Az országokonkénti hangulateloszlás feltárta a #Belarusprotest mozgalom visszajelzéseit, felfedve a nyugati országok részéről az eseményektől való bizonyos távolságtartást**

Az 1. klaszter a pozitív és negatív értékek ingadozása. A klaszter felhasználói pozitívan értékelték a különböző szeptemberi eseményeket, különösen 11-én, 15-én, 19-én és 21-én. Ezek többnyire politikai fejlemények voltak. Az EU elítélte az erőszakot, és követelte Fehéroroszországot, hogy bocsássa szabadon a bebörtönzött tüntetőket. A legkedvezőtlenebb tulajdonságok a legnagyobb fehéroroszországi tiltakozások és azok következményei voltak, mint például a hosszú szolidaritási sorok (augusztus 13.), valamint az, hogy három magas rangú ellenzéki szervező eltűnt egy nappal azután, hogy 10 000 fehérorosz Minszken átvonult. A tweetek figyelemre méltó hangsúlyt fektetnek a tiltakozó napokra, vasárnapokra és hétfőkre. A tiltakozó napok uralják a tweeteket. A hétfői pozitív számok többsége éjfél után következik be, ami tiltakozó felvonulásokat jelez. A felhasználók általában negatívak a közepén, különösen egy nappal a tiltakozások előtt, ami megmutatja, hogyan vélekednek a kormányról. Összességében az európai felhasználók elégedettek voltak az olyan politikai eredményekkel, mint a Koordinációs Tanács és az uniós szankciók, de semmi több. A 2. (keleti) klaszternek két negatív szélsősége van augusztus 13-án és szeptember 7-én. Az 1. klaszter hangulata drámaiabban változott, mint a 2. klaszteré. A 2. klaszter felhasználóit nagymértékben érintették a tiltakozások, míg a fehérorosz gyári munkások országos sztrájkját pozitívan értékelték. A 2. klaszter felhasználói a fehéroroszországi negatív incidensekre koncentráltak, például egy szmaljavicsai rendőrőrs közelében elkövetett öngyilkossági kísérletre, vagy amikor 390 nőt tartottak fogva fehérorosz katonák szeptember 19-én.

- **A legaktívabb klaszterek témamodelljeze feltárta a #Belarusprotest mozgalom fő mozgatórugóit, valamint politikai és korlátozó tényezőket**

Az 1. klaszter (Nyugat) fő témái a minszki tiltakozások, Alekszandr Lukasenko kapcsolata Vlagyimir Putyinnal, valamint a Fehéroroszország elleni uniós szankciók voltak. Összességében a vezetők a fehéroroszországi eseményeket elítélő tweetjei voltak a legfontosabbak. Emanuel Macron vagy Joe Biden megnyilvánulásai, utóbbi például megfeddte Trumpot, amiért nem ítélte el a despotát. Az 1. klaszter az európai politikára és Fehéroroszországra összpontosított. A 2. klaszter (keleti) felhasználóinak százalékos hozzájárulása magasabb, mint a nyugatiaké. Ezek a témák a tüntetések, az emberi jogok, a sztrájkok és az energiaunió. A Topic 4 reprezentatív tweetje az energia jelentőségére összpontosított Fehéroroszország politikai zűrzavarában, vagyis Oroszország azon törekvésére, hogy megkerülje a hagyományos tranzitországokat. A Jamal–Európa gázvezeték a Jamal-félszigetet és Nyugat-Szibériát köti össze Lengyelországgal és Németországgal Fehéroroszországon keresztül. Ausztria is beletartozik, mivel a gázt Fehéroroszországból és a Jamal–Európa gázvezetékéből szerzi. A 2. klaszter legnagyobb problémája a fehérorosz földgáztól való függés volt. Egy másik fontos téma a nők szerepe a tiltakozásban. A tweetek egy minszki demonstrációt tárgyaltak. A fehérorosz nők megállították a rendőri bántalmazást azzal, hogy virágokkal álltak a főútvonalak mentén. Összességében nincs interakció a két klaszter között, ami ugyancsak elszigetelégre utal.

**• A szlovákiai és a fehéroroszországi események összehasonlítása a konnektív és kollektív logika tükrében feltárta azokat a tényezőket, amelyek csak Szlovákiában vezettek sikerhez.**

Ugyan mindkét mozgalom használta a social médiát egy decentralizált hálózat létrehozására, ezek építőteglái alapjaiban különböztek. Szlovákia nagy, rugalmas hálózattal rendelkezett, korlátlan hozzáféréssel. A gyors növekedés elősegítette a kapcsolatokat és a mélyebb interperszonális kötelékeket. Fehéroroszországban a kormány részben felügyelte a hálózati hozzáférést így a kiterjedt, hálózat nem integrálódhatott. Ugyanakkor Szlovákia társadalmi normái legitimálták a csatlakozást azáltal, hogy másokat bátorított, hogy járuljanak hozzá a változáshoz. Fehéroroszország a társadalmi normák korlátozóak lehetnek. A személyes narratívákat a kollektív cselekvési keretekhez kell kapcsolni, mert az egyéni cselekvések a folyamatos találkozások és együttműködések révén kollektív cselekvési keretké válhatnak. Fehéroroszországban a társadalmi normák korlátzóan hatnak, mert hiányoznak a nők egyenjogúságának intézményi és törvényi alapjai. közigazgatásban. Ezért az országban egy férfi – női különbségtételre

építő politika nem indul egyenlő esélyekkel.

## A DOKTORI TANULMÁNYOK ALATT KÉSZÜLT PUBLIKÁCIÓK LISTÁJA

Kovács, T., Kovács-Györi, A., & Resch, B. (2022). #Belarus: Spatial Proximity—Classifying Protesters Based on Spatiotemporal and Sentiment Analysis of Twitter Data. *Sensors* (in progress) | WOS, SCI/SSCI Q1(Impact 3.576)

Kovács, T., Kovács-Györi, A., & Resch, B. (2021). #Allforjan: How Twitter Users in Europe Reacted to the Murder of Ján Kuciak—Revealing Spatiotemporal Patterns through Sentiment Analysis and Topic Modeling. *ISPRS International Journal of Geo-Information*, 10(9), 585. | WOS, SCI/SSCI Q1(Impact 3.05)

Kovács, T. (2020). A vizuális tartalom társadalom befolyásoló hatása. *Geopolitikai szemle*, 2(1), 221–233.

Kovács, T. (2020). Lengyelország felzárkózása a rendszerváltást követően. In Somkúti B. (Ed.), *Közép-Kelet Európa központi makrorégiója és a kapcsolódó államok gazdasági felzárkózásának sikerei és kudarcai 1990 után* (pp. 1-30) (accepted for publication).

Kovács, T. (2020). Bulgária felzárkózása a rendszerváltást követően. In Somkúti B. (Ed.) *Közép-Kelet Európa központi makrorégiója és a kapcsolódó államok gazdasági felzárkózásának sikerei és kudarcai 1990 után* (pp. 31-60) (accepted for publication).

Kovács, T. (2019). Euromaidan: The Route to the Crimean Crisis that went through Online. *Vojenská Reflexie*, 15(Mimoriadne číslo 1), 47–58.

Kovács, T. (2018). Recent #Martyrs. In *National and International Security 2018: Proceedings of the International Conference on National and International Security* (pp. 199–205). Armed Forces Academy of General Milan Rastislav Štefánik.