

SZAKDOLGOZAT

Szabó Gergő

2025

Budapesti Corvinus Egyetem

Adatelemzés és Informatika Intézet

A koronavírus regionális terjedési hálózatának modellezése Diebold-Yilmaz- féle indexekkel Közép-Kelet-Európában

Szabó Gergő

Gazdaságinformatikus BSc

2025

Szakszeminárium vezető: Kovács László

Tartalomjegyzék

Tartalomjegyzék	1
Nyilatkozat a Generatív MI használatáról	3
1. Bevezetés	5
2. Szakirodalmi áttekintés	7
2.1. Hálózatelméleti modellek	7
2.2. Kutatási módszerek	7
2.3. A saját kutatásom relevanciája	12
3. Az adatbázis	14
3.1. Adatforrás	14
3.2. Adatelőkészítés	15
3.2.1. A kiválasztott országok	15
3.2.2. Hiányzó adatok pótlása és a LOESS regresszió	15
3.2.3. A vizsgált időszak megválasztása	17
3.2.4. Kiugró és irreális adatok kezelése	18
3.3. A végső adatbázis	19
4. Kutatási módszerek	21
4.1. VAR modell	21
4.2. LASSO VAR modell	22
4.3. IRF és FEVD	24
4.3.1. Impulzus-válaszfüggvény	24
4.3.2. Becslési variancia dekompozíció	26
4.4. A Connectedness megközelítés	29
4.4.1. Összesített (TCI) index	29
4.4.2. Irányított spillover index	30
4.4.3. Nettó spillover index	30
4.4.4. Páronkénti spillover index	31
4.4.5. Az indexek hálózati interpretációja	31
5. A kutatás eredményei	33
5.1. Window-size érzékenységvizsgálat	33
5.2. Átterjedési (spillover) táblázat	35
5.3. TCI index alapú összekapcsoltsági elemzés	36
5.4. Nettó spillover indexek elemzése	39

5.5. Hálózati diagramok.....	40
5.5.1. Magyarország hálózati kapcsolatai.....	41
5.5.2. Magyarország és szomszédos régióinak hálózati kapcsolatai	44
5.5.3. Lengyelország hálózati kapcsolatai	48
5.5.4. A teljes modell hálózati kapcsolatai	51
6. Összefoglalás	53
7. Hivatkozásjegyzék.....	55
8. Mellékletek.....	57

Nyilatkozat a Generatív MI használatáról

Alulírott, Szabó Gergő hallgató (Neptun-kód: E46MDD) a Budapesti Corvinus Egyetem Etikai Kódexén alapuló és eszerint érvényesülő etikai felelősségem tudatában kijelentem, hogy a gazdaságinformatikus BSc képzési szakhoz kapcsolódó szakdolgozat elkészítéséhez generatív mesterségesintelligencia-rendszert vagy -szolgáltatást (a továbbiakban: GMI) az alábbiak szerint vettem igénybe.

1) Nem használtam fel GMI rendszert vagy szolgáltatást

2) GMI alapú rendszert használtam fel

a) Szakirodalom kutatásához, válogatásához

a rendszer neve, verziója: ChatGPT, GPT-5.1

b) Adatok előzetes elemzéséhez, értékeléséhez

a rendszer neve, verziója: ChatGPT, GPT-5.1

c) Számítógépes programkód írásához, fejlesztéséhez, ellenőrzéséhez

a rendszer neve, verziója: ChatGPT, GPT-5.1 és Gemini, Flash 2.5

d) Szöveges tartalom stilisztikai és nyelvtani javításához, parafrázáláshoz

a rendszer neve, verziója: ChatGPT, GPT-5.1

e) Képi vagy audiovizuális tartalom fejlesztéséhez, létrehozásához

a rendszer neve, verziója: nem használtam

f) Egyes szövegrészek fordításához (pl. magyar és idegen nyelv között)

a rendszer neve, verziója: ChatGPT, GPT-5.1

g) Egyéb módon, és pedig: alapvető ötletek generálására

a rendszer neve, verziója: ChatGPT, GPT-5.1

Kijelentem, hogy GMI-rendszer által előállított új tartalomelemeket az egyetemi GMI-szabályozás által megengedett módon használtam fel és a leadott munkában a szakos

szakdolgozati útmutatóban előírt módon (pl. lábjegyzetben „GMI által generált tartalom” megjegyzéssel és a terjedelem azonosításával) mindenhol megjelöltem.

Tudomással bírok arról, hogy az Egyetem az általam benyújtott munkákat mesterségesintelligencia-detektor alkalmazással átvizsgálhatja, s ennek eredményétől függően akár etikai eljárást is kezdeményezhet ellenem.

1. Bevezetés

A COVID-19 járvány 2019-es kirobbanása hatalmas hatást gyakorolt a világra, legyen szó gazdasági, szociális, egészségügyi, kulturális, technológiai vagy társadalmi szempontokról. A vírus az évek alatt több hullámban érte el Közép-Kelet-Európát, voltak hónapok, mikor az esetszámok nagyon megemelkedtek, máskor a vírus pangása okán a társadalom nagyobb szabadságban működhetett. Számos kutatás próbálta ezeket a hullámokat és megfertőződések modellezni, és feltárni a mögöttes okokat. Dolgozatomban azt vizsgálom, hogy Közép-Kelet-Európában kiemelten Magyarországon hogyan alakult a vírus terjedésének dinamikája térben és időben. A téma komplexitása abban rejlik, hogy az emberek mobilitása és a vírus terjedése is nehezen becsülhető, így már egy adott országon belül is nehéz ezeket a dinamikákat feltérképezni, nemhogy országokon átívelően. Ezért elengedhetetlen, hogy egy olyan megközelítést alkalmazzak, mely képes ezt a hálózatot feltárni. A szakirodalmi áttekintés során megvizsgálom a témában elkészített különböző publikációk erősségeit, de azok hiányosságait is, melyekre én a jelenlegi elemzésemben szeretnék megoldást találni.

Kutatásom során szeretnék rámutatni a következőkre.

1. Melyik ország mennyire alkot nyílt vagy zárt rendszert a vírus idején az esetszámokat tekintve.
2. Országon belül milyen irányúak a terjedések.
3. Országok között milyen kapcsolatok állnak fenn és hogy ezeknek mi lehet a mögöttes társadalmi vagy gazdasági oka.

A vizsgálat során egy hálózatalméleti modellt építek, melyekhez egy országokat régiókra bontó területi egységrendszert fogok használni, amely egész Európára egységesen adott.

A dolgozat további felépítése a következő. Először bemutatom az elemzéshez használt adatbázist és azt, hogy ezen milyen adatelőkészítési eljárásokat alkalmaztam, hogy egy konzisztens paneladattal tudjak dolgozni. Továbbá bemutatom az alkalmazott statisztikai és ökonometriai módszerek matematikai hátterét is, mint például a VAR modellt, az impulzus-válaszfüggvények és az előrejelzési variancia dekompozíció. Ezen túl ismertetem az elemzésem alapjául szolgáló Diebold-Yilmaz-féle indexeket és ezek alkalmazhatóságát. Végül az általam használt adatbázisra megbecsülöm ezt a modellt és különböző vizualizációk segítségével feltárom a közép-kelet-európai régiók által alkotott hálózati rendszer összefüggéseit.

Különböző alhálózatokat is vizsgálok, melyek az összehasonlíthatóságot is lehetővé teszik térben és időben egyaránt. A kutatás során szerzett tapasztalatok és megállapítások alapot adnak a jövőbeli elemzésekhez és további fejlesztésekhez.

2. Szakirodalmi áttekintés

2.1. Hálózatelméleti modellek

A COVID-19 világjárvány 2019 végén jelent meg, és annak gyors és globális terjedése különböző innovatív modellezési megoldások alkalmazását kívánta meg. A járvány dinamikájának megértéséhez már nem feltétlenül voltak elegendők a hagyományos epidemiológiai modellek, mint például a SIR - Susceptible-Infected-Recovered (Sun & Sathasivam, 2023) vagy a SEIR - Susceptible-Exposed-Infected-Recovered (Hu et al., 2021) megközelítések. Ahhoz, hogy ténylegesen átlássuk, hogy milyen tényezők befolyásolták a vírus terjedését, különböző gazdasági, társadalmi, mobilitási és kapcsolati struktúrák elemzésére is szükségünk lehet. A dolgozatomban egy hálózatelméleti modellt fogok alkalmazni. A hálózatok mindig úgy épülnek fel, hogy a csomópontok az egyének (ezek lehetnek ténylegesen személyek, de akár populációs vagy területi egységek is), míg az élek az ezen egyének közötti terjedési útvonalakat reprezentálják (Corcoran & Clark, 2022). Ezen modellek abban emelkednek ki, hogy velük különböző megelőzési vagy beavatkozási stratégiai döntéseket is előkészíthetünk. Hiszen egy-egy ilyen hálózatban azonosíthatunk központi csomópontokat (hubokat), melyek akár a kapcsolatok mennyisége vagy azok intenzitása miatt lehetnek kulcsszereplők.

2.2. Kutatási módszerek

A következőkben hálózati modellezést tartalmazó szakirodalmakat fogok bemutatni. Ennek összefoglalását táblázatos formában készítem el, mely lehetővé teszi a módszertanok gyors összehasonlítását. A táblázatban négy fő szempontra térek ki:

- a tanulmány címére,
- a vizsgálatához felhasznált adatforrásokra,
- az elemzési időszakra (amennyiben egyértelműen meg van határozva),
- valamint az alkalmazott módszertan rövid ismertetésére.

Célom, hogy megmutassam, mely hálózati alapú epidemiológiai megközelítések bizonyultak sikeresnek a covid terjedésének modellezése szempontjából.

Tanulmány	Adatforrás	Időszak	Módszertan
MAM: Flexible Monte-Carlo Agent based model for modelling COVID-19 spread	Hivatalos napi esetszámok és napi oltási arány Izraelben.	2020. december és 2022. február között	Egy dinamikus Monte-Carlo ügynökalapú modell, mely szimulált ügynökök által előidézett kimeneteket vizsgál sok ismétlésen keresztül. A különböző forgatókönyvekből származó eredményeket aztán összehasonlítja a valós esetszámokkal a becslés pontosságának értékeléséhez.
Structural modeling of COVID-19 spread in relation to human mobility	Napi fertőzöttségi és mobilitási adatok az USA különböző államaiban.	2020. május	Egy kétirányú strukturális regressziós modell, mely feltárja a kapcsolatot a járványterjedés és a humán mobilitás között. Mivel kétirányú, így egyszerre tudja vizsgálni nemcsak azt, hogy a mobilitás hogyan hat az esetszámokra, hanem a fertőzések számát a mobilitásra is. Így akár olyan rejtett változókat is megfigyelhetünk, mint például a lokális beavatkozások hatásai.
Modeling the complete spatiotemporal spread of the COVID-19 epidemic in mainland China	Hivatalos napi fertőzöttségi adatok, valamint mobileszközök LBS kéréseinek adatai a migráció jelzésére.	2020. január 20. és 2020. február 29. között	Egy alapvetően SEIR modell került megvalósításra egy SSH megközelítéssel kiegészítve, ami azt jelenti, hogy térségenként külön modellezték a járvány terjedését. Így a modell időben és térben is tudja becsülni a reprodukciós számokat és a lappangási időket.

Tanulmány	Adatforrás	Időszak	Módszertan
Modeling the spatiotemporal transmission of COVID-19	Járványadatok, utazási jellemzők, POI pontok, úthálózatok és népességi adatok Hongkong közigazgatási határain belül.	2022. január 1. és 2022. április 30. között	Egy ügynökalapú modell egy úgynevezett SERHR keretrendszerrel kiegészítve. A mozaikszóban a betűk azt jelentik, hogy az egyedek a következő szerepeket vehetik fel: fogékony, lappangó, jelentett fertőzött, rejtett fertőzött vagy már nem fertőző (gyógyult/elhunyt). A módszertan figyelembe veszi az egyének mozgását és a köztük lévő kontaktokat, ezek alapján mintázatokat állapít meg.
A lattice gas model for infection spreading: Application to the COVID-19 pandemic in the Mexico City Metropolitan Area	Napi igazolt fertőzöttségi adatok Mexikóvárosból és annak agglomerációjából.	2020. március 9. és 2020. augusztus 31. között	A Mexikóvárosban gyűjtött adatokra egy kétdimenziós rácsgáz modell lett felállítva. A modellben a városban rácspontokat jelöltek ki, és a rácspontokon tartózkodó egyének véletlenszerűen mozognak a szomszédos cellák felé, ezzel potenciális fertőzési láncot kialakítva. Ha egy fertőzött egyén egy fertőzésre fogékony másik egyeddel kapcsolatba lép, akkor valamekkora eséllyel bekövetkezik a fertőzés átadása. Iteratív lépésekben a modell folyamatosan frissíti az aktuális rácállapotot, így ez időben és térben is vizsgálható.

Tanulmány	Adatforrás	Időszak	Módszertan
Quantifying the Regional Disproportionality of COVID-19 Incidence Rates	Az Egyesült Államokban a napi megerősített COVID-19 esetszámok.	2020. január 21. és 2023. január 25. között	A tanulmány a Theil-index segítségével szándékozik a járvány előfordulásának aránytalanságait számszerűsíteni a különböző régiókban. Ezzel a módszerrel kimutatható, mely térségekben koncentrált a pandémia, és lehetővé teszi a területek közti egyenlőtlenségek kvantitatív leírását. Így megállapíthatóak forrópontok, mely intézkedéseket segíthet alátámasztani.
Deep learning based hybrid prediction model for COVID-19 cross-country spread	WHO által közölt napi eset- és halálozási számok a világ 10 legnépesebb országában.	2020. január 3. és 2022. augusztus 31. között	A kutatásban mélytanulási modelleket fejlesztettek ki, melyek előrejelzik a járvány terjedését és képesek megelőzési javaslatokat tenni. Több népszerű modellt is teszteltek, mint az LR, RF, SVM, MLP, CNN, GRU és LSTM megközelítések.
Methodology for modelling the new COVID-19 pandemic spread and implementation to European countries	Európai Unió országainak, továbbá Svájc és Egyesült Királyság fertőzöttségi arányai, és az egy főre jutó bruttó hazai GDP.	Nincsen konkrét időszak feltüntetve, 2021-ben készült a tanulmány.	A matematikai elemzéshez egy LPE-SG modellt alkalmaz a tanulmány, mely a paraméterek analitikus meghatározását teszi lehetővé. A modell outputja alapvető (R_0) és effektív (R_e) reprodukciós számok, továbbá a SIR-alapú megközelítéssel integrálva populációs küszöbértékek is becsülhetők.

Tanulmány	Adatforrás	Időszak	Módszertan
Impact of immobility and mobility activities on the spread of the COVID-19 pandemic in 29 European countries	29 európai ország hivatalos esetszámai, illetve a Google mobilitás napi idősoros mutatói (6 különböző mutató, pl.: parklátogatók, tömegközlekedési látogatók stb.).	2020. március 12. és 2021. augusztus 31. között	A tanulmányban a járvány terjedésének vizsgálatára egy PMG-ARDL nevű panelmodell került felépítésre. A napi esetszám volt az eredményváltozó, míg a többi mobilitási adat a magyarázóváltozók. A Pooled Mean Group becslő és a Granger-okságok segítségével sikerült rövid- és hosszútávú összefüggéseket meghatározni. Eredményként sikerült a mobilitási és egyéb szigorítási intézkedések hatását megvizsgálni a COVID-19 terjedésére vonatkozóan az európai országokban.
Nonlinear dynamic model for COVID-19 epidemics using small-world network and heterogeneous network structure	A Malajziai Egészségügyi Minisztérium hivatalosan közzétett napi esetlistája, napi halálozási listája és Malajzia egyes államainak népességlistája.	2020 és 2022 között	Egy úgynevezett Gauss-féle elosztott kisvilág-hálózati modellt fejlesztettek ki a tanulmány szerzői. A módszertan a klasszikus SIR modellt bővíti ki azzal, hogy a modell alapjául szolgáló hálózaton a fogékony/fertőzött/gyógyult egyedeket reprezentáló csomópontok közötti kapcsolatok számát Gauss-eloszlás alapján határozza meg. Ezen megközelítés kielégíti az emberi mobilitás elméletében szereplő frekvencia-távolság szabályt.

Ahogy a fenti kutatások is bizonyítják, rengeteg különböző megközelítés született a járvány struktúrájának feltérképezésére az elmúlt években. A leggyakorlatorientáltabb eredményeket az ügynök-alapú Monte-Carlo modellek (De-Leon & Aran, 2023) és a Hongkongban alkalmazott SERHR keretrendszerű modell (He et al., 2025) hozták, mivel lehetővé tették a vakcinációs stratégiák és a mobilitás-alapú beavatkozások hatásainak vizsgálatát, így közvetlenül képesek támogatni a döntéshozatalt. A térbeli különbségeket figyelembe vevő SEIR modellek, például Hu és munkatársai Kínára vonatkozó elemzése, valamint a rács-gáz modell Mexikóvárosban, jól szemléltetik, hogyan befolyásolja a térszerkezet és a lokális interakciók a járvány dinamikáját, ami hasznos lehet helyi intézkedések tervezésében. A Theil-index alapú elemzés az Egyesült Államok régióinak egyenlőtlenségeit mutatja ki, segítve a rendelkezésre álló erőforrások célzott elosztását. A mélytanulási hibrid modellek (Utku, 2023) előrejelzési kapacitása különösen a nemzetközi terjedés monitorozásában bizonyult hasznosnak, míg az európai országokra alkalmazott LPE-SG modell a reprodukciós számok és populációs küszöbértékek becslésével képes támogatni a stratégiai döntéshozatalt. Összességében a kutatások gyakorlati értéke abban rejlik, hogy lehetővé teszik a járvány térbeli és időbeli mintázatainak megértését, a beavatkozások tervezését, és a különböző intézkedések hatásainak előrejelzését.

2.3. A saját kutatásom relevanciája

Vannak kutatások, melyek inkább mikro szintű adatokkal (egyéni vagy lokális adatok), míg mások makroszinten aggregált adatokkal (országos vagy regionális számok) tesznek kísérletet a feltérképezésre. Több tanulmány próbál valamilyen gazdasági, társadalmi, mobilitási vagy egyéb adatok bevonásával kapcsolatokat feltárni és magyarázóerőt adni a modelljének.

A regionális különbségekre és az országok közötti kapcsolatokra explicit módon azonban kevés kutatás tud érdemi választ adni (Bouzouina et al., 2022). Jelen kutatásom főként ezt a hiányosságot ragadja meg. A Diebold-Yilmaz-féle hálózati összekapcsoltsági indexek alapvetően pénzügyi hálózatok elemzésében bevált módszertanok, viszont képesek a járvány modellezésére is. A pénzügyi területeken gyakran használják például volatilitás-terjedésének becslésére és a köztük lévő kapcsolatok feltérképezésére. A modell ezen előnyét használom ki, hogy képes különböző idősorok közötti késleltetett kereszt-korrelációk figyelembe vételével magyarázóerőt mutatni kvantitatív és vizuális módszerekkel egyaránt. A kutatás tehát kiegészíti

a korábbi hálózat alapú megközelítéseket azzal, hogy a fertőzés gyorsaságának és irányának elemzését is lehetővé teszi, mindezt az idő, mint fontos dimenzió bevonásával. Mindez gazdasági kontextussal és magyarázatokkal kiegészítve különböző kérdésekre is választ adhat: Mely országok regionális pozíciója befolyásolja leginkább Közép-Kelet-Európában a járvány terjedését? Hogyan alakul az országok közötti kapcsolat erőssége és iránya az idő előrehaladtával? A belföldi terjedés szempontjából melyik régió volt inkább terjesztő és melyik inkább befogadó? Milyen szerepet játszanak az egyes fővárosok, a határmenti régiók, vagy a nagyobb népességszámú gazdaságilag fontos centrumok? Az egyes hullámokban a kapcsolatok milyen hasonlóságokat vagy eltéréseket mutatnak? Mennyire összekapcsolt a teljes rendszer? További előnye a modellnek, hogy képesek lehetünk becslést adni kizárólag a régiók esetszámaiból kiindulva további változók bevonása nélkül.

3. Az adatbázis

3.1. Adatforrás

Ebben a fejezetben a panel adatbázis előkészítésének folyamatát mutatom be. Az adatbázis forrása, amivel dolgoztam, a COVID-19 European Regional Tracker, amely egy nyilvánosan elérhető GitHub repozitórium. A projekt tulajdonosa Asjad Naqvi (IIASA / WU Wien), aki különböző nemzeti forrásokból egy konzisztens és könnyen összehasonlítható panel adatbázist hozott létre. Az adatbázis 26 európai ország adatait tartalmazza, az adatgyűjtés 2020. január 15-én kezdődött, viszont ez országonként erősen eltér. A különböző kormányok más-más időpontban kezdték meg hivatalosan az esetszámok rögzítését, így vannak országok, ahonnan csak 2020. márciustól van elérhető adat. A következőkben ismertetem az eredeti adatbázis változóit:

- country: a megfigyeléshez tartozó ország neve
- nuts0_id: a megfigyeléshez tartozó ország kétbetűs rövidítése
- nuts2_id: a megfigyeléshez tartozó NUTS2 szintű régió kódja (ez a legtöbb megfigyelésnél üres, hiszen a legtöbb ország esetében NUTS3 szintű aggregálás történt, ez a mező kizárólag Lengyelország és Görögország esetében releváns)
- nuts3_id: a megfigyeléshez tartozó NUTS3 szintű régió kódja
- nuts_id: a megfigyeléshez tartozó NUTS2 vagy NUTS3 régió kódja (az előző kettő oszlop egyesítve)
- nuts_name: a megfigyeléshez tartozó nuts_id régió neve
- date: a megfertőződés hivatalos rögzítési dátuma
- population: a megfigyeléshez tartozó régió lakosságának száma az adott időpontban
- cases: a régióhoz tartozó kumulált esetszám az adatgyűjtés kezdetétől az adott dátumig
- cases_daily: a régióhoz tartozó napi esetszám
- cases_pop: cases/population hányados, vagyis a kumulált esetszám a populáció nagyságára vetítve
- cases_daily_pop: cases_daily/population hányados, vagyis a napi esetszám a populáció nagyságára vetítve

A GitHubon elérhető egy részletes leírás arról, hogyan történt a nyers adatok feldolgozása és homogenizálása. A régiók kialakítása NUTS alapú, ami lehetővé teszi a konzisztens összehasonlítást országokon átívelően, viszont ez az aggregálás nem volt egyértelmű. A NUTS rendszer 2013, 2016 és 2021 között több változáson is átesett. A szerző térinformatikai eszközökkel orvosolta a problémát. Több ország különböző közigazgatási egységeinek (pl. LAU-szintű adatok) határait átfedés-analízissel illesztette a NUTS 2016 régiókhoz. Így kis eltéréseket, például a határok eltolódását vagy a régiók átszervezését is kezelni tudta, és egy egységes, konzisztens térbeli keretrendszert hozott létre az EU-s COVID-adatokhoz.

3.2. Adatelőkészítés

3.2.1. A kiválasztott országok

Az elemzésem Közép-Kelet-Európára koncentrálok, így a fent bemutatott adatbázist első körben leszűrtem a releváns országokra, ezek a következők: Magyarország, Ausztria, Szlovákia, Horvátország, Szlovénia, Svájc, Olaszország, Németország és Lengyelország. A földrajzi elhelyezkedés és az országok közötti kapcsolat még indokolhatta volna Ukrajna, Szerbia és Románia elemzésbe történő bevitelét. Azonban az első kettő esetben nem volt elérhető adat, a panel adatbázis szerzője nem készítette el az adatok homogenizálását. Románia pedig a késői adatgyűjtés (2020. április 16-án volt az első hivatalos regisztrált esetszám az adatbázisban) miatt került ki a modellből.

3.2.2. Hiányzó adatok pótlása és a LOESS regresszió

A következő lépésben a hiányzó adatokat kezeltem. Az adatbázis készítője a konzisztencia megőrzése érdekében úgy járt el, hogy amelyik időszakban nem volt megbízható adatforrás, ott inkább üres értékeket vett fel. Így több olyan időpont is volt, ami vagy nem szerepelt az adatbázisban, vagy NaN érték állt a helyén. Ezt olyan módon kezeltem, hogy a hiányzó értékek utáni első elérhető napi esetszámot egyenletesen visszaosztottam a megelőző napokra. Tehát ha például az egyik régióban 2020. június 1. és 2020. június 5. között nincs adat a `cases_daily` oszlopban, de 2020. június 6-án már van, akkor ezt elosztom 6-tal, és egyenletesen a hiányzó napokhoz rendelem. Itt két fontos feltételezéssel élek. Egyrészt, hogy egyenletesen oszlanak

meg az esetszámok a hiányzó napokon, ami természetesen csak egy közelítés, de ezzel az adatpótlási módszerrel követem el várhatóan a legkisebb mértékű hibát. Másrészt azt is feltételezem, hogy a hiányzó napok esetszámainak összevonva leközik a következő nem üres napok számai. Ezt arra alapozom, hogy ezek az első nem üres napok esetszámai rendszeresen a környező időpontokban felfelé kilógó nagyságú értékek, mely jelezheti azt, hogy itt a hiányzó napok aggregálása történt meg.

Egy másik adatminőségi hiba is felmerült, ami a különböző országok eltérő adatgyűjtéséből következett. Ugyanis az általam az elemzésbe bevett országok más-más időpontban kezdték el regisztrálni az esetszámokat. Mivel a későbbiekben kereszt-korrelációk vizsgálatával szeretném az összefüggéseket feltárni, így a hiányos adatok torzított eredményekhez vezetnének. Ha a hálózati modellem egyes csúcspontjai hiányosak bizonyos időpontokban, ezzel fontos ok-okozati összefüggéseket veszíthetek el. Ebben az esetben egy úgynevezett backcasting eljárást használtam (Hamilton, 2020), amit a LOESS regressziós módszerrel kiegészítve becslést tudtam készíteni a múltban hiányzó adatokra.

A probléma abban rejlik, hogy az esetemben három ország (Magyarország, Horvátország, Lengyelország) első napi esetszáma az adatbázisban a hivatalos első észlelt esethez képest hetekkel későbbi. Ezért ez az első érték kilógóan nagy, hiszen az előző napok összes esetszámát abban az első napban tárolták el, ahonnan az adatbázis kezdődik. Ezt a helyzetet kezeltem a fent említett LOESS módszerrel (Cleveland et al., 1992). Ez annyiban tér el a klasszikus lineáris regressziótól, hogy minden egyes aktuális t értéknek, csak egy előre meghatározott s százaléknyi környezetét használjuk fel az OLS becsléshez. Ez azt jelenti, hogy minden egyes t értékhez más adathalmazt használunk fel a predikcióhoz, csak azt, ami attól az időponttól nincsen túlságosan messze. Erre azért van szükség az esetünkben, hiszen az adatrögzítés első pár értékére tett backcasting becslésemnél nem lenne jó, ha a hónapokkal későbbi trendek befolyásolnák azt. Viszont ha csak az adatbázis elejét veszem figyelembe, akkor láthatom, miként indult a terjedés, vagyis milyen módon van értelme az első rögzített esetszámot 'megosztani' a hiányzó dátumok között.

Egy egyszerű OLS trend illesztésnél a modell négyzetes eltérését, azaz az SSE-t szeretném minimalizálni:

$$\text{MIN: SSE} = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - (\beta_1 * t + \beta_0))^2$$

Ehhez képest a LOESS becslésnél minden t -edik időpontra meg kell ezt nézmem valamilyen s százalékban, így ez az összeg dinamikusan változik minden időpontban. Például a 10. megfigyelésre illesztett becslés a következőképpen nézne ki (ahol $0 < s < 1$):

$$\text{MIN: } \text{SSE}_{10} = \sum_{t=10-s*10}^{10+s*10} (\hat{Y}_t - Y_t)^2 = \sum_{t=10-s*10}^{10+s*10} (Y_t - (\beta_1 * t + \beta_0))^2$$

A backcasting rész azt jelenti, hogy időben nem előre, hanem visszafelé teszek becsléseket, tehát a múltban hiányzó adatokat próbálom közelíteni. A saját LOESS regressziómban 10%-os környezettel dolgoztam. Gyakorlatban tehát ez azt jelenti, hogy adott az első dátum, ahol tudjuk, hogy voltak esetszámok, csak az adatbázisban nem szerepel. Erre a dátumra az első 2-3 hónap alapján készítek egy becslést, majd megyek tovább a következő hiányzó dátumra, ahol a becslésre használt adathalmazt is 'görgetem' eggyel, és az új becslült értéket is felhasználom a predikcióhoz.

Miután megkaptam a becslült értékeket, egyetlen feladatom, hogy arányosítsam őket az első adatbázisban szereplő 'halmazott' esetszámhoz. Tehát ha például 30 napra visszamenőleg készítettem a LOESS becslést, és ennek a 30 napnak az összes esetszáma 100 a becslés alapján, de nekem az első értékem az adatbázisban csak 50, akkor minden napra a becslült értéket leosztom 100-zal és megszorozom 50-el. Ebbe az arányosításba az első 'halmazott' esetszámot is belevettem, természetesen nem szeretném, hogy az a torzított érték továbbra is ott legyen. Így egy konzisztens adatbázist kapok, hiszen a becsléseim összességében kiadják a kumulált esetszámot az első napig.

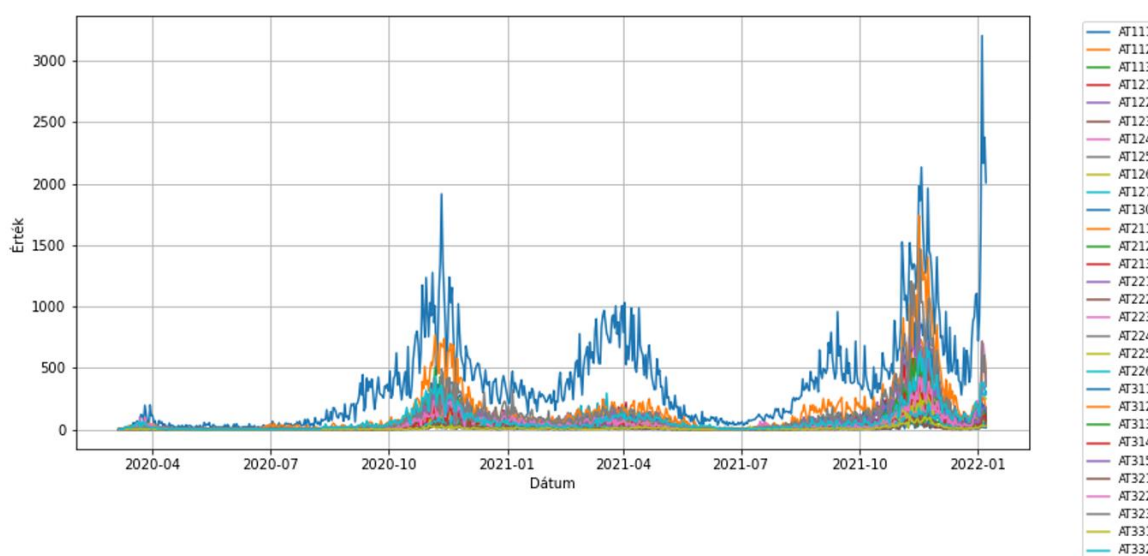
3.2.3. A vizsgált időszak megválasztása

Miután kezeltem a hiányzó adatokat, egy egységes időintervallumot határoztam meg az összes ország számára. Az elemzésembe bevett 9 közép-kelet-európai ország közül Szlovákiában rögzítették legkésőbb az első esetszámot, 2020. március 6-án. Ezt még nem találtam olyan késői kezdésnek, így ezt határoztam meg a kezdődátumnak. Hasonló megfontolások mentén megnéztem, melyik ország esetében ér véget legkorábban az adatok gyűjtése, ez Lengyelország

2022. január 7-én. Így tehát az egész adatbázist inntől 2020. 03. 06. – 2022. 01. 07. között kezeltem.

3.2.4. Kiugró és irreális adatok kezelése

Az így kapott adatbázis már közel áll ahhoz, hogy elvégezzem rajta a terjedési hálózat modellezését. Végeztem egy outlier vizsgálatot, hogy megnézzem, maradtak-e irreális adatok a táblázatban. Ehhez egyszerűen kirajzoltam az idősort:



1. ábra: Ausztria első 30 NUTS3 régiójának nyers esetszámai

Az 1. ábra például a betűrend szerinti első 30 osztrák NUTS3 régióiban a napi esetszámokat mutatja. Láthatjuk, hogy a késsel jelölt AT130 régió idősorának végén, 2022. januárjában van egy hatalmas kiugrás, ami nem tekinthető reálisnak. Illetve 2020. végén is látunk egy gyanúsan magas értéket. A többi diagramot is vizsgálva, egy-egy régió esetében láthatunk ilyeneket, sőt nagyon ritkán (az adatok 0,3%-a) negatív értékeket is (l. 1. melléklet). Ennek az az oka, hogy vannak dátumok, amikor valamiféle adatpótlás, vagy adatkorrigálás történt. Esetenként vannak egymás utáni időpontok, mikor egyszer egy nagyon magas érték látszik a napi esetszámban, majd másnap azt negatívban rögzítik, ez az inkonzisztens adatgyűjtés miatt lehet.

Én ezeket az értékeket kilógó értéként kezelem, és próbálom őket korrigálni. Egy kvantilisekre épülő kritikus felső értéket határozok meg, aminél ha magasabb az esetszám, azzal kezdenem

kell valamit. Veszem minden régiónak a felső és alsó kvartilisét, ezek különbsége az interkvartilis terjedelem. Majd ezt háromszor felmérem a felső kvartilisre, képezve ezzel egy felső kerítést, ami a kritikus értéként szolgál (Hunyadi & Vita, 2002). Ha ennél nagyobb egy napi esetszám, akkor helyettesítem a 3-3 szomszédjának átlagával, ezzel korrigálva az irreális kiugrást. Ugyanezt teszem, ha egy érték negatív. Tehát például ha találok egy outliert 2020. június 10-én, akkor átlagolom a 2020. június 7-9 és 2020. június 11-13 esetszámokat, így egy, az idősorba simuló reális értéket kapok.

Az összes 645 régióra elvégezve ezt a folyamatot, azt kaptam, hogy összesen 14298 ilyen kiugró érték van. Mivel 645 régióm van és 673 nap a megfigyelések száma, így az outlierek aránya csupán $14298/(645*673) = 3,29\%$. Ez nem vésszesen sok adat, elvégzem a fent említett logika mentén ezen értékek korrigálását.

Egyetlen régiót muszáj volt törölnöm az adatbázisból, ez a CH053 kódot viselő svájci Appenzell Ausserrhoden régió. Ennek az az oka, hogy valamiért mindössze 2021. 01. 22. dátummal bezárólag állnak rendelkezésre itt adatok, ez 1 év adatvesztés a többi régióhoz képest, és az ekkora távra adott becslés nem tekinthető megbízhatónak.

3.3. A végső adatbázis

Ezen a ponton tehát van egy adatbázisom, benne a korábban említett 9 ország összesen 645 régiójával és egységesen 2020. 03. 06. és 2022. 01. 07. között egyetlen hiányzó érték nélkül.

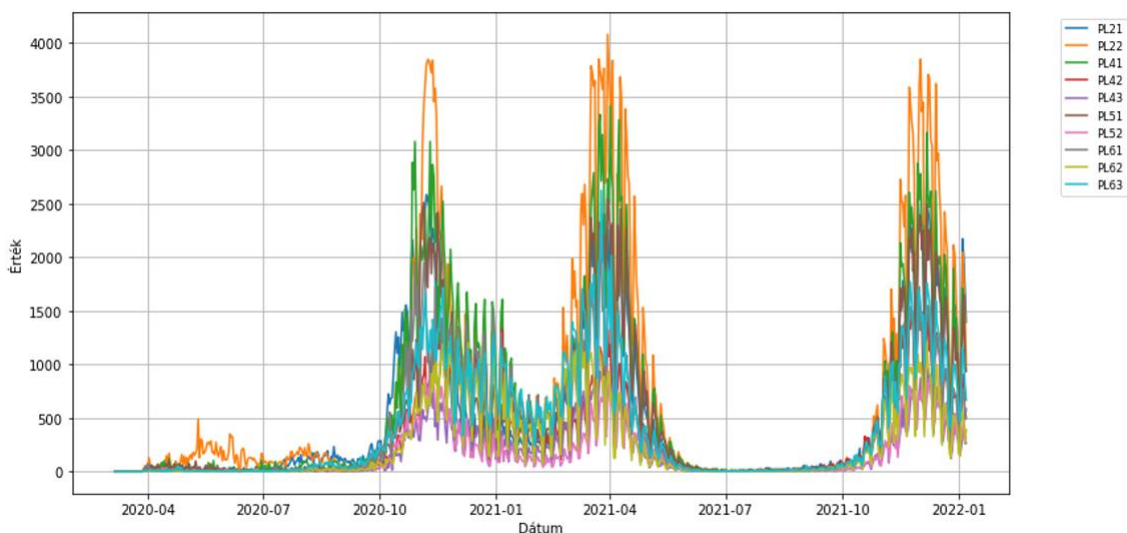
Ahogy a *3.1. fejezetben* említettem, voltak országok, ahol alapvetésként a NUTS3 régió szintjén történt az aggregálás, viszont pár országnál ez NUTS2. Az általam elemzett országok közül csak Lengyelország adatai NUTS2 szintűek. Ezt szeretném egységesíteni, így tehát azt kell eldöntenem, hogy NUTS2 vagy NUTS3 legyen az egységes szint. A döntésem alapja a megfigyelt egyedek számában van, anélkül, hogy Lengyelországot további régiókra bontanám, így is majdnem annyi változóm van, amennyi megfigyelésem.

Mivel a későbbiekben a modellem építésénél OLS regressziót szeretnék használni, nem szerencsés ez a nagyon sok változó. A túlparaméterezettség könnyen túlillesztéshez vezethet, mely azt jelenti, hogy a modell túlzottan a tanuló adathalmazra optimalizál, és nem mintázatokra tanul rá, hanem a konkrét értékekre. Ráadásul, ha több a dimenzió, mint a megfigyelések száma, matematikailag nincs megoldása a regressziónak.

Ezért úgy döntöttem, végül a NUTS2 szinttel megyek tovább, így Lengyelországon kívül az összes többi országot a régiók kódjai alapján aggregáltam. Végül így alakultak a NUTS2 régiók számai az egyes országokban:

- Magyarország: 8 db
- Ausztria: 9 db
- Szlovákia: 4 db
- Horvátország: 2db
- Szlovénia: 2 db
- Svájc: 7 db
- Olaszország: 21 db
- Németország: 38 db
- Lengyelország: 17 db

Ez összesen 108 db NUTS2 régió, ami már kezelhető mennyiség a 673 megfigyelés számához képest.



2. ábra: Lengyelország első 10 NUTS2 régiójának esetszámai az adattisztítás után

Az aggregálás után például az első 10 lengyel régió idősorai a 2. ábrán látható módon néznek ki. Ez egy kilógó értékektől, negatív esetszámoktól és üres adatoktól mentes adatbázis, amivel el tudom kezdeni az elemzést.

4. Kutatási módszerek

4.1. VAR modell

Az adatelőkészítés után tehát szeretném modellezni az előállt idősoraimat. Nagyon intuitív gondolat, hogy egy adott régióban egy adott napi covid esetszám nagyban függ az ugyanabban a régióban az azt megelőző időszakok esetszámától. Máshogyan fogalmazva tehát egyrészt szeretném felhasználni egy adott változó saját múltbeli értékeit a modellezéshez. Másfelől mivel az elemzésem eredményeiben főleg a Közép-Kelet-Európa régiói által alkotott rendszer összekapcsoltságáról szeretnék egy átfogó képet kapni, így megvizsgálom azt is, hogy egy régió esetszámait nem csupán a saját, hanem más régiók múltbeli esetszámai hogyan befolyásolják. Ezt a kettő megközelítést képest integrálni a VAR (Vector Autoregressive) modell (Koop et al., 1996). Az elnevezés onnan ered, hogy a modell az idősorokat vektor (innen a 'V rész') formában kezeli, vagyis egyszerre több és egymással összefüggő változót ír le egy közös rendszerben. Az 'AR rész' pedig az autoregresszív módszerre utal, ami olyan modelleket jelent, melyek a változó jelenlegi értékét az összes többi változó múltbeli értékével magyarázzák.

Az említett VAR modell két idősor (Y_t , X_t) esetén a következőképpen néz ki:

$$\begin{aligned} Y_t &= \beta_0 + \beta_{11} * Y_{t-1} + \beta_{12} * Y_{t-2} + \dots + \beta_{1k} * Y_{t-k} \\ &+ \beta_{21} * X_{t-1} + \beta_{22} * X_{t-2} + \dots + \beta_{2k} * X_{t-k} + u_t \\ X_t &= \alpha_0 + \alpha_{11} * Y_{t-1} + \alpha_{12} * Y_{t-2} + \dots + \alpha_{1k} * Y_{t-k} \\ &+ \alpha_{21} * X_{t-1} + \alpha_{22} * X_{t-2} + \dots + \alpha_{2k} * X_{t-k} + v_t \end{aligned}$$

Az egyenletekben azt láthatjuk, hogy a konstans (β_0) mellett egy idősor összes saját késleltetése kap egy β paramétert. És a rendszerben szereplő összes többi változó (idősor) összes késleltetése szintén kap egy külön β paramétert. Az együtthatók becslése itt is OLS módon, tehát a legkisebb négyzetek módszerével történik.

A lag-ot, vagy magyarul késleltetést különböző módszerekkel lehet megválasztani, például információs kritériumok alapján, de lehet a kontextust átgondolva a saját logikus megítélésünkre is hagyatkozni. Itt a megfigyelések számához képest a nagyon sok dimenzió

(változó) miatt a leggyakrabban használt információs kritériumok, az AIC (Akaike), HQ (Hannan-Quinn) és BIC (Bayes-Schwarz) nem adnak megfelelő javaslatot a késleltetés megválasztására. Így önkényesen a $k=8$ lag-ot választom, úgy vélem, hogy nagyjából 1 hetet érdemes visszatekinteni a múltba az esetszámok tekintetében. A magas dimenziószám egyéb problémákat is eredményez, ezt a következő alfejezetben mutatom be.

4.2. LASSO VAR modell

Amennyiben utánaszámolunk, minden idősor t -edik elemének becslésére így összesen 864 ($108 \cdot 8$, hiszen minden idősort felhasználok 8 nappal visszamenőleg) változót használok. Ezzel van egy kritikus probléma a VAR modellben, mert ez több, mint ahány megfigyelésem összesen van (673).

Ha a 4.1. fejezetben említett OLS becslést használnánk a β paraméterek megállapításához, akkor a következő zárt alakot kellene alkalmaznunk:

$$\hat{\beta} = (X^T * X)^{-1} * X^T * y,$$

ahol X a megfigyelések és a becslésre használt változókból alkotott mátrix, y pedig a becsülni kívánt célváltozó. Ez az egyenlet csak akkor értelmes, ha az $(X^T * X)$ mátrix, ahogyan a fenti képletben is látszik, invertálható. Viszont ha $p > n$, vagyis a predikcióra felhasznált változók száma nagyobb, mint a megfigyelések, esetünkben a napok száma, akkor X mátrix nem teljes oszloprangú. Ez máshogy mondva azt jelenti, hogy az oszlopai nem lineárisan függetlenek egymástól, tehát legalább az egyik oszlop biztosan előállítható a többi oszlopok valamilyen lineáris kombinációjaként. Ha ez igaz, akkor az is következik ebből, hogy:

$$\det(X^T * X) = 0$$

Az inverzszámításnál pedig éppen a mátrix determinánsával kell osztani, így ennek következtében ez az inverz nem létezik.

Egyébként, még csak az sem kell, hogy a magyarázóváltozók száma meghaladja a megfigyelések számát, ha p kellően nagy és közel van n -hez, akkor matematikailag ugyan megoldható az OLS becslés, viszont nagyon könnyen instabillá válik a VAR modell az úgynevezett overfitting (vagy magyarul túlillesztés) miatt. Ilyenkor ugyanis az OLS közel nullára tudja csökkenteni a reziduális hibát, ami ugyan jónak tűnhet elsőre, de ez valójában nem

tanulást eredményez, sokkal inkább memorizálást. Nem azt szeretnénk, hogy minél jobban illeszkedjen a tanuló adathalmazra, hanem hogy a megfigyeléseken túli világban is jól alkalmazható, azaz általánosítható modellünk legyen.

A probléma megoldásához egy másik modellhez kell folyamodnom, az úgynevezett LASSO - Least Absolute Shrinkage and Selection Operator (Han et al., 2015) modellhez. Ezt a módszert kifejezetten a nagy dimenziójú adatbázisok kezelésére találták ki, ahol szükség van a modell komplexitásának kontrollálására és az optimális változók kiválasztására. A módszer 1996-ban jelent meg Robert Tibshirani munkájában, aki a klasszikus lineáris regressziót egy szabályozó mechanizmussal egészítette ki. Formálisan a következőképpen néz ki a LASSO célfüggvénye:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \sum_{i=1}^n (y_i - x_i^T * \beta)^2 + \lambda * \sum_{j=1}^p |\beta_j| \right\}$$

A módszer matematikai alapja abban rejlik, hogy az egyenlethez hozzáadunk egy büntetőtagot, mely azt vizsgálja, mekkorák az együtthatók abszolútértékeinek összege. Az első tag a klasszikus OLS esetén már megismert SSE, tehát maga a reziduális négyzetösszeg. Ami az esetünkben újdonság, az a második tag, amit hivatalosan L_1 regularizációs tagnak hívunk. Ennél kulcsfontosságú szerepe van a λ paraméter megválasztásának:

- ha $\lambda \geq 0$, ez meghatározza, hogy mekkora súllyal számít majd bele a büntetés a célfüggvénybe.
- ha $\lambda = 0$, a modell visszavezethető a klasszikus lineáris regresszióra, hiszen a büntetőtag mindenhol 0 értéket vesz fel.
- ha $\lambda \rightarrow \infty$, az összes paraméter 0-vá válik, a modell leegyszerűsödik pusztán a konstans (β_0) tagra.

A λ növelésével tehát egyre több β paraméter válik nullává, vagy ahhoz közeli kicsi értéké, ami csökkenti a modell varianciáját, azaz a különböző mintákon elvégzett becslésből eredő átlagos hibákat. Ez azért nagyon fontos, mivel ezzel éppen a fent említett a matematikai korlátot és a túlillesztést tudjuk áthidalni. Cserébe viszont a tényleges, sokaságban is fennálló kapcsolatok egy részét 'elveszthetjük', ha a paramétereket túlságosan 'összenyomjuk', így akár fontos összefüggéseket is kihagyhatunk a modellből. A λ paraméter megválasztása tehát nagyon fontos, hiszen a gépi tanulásban egy alapvető problémához jutunk el ezzel, hogy kompromisszumot kell kötnünk az illeszkedés pontossága és a túltanulás között.

Fontos továbbá megemlíteni, hogy a LASSO-ra nincsen zárt alakú megoldás. A klasszikus OLS esetén olyan minimumpontokat határoztunk meg, ahol a célfüggvény deriváltja 0, a LASSO regresszió esetében viszont a büntetőtag abszolútértékben van, így ez nem deriválható. Ezért egy másik megközelítéssel becsüljük ezt a tagot, az úgynevezett koordináta-ereszkedési algoritmussal (Fraleigh & Hesterberg, 2009). Ez egy szubdifferenciálásra épülő megoldás, ahol egyszerre csak egy dimenzió mentén optimalizálunk, vagyis egyidejűleg csak egy paramétert frissítünk. Ez számítási kapacitásban is előnyös lehet, illetve nagy dimenziószámú problémáknál hatékony, ahol nem differenciálható a célfüggvény.

Ez a LASSO módszer tehát képes lesz VAR modellt becsülni a covid esetszámokra a megfigyelések számánál több magyarázóváltozó ellenére is. Ez pedig így már megalapozza az elemzésem legfontosabb részét, mégpedig a régiók közötti kapcsolatok feltárását.

4.3. IRF és FEVD

Legyen szó bármilyen pénzügyi vagy gazdasági rendszerről, azok feltáró elemzése során az egyik legfontosabb kérdés az, hogy a különböző változók hogyan hatnak egymásra. Ha például ez egyik változóban bekövetkezik egy sokk, azaz valamilyen hatás éri a rendszert, kritikus megvizsgálni azt, hogy az időben és térben hogyan terjed tovább a rendszer többi elemére. Ezen függőségek feltárására vezette be Diebold és Yilmaz (2012) a *Connectedness Approach* nevű módszertant, mely azóta széles körben az egyik leggyakrabban használt megközelítés pénzügyekben volatilitásterjedési hálózatok elemzésére. A következőkben bemutatom, hogy milyen további módszertanok és statisztikai megoldások szükségesek ahhoz, hogy ez a modell felépíthető legyen.

4.3.1. Impulzus-válaszfüggvény

A Connectedness megközelítés a VAR modellekből levezethető impulzus-válaszfüggvényekre (Impulse-Response Function: IRF) és az abból számolható előrejelzési hiba variancia dekompozícióra (Forecast Error Variance Decomposition: FEVD) épül. Ugyanis ezekből meghatározható, hogy az egyes változók előrejelzési hibájának mekkora százalékát

magyarázzák a saját, vagy a többi változóban bekövetkezett múltbeli sokkok. Máshogyan fogalmazva megbecsülhető, hogy mekkora a sokkterjedés relatív mértéke (spillover).

Az előző fejezetben bemutatott VAR vagy LASSO VAR modell átírható egy úgynevezett végtelen rendű mozgóátlag reprezentációra:

$$y_t = \sum_{i=0}^{\infty} \phi_i * \varepsilon_{t-i}$$

A VAR(p) modell a múltbeli változók valamely lineáris kombinációjaként írja le a jelenlegi értéket, ahogy ezt korábban bemutattam. Ezzel szemben ez az MA(∞) formula a múltbeli hibák, vagyis sokkok lineáris kombinációját használja fel a jelen érték becslésére. Ez azért fontos az esetünkben, mert a megfelelő feltételek mellett ezek az ε_t értékek nem korrelálnak egymással, így a rendszer válaszai ezekre a hibákra izoláltan vizsgálhatók.

A képletben a ϕ_i együtthatómátrixok azt mutatják meg, hogy egy i -edik késleltetésben bekövetkezett sokk mekkora hatást gyakorol az idősor aktuális értékére. Az ε_{t-i} maga az exogén sokk, ami $t-i$ -edik időpontban éri a rendszert. Ezzel az MA átalakítással intuitív módon értelmezhető a sokkok mechanizmusa, hiszen láthatjuk, hogy egy adott időpontban bekövetkezett sokk nemcsak azonnali hatást fejt ki, hanem hosszútávú hatása lehet a rendszerre. A mozgóátlag-forma egyik legfontosabb tulajdonsága, hogy képes időben kiterjeszteni a VAR-modell alapvetően rövid távra fókuszáló autoregresszív struktúráját.

A ϕ_i együtthatómátrixok segítségével tehát impulzus-válaszfüggvény elemzés végezhető a következő módon (Pesaran, 2015):

$$IRF_{ij}(h) = e_i' * \phi_h * e_j$$

A képletben a h az időhorizontot mutatja, vagyis azt, hogy legfeljebb hány periódussal a sokk bekövetkezte után szeretnénk azt megmérni. Az e_i és e_j egységvektorok, melyek az i -edik és j -edik változót képesek a rendszerben azonosítani. És továbbra is a ϕ_h mátrix a h időtávon a sokk terjedésének intenzitását és irányát írja le.

Korábban már említettem, hogy a mozgóátlag-reprezentáció a modell hibáin keresztül határozza meg az eredményváltozó értékét, az IRF ezt számszerűsíti. Máshogyan mondva, ez a kifejezés azt mutatja meg, hogy ceteris paribus, egy egységnyi sokk a j -edik idősorban milyen hatást fejt ki az i -edik idősorra h periódussal később. Ezen a ponton négy fontos megjegyzést tennék az IRF kapcsán, amit az elemzések során figyelembe kell venni:

- 1) Ezek a sokkhatások változásai ceteris paribus, azaz minden egyéb változatlansága mellett értelmezhetőek, ez azért kiemelten fontos, mert a rendszer kapcsolatainak közvetlen elemzését teszi lehetővé.
- 2) Ezeknek a impulzus-válasz kapcsolatoknak iránya van. A képletben a j -edik idősort sokkoljuk, és vizsgáljuk, hogy hogyan reagál rá az i -edik idősor. Vannak helyzetek, amikor nem szeretnénk a kapcsolat irányára, csak nagyságára és lecsengésére fókuszálni, erre létezik egy módszer, melyet a 4.3.2. fejezetben be fogok mutatni.
- 3) A ϕ_h elemei tehát a rendszer dinamikus multiplikátorai, melyek segítségével a sokkok számszerűsíthetőek és azok időbeli lecsengése is megfigyelhető. Nagyon fontos megfigyelni, hogy ha például $\phi_h(i, j) > 0$, akkor egy pozitív sokk a j -edik idősorban h időszak múlva növeli az i -edik idősor értékét. Ugyanígy, ha negatív az előjel, az fordított kapcsolat fennállását jelzi a két komponens között.
- 4) Említést érdemel, hogy az IRF grafikus formában is rendkívül jól értelmezhető. Ezek az impulzus-válaszgörbék az idő függvényében diagramon ábrázolják, hogy hogyan reagál az egyik idősorban bekövetkezett változásra a másik idősor, és ezek mértéke, iránya és lecsengése intuitív módon értelmezhetőek.

4.3.2. Becslési variancia dekompozíció

Az előrejelzési hiba variancia dekompozíció elsődleges célja, hogy feltárjuk vele a rendszer jövőbeli bizonytalanságának ok-okozati kapcsolatait (Lütkepohl, 2005). A FEVD (Forecast Error Variance Decomposition) ugyanis képes arra, hogy visszavezesse, hogy egy adott idősor előrejelzési hibájának mekkora hányadát teszik ki a saját sokkjai, és mekkora hányadát a többi idősor exogén sokkjai. Formálisan a FEVD is a 4.3.1. fejezetben bemutatott mozgóátlag-reprezentációból vezethető le:

$$\theta_{ij}(H) = \frac{\sum_{h=0}^{H-1} (e'_i * \phi_h * \sum e_j)^2}{\sum_{h=0}^{H-1} (e'_i * \phi_h * \sum \phi_h' * e_i)}$$

A H az előrejelzési horizontot, vagyis azt az időtávot jelenti, amelyre a modell előrejelzéseit és hibáit vizsgálni szeretnénk. Az e_i és e_j pontosan ugyanazok az egységvektorok, mint az IRF

esetében, feladatuk, hogy azonosítsák az i -edik és j -edik idősort. A ϕ_h a h -adik késleltetéshez tartozó impulzus-válaszfüggvény értékeinek mátrixa.

Ha szeretnénk megérteni a törtet, érdemes a számlálót és a nevezőt külön megvizsgálni. A számláló azt fejezi ki, hogy az i -edik változó előrejelzési hibájának mekkora részét teszi ki a j -edik változóban h perióduson keresztül bekövetkezett sokk. A nevező ezzel szemben a teljes előrejelzési hiba varianciát adja meg a i -edik változóban az összes többi idősort figyelembe véve, nemcsak a j -edik idősort. Ez tehát már nem csak egy-egy kapcsolatban értelmezi egy idősor becslési varianciáját, hanem a teljes rendszer hatása benne van a nevezőben. Így már könnyebb látni, hogy mindez végső soron egy arányt fejez ki, mely azt mutatja, hogy az i -edik idősor jövőbeli bizonytalanságának hány százalékát teszi ki egy j -edik idősor exogén sokkja. Tehát ha például $\theta_{ij}(H) = 0,2$ adódik, akkor ez azt jelenti, hogy az i -edik idősor előrejelzési hiba varianciáját 20%-ban magyarázzák a j -edik idősorból eredő sokkhatások. Ilyen százalékok vizsgálatával már egyre közelebb kerülünk ahhoz, hogy a rendszer összekapcsoltságáról fogalmazzunk meg állításokat. Mert ha $\theta_{ij}(H)$ egy alacsony százalék egy idősor esetében, akkor egy zárt rendszert feltételezhetünk, hiszen az adott idősor jövőbeli bizonytalanságáért nagyrészt a saját múltjai felelősek. Ezzel szemben ha $\theta_{ij}(H)$ magas, akkor erős kereszthatásokat feltételezhetünk a rendszerben, az i -edik idősor viselkedése ekkor jelentősen függ a j -edik idősortól.

A hagyományos, fent bemutatott előrejelzési hiba variancia dekompozíció módszer egyik fontos korlátja, hogy annak értelmezése és alkalmazása függ a változók sorrendjétől. Ennek röviden az az oka, hogy a számításhoz használt sokkokat először valamilyen mesterséges módszerrel ortogonizáljuk. Ez azt is jelenti, hogy a változók közötti korrelációt valamilyen mesterséges módszerrel megszüntetjük, például a Cholesky-felbontás módszerével, így a hibatagok függetlenek lesznek egymástól. Ez a módszer azonban szekvenciális módon zárja ki az egymásra gyakorolt hatásokat, aminek súlyos következménye, hogy a VAR modell különböző sorrendű magyarázóváltozók esetén különböző FEVD eredményeket adhat. Tehát míg egy az egy kapcsolatoknál nagyon jól értelmezhető a klasszikus FEVD, addig a rendszer teljes összekapcsoltságának vizsgálatakor más megközelítést alkalmazunk.

Erre a problémára az úgynevezett generalizált előrejelzési hiba variancia dekompozíció (GFEVD) áll a rendelkezésünkre, mely kiküszöböli ezt a sorrendérzékenységet.

Formálisan a GFEVD képlete a következő módon írható fel (Sayed, 2020):

$$\tilde{\theta}_{ij}(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e'_i * \phi_h * \sum e_j)^2}{\sum_{h=0}^{H-1} (e'_i * \phi_h * \sum \phi_h' * e_i)}$$

Amint láthatjuk, a képlet nevezője teljesen változatlan, míg a számláló kap egy σ_{jj}^{-1} korrigációs tagot. A σ_{jj} a j -edik hibatag varianciája, vagyis:

$$\sigma_{jj} = e'_j * \sum e_j$$

Mindez egyfajta skálázást tesz lehetővé, mert az egyes komponensek akkor is összehasonlíthatóvá válnak, ha a sokkok nagysága eltérő a változók között. Hiszen a j -edik hibatag hatását így a saját varianciájához viszonyítva súlyozza, így elkerülhető, hogy a nagyobb varianciájú változók túl nagy hatást fejtsenek ki, míg a kisebb varianciájú komponensek hatását alábecsüljük. Ez a kvázi standardizálás tehát minden sokkhatást egységnyi intenzitásra hoz, így azok összehasonlíthatóak és értelmezhetőek.

Mindez a már korábban említett Diebold-Yilmaz-féle Connectedness megközelítés és a spillover indexek számításának alapját képezi. Egyetlen korrekcióra van még szükség, hogy a fentiek eredményeképpen felépítsem ezt a Connectedness indexet. A GFEVD komponensek összege nem feltétlen egyezik meg 1-gyel, mivel a hibatagok korreláltak, így lehetnek bennük átfedések. Ahhoz tehát, hogy a GFEVD automatikusan az i -edik idősor teljes előrejelzési hiba varianciájának arányaiban mérje a j -edik idősor sokkjainak hatását, még ezeket az értékeket Diebold és Yilmaz az alábbi módon javasolja korrigálni:

$$\tilde{\theta}_{ij}(H) = \frac{\tilde{\theta}_{ij}(H)}{\sum_{j=1}^N \tilde{\theta}_{ij}(H)}$$

Ez a klasszikus normalizálás tehát elsősorban az értelmezést segíti. Megmutatja, hogy egy adott változó előrejelzési hibájának hány százalékát magyarázza a saját múltja, és hány százalékot a rendszer többi változóiban előidézett sokkhatások. A GFEVD tehát hasonlóan értelmezhető, de három kulcsfontosságú előnyt is kínál a hagyományos FEVD-hez képest: sorrendfüggetlen, a kovarianciastruktúra torzítatlan és százalékos formában értelmezhető. A következő fejezetben a Connectedness modellt mutatom be, amivel a végső elemzésemet is végzem.

4.4. A Connectedness megközelítés

A Diebold-Yilmaz-féle Connectedness alapvető célja, hogy kvantifikálja egy adott rendszerben a sokkterjedés irányát és mértékét. Ahogyan említettem a 4.3. fejezetben, elsődlegesen pénzpiacok elemzésénél használják, ahol például az egyes részvények árfolyamainak egymásra gyakorolt hatása kulcsfontosságú. De értelmezhető 'fertőzés átadás' kontextusban is, kiválóan mutatja meg, hogy mely régiók milyen szerepet töltenek be a világjárvány terjedésének dinamikájában. Mindez a korábban levezetett GFEVD mutatókra épül, ebből különféle spillover indexek számolhatók, melyekkel a rendszer egésze, de akár konkrét régiók kapcsolatai is leírhatók. A következő alfejezetekben sorra bemutatom a Connectedness által kínált különböző indexeket, hogy melyik mit jelent, hogyan számolhatók és mi a jelentőségük (Diebold & Yilmaz, 2014).

4.4.1. Összesített (TCI) index

A rendszer egészére kiterjedő legfontosabb mutató a Total Connectedness Index (TCI). Ez az egész modellre kiterjedő sokkterjedést méri, és megmutatja, hogy mennyire összekapcsolt a rendszer valójában. Matematikailag a következő módon írható fel a TCI:

$$TCI(H) = \frac{\sum_{i=1}^N \sum_{j=1, j \neq i}^N \tilde{\theta}_{ij}(H)}{\sum_{i=1}^N \sum_{j=1}^N \tilde{\theta}_{ij}(H)} * 100$$

Megfigyelhető, hogy egyetlen különbség van a számlálóban és a nevezőben, mégpedig az, hogy a számlálóban az egy adott idősor múltjából önmagára terjedő hatást nem vizsgáljuk. Így a tört számlálója tehát kizárólag a kereszt-hatásokat összegzi, az összes olyan hatást, ami az egyik változóból a másikra terjed. Eközben a nevező az összes hatást vizsgálja, ez a rendszer teljes előrejelzési hiba varianciája. A tört így egy arányt fejez ki, mely azt mutatja, hogy az előrejelzési hiba variancia hány százaléka származik más változókból eredő sokkokból. A tört természetesen 0 és 1 közötti értékeket vehet fel, a TCI index pedig így 0% és 100% közé skálázódik. Minél közelebb vagyunk a 100%-hoz, a variancia annál nagyobb hányadát magyarázzák más változók, a rendszer tehát annál összekapcsoltabb. A 0%-hoz közeli érték pedig alapvetően egy izolált rendszert feltételez, ebben az esetben nem jelentős a fertőzési dinamika.

4.4.2. Irányított spillover index

A TCI tehát a nagy képet mutatja, a rendszer egészére ad egy számszerűsített kapcsoltsági mutatót. Azonban ha szeretnénk a kapcsolati hálóban mélyebb elemzéseket végezni, erre használhatjuk az irányított spillover mutatókat. Sok esetben az is nagyon fontos, hogy ezek a sokkok honnan hová terjednek. Diebold és Yilmaz erre a vizsgálati szempontra is vezettek be mutatókat:

- Kimenő (To) spillover index: azt mutatja meg, hogy egy adott változó sokkjai mekkora hatást fejtenek ki a rendszer többi változóra.
- Bejövő (From) spillover index: megmutatja, hogy egy adott változóra mekkora hatással van a rendszer többi változójának összes sokkja. Máshogyan fogalmazva, mekkora mértékben magyarázzák az adott változó előrejelzési hiba varianciáját a többi változóban bekövetkezett sokkok.

Képletekkel mindez a következő módon írható le:

$$C_{i \rightarrow \cdot}^{(H)} = \frac{\sum_{j=1, j \neq i}^N \tilde{\theta}_{ji}(H)}{\sum_{j=1}^N \tilde{\theta}_{ji}(H)} * 100$$

$$C_{\cdot \rightarrow i}^{(H)} = \frac{\sum_{j=1, j \neq i}^N \tilde{\theta}_{ij}(H)}{\sum_{j=1}^N \tilde{\theta}_{ij}(H)} * 100$$

A felső képlet a kimenő mutató, míg az alsó a bejövő index képletét mutatja. A képletből látszik, hogy mindig a megfelelő GFEVD értékek kerülnek összegzésre, figyelve arra, hogy az önmagára gyakorolt hatást kihagyjuk a számlálóból. Ezek a To és From indexek egy újabb fontos mutatóhoz vezetnek.

4.4.3. Nettó spillover index

A fent bemutatott két érték különbségével jutunk el a nettó spillover indexhez. Ez az index az i -edik idősor esetében a bejövő és kimenő kapcsolatok közötti eltérést mutatja:

$$C_i^{NET}(H) = C_{i \rightarrow \cdot}^{(H)} - C_{\cdot \rightarrow i}^{(H)}$$

A mutató előjeléből fontos következtetéseket vonhatunk le:

- Ha $C_i^{NET} > 0$, az azt jelenti, hogy a rendszerben az adott változó nettó kibocsátó, vagyis több hatást fejt ki a többi változóra, mint amennyit maga kap a többi változótól.
- Ha $C_i^{NET} < 0$, az azt jelenti, hogy a rendszerben az adott változó nettó befogadó, vagyis a többi változóból érkező sokkhatások dominálják az ő saját bizonytalanságát.

A mutatóval tehát a sokkok iránya és a rendszerben jelenlévő idősorok szerepkörei is meghatározhatóak. Mindez nagyon hasznos a kulcsfontosságú szerepkörök meghatározására.

4.4.4. Páronkénti spillover index

Joggal merülhet fel az igényünk, hogy páronként is szeretnénk vizsgálni a rendszer elemeit. Ha szeretnénk két konkrét változót kiragadni, azok között is tudok információáramlást vizsgálni a páronkénti spillover index segítségével:

$$C_{ij}^{(H)} = (\tilde{\theta}_{ij}(H) - \tilde{\theta}_{ji}(H)) * 100$$

Ez megmutatja, hogy kizárólag az i -edik és a j -edik idősor között milyen irányú kapcsolat áll fent. Pozitív érték esetén a terjedés inkább az i -edik idősorból a j -edik idősor irányába mutat, negatív esetben pedig a j -edik idősorból az i -edik irányába domináns.

4.4.5. Az indexek hálózati interpretációja

A fent bemutatott mutatókon kívül további vizuális megoldások is a rendelkezésünkre állnak, melyekkel további összefüggéseket mutathatunk ki, és akár különböző időpontokban is vizsgálódhatunk. Ezeket a dolgozatom 5. fejezetében mutatom be. Mindezek segítségével a hálózat dinamikáját tudjuk feltérképezni, és fontos megfigyeléseket tehetünk. Legyen szó pénzügyi, gazdasági vagy bármilyen egyéb idősorról, azok egy csomópontként értelmezhetőek. A köztük megjelenő spillover-hatások pedig irányított élek, melyek súlyát a korábban bemutatott $\tilde{\theta}_{ij}(H)$ vagy $C_{ij}^{(H)}$ értékek adják meg. Ezeket a gráfokat ábrázolva megfigyelhetjük a hálózat következő tulajdonságait:

- A nagy számú kimenő kapcsolattal rendelkező csomópontok a hálózat fertőzésforrásai, melyek nagy hatással bírnak a rendszerre.
- A nagy számú bejövő kapcsolattal rendelkező csomópontok a legfőbb befogadók, melyek így a rendszer gyengepontjaiként értelmezhetők.
- A vastagabb élek (ha ezt is ábrázoljuk) erősebb kapcsolatot jelentenek, így azonosíthatjuk a páronként nagyon összekapcsolt idősorokat.
- Gördülőablakos VAR modellillesztéssel az időbeni változás is megfigyelhető, ezt az *5.1. fejezetben* be is mutatom.
- Megfigyelhetünk akár alrendszereket, klasztereket alkotó vagy izolált változókat, melyek különleges szerepet játszhatnak a rendszerben.
- Ha a hálózatban néhány csomópontot látunk dominánsnak, ez magas centralizációt feltételez, ellenkező esetben decentralizált rendszerrel állunk szemben.

Számos, a fentiekhez hasonló hasznos következtetés vonható le a Connectedness modellek alkalmazásával. A következő fejezetben bemutatom a kutatásom eredményeit, melyeket a fent bemutatott módszertanokra alapoztam.

5. A kutatás eredményei

5.1. Window-size érzékenységvizsgálat

Ahogy a 4.4.5. fejezetben említettem, egyik nagy előnye a Diebold-Yilmaz megközelítésnek, hogy a hálózatot különböző időpontokban is le tudjuk írni. Én egy paneladattal dolgozom, melynek nem csak térbeli, hanem időbeli kiterjedése is van. Ez azt jelenti, hogy a COVID-19 világjárvány fertőzöttjeinek számát nem csupán a különböző régiókban, de majdnem 2 éven keresztül a különböző napokban is tartalmazza. És természetesen nagy hiányossága lenne az elemzésemnek, ha azzal a feltételezéssel élnék, hogy a rendszer ebben az időtartamban minden nap ugyanolyan összekapcsoltsági struktúrával rendelkezett. Hiszen többek között a különböző kormányzati intézkedések önmagukban nagyon nagy hatással voltak a fertőzés terjedésére. Előzetesen azt várjuk, hogy a kijárási korlátozások idején a kevesebb fizikai kontakt és utazás miatt a rendszer zártabb, így kevésbé összekapcsolt, míg a mobilitás kisebb mértékű korlátozása idején szorosabb. Ezen felül természetesen az oltások megjelenése, annak elterjedése, a járvány természetes lecsengése és számos egyéb tényező befolyásolhatja mind országon belül, mind országok között az összefüggéseket.

Ahhoz tehát, hogy meg tudjuk figyelni az időbeli változásokat, egyszerre az adatoknak csak egy bizonyos részét elemezzük. A rolling-window, vagy magyarul gördülőablakos módszer integrálása pont ezt a célt szolgálja (Demirer et al., 2018). Ez nagyon hasonlít a korábban már bemutatott LOESS regresszió elméletéhez. Az ablakméret paraméter függvényében, az első n darab megfigyelés alapján becsülünk LASSO regressziókat. Ez azt jelenti, hogy ha például $n = 100$, akkor az 1. és a 100. közötti megfigyelések alapján fog prediktálni a modell, melyet a 100. dátumhoz fog rendelni. Így ugyan meg tudjuk figyelni a regressziós együtthatók, és ezen keresztül a GFEVD alapú hálózati struktúra időbeli változását, viszont ez adatvesztést is eredményez. Ha az ablakméretet most w -vel jelöljük, akkor az elemzésre használható dátumok a következőképpen alakulnak:

$$N_w = N - w + 1$$

Továbbá említettem, hogy $k=8$ késleltetéssel szeretném a LASSO VAR modellt felépíteni. Ez is közre fog játszani abban, hogy végül mennyi megfigyeléssel számolhatok tovább. Ez ugyanis azt jelenti, hogy az ablakon belül csak $(w - lag)$ tekinthető használhatónak, és az R -ben

implementált függvény is ezt a logikát követi. A késleltetések számát is figyelembe véve az elemszám továbbá így alakul:

$$N_{w,lag} = N - (w - lag) + 1$$

Összesen 673 megfigyelésem van, így ha például 100-as ablakmérettel dolgozom, akkor $673 - (100 - 8) + 1 = 582$ dátumom marad, melyre az összekapcsoltságot vizsgálhatom. Ez az adatvesztés elkerülhetetlen, ha időben szeretném vizsgálni a kapcsolatokat, mely kulcsfontosságú része az elemzésemnek.

A megfelelő ablakméret kiválasztását érzékenységvizsgálattal végeztem. Mivel egyetlen ablakméret mellett is a modell futási ideje nagyjából 10 óra volt, így a számítási kapacitás miatt csak pár különböző értéket volt lehetőségem megvizsgálni. Az elemszám figyelembe vételével a következő méretek mellett futtattam a modellt: 100, 150, 200. Annak eldöntésére, hogy melyiket használjam, egy egyszerű átlagos abszolút eltérést használtam. Vettem a 3 Connectedness modell által kapott TCI indexek átlagát, majd megnéztem, hogy melyik esetben van a legkisebb eltérés. Az eredmények alapján a teljes hálózatra, mely az összes régiót tartalmazza, a 150-es ablakméret bizonyult a legreprezentatívabbnak.

A fenti modellen kívül, mely tehát mind a 108 db régiót tartalmazza, további kisebb hálózatokat is megvizsgáltam. Magyarországot önmagában is szerettem volna elemezni, így külön a 8 db magyar régióra is futtattam ugyanezt a modellt.

Emellett Magyarországot a vele közvetlenül szomszédos régiók viszonylatában is érdekesnek találtam, hiszen így a közvetlen külföldi mozgásokat is megfigyelhetjük. Így két fontos ország, Ausztria és Szlovákia szomszédos régióit is bevontam az elemzésbe, előbbiből 4 db, utóbbiból 3 db került bele ebbe az új, harmadik modellbe.

Végül egy másik ország belföldi dinamikáit is szerettem volna feltárni, erre Lengyelországot választottam, melynek 17 db régiójára külön szintén futtattam egy modellt.

A fentebb említett három különböző részmodell esetében is megvizsgáltam ugyanezen abszolút eltérés elvén az ablakméreteket. Mindegyik esetben a 200-as paraméterérték esetében kaptam a legkisebb átlagos eltérést. Így az elemzésre használható dátumok a különböző modellekben a következőképpen alakultak át:

- Teljes modell (150-es ablakméret): 2020. 07. 25. – 2022. 01. 07.
- Magyarország modell (200-as ablakméret): 2020. 09. 13. - 2022. 01. 07.

- Magyarország és szomszédjai modell (200-as ablakméret): 2020. 09. 13. - 2022. 01. 07.
- Lengyelország modell (200-as ablakméret): 2020. 09. 13. - 2022. 01. 07.

5.2. Átterjedési (spillover) táblázat

Az első és legegyszerűbb megközelítés a terjedés vizsgálatára az úgynevezett spillover táblázat. Az 1. táblázatban a Magyarország régiói közötti átterjedéseket láthatjuk, mely a teljes időszakon illesztett LASSO-VAR modell eredménye. Ez a teljes modellre is kinyerhető, viszont a táblázatnak annyi oszlopa és sora van (a FROM és TO értékeken kívül), ahány régió, így a teljes 108 db NUTS2 régió egyszerre nehezen lenne átlátható táblázatos formában.

Régió	HU11	HU12	HU21	HU22	HU23	HU31	HU32	HU33	FROM
HU11	33.29	19.48	11.41	6.91	7.73	7.78	6.58	6.83	66.71
HU12	18.73	28.71	12.10	6.65	8.48	8.62	8.51	8.20	71.29
HU21	11.66	12.69	28.53	8.75	10.87	9.75	8.63	9.13	71.47
HU22	8.56	8.90	10.77	32.09	9.61	10.59	9.14	10.34	67.91
HU23	8.43	9.62	11.68	8.33	34.62	8.69	8.98	9.66	65.38
HU31	8.46	9.65	10.40	8.91	8.73	32.34	11.20	10.30	67.66
HU32	8.00	10.48	9.80	8.37	9.27	12.78	30.35	10.95	69.65
HU33	7.54	9.06	9.41	8.60	9.96	10.94	10.62	33.88	66.12
TO	71.37	79.88	75.57	56.52	64.64	69.14	63.67	65.41	546.20

1. táblázat: Magyarország NUTS2 régióinak spillover mátrixa a vizsgált időszakban

A mátrix a régiók közötti dinamikus kapcsolatok súlyát mutatja, mely azt mutatja meg, hogy az egyes régiók milyen mértékben járulnak hozzá a többi régió előrejelzési hibájához. A táblázat sorai a „FROM” hatásokat mutatják, vagyis az egyes régiók által kapott hatásokat, míg a táblázat oszlopai a „TO” értékek, melyek az átadott hatásokat mutatják. Ahogyan látható, a

főatlóban a változók saját magukra gyakorolt önhatásai jelennek meg, míg a nemdiagonális elemek a kereszt-hatásokat reprezentálják.

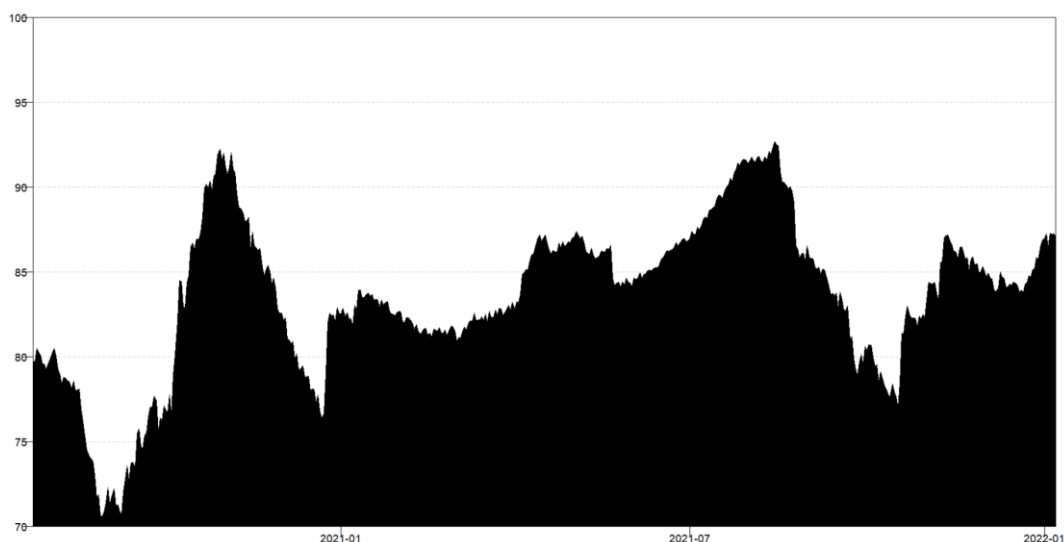
Ha a táblázat egy elemét kiragadjuk és szeretnénk értelmezni, akkor ezt is megtehetjük. Például a HU22 sorában és HU21 oszlopában 10.77-es érték található. Ez azt jelenti, hogy a HU21-es régió (Közép-Dunántúl) 10.77%-ban járul hozzá a HU22-es régió (Nyugat-Dunántúl) variáciájához. Máshogy fogalmazva a nyugat-dunántúli COVID-19 esetszámok változékonyságát 10.77%-ban magyarázzák a közép-dunántúli fertőzések korábbi értékei. A 4.3.2. fejezetben bemutattam az időhorizont szerepét a GFEVD esetében, a modellekben $H = 4$ értéket használtam.

Összességében a „TO” értékek alapján azt a következtetést vonhatjuk le, hogy Budapest (HU11), Pest (HU12) és Közép-Dunántúl (HU21) befolyásolják legnagyobb mértékben a hálózatot és közben erősen magyarázzák saját variációjukat is. Ez magyarázható Magyarország alapvetően centralizált jellegével. Pest vármegyében és Budapesten együttesen körülbelül 3 millió ember él, tehát a mindennapi közlekedés és ingázás miatt ez a térség a személyes kontaktok szempontjából kulcsfontosságú. Emellett az úthálózat is sugaras szerkezetű Budapestre irányuló tengelyekkel, és a forgalmas reptér is szerepet játszhat.

A „FROM” értékek 66,12% és 71,47% között mozognak, mely azt jelzi, hogy a fertőzés terjedése erősen hálózatos jellegű, a rendszer alapvetően erősen összekapcsolt.

5.3. TCI index alapú összekapcsoltsági elemzés

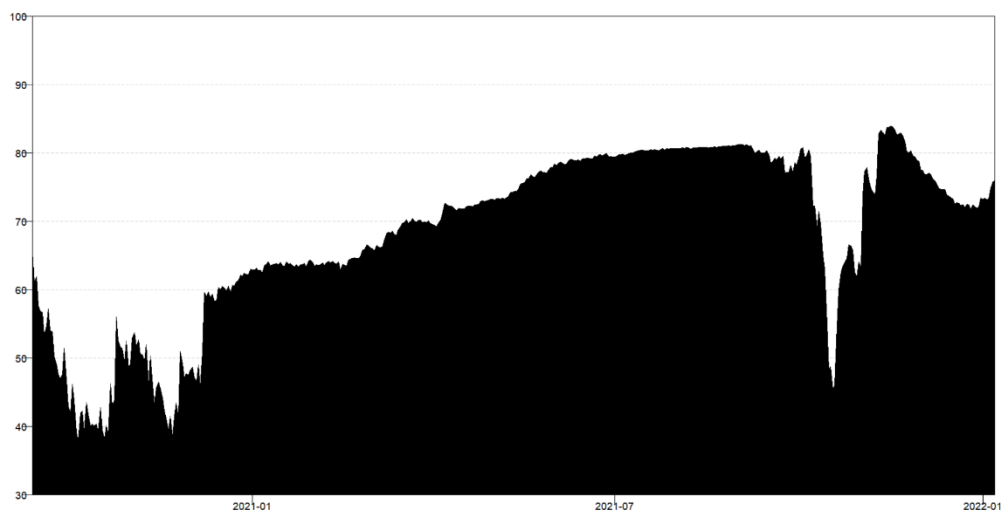
A 4.4.1. fejezetben bemutatott TCI index tehát egy százalékos érték, mely a rendszer összekapcsoltságát képes kvantifikálni. A gördülő ablakos módszerrel ezek az értékek időben is összehasonlíthatóak. A 3. ábrán az összes régiót tartalmazó modell TCI indexeinek változása látható 2020. 07. 25. – 2022. 01. 07. között:



3. ábra: A teljes rendszer TCI indexeinek időszora a vizsgált időszakban

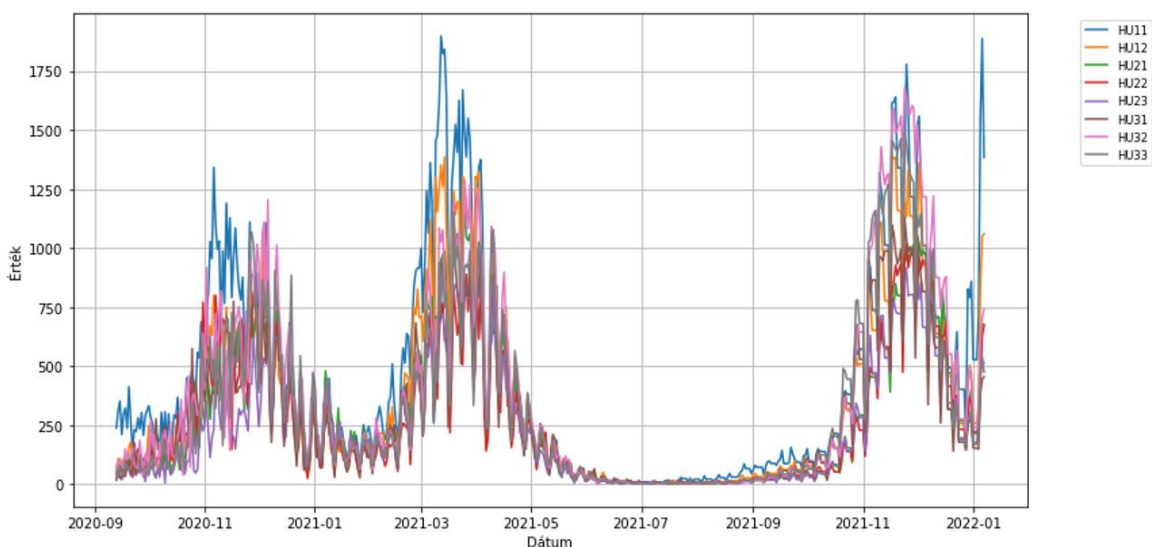
A függőleges tengelyen a TCI index százalékos értéke, míg a vízszintes tengelyen a fent említett dátumok láthatók. Az y tengely 70%-tól indul, hogy jobban láthatók legyenek a mozgások. Azt láthatjuk, hogy minden egyes nap 70% fölötti ez a TCI, vagyis az előrejelzési hiba variancia jelentős része más változókból származik. Vannak időpontok, ahol egészen 92% fölé is emelkedik ez az érték.

Az összekapcsoltság és a szigorítások között intuitívan egy fordított kapcsolatot várunk. Hiszen amikor nagyobbak a szigorítások, akkor a kisebb mobilitás és a kevesebb fizikai érintkezés egy izoláltabb rendszert eredményez. A szigorúbb szabályozások pedig a járvány fellendülésekor várhatóak, így azt várjuk, hogy közvetetten a magasabb esetszámok és a hálózat összekapcsoltsága negatívan korreláljanak. Mindezt az átláthatóság kedvéért kizárólag a magyarországi régiók esetében vizsgálom (4. ábra):



4. ábra: Magyarország TCI indexeinek idősora a vizsgált időszakban

Az 5. ábrán pedig egyszerűen a 8 magyar régió esetszámainak idősorai láthatók:



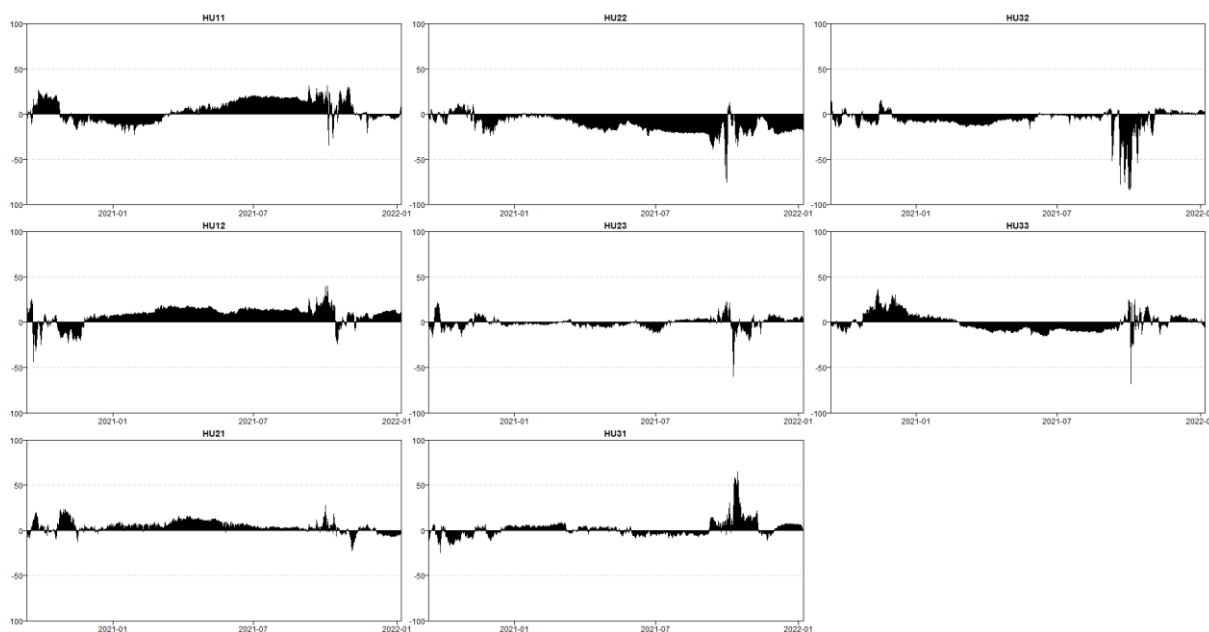
5. ábra: Magyarország NUTS2 régióinak esetszámai a vizsgált időszakban

A három hullám közül a 2020. novemberi és a 2021. év végi hullámokat jól leköveti az összekapcsoltság. Utóbbi esetben látható a legnagyobb beesés a TCI indexek idősorában, ebből logikus következtetés lehet, hogy ilyenkor voltak Magyarországon a legszigorúbb kijárási korlátozások. A 2021. márciusban induló hullám esetében csak stagnál az összekapcsoltság, viszont a hullám lecsengése után egyértelműen megfigyelhető a hálózat elemeinek egymásra gyakorolt hatásának növekedése. A teljes modellhez képest itt az alacsony régiószám miatt nem

olyan magasak ezek a TCI értékek, viszont az 50%-60% közötti átlagos értékek továbbra is egy szoros rendszert jelentenek.

5.4. Nettó spillover indexek elemzése

A 4.4.3. fejezetben bemutatott nettó spillover indexek idősoros ábrázolásának értelmezése nagyon intuitív: amennyiben egy régió indexe a vízszintes tengely 0 pontja alatt van, abban az időszakban inkább fogadó, ahol fölötté, ott pedig terjesztő. A 6. ábra (ismételten az átláthatóság kedvéért) a magyarországi régiókhoz tartozó értékeket mutatja be:



6. ábra: A magyarországi NUTS2 régiók nettó spillover indexei a vizsgált időszakban

Minden régiót egy időszorként ábrázolok, ahol minden időpontban megvizsgálom a régió összes hatását a rendszerre és kivonom belőle az összes hatást, ami a régiót éri a rendszer által. Ahogyan az 5.2. fejezetben vizsgáltam, a Budapest (HU11), Pest (HU12) és Közép-Dunántúl (HU21) régiók mutatták a legnagyobb „TO” értékeket, és ez összhangban van az itt látott nettó értékekkel. Az első oszlopban vannak ezek a régiók feltüntetve, és a teljes időtartam jelentős részében az x tengely felett pozitív értékeket mutatnak.

Elsődlegesen a HU22 (Nyugat-Dunántúl) és HU32 (Észak-Alföld) régiók tűnnek a vizsgált időszak legnagyobb részében fogadónak. A HU23 (Dél-Dunántúl), HU33 (Dél-Alföld) és HU31 (Észak-Magyarország) NUTS2 kódú régiók pedig inkább neutrálisan viselkednek, vannak időintervallumok, amikor inkább fogadnak, és van, amikor inkább terjesztenek több varianciát a rendszeren belül.

5.5. Hálózati diagramok

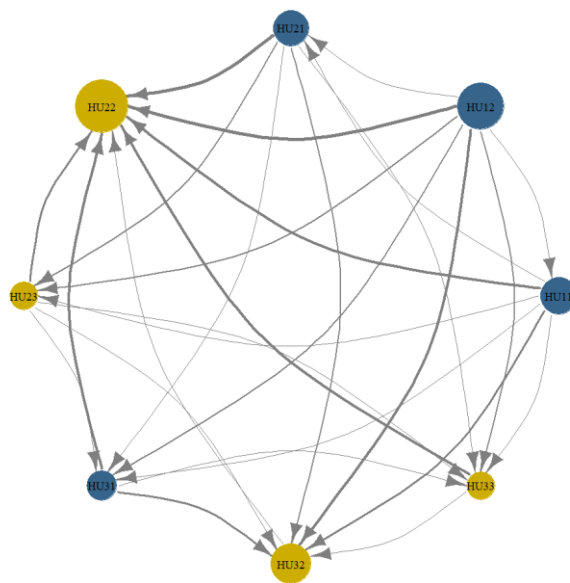
Ebben a fejezetben bemutatom a teljes rendszer és az általam választott alrendszerek hálózati diagramjait. Ezen vizuális megjelenítés lényegében egyesíti mindazt, amit fent bemutatam a különböző indexek kapcsán. A hálózati diagramon a NUTS2 régiók a csúcspontok, és a köztük lévő kapcsolatok irányát és erősségét is megvizsgálhatjuk. Minél vastagabb a kapcsolatot jelölő nyíl a két csúcspont között, annál nagyobb a variancia terjedése és a vonal irányából a terjedés iránya is leolvasható. Ezen felül szinkódolva is vannak a régiók: a sárga színű régió az adott időpontban vagy időszakban inkább fogadó (vagyis a hozzá tartozó nettó spillover index értéke negatív), míg a kék színnel jelölt régiók inkább terjesztők (a nettó spillover index pozitív). A régiókat reprezentáló csúcspontok nagysága is hordoz információt, minél nagyobb, annál jelentősebb fogadó/terjesztő a rendszerben („TO” és „FROM” értékek alapján). Az alfejezetekben a különböző modelleknek a hálózati diagramjait mutatom be és vonok le belőlük következtetéseket. Mivel a gördülőablakos módszer segítségével számított indexek lehetővé teszik az időben különböző időpontokban való elemzést, ezt ki is használok, és a teljes intervallumon kívül további négy különböző fontos időpontban vizsgálom a rendszert. Ezekre kiválasztottam egy-egy reprezentatív napot, ezeknek a hálózati diagramjait is be fogom mutatni:

- 2. hullám: a rendelkezésre álló időszakban az első nagyobb hullám, ahogyan az 5. ábrán is látszik, 2020. november és 2021. január között figyelhető meg. Ez hivatalosan a második hullám, az erre kiválasztott reprezentatív nap: 2020. 12. 01.
- 3. hullám: a következő időszak, mikor a megfertőzések száma erősen megemelkedett, ez 2021. március és 2021. május között figyelhető meg (l. 5. ábra). Az erre kiválasztott reprezentatív nap: 2021. 04. 01.
- 2021 nyara: megvizsgálom a kapcsolatokat egy olyan időpontban is, mely során lappangott a vírus, ez 2021. június és 2021. szeptember között látszik (l. 5. ábra). Az erre kiválasztott reprezentatív nap: 2021. 08. 01.

- 4. hullám: végül egy újabb fellendülés figyelhető meg 2021. november és 2022. január között, mely a teljes Közép-Kelet-Európa viszonylatában az egyik legsúlyosabb esetszámokat okozta (l. 5. ábra). Az erre kiválasztott reprezentatív nap: 2021. 12. 01.

5.5.1. Magyarország hálózati kapcsolatai

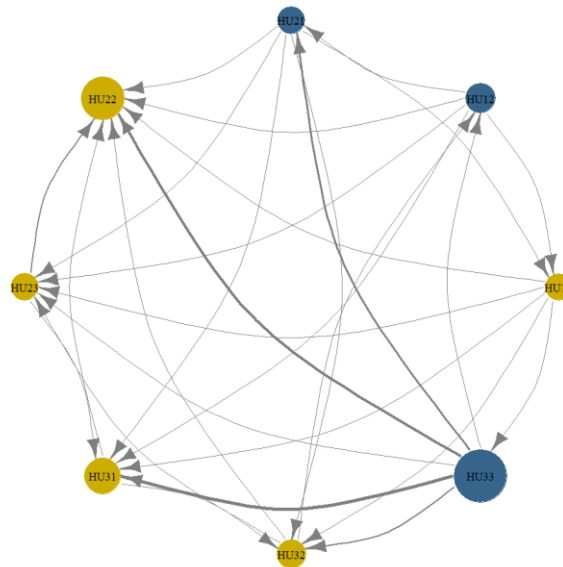
A 7. ábrán a teljes általam vizsgált időszakra vonatkozó, csak a magyar régiókból álló hálózat látható:



7. ábra: Magyarország NUTS2 régióinak hálózati kapcsolatai a teljes vizsgált időszakban

Megfigyelhető, hogy ez a vizuális megjelenítés megfelel a nettó spillover indexek megjelenítésére használt táblázattal, de sokkal jobban értelmezhető. A Nyugat-Dunántúl (HU22) régió tűnik a legfontosabb fogadónak, minden egyes régióval szemben negatív a páronkénti átterjedési indexe. Míg a Pest (HU11) és Budapest (HU12) régiók a legfontosabb terjesztők, egymáson kívül minden további régió irányába több információt adnak át, mint amennyit kapnak.

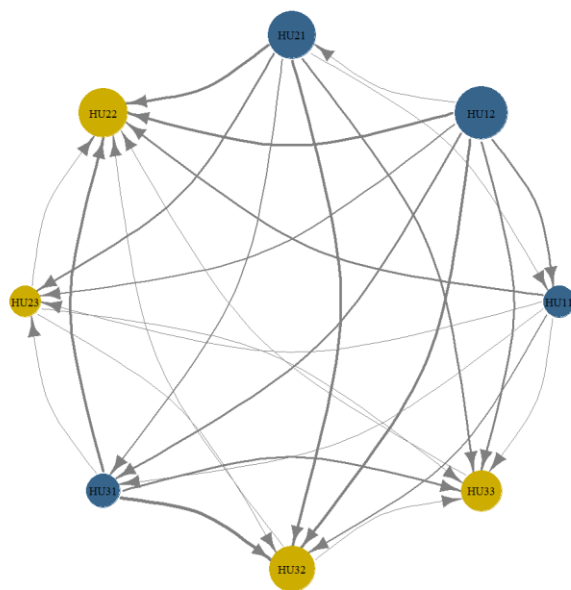
A 8. ábra a 2. hullám kapcsolatait mutatja be:



8. ábra: Magyarország NUTS2 régióinak hálózati kapcsolatai 2020. 12. 01-én

Szembetűnő különbségek figyelhetőek meg, a Nyugat-Dunántúl (HU22) továbbra is a legjelentősebb fogadó, viszont a Dél-Alföld (HU33) vált a legerősebb terjesztővé. Kijelenthető a modell alapján, hogy ebben az időszakban déli irányból indult a fertőződés. Lehetséges magyarázat erre a megfigyelésre, hogy a déli régiókban ebben az időszakban erősebb volt a határ menti átjárás (Szerbia vagy Románia felől), illetve ezekben a térségekben hamarabb jelenhettek meg olyan fertőzési gócpontok, amelyek gyorsabban indították el a variancia terjedését.

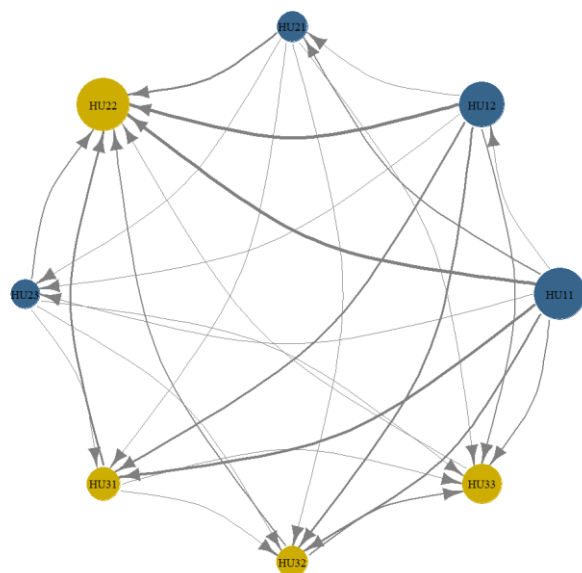
A 9. ábrán a 3. hullám hálózati kapcsolatai láthatók:



9. ábra: Magyarország NUTS2 régióinak hálózati kapcsolatai 2021. 04. 01-én

A 2021. április 1. alapján megvizsgált kapcsolatok nem mutatnak jelentős eltérést a teljes időszakra vonatkozó dinamikához képest, ugyanazok a régiók számítanak terjesztőnek és fogadónak, mint amit a 7. ábrán látunk.

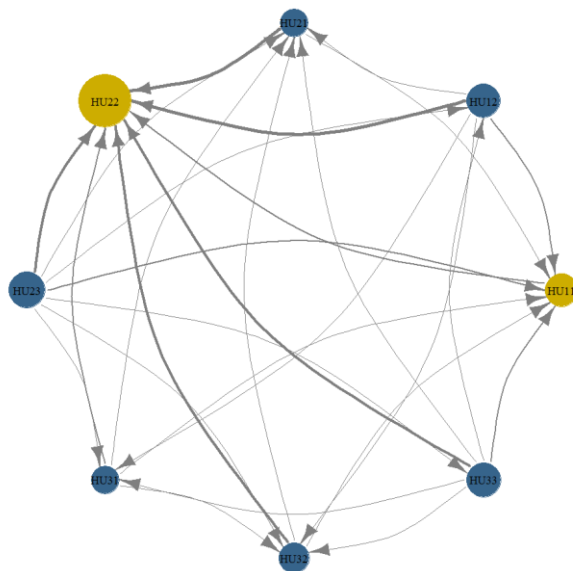
Megvizsgálom a „nyári szünetre” vonatkozó kapcsolatokat is:



10. ábra: Magyarország NUTS2 régióinak hálózati kapcsolatai 2021. 08. 01-én

Egyetlen különbséget detektáltam, miszerint a Dél-Dunántúl (HU23) és Észak-Magyarország (HU31) szerepet cseréltek, ebben az időszakban utóbbi inkább terjesztő, előbbi inkább fogadó.

Végül bemutatom a 4. hullámot is a következő diagramon:



11. ábra: Magyarország NUTS2 régióinak hálózati kapcsolatai 2021. 12. 01-én

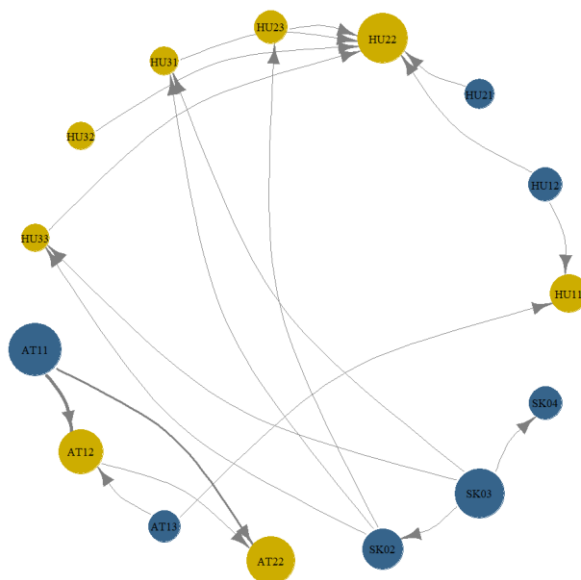
Ebben az esetben szignifikáns eltérést tapasztalunk a rendszer dinamikájában. Bár Nyugat-Dunántúl (HU22) szerepe változatlanul erősen fogadó, Az eddig inkább terjesztő Pest (HU11) régió vált a rendszer másik fogadjává. Ennek hátterében az állhat, hogy a negyedik hullámban Pest régióban a fertőzés már korábban lefutott, így a variancia inkább beérkező, mint kiinduló irányba mutat.

5.5.2. Magyarország és szomszédos régióinak hálózati kapcsolatai

Ebben az alfejezetben az előzőhöz hasonló módon fogom bemutatni az érdekesebb hálózati diagramokat, azzal a különbséggel, hogy a már korábban említett szomszédos régiókat is bevonom az elemzésbe. Mivel itt már jelentős mennyiségű régió szerepel, ez a vizuális módszer egyre nehezebben átlátható. Ezért egy olyan szűrést alkalmazok, hogy a diagramon csak egy bizonyos értéknél erősebb kapcsolatok jelenjenek meg. Így ugyan a kevésbé intenzív kapcsolatokat nem láthatjuk, viszont koncentrálhatunk a fontos kapcsolatokra. Ebben az

esetben 0,2-nek választottam meg ezt a küszöbértéket, ez azt jelenti, hogy a hálózatba kerülő nyilak csak akkor jelennek meg, ha a páronkénti spillover-hatás abszolút értéke legalább 0,2.

Először szintén a teljes időszakot leíró diagramot mutatom be (12. ábra):



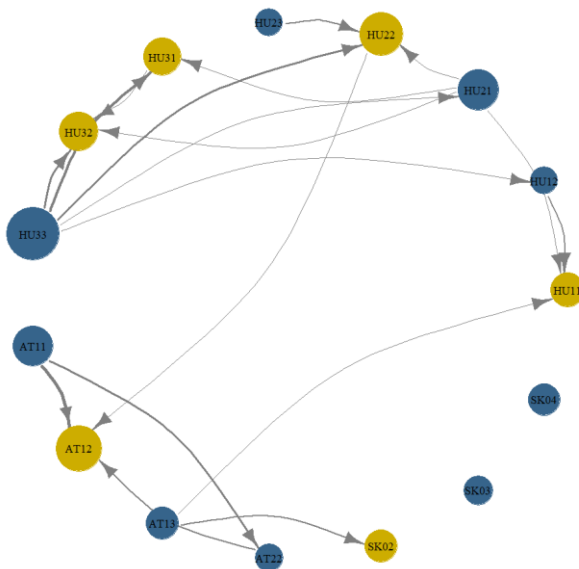
12. ábra: A magyarországi és szomszédos kiválasztott NUTS2 régióknak hálózati kapcsolatai a teljes vizsgált időszakban

A magyarországi belföldi kapcsolatokról már az előbbieken beszéltem, így most inkább a külföldi régiók hatásaira koncentrálnék. Természetesen Ausztriában és Szlovákiában is az ország régiói között látunk szignifikáns kapcsolatokat. A Szlovákia középső részét lefedő (SK03) régió terjesztőként fogható fel az ország keleti (SK04) és nyugati (SK02) régiói irányába. Ausztriában Burgenland (AT11) ad át sok variációt mind az ország észak-keleti (AT12), mind a dél-keleti (AT22) régióinak. Ennél érdekesebb az országokon átívelő kapcsolatok vizsgálata. Ausztrián belül Bécs (AT13) áll szorosabb kapcsolatban Magyarországgal, egész pontosan Budapesttel (HU11). A két főváros közül Bécsből Budapest felé terjed az információ, tehát megállapítható, hogy inkább az osztrák fővárosból jött be a járvány Budapestre a teljes vizsgált időszakban. Ez a két város közötti intenzív nemzetközi mobilitásnak (ingázás, turizmus, üzleti utazások stb.) tudható be, melyek esetében a Bécsből Budapestre irány volt a domináns.

A szlovák régiókkal is van kapcsolatunk, amelyben a magyar régiók vannak fogadó szerepben, így tehát az országba befelé mozgás volt Szlovákia felől a jelentősebb. A nyugat-szlovák

(SK02) és közép-szlovák (SK03) régiók állnak kapcsolatban több magyar régióval is, melyek inkább az ország keleti régiói.

A 2. hullám esetén a következőképpen alakult a rendszer (13. ábra):

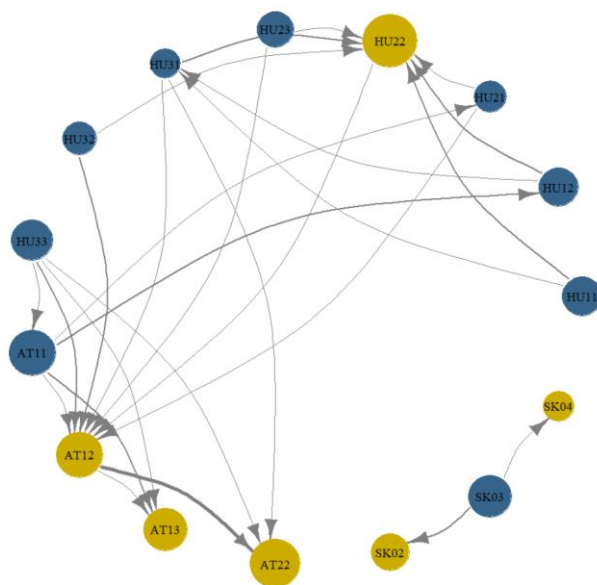


13. ábra: A magyarországi és szomszédos kiválasztott NUTS2 régióknak hálózati kapcsolatai 2020. 12. 01-én

Azt láthatjuk, hogy a rendszer az országok közötti kapcsolatok tekintetében sokkal zártabb, a szlovák régiókkal való szignifikáns kapcsolatok eltűntek. A Bécs-Budapest korábban megfigyelt kapcsolata továbbra is fennáll és az iránya is változatlan.

A harmadik hullám nagyon hasonló dinamikát mutat, a kapcsolatok gyengék és még a Bécs-Budapest információáramlás sem éri el a küszöbértéket (1. 2. melléklet).

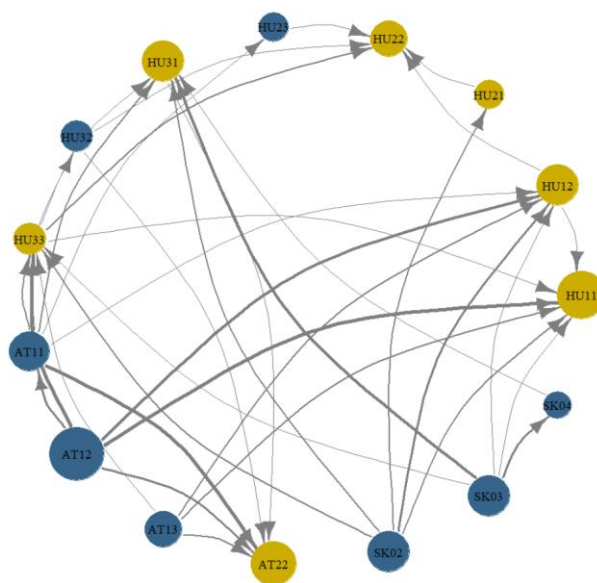
Ami ennél sokkal érdekesebb, az a 2021. „nyári szünetet” reprezentáló kapcsolati háló (14. ábra):



14. ábra: A magyarországi és szomszédos kiválasztott NUTS2 régióknak hálózati kapcsolatai 2021. 08. 01-én

A mobilitásra vonatkozó megszorítások csökkenésével egyértelműen megfigyelhető az országok közötti kapcsolatok erősödése. A szlovák régiók továbbra is zárt rendszert alkotnak, viszont Ausztria és Magyarország között számos szignifikáns információáramlás alakult 2021. augusztus 1-én. A színkódolásból láthatjuk, hogy az osztrák régiók nagyrésze a rendszer erős fogadó szereplőivé váltak. Ezzel szemben a már korábban is eltérően viselkedő Nyugat-Dunántúl (HU22) régió kívül mindegyik régió terjesztővé vált. Ez arra utal, hogy a magyarok ebben az időszakban sokat utaztak Ausztriába, akár üdülési, akár munka célzattal, mindenesetre az utazások egyértelműen Magyarországról Ausztriába irányultak.

Végül ebben a fejezetben is bemutatom a 4. hullámra vonatkozó diagramot (15. ábra):



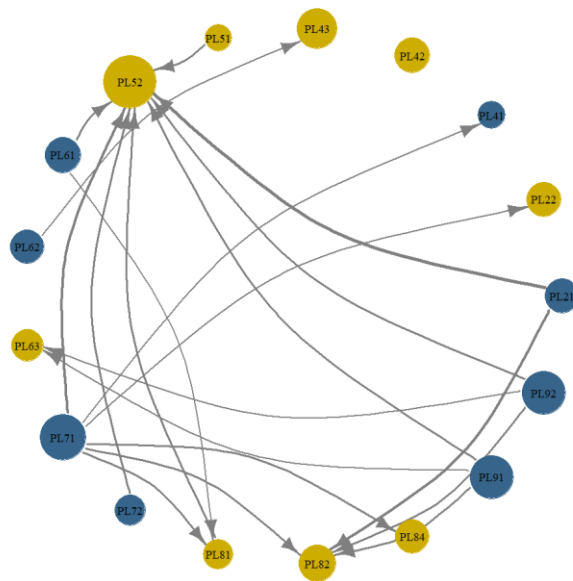
15. ábra: A magyarországi és szomszédos kiválasztott NUTS2 régióknak hálózati kapcsolatai 2021. 12. 01-én

Ebben a hullámban nagyon erőteljesek a kapcsolatok a teljes rendszerben. Erre magyarázat lehet, hogy ebben az időszakban történt a legtöbb fertőzés a rendszerben. Az irányok megfordultak a nyári időszakhoz képest Ausztria és Magyarország viszonylatában, már inkább külföldről érkezik a járvány hazánkba. Szlovákia is határozottan terjesztő, mely Ausztriával nem, de Magyarországgal erős kapcsolatokat mutat. Budapest (HU11), Pest (HU12), Nyugat-Dunántúl (HU22) és Észak-Magyarország (HU31) mind jelentős fogadók a rendszerben. Így a modellben az jelenik meg, hogy a variancia a vizsgált országokból érkezik be Magyarországra 2021 végén.

5.5.3. Lengyelország hálózati kapcsolatai

Lengyelországot elsősorban a mérete miatt választottam, egy nagyobb országban máshogy alakulhatnak a kapcsolatok, mint Magyarországon, illetve fontosabbnak bizonyulhat, hogy mely régiók helyezkednek el egymáshoz közel földrajzilag. Az előző modellhez hasonlóan az a probléma fennáll, hogy az összes kapcsolat ábrázolása esetén a diagram átláthatatlan. Így itt is választottam egy küszöbértéket, ez 0,5-re adódott.

Alább láthatjuk a teljes időszakra vonatkozó diagramot (16. ábra):

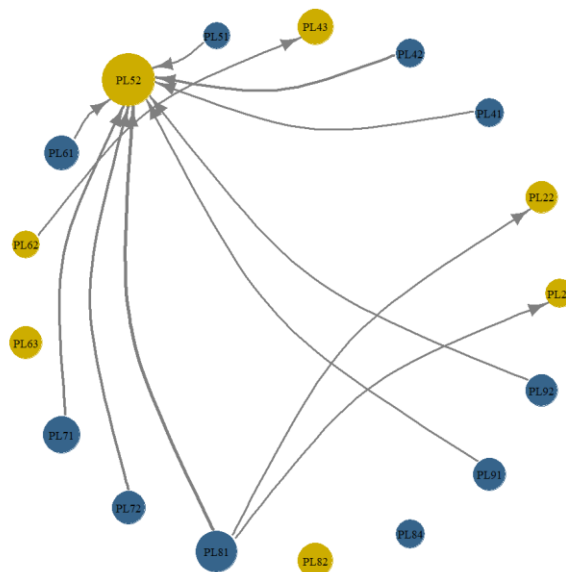


16. ábra: Lengyelország NUTS2 régióinak hálózati kapcsolatai a teljes vizsgált időszakban

Magyarország esetében is láthattuk, hogy a főváros a legnagyobb terjesztő, mindez megfigyelhető Lengyelország esetében is. Az ország kevésbé centralizált lakossági és gazdasági szempontból egyaránt. Az ország három legnagyobb városa a következő: Varsó (PL91), Krakkó (PL21) és Lodz (PL71). Ezzel tökéletesen összhangban van a hálózati diagram, ezen három régió a legnagyobb terjesztője a rendszernek. Látványos fogadó a PL52-es kódú Opolskie régió, mely az ország déli részén helyezkedik el. Ennek oka lehet, hogy mivel nem fontos központ ez a régió, így nem generál sok varianciát, viszont a déli utazások folyosójaként szolgálhat, így befogadhat információkat. A 0,5-ös küszöbérték mellett is sok kapcsolatot mutat a diagram, mely egy összekapcsoltabb rendszerre utal.

A második és a harmadik hullám hálózati dinamikája nagyon hasonló Lengyelországban, mint a Magyarország és szomszédos régióinak modellje, így ezt külön nem részletezném (l. 3. melléklet és 4. melléklet). A szerepek érdemben nem változnak ezekben a hullámokban, csak a kapcsolatok száma és erőssége, a második hullám itt is a leginkább zárt rendszer.

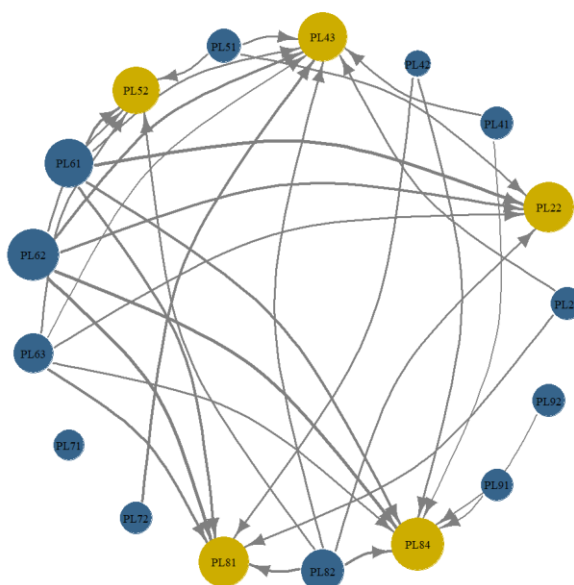
A „nyári szünet” időszakára vonatkozó diagram alább látható (17. ábra):



17. ábra: Lengyelország NUTS2 régióinak hálózati kapcsolatai 2021. 08. 01-én

Különbséggként megemlíthető, hogy a PL81-es kódú Lubelskie kelet-lengyelországi régió a legnagyobb terjesztővé vált. Ez akár a kelet felől érkező turizmusnak és utazásoknak tudható be. Emellett Opolskie (PL52) továbbra is a legfontosabb gócpont Lengyelországban 2021. augusztus 1-én is.

A 4. hullám Lengyelországban is sokkal erősebb kapcsolatokat mutat, ez az alábbi ábrán látható:

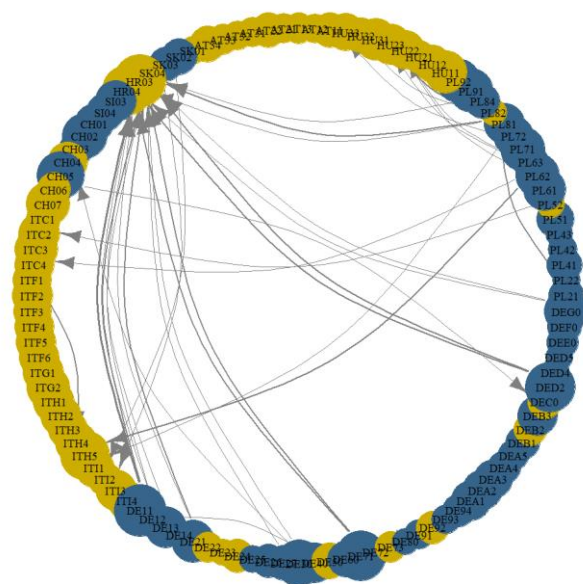


18. ábra: Lengyelország NUTS2 régióinak hálózati kapcsolatai 2021. 12. 01-én

A megszokott Opolskie (PL52) mellett Slaskie (PL22), Lubuskie (PL43), Podlaskie (PL84) és Lubelskie (PL81) váltak a rendszer varianciájának befogadóivá. Ezek rendre határmenti régiók, főleg az ország déli és keleti részein. A két legfontosabb terjesztő a Kujawsko-Pomorskie (PL61) és Warminsko-Mazurskie (PL62), melyek az északi határi régiók. Ezek alapján tehát egy alapvetően északról dél-keletre menő terjedési irányt állapíthatunk meg Lengyelországon belül a 2021-es év végén.

5.5.4. A teljes modell hálózati kapcsolatai

A mind a 108 kelet-európai régiót tartalmazó modell esetében már a hálózati diagram nagyon nehezen értelmezhető a sok csomópont és a még több kapcsolat miatt. Így ebben az esetben már csak a teljes vizsgált időszakra vonatkozó modellt mutatom be (19. ábra). Itt is az irányított élek átláthatósága érdekében egy 0,3-as küszöbértéket határoztam meg:



19. ábra: A teljes modell 107 db NUTS2 régiójának hálózati kapcsolatai a teljes vizsgált időszakban

Az áttekinthetőségi problémák mellett azért láthatjuk a dinamikákat. Az olaszországi, magyarországi és osztrák régiók sárga színe azok befogadó jellegére utal, míg Németország, Lengyelország és Svájc inkább a járvány terjesztésében játszanak szerepet. Ha az országokon átívelő kapcsolatokat nézzük, Horvátország és Szlovákia egyes régiói nagyon sok külföldi (elsődlegesen német és lengyel) régióktól kapnak információt. Emellett Lengyelország-

Olaszország, Németország-Svájc és Lengyelország-Magyarország irányú kapcsolatokat is megfigyelhetünk. Horvátország és Olaszország turizmus szempontjából kiemelkedik, ez is magyarázhatja azok több külföldi régióval szembeni befogadó szerepét.

6. Összefoglalás

A vizsgálat során különböző elemzések segítségével sikerült feltárnom összefüggéseket a COVID-19 járvány kelet-közép-európai terjedésében. A kutatásom megerősítette azt a hipotézisemet, hogy a kiválasztott országok NUTS2 szintű régióira hálózati modell építhető, mely képes feltárni a rendszer komplexitását és dinamikáit. Az eredményekből megállapítható, hogy Magyarország, Ausztria, Szlovákia és Lengyelország régiói között stabil és erős információáramlási kapcsolatok léteznek, amelyek a járvány alakulásának különböző szakaszaiban eltérő mintázatokat mutatnak. Emellett sikerült azt is kimutatni, hogy a járvány hullámai, a mobilitási korlátozások és a fertőzési dinamika között összefüggés áll fenn. A szigorúbb intézkedésekkel jellemezhető időszakok jelentősen csökkentik a hálózat összekapcsoltságát, míg a korlátozások feloldása erőteljesebb országokon átívelő terjedést eredményez a vizsgált régiókban.

Megfigyelhető továbbá, hogy a különböző fővárosok vagy nagyobb centrumok (például Budapest, Bécs, Varsó vagy Krakkó) a rendszer legfontosabb terjesztő régiói, mind a belföldi, mind pedig a nemzetközi vonatkozásban. Ezek a központok koncentrálják a mobilitást, a gazdasági aktivitást és a mindennapi személyes interakciókat, így természetes módon válhatnak a varianciaátadás gócpontjaivá. Ugyanakkor a vizsgálat feltárta azokat a kevésbé intuitív jelenségeket is, mint például a magyarországi Nyugat-Dunántúl vagy a lengyelországi Opolskie tartósan fogadó szerepe, melyek jelentős közlekedési csomópontban lévő elhelyezkedésüknek köszönhetően válnak a fertőzések fő fogadóivá. Ezen kívül a negyedik hullám során megfigyelhetőek irányváltások az Ausztria–Magyarország és Szlovákia–Magyarország relációkban a nyári időszakhoz képest, amely a nyaralások járványterjedési mintázatainak hatásaira világít rá. Ezekből az eredményekből következtethetünk arra is, hogy a járvány természete nem statikus, így a térbeli dimenzió mellett a régiók kapcsolatainak időbeli elemzése is kritikus része az epidemiológiai elemzéseknek.

Összességében elmondható, hogy a vizsgálatom hozzájárul a járványok regionális terjedési mechanizmusainak mélyebb megértéséhez. Az alkalmazott módszertan alkalmas arra, hogy a jövőben más fertőző betegségek, mobilitási jelenségek vagy gazdasági sokkok terjedési pályáit is elemezze. A dolgozatom eredményei nemcsak elméleti szempontból relevánsak, hanem hozzájárulhatnak a járványügyi felkészüléshez, határmenti együttműködések tervezéséhez, valamint különböző döntéstámogató rendszerek kialakításához. A kutatásom során szerzett

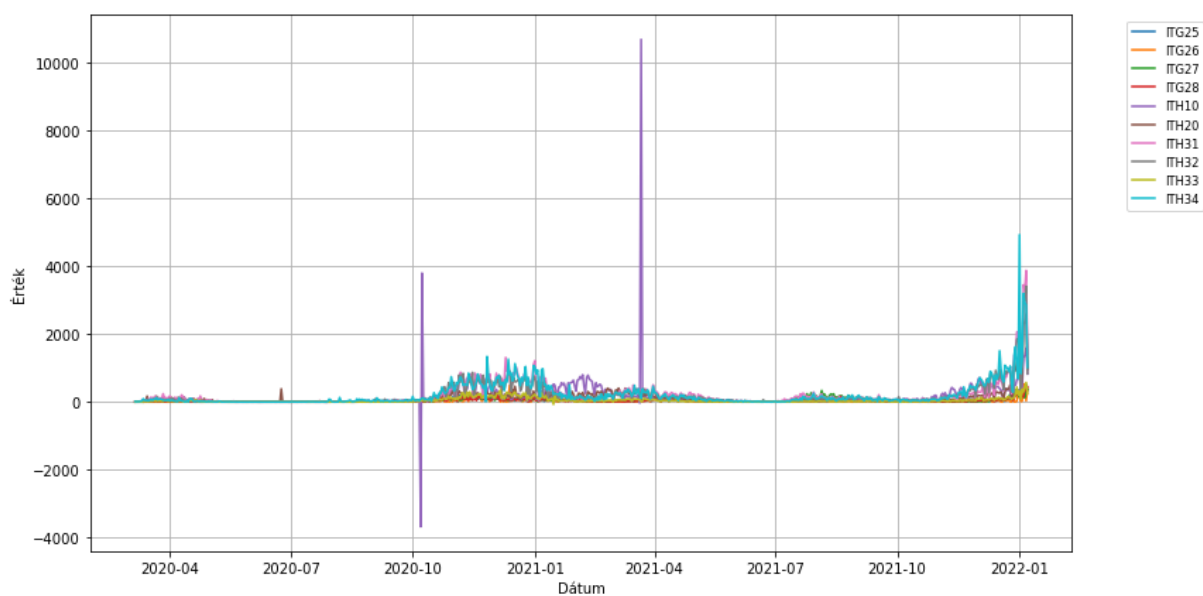
megállapítások egyúttal alapot biztosítanak további kutatások számára is, különösen a hálózat időbeli instabilitásának, a mobilitási mintázatok mélyebb rétegű bevonásának és a finomabb térbeli felbontású modellek alkalmazásának irányába.

7. Hivatkozásjegyzék

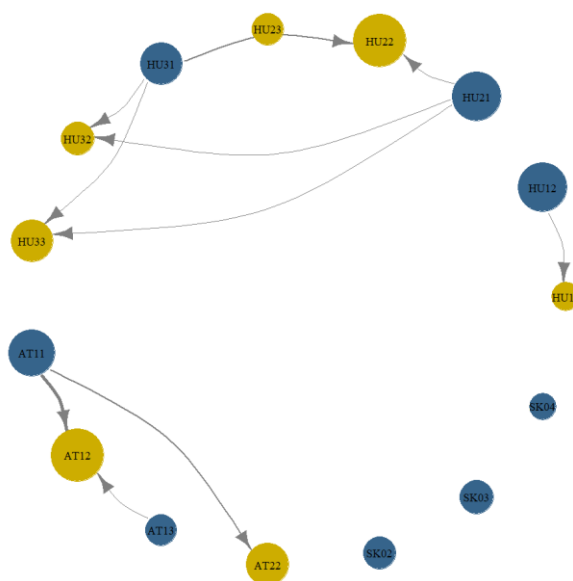
- 1) De-Leon, H., & Aran, D. (2023). *MAM: Flexible Monte-Carlo Agent based model for modelling COVID-19 spread*. *Journal of Biomedical Informatics*, 141, 104364. <https://doi.org/10.1016/j.jbi.2023.104364>
- 2) Rafiq, R., Ahmed, T., & Uddin, M. Y. S. (2022). Structural modelling of COVID-19 spread in relation to human mobility. *Transportation Research Interdisciplinary Perspectives*, 13, 100528. <https://doi.org/10.1016/j.trip.2021.100528>
- 3) Hu, B., et al. (2021). Modeling the complete spatiotemporal spread of COVID-19 in the world. *International Journal of Infectious Diseases*, 110. <https://doi.org/10.1016/j.ijid.2021.07.045>
- 4) He, J., et al. (2025). Modeling the spatiotemporal transmission of COVID-19 epidemic by coupling the heterogeneous impact of detection rates: A case study in Hong Kong. *Health & Place*, 92, 103422. <https://doi.org/10.1016/j.healthplace.2025.103422>
- 5) Salcido, A. (2020). A lattice gas model for infection spreading: Application to the COVID-19 pandemic in the Mexico City Metropolitan Area. *Results in Physics*, 20, 103758. <https://doi.org/10.1016/j.rinp.2020.103758>
- 6) Sasaki, K., et al. (2025). Quantifying the regional disproportionality of COVID-19 spread: Modeling study. *JMIR Formative Research*, 9, e59230. <https://doi.org/10.2196/59230>
- 7) Utku, A. (2023). Deep learning based hybrid prediction model for predicting the spread of COVID-19 in the world's most populous countries. *Expert Systems with Applications*, 231, 120769. <https://doi.org/10.1016/j.eswa.2023.120769>
- 8) Maltezos, S. (2021). Methodology for modelling the new COVID-19 pandemic spread and implementation to European countries. *Infection, Genetics and Evolution*, 91, 104817. <https://doi.org/10.1016/j.meegid.2021.104817>
- 9) Bouzouina, L., et al. (2022). Impact of immobility and mobility activities on the spread of COVID-19: Evidence from European countries. *Regional Science Policy & Practice*, 14(S1), 6–21. <https://doi.org/10.1111/rsp3.12565>
- 10) Sun, J., & Sathasivam, S. (2023). Nonlinear dynamic model for COVID-19 epidemics using the Gaussian distributed wiring small-world network technique. *Alexandria Engineering Journal*, 81(15 October), 243–255. <https://doi.org/10.1016/j.aej.2023.07.039>
- 11) Naqvi, A. (2022). *COVID19-European-Regional-Tracker* [GitHub repository]. GitHub. <https://github.com/asjadnaqvi/COVID19-European-Regional-Tracker>

- 12) Corcoran, C., & Clark, J. M. (2022). Adaptive network modeling of social distancing interventions. *Journal of Theoretical Biology*, 546, 111151. <https://doi.org/10.1016/j.jtbi.2022.111151>
- 13) Hamilton, J. D. (2020). *Time series analysis*. Princeton University Press.
- 14) Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S*. Chapman & Hall.
- 15) Hunyadi, L., & Vita, L. (2002). *Statisztika közgazdászoknak*. Központi Statisztikai Hivatal.
- 16) Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1), 119–147. [https://doi.org/10.1016/0304-4076\(95\)01753-4](https://doi.org/10.1016/0304-4076(95)01753-4)
- 17) Han, F., *et al.* (2015). A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16(97), 3115–3150.
- 18) Fraley, C., & Hesterberg, T. C. (2009). Least-Angle Regression and LASSO for Large Datasets. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 1(4), 251–259. <https://doi.org/10.1002/sam.10021>
- 19) Pesaran, M. H. (2015). Impulse response analysis. In *Time Series and Panel Data Econometrics* (pp. 584–608). Oxford University Press.
- 20) Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.
- 21) Sayed, R. (2020). *Measuring Volatility Spillovers and Connectedness*. RPubS. <https://rpubs.com/rsayed/573439>
- 22) Diebold, F. X., & Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *[Working Paper]*. <https://www.sas.upenn.edu/~fdiebold/papers/paper75/DY2final.pdf>
- 23) Diebold, F. X., & Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1), 57–66. <https://doi.org/10.1016/j.ijforecast.2011.02.006>
- 24) Diebold, F. X., & Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1), 119–134. <https://doi.org/10.1016/j.jeconom.2014.04.012>
- 25) Demirer, M., Diebold, F. X., Liu, L., & Yilmaz, K. (2018). Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1), 1–15. <https://doi.org/10.1002/jae.2585>

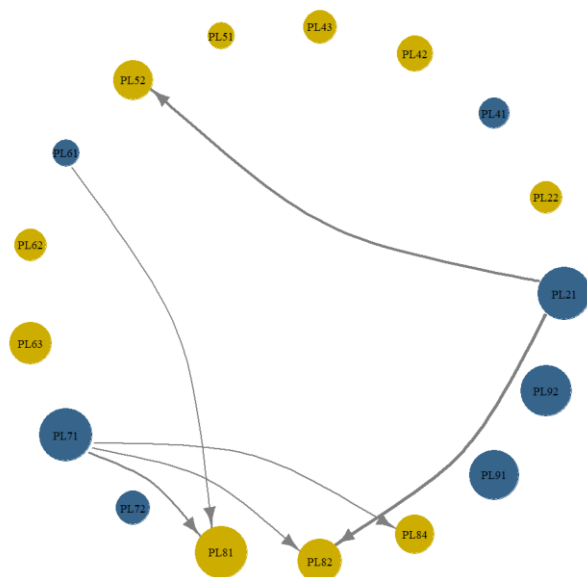
8. Mellékletek



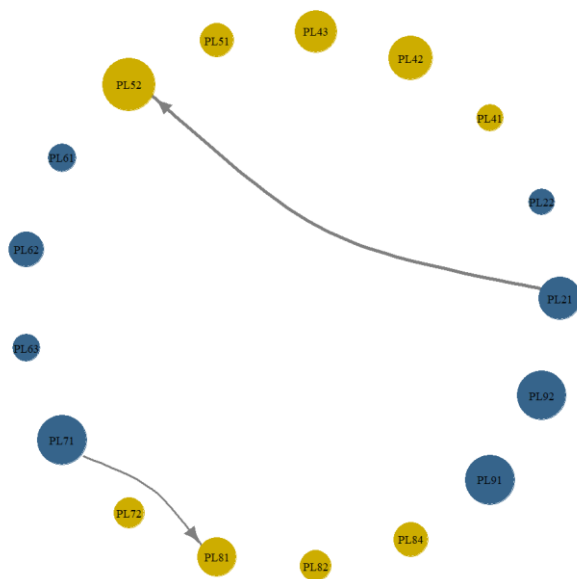
1. melléklet: Például 10 olaszországi régió időszora, melyek között negatív érték is szerepel



2. melléklet: A magyarországi és szomszédos kiválasztott NUTS2 régióknak hálózati kapcsolatai 2021. 04. 01-én



3. melléklet: Lengyelország NUTS2 régióinak hálózati kapcsolatai 2020. 12. 01-én



4. melléklet: Lengyelország NUTS2 régióinak hálózati kapcsolatai 2021. 04. 01-én