

# Kontaktkutató, vezetői információs rendszer

Bokányi Eszter<sup>1,2</sup>, Pollner Péter<sup>3,4\*</sup>, Joó Tamás<sup>4</sup>

<sup>1</sup>NETI Lab, CIAS, Budapesti Corvinus Egyetem, Budapest, Magyarország

<sup>2</sup>ANET Lendület Csoport, KRTK, Budapest, Magyarország

<sup>3</sup>MTA-ELTE Statistikus és Biológiai Fizika Kutatócsoport, ELKH, Biológiai Fizika Tanszék, ELTE, Budapest, Magyarország

<sup>4</sup>Egészségügyi Menedzserképző Központ, Semmelweis Egyetem, Budapest, Magyarország

Beérkezett: 2021. 05. 08.; Elfogadva: 2021. 06. 10.

## Összefoglalás

Tanulmányunkban bemutatjuk a hazai COVID-járvány első hulláma során kidolgozott informatikai megoldást, amely a kontaktkutató hálózattudományi megközelítés alapján segítette, és hozzájárult az első hullám sikeres megfékezéséhez.

A kifejlesztett vizuális reprezentációs technika látványos és részletekbe menő megértést, problémafeltárást képes biztosítani a járványügyi szakemberek számára. A grafikus elemek segítenek a gyors megértésben, a különböző hálózati elrendezések bizonyos jelenségekre, például gócpontokra, fertőzési klikkekre vagy a földrajzi terjedésre irányítják a figyelmet. A böngészőből történő futtatás alacsony technológiai belépési küszöböt biztosít a társterületeken kutatók számára, nekik így nem szükséges a problémafeltáráshoz külön szoftvereket telepíteni. Az adatbázis SQL-alapú szűrése a vizualizációs felületről lehetőséget biztosít összetettebb kérdések megfogalmazására is.

**Kulcsszavak:** kontaktkutató, hálózatelemzés, döntéstámogatás, COVID-19

## Contact tracing, management support system

Eszter Bokányi<sup>1,2</sup>, Péter Pollner<sup>3,4</sup>, Tamás Joó<sup>4</sup>

<sup>1</sup>NETI Lab, CIAS, Corvinus University, Budapest, Hungary

<sup>2</sup>Agglomeration and Social Networks, MTA Lendület Research Group Centre for Economic and Regional Studies, Budapest, Hungary

<sup>3</sup>MTA-ELTE Statistical and Biological Physics Research Group, ELKH, Dept. of Biological Physics, ELTE, Budapest, Hungary

<sup>4</sup>Health Services Management Training Centre, SE, Budapest, Hungary

## Summary

In our study, we present an IT solution developed during the first wave of the domestic COVID epidemic. This tool served as an aid for contact tracing. The development focused on the network scientific aspects and contributed to the successful handling of the first wave.

In case of absence of effective drugs or vaccines, controlling a contagious disease can only be achieved by preventing its spread. To this end

- infectious individuals must be identified,
- patients, exposed to the infection must be identified,
- the epidemic branching points that cause the greatest infection must be uncovered,
- information on the course of the disease must be collected,
- temporal and efficacy parameters must be determined, and
- potential cases of infection must be described.

One possible way to accomplish these tasks is achieved by contact tracing. Classical contact tracing is carried out by personal data collection, during which the commissioned epidemiologist has to fill in a questionnaire. The questionnaire basically includes data used to identify the infected person, as well as the data of the persons who were in contact with the infected person, i.e. in contact with them. The effectiveness of the research is also enhanced if the

questionnaire records disease-related parameters (e.g., symptoms, timing-related times, etc.) as well. Once the disease is known, questionnaires can be designed according to a definite template format, the organization of data collection groups and the associated costs can be planned in advance. However, in the case of a new, unknown disease, flexibility and the ability to adapt quickly during data collection are of paramount importance.

The developed visual representation technique is able to provide spectacular and detailed understanding and a problem-solving user interface for epidemiologists. Graphical elements help in quick understanding, different network layouts can direct the attention to certain phenomena such as focal points, infectious cliques, or geographical spreading patterns. Running from a browser provides a low technology entry threshold for researchers in other scientific fields, so they don't need to install separate software. The SQL-based filtering of the database on the visualization interface also provides an opportunity to study more complex questions.

Thus, with the help of the presented computer system, a relational database can be generated from the initially unstructured data of the contact research protocols through several steps. The relational database is made available to analysts and decision-makers.

As the final balance of the first wave of COVID-19 in Hungary showed, data from well-organized contact research and processed in appropriate analytical tools can provide important information for controlling the epidemic and saving lives.

**Keywords:** contact tracing, network analysis, decision support, COVID-19

## Bevezető

Az emberi civilizáció fejlődésével számos esetben alakult ki olyan körülmények, amelyekben az egymáshoz közel élők között valamilyen betegség terjedt el. A nagy kiterjedésű, monokulturális növényi ültetvények vagy a nagy tömegben tartott háziállatok kártevői és betegségei már évezredek problémák. Az egymással szoros kontaktusban élő emberek között is előfordultak járványok. Ezek mérete, terjedési sebessége, a fertőzések által okozott emberi és gazdasági veszteségek a gyors közlekedés és a hálózatos közösségi kapcsolatok kiépülésével napjainkban váltak először a teljes Földre kiterjedő, globális problémává.

2019-ben egy új típusú koronavírus terjedt el, amelynek pontos hatásmechanizmusa, terjedésének módja és hatékonysága ismeretlen volt az egészségügyi hatóságok számára. Először Kínából jelentettek tömeges megbetegedéseket, így mire Európába ért a járvány, már részben ismert volt, hogy légúti fertőzéssel terjed emberről emberre kontaktusok útján. A közös légtérben tartózkodást azonosították elsődleges fertőzési lehetőségként, de a tárgyak felületei, illetve az ételek és italok is felmerültek potenciális fertőző közegeként. A járvány okozta megbetegedés olyan komoly tünetekkel és – különösen a kezelt betegek esetén – olyan magas halálozási rátával járt, hogy azonnal fel kellett lépni a megfékezése érdekében.

Ahogy egyre több tapasztalat gyűlt a vírussal kapcsolatban, úgy vált egyértelművé, hogy a betegség gyógyítása reménytelenül nagy terhelést okozna az egészségügyi ellátórendszer számára. Vakcinák hiányában a fertőző betegségekre évezredek óta bevált módszert lehetett alkalmazni: meg kellett szüntetni a fertőzőképes személyekkel azokat a kapcsolatokat, amelyek során át tudták adni a betegséget. Ennek a kapcsolatmegszüntető módszernek az alapvető eszköze a kontaktuskutatás.

## A kontaktuskutatás célja és eszközei

Egy ragályos betegség megfékezése hatékony gyógyszerek vagy oltások hiányában csak a terjedés megakadályozásával történhet. Ennek érdekében

- meg kell találni a fertőző személyeket,
- azonosítani kell a fertőzésnek kitett egyéneket,
- fel kell kutatni azokat a járványgócokat, amelyek a legnagyobb fertőzést okozzák,
- információt kell gyűjteni a betegség lefolyásáról,
- meg kell határozni a fertőzőképesség időbeli és hatékonysági paramétereit,
- minimalizálni kell a potenciális fertőzési alkalmakat.

Ezeknek a feladatoknak az elvégzésére egy lehetséges módszer a kontaktuskutatás. A klasszikus kontaktuskutatás személyes adatfelvétellel történik, melynek során az erre megbízott járványügyi biztosnak egy kérdőívet kell kitöltenie. A kérdőív alapvetően a fertőzött személy azonosítására szolgáló adatokat veszi fel, illetve azoknak a személyeknek az adatait, akik a fertőzés átadására lehetőséget teremtő kapcsolatban, azaz kontaktusban voltak a fertőzött személlyel. A kutatás eredményességét növeli, ha a kérdőív rögzíti a betegséggel kapcsolatos paramétereket (pl. tünetek, a lefolyáshoz kapcsolódó időpontok stb.) is. Amikor egy betegség már ismert, a kérdőíveket határozott sablonforma szerint lehet alakítani (*World Health Organization 2014*), az adatfelvételt végző csoportok szervezését és a kapcsolódó költségeket előre lehet tervezni. Egy új, ismeretlen betegség esetén azonban kiemelkedően fontos a rugalmasság és gyors alkalmazkodási képesség az adatfelvétel során (*Koetter et al. 2020*).

Tanulmányunkban bemutatjuk a hazai COVID-járvány első hulláma során kidolgozott informatikai megoldást, amely a kontaktuskutatást hálózattudományi megközelítés alapján segítette, és hozzájárult az első hullám sikeres megfékezéséhez.

## Adatok és módszerek

A kontaktkutató során nagy mennyiségben rögzülnek személyes adatok. Az ezek kezelésére jogosított civil személyek száma igen korlátozott, ezért az alapvető adatokat a rendőrség segítségével hatósági jegyzőkönyvek formájában rögzítették. Ez a szabad szöveges formátum maximális rugalmasságot biztosított a pandémia első hulláma során, amikor a megoldandó feladatok között szerepelt a rögzítendő adatok körének meghatározása is. Ezeket a szöveges dokumentumokat adatrögzítők dolgozták át részben strukturált, táblázatos formába. Egy táblázatkezelőbeni nézetben minden munkalap egy fertőzött személy kontaktkutatói eredményeit tartalmazta.

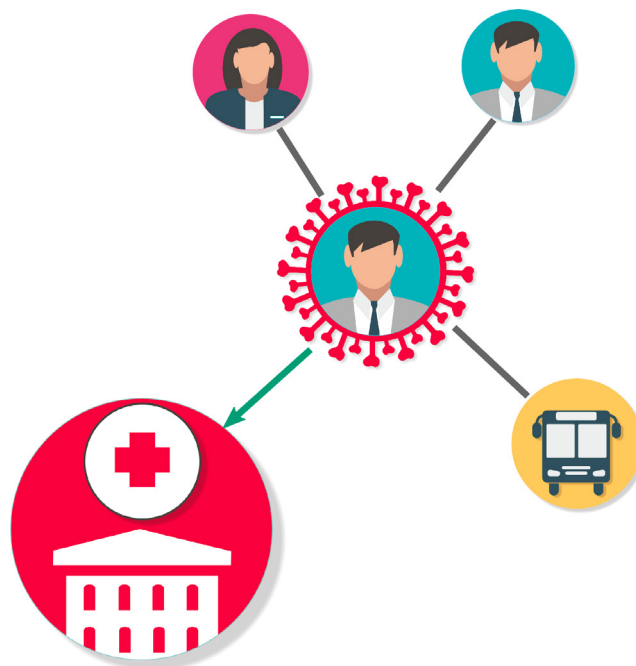
Az első szöveges dokumentumok feldolgozását a táblázatos sablon kialakítására használta az adatrögzítő csoport. Több különböző verzió tesztelése után véglegesítették a táblázat struktúráját és a benne szereplő adatokat. Az informatikai fejlesztés során külön kihívás volt a személyes adatok folyamatos védelme, azaz annak biztosítása, hogy a pandémiás körülmények között a fejlesztés számára elérhetőek legyenek megfelelő, tesztelésre alkalmas példák, de a valódi adatok a kezelésükre jogosultak körén túl ne legyenek megismerhetők.

## Az adatbázis felépítése

A munkalapon a következő információkat tüntették fel:

- a pozitív teszttel rendelkező *fertőzött* személyt,
- az elmúlt két héttel *kontaktszemélyeket*,
- a fertőzött személyhez köthető *intézmények*, helyszínek közül a három legfontosabbat (pl. kórház, idősotthon, bolt, iskola), illetve
- a fertőzött személy által igénybe vett *közlekedési eszközök* közül a három legfontosabbat (pl. ingázáshoz használt busz, távolsági vonatjárat).

A négy különböző kategóriához soronként szerepeltek az azonosításra szolgáló alapadatok (pl. név, születési dátum, lakcím, intézmény neve, címe), illetve egyéb, a fertőzés szempontjából releváns adatok (tünetek, külföldi utazás stb.). A táblázatkezelő-munkalap struktúrája az adatrögzítési periódus során – főleg az első időszakban – többször megváltozott a feldolgozó visszajelzések, illetve újonnan felmerülő járványügyi szempontok alapján. Ez részben azt mutatja, hogy egy ilyen technikájú rögzítésben nagyon könnyű alakítani a szerkezetet, másrészt a további adatfeldolgozás során külön nyomon kell követni a lehetséges elrendezéseket, adattípusokat, a kérdések változását. Az 1–2. táblázat tartalmaz egy fiktív adatokkal kitöltött lehetséges táblázatkezelő munkalapot egy lehetséges fejléccel együtt. A szabad szöveges állományokból kivonatolt táblázatok csak részben strukturált adatgyűjtésnek nevezhetjük, mert az egyes cellák tartalma általános szöveges információ volt. Azaz adattörzsek, kapcsolati kódok nem voltak használatban a feldolgozásnak ebben a szakaszában.



1. ábra

Az 1–2. táblázatban szereplő táblázatkezelő munkalapból, illetve a 3. táblázatban szereplő adatbázistábla-részletekből készített példahálózat, melyben az ikonok jelzik a csúcs típusát (férfi/nő, fertőzött, közlekedési eszköz, kórház). Az irányított élen szerepel nyíl is, az irányítatlan él nyílhegy nélküli. Az él kék színe jelzi, hogy a személy fertőzöttként került kapcsolatba az intézménnyel.

Az 1. ábrán látható az 1–3. táblázatok fiktív adataiból készített hálózati reprezentáció, melyben 5 csúcs (3 ember, 1 intézmény, 1 közlekedés) szerepel 4 éllel. Minden táblázatkezelő munkalap egy-egy ilyen csillag alakú hálózatnak felel meg, melynek közepén van a fertőzött. Az adatfeldolgozás során ezek a munkalapok külön rétegekben rögzültek, mert így utólag is követhető, hogy egy-egy adatgyűjtőhöz mennyi személyes információ kerül, illetve az adatrögzítési és adattisztítási lépések elválaszthatók egymástól (*WHO Ebola Response Team 2014*).

A teljes értékű elemzéshez ezeket a diszjunkt hálózatokat a különböző táblázatkezelő munkalapokon szereplő azonos objektumok összekapcsolásával a feldolgozás következő rétegeiben lehetett összefűzni.

Itt említjük, hogy az adattisztítási folyamat már ebben a lépésben elkezdődött. A fertőzött, illetve a kontaktszemélyek nemét az adatrögzítés során nem vették fel. Ezért ezt az információt a nevek alapján kellett meghatározni. Magyar nevek esetén vagy a férjezett névből, vagy az utónévből lehetett következtetni erre az adatra. Az utónévek nemekhez rendelése az MTA Nyelvtudományi Intézete által anyakönyvi bejegyzésre alkalmasnak minősített utónévek jegyzéke alapján, algoritmikus meghatározással történt (*MTA Nyelvtudományi Intézet 2021*). Külföldi, illetve elgépelte nevek esetén emberi adatrögzítők bevonására volt szükség.

Az adatrögzítési folyamat következő lépésében egy SQLite relációs adatbázisba kerültek az információk

(SQLite Database 2021). Ennek sémáját az táblázatkezelő munkalapok struktúrája határozta meg. Az adatbázis-motort a várható adatmennyiség és a könnyű hordozhatóság szempontja alapján választottuk.

Az adatbázisban egy táblába (*nodes*) került a következő három típusú adat:

- személy típusú adat (*fertőzött és kontaktszemély*),
- intézményi adat,
- közlekedési adat.

Az adatbetöltés során minden táblázatkezelő munkalap minden rekordja egyedi azonosítót kapott, és az adat

típusa is rögzítésre került. Ebben a táblában szereplő rekordok alkotják a kontaktusok hálózatában a csúcsokat.

A rekordok közötti kapcsolatokat, a hálózat éleit, az *edges* nevű tábla tárolta. Minden táblázatkezelő fájlban a fertőzött személy és a kontaktszemélyei, illetve a fertőzött személy és az intézmények, valamint a fertőzött személy és a közlekedési eszközök között jött létre egy-egy irányított él. Az irányítást a későbbiekben a „Felhasználói felület komplex lekérdezésekhez és vizualizációhoz” fejezetben írjuk le részletesen.

1. táblázat | Táblázatkezelő munkalap fertőzött és kontaktdatokra vonatkozó fejléce mintaadatokkal kitöltve

Dátum: 2020. 04. 10.			
COVID19 pozitív személy kontaktjai			
E-mail érkezésének időpontja: 2020. 04. 12. 18:20	Fertőzött	Kontakt személy	Kontakt személy
Név:	Kovács János	Kovács Jánosné	Kovács Márton
Születési hely:	Zalaegerszeg	Szeged	Budapest
Születési dátum:	15/08/1940	01/05/1946	13/04/1976
Édesanyja leánykori neve:	Tóth Mária	Nagy Katalin	Kiss Jolán
Állandó lakcím:	8900 Zalaegerszeg, Kossuth u. 2.	8900 Zalaegerszeg, Kossuth u. 2.	1097 Budapest, Tóth Kálmán u. 4.
Tartózkodási hely:	1094 Budapest, Ferenc krt. 33.	1094 Budapest, Ferenc krt. 33.	1097 Budapest, Tóth Kálmán u. 4.
Telefonszám (pl. xx/xxxxxxx):		20/1111222, 92/311222	
Foglalkozás:		nyugdíjas	pedagógus
Járt-e külföldön az elmúlt 14 napban?	nem	nem	nem
Ha igen, hol, mettől-meddig? (Ország, város)		nem	nem
Ha igen, mikor érkezett Magyarországra		nem	nem
Érintkezés dátuma:		folyamatos	múlt héten egy nap
Egy háztartásban él a Fertőzött személlyel (igen/nem)		igen	igen
Közvetlen fizikai kapcsolat volt-e (igen/nem)		igen	igen
Védőeszköz használata nélküli kontaktus volt-e (igen/nem)		igen	igen
Személyes kapcsolat volt-e (igen/nem) (2 m-en belül 15 percnél hosszabb ideig)		igen	igen
Zárt légtérben együtt tartózkodott? (igen/nem) (2 m-en belül 15 percnél hosszabb ideig)		igen	igen
Légúti tünetei vannak? (láz, köhögés, nehézlégzés)		nem	nem
Egyéb helyszínek:			
Egészségügyi intézményben járt-e az elmúlt 14 napban? (igen/nem)	igen	nem	nem
Ha igen, hol, milyen rendelésen?		nem	nem
Kapcsolat típusa (férje, felesége, lánya, fia, szomszédja, munkatársa, fia felesége stb.)		felesége	fia
Munkahely neve, címe (ha van)			
Krónikus betegség (ha van)	vesebetegség		asztma
Kár korábban igazolt fertőzött – csak igen válasz esetén kitöltendő			

A 3. táblázat mutatja, hogy egy példátáblázatból milyen adatbázisrekordok (csúcsok és élek) készültek a *nodes* és az *edges* táblában, illetve az 1. ábrán illusztráljuk a csillag alakú irányított hálózatot, amely a táblában szereplő relációkat vizualizálta.

2. táblázat | Táblázatkezelő munkalap intézményi és közlekedési adatokra vonatkozó része mintaadatokkal

Intézményi adatok	Eset intézmény
intézmény jellege	kórház
intézmény neve	Szent Rafael Kórház; 8900 Zalaegerszeg, Zrínyi u. 1.
osztály, csak kórház esetén	
kapcsolat jellege	betegség (kezelés esetén)
egyéb (időpont stb.)	hétfőn, szerdán és pénteken dialízis; 2020. 03. 08. óra kórházban
milyen panasszal került be, csak kórház esetén	nehézlégzés

Közlekedési adatok	Eset közlekedés
típus	autóbusszal
időpont	08/04/2020
egyéb	

3. táblázat | Az 1–2. táblázat néhány kiválasztott adata a relációs adatbázis csúcsokat (*nodes*), illetve éleket (*edges*) tartalmazó táblájában. A *nodes.global\_id* nevű oszlop tartalmazza a csúcsok elsődleges azonosítóját. Az *edges.source*, illetve az *edges.target* oszlopok is ezt az azonosítót tartalmazzák, így a hálózat állítójaként lehet erre a táblára tekinteni. Az *edges.directed* oszlop megadja, hogy az él irányítható-e (TRUE: van irányításhoz információ az élen, FALSE: nincs irányításhoz információ az élen). A *nodes.global\_id* csúcsazonosító első karaktere megadja a csúcs típusát (P – person: fertőzött személy / kontaktszemély, I – institute: intézmény, T – transport: közlekedés)

nodes		
global_id	name	infected
P_00001	Kovács János	TRUE
P_00002	Kovács Jánosné	FALSE
P_00003	Kovács Márton	FALSE
I_00001	Szent Rafael Kórház; 8900 Zalaegerszeg, Zrínyi u. 1.	None
T_00001	autóbusz	None

edges			
source	target	type	directed
P_00001	P_00002	ember-ember	FALSE
P_00001	P_00003	ember-ember	FALSE
P_00001	I_00001	ember-intézmény	TRUE
P_00001	T_00001	ember-közlekedés	FALSE

Nemcsak a hálózat csúcsaihoz, hanem az egyes élekhez is rögzültek metaadatok, például szociális intézmények, illetve kórházak esetén a fertőzött személy és az intézmény viszonya is bekerült az adatbázisba: bentlakóként/betegként vagy dolgozóként kapcsolódik-e az adott intézményhez. A fertőzés terjedésére utaló élcímek a feldolgozási lépés során kerültek rá a hálózatra.

Az adatrögzítésnek ebben a rétegében már megjelentek az adatkapcsolási aspektusok is. Az *edges* tábla *source* és *target* mezői a *nodes* tábla *global\_id* mezőjének idegen kulcsai voltak. A 3. táblázat mutatja, hogy milyen rekordok készültek az 1–2. táblázatban szereplő fiktív adatokkal kitöltött táblázatkezelő-munkalapból.

Az adatbázisban ezen a két fő táblán (*nodes*, *edges*) kívül még adminisztrációs adatok rögzültek a betöltött fájlokról (*records*), a betöltési műveletekről (*runs*), illetve az adatfeldolgozás során fellépő különböző típusú hibákról (*error\_fread*, *error\_xls*, *error\_html*).

### Adattisztítás és a fertőzési láncok kapcsolása

A kontaktkutató akkor lesz igazán hasznos eszköz, ha hálózatos módszertana segítségével láthatóvá teszi a fertőzési láncokat (Faye et al. 2015). Az előre kutató módszertan a potenciálisan fertőzésnek kitett és megbetegedéssel veszélyeztetett személyeket gyűjti. A visszamenőlegesen kutató módszertan a fertőzés kiindulópontjait, az esetleg fel nem fedezett oldalakat és a tömeges megfertőződést kirobbantó gócpontokat fedi fel. Amennyiben utánkövetésre is van lehetőség, akkor azok a folyamatok is felfedhetők, hogyan lesz egy korábbi kontaktszemélyből pozitív teszttel rendelkező fertőzött, illetve a fertőzés lefolyása és az esetleges tünetek megjelenése milyen időrendet követ. A szokásos módszertanon túllépve ebben a kutatásban nemcsak a személyek, hanem a különböző intézmények, helyszínek szerepe is követhető.

Az egyedi adatfelvételekből származó csillag alakú hálózatokat összekapcsolva megkaphatók a fertőzési láncok. Ehhez viszont meg kell mondani, hogy az egyik táblázatkezelő fájl kontaktszemélye ugyanazt az embert reprezentálja-e, mint a másik fájl fertőzött személye, illetve egy-egy intézményen belül a különböző fertőzöttek kontaktjai között is szerepelhet ugyanaz a személy többször. Az adatkapcsolást megnehezítették az utólagos tisztításból származó duplikációk is, ugyanis a javított táblázatok az adatrögzítők munkájának nyomon követése érdekében új, független adatként jelentek meg. Emiatt akár ugyanaz a fertőzött személy ugyanazokkal a kontaktszemélyekkel többször is szerepelhetett a csillaghálózatokat tartalmazó adatbázisban. Ezen kívül a szabad szöveggel bevitt intézményneveket is azonosítani kellett, és összevonni a hozzájuk tartozó különböző névvariációkat. Végül fontos megemlíteni a hiányzó, elgépelt, köznyelvi elnevezéssel jelölt stb. eseteket, melyeket



szintén kezelni kellett, illetve a megfeleltetéseket meg kellett találni.

A csillaghálózatok csúcseinak összevonását a „*union and find*” algoritmus végezte (Cormen et al. 2009). Ennek során több különböző rekord is ugyanazt a személyt/intézményt reprezentálja. Ezekből a rekordokból egy halmazt kell képezni, amiből egy reprezentatív elem azonosítja később az összes többi rekordot. Az összetartozó rekordok halmazát az összevonható rekordokból épített irányított gráf-erdő összefüggő komponensei adják meg. Egy összefüggő komponensen belül az irányított fa gyökere lesz a reprezentatív elem. A kiindulási állapotban minden rekord reprezentánsa, ún. *főnöke* önmaga. Ez felel meg annak, amikor még minden rekordot önálló entitásként kezelünk. Ezeket a kezdeti fáéleket eltávolítjuk. Ezután végigfutunk az összes lehetséges páron, amiket össze szeretnénk vonni, majd megállapítjuk egy rendezési feltételrendszerrel, hogy az adott párból melyik elem reprezentálja a másikat. A következő lépésben átállítjuk mindkét elem főnökét arra az elemre, amelyik a vizsgált párból a reprezentáns volt (*union* lépés). Vegyük észre, hogy ha egy elem a főnöke volt már más elemeknek, akkor a fában az összes alatta levő elem új gyökérhez, tehát új főnökhöz fog tartozni. Majd végigfutunk a főnökök mutatólistáján, minden rekordhoz rekurzívan megkeressük a gyökeret, és átállítjuk a főnökmutatóját erre a reprezentáns elemre (*find* lépés).

Az algoritmus végeredményeképpen kapott irányított fáéleket a *merge\_dict* nevű tábla tárolja a csúcsok globális azonosítóinak segítségével. A csillaghálózatok összefűzésekor ez alapján a tábla alapján kellett összevonni a csúcsokat. Mivel az összevonási táblázat külön, önállóan szerepel az adatbázisban, bármikor ellenőrizhető, hogy az összevonások helyesek-e, illetve bármely téves összevonás gyorsan korrigálható.

### Személyek és intézmények kapcsolása

A személyek esetén a reprezentációban használt rendezési elv a következő volt. Ha felmerült a lehetséges egyezés, akkor két fertőzött vagy két kontaktszemély között az újabb módosítási dátumú (adattisztításból vagy adatpótlásból származó táblázat) fájlból származót tettük meg *főnöknek*. Ha pedig az egyikük fertőzött, a másikuk kontaktszemély volt, akkor automatikusan a fertőzött rekordot tettük meg *főnöknek*, mert a kontakt személyekről csak a megkérdezett fertőzött vagy annak családtagja/képviselője adott adatot, míg a fertőzöttről hatósági információk álltak rendelkezésre. Hét különböző mező alapján lehetett megállapítani az egyezéseket:

1. név,
2. lakhely irányítószáma,
3. lakhely városa,
4. születési hely,
5. születési dátum,
6. anyja neve,
7. neme.

Ezek közül az adatok közül bármelyik lehetett hibás, illetve bármelyik hiányozhatott. Így először a kitöltött mezők egyezésére (*match\_flags*), illetve a legalább az egyik személynél üres mezőkre (*missing\_flags*) készítettünk egy-egy hételemű igaz/hamis sorozatot. Ezekből akkor állapítottunk meg egyezést, ha

- a fenti változósorozat első 5 eleméből csak 1 kitöltött
  - és a név mező a kitöltött, és az megegyezik,
- a fenti változósorozat első 5 eleméből 2, 3 vagy 4 kitöltött
  - és a név és a születési dátum megegyezik,
- a fenti változósorozat első 5 eleméből 4 kitöltött, és
  - vagy a név, vagy a születési dátum megegyezik,
  - a nem (7.) vagy hiányzik, vagy megegyezik,
  - az anyja neve (6.) vagy hiányzik, vagy megegyezik,
- a fenti változósorozat első 5 eleméből 5 kitöltött és
  - minden kitöltött megegyezik.

Ezzel a módszerrel meg lehetett találni azokat a duplikált személyeket is, ahol a nevet különbözőképpen rögzítették (pl. elgépelés, doktori cím, lánykori név és férjzett név különböző keverékei, két keresztnév), de a többi adat többségében kitöltött és többségében megegyező volt. Természetesen ez a módszertan sem talál meg minden egyezést, illetve tévesen összevonhat különböző személyeket. A potenciálisan téves összevonások az alábbi lehetséges kategóriáinként kerültek egy külön értesítési listára, amit az adattisztítók további egyeztetésre vettek át. A figyelt kategóriák a következők voltak:

- gyakori nevek és sok hiányzó mező.
- azonos néven, azonos lakcímen lakó apa és fia, anyja és lánya, ha hiányzik a születési dátum, vagy hiányzik az anyja neve. Anya-lánya esetben a férjzett név segíthet a megkülönböztetésben, de nem mindig.
- azonos nemű, otthon lakó ikrek, akiknek a felsorolt mezők közül csak a neve különbözik.

Ez a három jelenség azonban csak kis hibaszázalékot adott a rekordok azonosítása során.

Az intézmények esetén a rendezési elvnel a NEAK által nyilvántartott hivatalos kórházi illetve szociális otthoni lista elemei voltak a reprezentánsok az adatbázis *external\_nodes* nevű táblájában. Itt csak az intézmény neve alapján volt lehetőség a rekordok között egyezéseket találni. A végeredményben minden egyes intézményi rekord *főnöke* az *external\_nodes* táblából származott, adatmezőből származó *főnök* nem volt lehetséges.

A nevek összehasonlítását a SimString algoritmus végezte (Okazaki-Tsuji 2010). Az algoritmus a karakter-sorozatokat adott hosszúságú részsorozatokra bontja, majd a részsorozatokat egyezőségét méri. Az intézménynevek esetén egy előfeldolgozási lépésként el kellett távolítani a leggyakoribb szóelemeket, melyek túl sok egyezést generáltak volna. Ilyen, eltávolított szavak például: kórház, szent, megyei, idősok, otthon stb. Intézménynevek esetén a gyorsabb programfutás érdekében a hivatalos nevek mellett a már azonosított, kézi rögzítésű

**4. táblázat** | A táblázat egy példát mutat az intézményi nevek variációira, amik között az egyezést meg kellett találni, és a hivatalos intézménynévhez társítani. A példában a zalaegerszegi kórház – melynek hivatalos neve: Zala Megyei Szent Rafael Kórház – szerepel

---

Zalaegerszegi kórház
Zalaegerszegi Szent Raffael Kórház
Zalaegerszegi Kórház
Zala megyei Kórház központ
Zala megyei Kórház
Zala m. Szt Raffael kh.
Zala Megyei Szent Raffael kórház
Zala Megyei Szent Raffael Kórház
Zala Megyei Szent Rafael Kórház I. Sz. Belgyógyászat
Zala Megyei Szent Rafael Kórház Diaverum Mvese Állomás
Zala Megyei Szent Rafael Kórház (Zalaegerszeg)
Zala Megyei Szent Rafael Kórház
Zala Megyei Szent Rafael Kórház
Zala Megyei Szent Farael Kórház
Zala Megyei Kórház Vese Dialízis Központ
Zala Megyei Kórház
ZM Szent Rafael Kórház
Szent Raffael Kórház; Zalaegerszeg, Zala megye
Szent Raffael Kórház
Szent Rafael Kórházba
Szent Rafael Kórház; Zalaegerszeg
Szent Rafael Kórház; 8900 Zalaegerszeg, Zrínyi u. 1.
Diaverum Dialízis Központ, Zalaegerszeg, Zrínyi u. 1
Zala Megyei Szent Rafael Kórház

---

névváltozat is bekerült a SimString keresősztárába. Így lehetővé vált, hogy sok-sok elgépelés-, illetve eltérésláncan keresztül is azonosítani lehessen egymással az intézményeket.

Amennyiben algoritmikusan nem sikerült standard intézménynevet rendelni a rekordhoz, akkor az adott intézményi rekord néhány más adattal együtt (melyik táblázatból származik, az intézményhez milyen postai címet rögzítettek stb.) bekerült az *error\_html* adatbázistáblába. Ebből a táblából egy HTML-úrlap készült, melyen legördülő menüvel tudták az adattisztítók kiválasztani a megfelelő standard nevet az adott rekordhoz. Ezek a tisztított adatok feltöltés után, automatizálva kerültek be a *merge\_dict* táblába.

Megjegyezzük, hogy a rekordazonosítási lépéseket megkönnyíti, ha akár az adatrögzítési, akár az adatfeldolgozási lépésben törzsadatokkal (név, hivatalos címjegyzék, irányítószám-jegyzék) lehet összevetni a bevitt adatokat. Kitisztított intézményi adatok esetén például egy ilyen adatbázis értékes információval szolgálhat a fertőzési valószínűségekről, melyre a későbbiekben fontos döntéseket lehet alapozni (pl. oltási sorrend, elkülönítési

protokollok, látogatási tilalom). A rögzített alapterületek, akut panaszok is könnyebben értelmezhetők, ha a standardizálás során szakemberek által hitelesített egyedi hozzárendelések, vagy akár gépi tanulás segítségével meghatározott BNO-kódok kerülnek a végleges adatbázisba. A kérdezőbiztosok számára azonban továbbra is a szabad szavas rögzítés ajánlott olyan esetben, ahol új típusú járvánnyal kapcsolatban kell információt gyűjteni.

Végül megemlítiük, hogy két technikai adminisztrációs tábla szolgálta a program futásának nyomonkövetését. Az *error\_fread* tábla azon táblázatkezelő fájlok listáját tartalmazta hibáüzenettel együtt, melyeket nem sikerült beolvasni az adatbázisba, az *error\_xls* tábla pedig az adatrögzítőknek ad visszajelzést a kulcsfontosságú hiányzó mezőkről (például ha hiányzik a megfertőződési dátum a munkalapról). Az adatbázis teljes sémadiagramját a 2. ábra mutatja.

A megfelelő rekordok összevonása után a kontaktkutatósi jegyzőkönyvekből fel lehet építeni az időben növekedő fertőzési láncok hálózatát, amellyel részben emberek közötti kontaktusokat, részben pedig az egyes intézmények fertőzési láncolatokban betöltött szerepét is azonosítani lehet. Az időcímkék alapján felépíthető dinamikus gráf a terjedési útvonalak, a góccok kialakulása és az ország egyes területei közötti járványterjedés időbeli fejlődéséről is részletes képet ad. Érdekes megjegyezni, hogy viszonylag kis időtávon is kialakul egy országos, szinte teljes földrajzi lefedettséget mutató komponens, ami jelzi, milyen hatékonyan terjedt el még kijárási korlátozások mellett is a fertőzés az ország minden részébe. Adataink szerint április 17-re érte el a fertőzési gráf a domináns geográfiai lefedettséget, bár még csak az első hullámban megbetegedettek számának 30%-a szerepel csak ebben a komponensben. A kontaktkutató hatékonyágát is mutatja, hogy a gyors terjedés ellenére mindig sikerült izolálni a fertőzötteket. Így az első hullámban világviszonylatban is sikeresnek mondható járványkezelési eredményeket lehetett elérni.

### *Felhasználói felület komplex lekérdezésekhez és vizualizációhoz*

Mivel a fertőzési láncok, hálózatok földrajzi elhelyezkedése, a fertőzés földrajzi terjedése fontos információ, a fertőzöttek, kontaktszemélyek, illetve intézmények geolokációja is megtörtént. A beérkező adatokban a személynevekhez kapcsolt címek szabadszavas mezőként voltak rögzítve, ezért az adattisztítási folyamathoz kellett informatikai megoldást biztosítani a településnevek és az irányítószámok azonosításához. Ezt a feladatot a KSH standard település-adatbázisa (*KSH 2021*) segítségével lehetett elvégezni, amely során az elgépelést tartalmazó településneveket, illetve pontatlan irányítószámokat lehetett javítani. Az intézményi reprezentáns rekordok pontos címe ismert volt, ezért itt elegendő volt csak a deduplikációs és adatkapcsolási lépést megtenni.



2. ábra | Az SQLite adatbázis sémadiagramja. Kék alapon a táblanevek, alatta a mezők nevei és adattípusai, félkörrel az elsődleges kulcsok vannak kiemelve. A kapcsolható mezők között futnak az élek

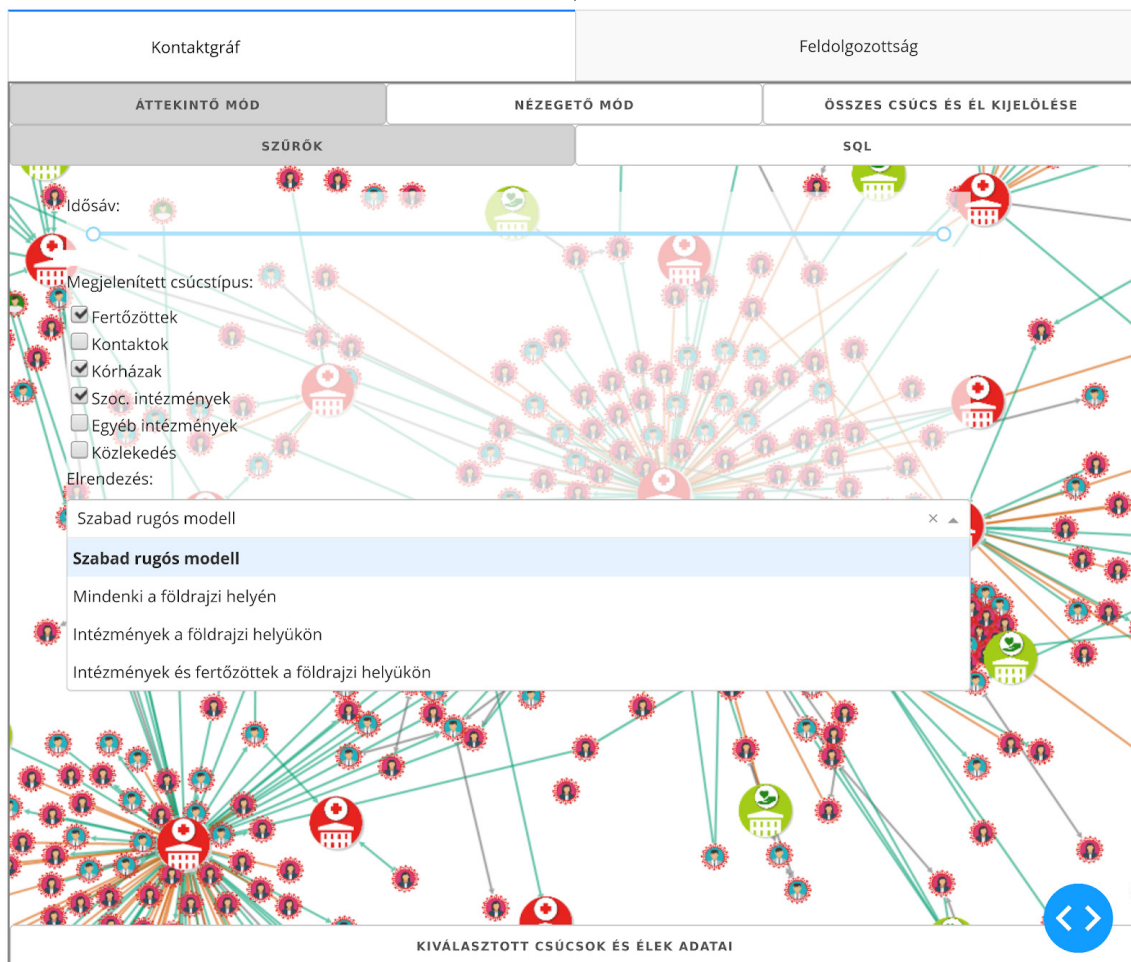
Az adatok tisztítása után az egyes entitások települési szintű koordinátákat kaptak. Így az időbeli dimenzió mellett térképes megjelenítés is lehetségessé vált, ami a fertőzés térbeli-időbeli terjedési mintázatainak követésére és vizuális tanulmányozására is alkalmat adott.

Az Operatív Törzs munkakörnyezetéhez illesztve egy interaktívan használható, böngészőből futó hálózati vizualizáció készült, mely backendként használja a létrehozott SQLite adatbázist (3. ábra). A vizualizáció a Python Dash (Dash 2021) keretrendszerén alapul. A böngészőben futás előnye, hogy számítógépes eszközök igen széles körével kompatibilis.

A vizualizációs felületen a hálózat csúcsait grafikus által tervezett, a csúcs által reprezentált entitás tulajdonságait megjelenítő piktogramok jelölik. A főbb típusokra az 1. ábra mutat néhány példát, a 3. ábrán pedig egy nagyobb hálórészlet látható a szűrési opciókkal. A hálózat élének irányítása fontos járványterjedési adatot jelöl (a fertőzés terjedésének irányát), ezért az irányítottságot az alábbi szempontok határozták meg:

- Az élek alapértelmezetten irányítatlanok.
  - Két ember között akkor került irányítás az élre, ha mindketten fertőzöttek és mindkettőnél ismert a pozitív PCR-teszt dátuma. Ekkor a régebben tesztelt felől az újabban tesztelt felé mutat az él.
  - Kórházak esetén az adatfelvételi úrlapon rögzítették azt a tünetlistát, mellyel az adott fertőzött a kórházba bekerült. E lista alapján minden kapcsolat, amely egy fertőzött személy és egy kórház között volt, a következő kategóriák szerint lett irányítva:
    - a fertőzött COVID-19-cel vagy COVID-19-gyanúval került be a kórházba,
    - biztosan *nem* voltak COVID-19 tünetei,
    - nem eldönthető / nem volt adat.
- Ha a beteg COVID-dal, vagy annak gyanújával került kórházba, akkor az él a kórház irányába mutatott, ha pedig biztosan nem COVID-dal került be, akkor az él a kórházból mutatott a személy felé.
- Ennek az élírányítási definíciónak köszönhetően követhető a fertőzés terjedési útja. Ez különösen fontos





3. ábra | A zárt rendszerű helyi hálózatban, web-böngészővel megjelenített interaktív felület a különböző szűrési (idő szerinti, csúcstípus szerinti) és hálózati elrendezési opciókkal

egy egészségügyi intézmény esetén, ahol láthatóvá válik, hogy az intézmény azért kapcsolódik-e sok fertőzötthöz, mert a megbetegedetteket látja el, vagy pedig lehetséges-e, hogy az adott beteg az intézményben kapta el a fertőzést.

Az élek irányítása mellett a csúcsok elhelyezésével is lehet információkat megjeleníteni. Ehhez az interaktív vizualizációban különböző algoritmikus elrendezéseket lehet választani (3. ábrán látható legördülő menü).

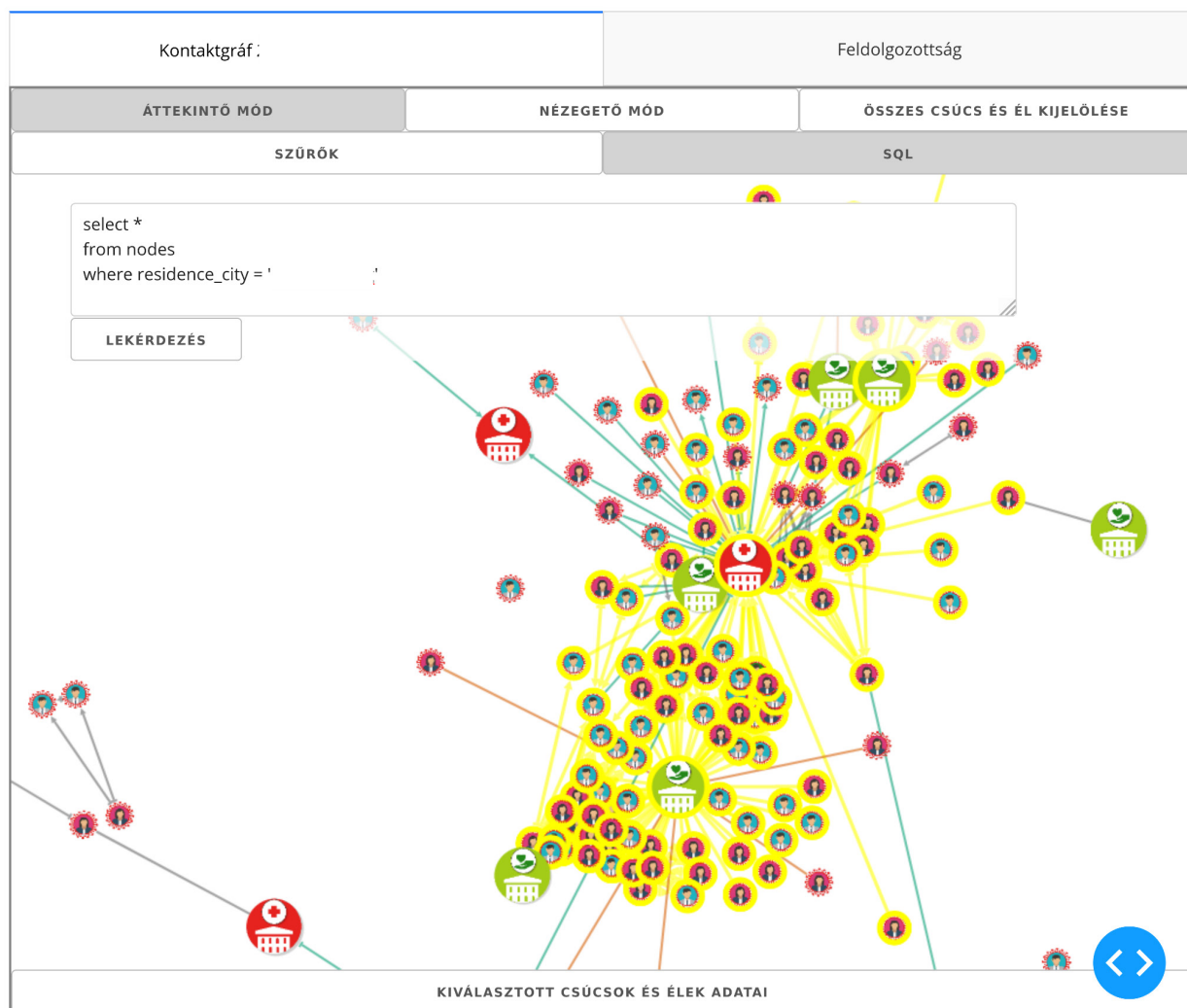
Három fő megjelenítési stratégiát érdemes megemlíteni:

- Hálózat-topológiai megjelenítés során egy rugós, azaz sfdp (*graph\_tool.draw 2021*) algoritmussal a hálózat sűrűn összekapcsolt részei kerülnek közel egymáshoz. Ilyenkor csak a terjedés hálózatos viszonyairól jelenik meg információ, a földrajzi elhelyezkedés eltűnik.
- A kombinált elrendezésekben lehetőség nyílik a különböző típusú csúcsok településekre történő elhelyezésére is. Például a fertőzöttek településre helyezésével, de az intézmények rugós modellen alapuló elrendezésével vizuálisan is információt kaphatunk arról, hogy egyes földrajzi régiókban mely intézmények

vesznek részt a fertőzöttek ellátásában, vagy mely intézmények vehetnek részt a fertőzési láncok országos méretűvé terebélyesedésében.

- Ha az összes csúcstípust a földrajzi helyére tesszük le, láthatóvá válnak a térbeli terjedési mintázatok.

A hálózati csúcsok elrendezése mellett egy-egy alkalmasan választott szűrési szempont is segíthet a látványos jelenség kiemelésében. Az interaktív felületen kattintással kiválaszthatók a megjelenítendő csúcsok (elrejtethők a kontaktok, intézmények stb.), illetve a megjeleníteni kívánt időintervallum is állítható. Mivel a csúcsok nagy száma miatt a hálózati elrendezések koordinátáinak kiszámítása sok időbe telik, ezért ezeket a koordinátákat a legfontosabb elrendezésekre és csúcstípusi kombinációkra a napi adatfrissítés során a rendszer kiszámolta. A pozíciókat az adatbázis *graph* nevű táblája tárolta. A haladó szintű felhasználók számára, akik összetett kérdésekre szeretnének választ kapni, az interaktív felületen lehetőség van SQL-lekérdezések beküldésére (4. ábra), és a lekérdezés eredményeképpen visszaadott csúcsok kijelölésére, vagy csak az így kapott csúcsok által feszített részgráf vizsgálatára.



4. ábra | Az interaktív lekérdezési felület SQL-szűrési opciójának szemléltetése: a sárgával jelölt fertőzött és intézményi csúcsok ugyanazon SQL-lekérdezéssel kiválasztott városhoz tartoznak

## Diszkusszió

Az eddigiekben bemutatott számítógépes rendszer segítségével tehát a kontaktkutatási jegyzőkönyvek eredetileg strukturálatlan adataiból több lépésen keresztül egy relációs adatbázis állítható elő. A relációs adatbázist az elemzők és döntéshozók számára az interaktív vizualizációs felület teszi használhatóvá, mellyel a kontakthálózat megértésére, problémás jelenségek feltárására, vezetői rendszerekbe integrálásra, döntéshozatali folyamatokban való támogatásra nyílik rugalmas lehetőség.

A jól szervezett kontaktkutatás révén előálló, és megfelelő elemzőeszközökkel tanulmányozott adatok a járvány megfékezése szempontjából fontos információkat tudnak szolgáltatni. Például az elszigetelt eseteket, a létrejövő fertőzési láncokat, a problémás intézményeket, szuperterjesztőket azonosítani lehet, esetleg a kontaktkörében előre lehet jelezni a terjedést. Azonban a módszer alkalmazhatósági korlátaival, más kontaktkutatási módszerekhez viszonyított előnyeivel és hátrányaival is érdemes tisztában lenni.

## Néhány hátrány

A beérkező táblázatkezelő-fájlok feldolgozása az ismert rendszerben teljesen automatikus, tehát egy kialakult táblázatstruktúra mellett az adatok bővítése csupán számítógépes időt vagy erőforrást vesz igénybe. Azonban az eredeti, folyó szöveggel rögzített jegyzőkönyvek táblázatos formába történő elkódolása mindenképpen szűk keresztmetszetet jelent. Mivel a jegyzőkönyvek igazolt megbetegedési esetre készülnek (a járvány korai fázisában egyértelmű klinikai diagnózisok, később pozitív PCR-teszt alapján), az adatfelvétel és a fertőzés között mindenképpen van néhány nap késleltetés. Ez az időeltolódás nagyon ronthatja a megelőzés hatékonyságát (Kretzschmar et al. 2020). A kontaktkutatásnak ez a kérdőíves módszere prevencióra csak nagyon korlátozott esetekben használható. Ilyen esetek például a karanténrendelések, ahol a kontaktkutatás felderítheti, hogy idősök vagy egyéb veszélyeztetett személyek lennének kitéve az összeczártság miatt intenzívebb fertőzésnek. Általános élethelyzetben azonban, azaz nem a kijárási korlátozás

idején, a lassú kontakt kutatás már nem alkalmas a terjedés megállítására, hiszen az esetlegesen megfertőződött kontaktok kontaktjai is megfertőződhetnek a kutatást követő intézkedések előtt (Kwok et al. 2019).

Magának a jegyzőkönyvnek a felvétele és a kontaktok felkeresése is sok idő és sok emberi munka, emiatt ez a kontakt kutatási módszer csak adott napi esetszám alatt lehet hatékony (Dhillon–Srikrishna 2018). Kijárási korlátozások és egyéb önkéntes vagy rendeletek útján szabályozott kontaktuscsoökkentő intézkedések esetén magasabb esetszám mellett is használható marad még a módszer. Ilyenkor ugyanis az egy emberre jutó átlagos kontaktusszám kevesebb. Ennek eredményeként nemcsak kevesebb embert kell felkeresni, hanem a kérdezett alanyok maguk is könnyebben felsorolják a kontaktjaikat, kevésbé torzul az adat a pontatlan visszaemlékezések miatt. Zsúfolt nagyvárosokban, különösen ahol egyes családoknak tartalékok híján nem áll módjukban hosszú ideig elzártan élni, még korlátozások idején sem lehet teljes körűnek tekinteni a kontakt kutatási adatokat (Faye et al. 2015).

Fontos megjegyezni a módszer használhatóságára vonatkozóan, hogy a meglátogatott helyek, intézmények listáját, illetve a használt járműveket csak kijárási korlátozás, illetve jócskán lecsökkent mobilitás mellett van értelme rögzíteni. Csökkent mobilitás mellett ugyanis könnyen beazonosíthatók azok az emberek, akik például ugyanazzal a hajnali vonatjáráttal mentek munkába, vagy akik ugyanazt a menetrend szerinti buszjáratot használják. Ugyanígy lehet megállapítani az intézmények esetleges szerepét a terjedésben, mert sok meglátogatott helyszín mellett már nehezebb a gócpontokat azonosítani.

### Néhány előny

A bemutatott informatikai alkalmazáshoz használt adatgyűjtési módszertannak számos korlátja mellett több előnye is van. Mivel központosított az adatfelvétel, és képzett, fegyelmezett alkalmazottak végzik az adatok gyűjtését, nincs elkallódott úrlap, tehát ami rögzítésre kerül, abból minden adat bekerül a hálózati reprezentációba. További előny, hogy a telefonos vagy élő kérdés miatt lehetséges a válaszadók által adott adatokra vissza-kérdezni, azokat pontosítani.

Sokszor felmerülő kritika az adatgyűjtési módszertannal szemben, hogy jelentős emberi munkaigénnyel jár. Ezzel szemben az automatizált vagy okos eszközökre alapozott módszerek a fejlesztésen túl már elsősorban a lekutatóktól igényel időt. Megjegyezzük, hogy a központi kérdéssor és központilag irányított kérdező személyzet miatt bármilyen változtatás egységesen kerül át az összes adatfelvételbe. Ezzel ellentétben például webes vagy mobilalkalmazások esetén az aszinkron frissítésekkel, verziókövetésekkel állandó probléma van, a beérkező válaszok egy-egy frissítés vagy változtatás után hosszú ideig a korábbi verziókkal keveredve érkeznek meg.

A teljesen automatizált kontakt kutatási módszerekkel szemben, amelyek valamilyen szenzorokból gyűjtenek rendszeres információt, a felmérésben keletkező adatmennyiség kis mennyiségűnek számít. Az adatok feldolgozása, tárolása és a vizualizáció informatikai erőforrás-igénye, szoftverigénye kicsi, akár egy modernebb asztali számítógép is el tudja látni a szükséges feladatokat.

Hasonlóan egyéni szintű, de tökéletesen anonim kontakt kutatást tennének lehetővé a decentralizált bluetooth-technológián alapuló okostelefonos applikációk. A kérdőíves kontakt kutatással szemben ezeknél komoly hátrány (Anglemeyer et al. 2020), hogy gyakran éppen a legsérülékenyebb populációk nem rendelkeznek az applikációk telepítéséhez szükséges technológiákkal, például az idősek otthonában gondozottak körében nem elég elterjedtek az okostelefonok. Még az okostelefonnal rendelkező lakosság körében is kérdéses, hogy egy ilyen applikáció mennyire képes elterjedni: ez a módszer egy bizonyos adaptálási ráta fölött képes csak csökkenteni a terjedést, ezzel szemben nincsen esetszámkorlátja. Az anonimitásért cserébe terjedési statisztikákat, rizikókat, utólagos elemzéseket ilyen applikációk adataiból nem lehet végezni, mint ahogy a megfertőződött korát, nemét, betegségét, földrajzi helyét sem lehetséges így rögzíteni. A bluetooth-jel azt sem képes érzékelni, a két emberen van-e maszk, így szükségtelen esetekben akár felülbecsülheti a megfertőződési valószínűséget.

Mobilcella-alapon, illetve GPS-jel segítségével is lehetséges megbecsülni, hogy két ember tartózkodott-e annyira ideig olyan közelségben, ami már rizikót jelenthet a fertőzés átadására. Sajnos itt hatványozottan igaz az, ami a bluetooth-jelnél, hogy lehetetlen megmondani, hogy a két ember között volt-e maszk, sőt, hogy volt-e esetleg közöttük épületfal, plexilap stb. Emellett az ilyen módszer feltételezi a teljes populáció aktuális lokációjának monitorozását, ezen adatok feldolgozását, ami erőforrás-igényes, illetve adatvédelmi szempontokból is aggályos lehet.

Megjegyezzük még, hogy a kontakt kutatás részlegesen elvégezve, azaz fertőzöttek és kontaktjaiknak csak egy kis részét kikérdezve is nagyon hatékony lehet a fertőzési góccok felderítésében. A fertőzési lánc mentén visszamenőlegesen gyűjtve az adatokat, azaz a fertőzöttek kontaktjai között azokat felmérve, akik korábban voltak fertőzöttek, a hálózat általános matematikai tulajdonságai miatt a legveszélyesebb szuperterjesztőket nagy valószínűséggel meg lehet találni. Ez a megtalálási valószínűség ráadásul a visszamenő lépések számával exponenciálisan nő (Kojaku et al. 2021; Endo et al. 2021). Ha a megkérdezetteket az úgynevezett gyakorisági súlyozás szerint választják ki, akkor még kevesebb megkérdezettel is meg lehet találni a szuperterjesztőket. A részleges mintavételezésen alapuló eljárásokat azonban alapos lakossági tájékoztatásnak kell kísérnie. Ugyanis az egyén számára nem átlátható, nem mindenre kiterjedő adatfelvétel megrendítheti a lakosság bizalmát, ami a kutatás eredményességét és a szolgáltatott adatok pontosságát is rontja.



A személyes találkozásokon alapuló kontaktkutatás megfelelő törvényi háttérrel ötvözve a nehéz szociális helyzetben lévők segélyezésében, vagy a központilag biztosított elkülönítőhelyek biztosításában is hatékony lehet. Ezek a keretek jelentős sikereket tudtak felmutatni Kínában, illetve az 1900-as évek elején az Egyesült Királyságban (Mooney 2020).

### Adatkezelés felelőssége

Mivel egy kérdőíves kutatásban rendkívül szenzitív adatok kerülnek adatbázisba, kiemelten kell ezek kezelésére figyelni (Dubov–Shoptawb 2020). A kontaktkutatási feladathoz elengedhetetlenek ezek a szenzitív adatok, és a társbetegségek, életkor, kórházi tartózkodások fontos információt tartalmaznak a betegség súlyosságára vonatkozóan. Azonban az egyes rekordok azonosítása után, amikor a lehetséges hálózati struktúrák elkészültek, számos kérdés vizsgálatához már nem szükségesek az adatok részletei. Ezekben az esetekben a vizualizációt, elemzési lehetőségeket anonim globális azonosítók segítségével szélesebb szakmai körben is vizsgálhatóvá lehet tenni. Sajnos a teljes anonimitás elérése csak nagyon aggregált szinten lehetséges, hiszen már egy kisebb lakosság település elhelyezkedéséből, vagy egy idősotthon, illetve kórház földrajzi helyéből is lehet következtetni az anonimizált személy vagy intézmény valódi adataira.

Egy lehetséges megoldás a teljes anonimitás és a részletes időbeli, térbeli és korosztálybeli felbontás között a csupán a kontaktusszámokra rákérdező reprezentatív, illetve online felmérés, amely Magyarországon a járvány kezdete óta fut (MASZK 2021). Egy ilyen adatban a részletes hálózati mintázatokat nem figyelhetjük meg, tehát a kontaktkutatás elsődleges feladatait (pl. góccok azonosítása, karantén elrendelése, nagy kockázatú veszélyeztetettek elkülönítése) nem helyettesítheti. Járványmodellezéshez, illetve döntéshozatal előkészítéséhez, intézkedések hatásának monitorozásához a lakosság megfelelő technológiai és eszközhozzáféréssel rendelke-

ző, önkéntes adatszolgáltatásra nyitott csoportját tekintve jobb lehet. Meg kell azonban jegyezni, hogy egy fertőzött vagy karanténban lévő, egyébként is nehéz helyzetben, pszichés nyomás alatt álló egyén számára a sok, párhuzamos adatfelvételi megkeresés tovább csökkenti az együttműködést.

A fentiek alapján megállapíthatjuk, hogy a központilag szervezett kérdőívre, illetve az önkéntes alapú módszertanokra vagy technikai megoldásokra mint egymás kiegészítőire és nem mint egymást helyettesíteni képes lehetőségekre kell gondolni (lásd 5. táblázat).

### Összefoglaló

A kifejlesztett vizuális reprezentációs technika látványos és részletekbe menő megértést, problémafeltárást képes biztosítani a járványügyi szakemberek számára. A grafikus elemek segítenek a gyors megértésben, a különböző hálózati elrendezések bizonyos jelenségekre, például gócpontokra, fertőzési klikkekre vagy a földrajzi terjedésre irányíthatják a figyelmet. A böngészőből történő futtatás alacsony technológiai belépési küszöböt biztosít a társterületeken kutatók számára, nekik így nem szükséges a problémafeltáráshoz külön szoftvereket telepíteni. Az adatbázis SQL-alapú szűrése a vizualizációs felületről lehetőséget biztosít összetettebb kérdések megfogalmazására is. Ilyen kérdések lehetnek például a kialakult fertőzési láncok hálózatának összehasonlítása különböző referencia-hálózatokkal, például véletlen modellekkel vagy szociológiai változókkal súlyozott hálózat-növekedési modellekkel.

### Irodalomjegyzék

- Anglemyer, A., Moore, T., Parker, L., Chambers, T., Grady, A., Chiu, K., Parry, M., Wilczynska, M., Fleming, E., & Bero, L. (2020) Digital contact tracing technologies in epiDemics: A rapid review. *The Cochrane Database of Systematic Reviews*, Vol. 41. No. 9. p. 1028. <https://doi.org/10.1002/14651858.CD013699>
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*, Third Edition (3rd ed.). The MIT Press
- Dash (2021) Dash Overview. <https://plotly.com/dash/> [Letöltve: 2021. 05. 06.]
- Dhillon, R. S., & Srikrishna, D. (2018) When is contact tracing not enough to stop an outbreak? *The Lancet Infectious Diseases*, Vol. 18. pp. 1302–1304. [https://doi.org/10.1016/S1473-3099\(18\)30656-X](https://doi.org/10.1016/S1473-3099(18)30656-X)
- Dubov, A., & Shoptawb, S. (2020) The Value and Ethics of Using Technology to Contain the COVID-19 Epidemic. *The American Journal of Bioethics*, Vol. 20. No. 7. pp. W7–W11. <https://doi.org/10.1080/15265161.2020.1764136>
- Endo, A., Leclerc, Q. J., Knight, G. M., Medley, G. F., Atkins, K. E., Funk, S., & Kucharski, A. J. (2021) Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks. *Wellcome Open Research*, Vol. 5. p. 239. <https://doi.org/10.12688/wellcomeopenres.16344.3>
- Faye, O., Boëlle, P.-Y., Heleze, E., Faye, O., Loucoubar, C., Magassouba, N., Cauchemez, S. et al. (2015) Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *The Lancet Infectious Diseases*, Vol. 15.

5. táblázat | Ajánlás jövőbeni járványok kontaktkutatását hatékonyabbá tévő módszerekre

Adatok rögzítése egységes, központi törzsadatok alapján (személyek, lakcímek, betegségsnevek, intézménynevek)
Adatok rögzítése űrlapban szabad szöveg helyett
Köznapi és törzsszállományban szereplő elnevezések párosítására gépi intelligencia alkalmazásával folyamatosan frissülő szinonimaszótár építése
Tájékoztató kampányok szervezése a kontaktkutatás módszereinek ismertetése érdekében
Adatszolgáltatási lojalitás növelése a párhuzamos adatgyűjtések összefogásával
Más technológián alapuló kontaktkutatási módszerek adatainak összekapcsolása a felmérésen alapuló adatokkal



- No. 3. pp. 320–326. [https://doi.org/10.1016/S1473-3099\(14\)71075-8](https://doi.org/10.1016/S1473-3099(14)71075-8)
- graph\_tool.draw (2021) Graph drawing and layout — graph-tool 2.37 documentation. [https://graph-tool.skewed.de/static/doc/draw.html#graph\\_tool.draw.sfdp\\_layout](https://graph-tool.skewed.de/static/doc/draw.html#graph_tool.draw.sfdp_layout) [Letöltve: 2021. 05. 06.]
- Koetter, P., Pelton, M., Gonzalo, J., Du, P., Exten, C., Bogale, K. et al. (2020) Implementation and Process of a COVID-19 Contact Tracing Initiative: Leveraging Health Professional Students to Extend the Workforce During a Pandemic. *American Journal of Infection Control*, Vol. 48. No. 12. pp. 1451–1456. <https://doi.org/10.1016/j.ajic.2020.08.012>
- Kojaku, S., Hébert-Dufresne, L., Mones, E., Lehmann, S., & Ahn, Y. Y. (2021) The effectiveness of backward contact tracing in networks. *Nature Physics*, 17. pp. 652–658. <https://doi.org/10.1038/s41567-021-01187-2>
- Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C. J., van Boven, M., van de Wijgert, J. H. H. M., & Bonten, M. J. M. (2020) Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health*, Vol. 5. No. 8. pp. e452–e459. [https://doi.org/10.1016/S2468-2667\(20\)30157-2](https://doi.org/10.1016/S2468-2667(20)30157-2)
- KSH (2021) Magyarország helységnévtára. [https://www.ksh.hu/apps/hntr.egyeb?p\\_lang=HU&p\\_sablon=LETOLTES](https://www.ksh.hu/apps/hntr.egyeb?p_lang=HU&p_sablon=LETOLTES) [Letöltve: 2021. 05. 06.]
- Kwok, K. O., Tang, A., Wei, V. W. I., Park, W. H., Yeoh, E. K., & Riley, S. (2019) Epidemic Models of Contact Tracing: Systematic Review of Transmission Studies of Severe Acute Respiratory Syndrome and Middle East Respiratory Syndrome. *Computational and Structural Biotechnology Journal*, Vol. 17. pp. 186–194. <https://doi.org/10.1016/j.csbj.2019.01.003>
- MASZK (2021) <https://covid.sed.hu/tabs/response> [Letöltve: 2021. 05. 06.]
- Mooney, G. (2020). “A Menace to the Public Health” — Contact Tracing and the Limits of Persuasion. *New England Journal of Medicine*, Vol. 383. No. 19. pp. 1806–1808. <https://doi.org/10.1056/NEJMp2021887>
- MTA Nyelvtudományi Intézet (2021) Az MTA Nyelvtudományi Intézete által anyakönyvi bejegyzésre alkalmasnak minősített utónevek jegyzéke. <http://www.nytud.hu/oszt/nyelvmuvelo/utonevek/> [Letöltve: 2021. 05. 06.]
- Okazaki, N., & Tsujii, J. (2010) Simple and efficient algorithm for approximate dictionary matching. *Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference*, Vol. 2, 851–859.
- SQLite Database (2021) <https://www.sqlite.org/index.html> [Letöltve: 2021. 05. 06.]
- WHO Ebola Response Team (2014) Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine*, Vol. 371. No. 16. pp. 1481–1495. <https://doi.org/10.1056/nejmoa1411100>
- World Health Organization (2014) Contact Tracing During an Outbreak of Ebola Virus Disease. <http://www.who.int/csr/resources/publications/ebola/contact-tracing-during-outbreak-of-ebola.pdf>

A cikk a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>) feltételei szerint publikált Open Access közlemény, melynek szellemében a cikk bármilyen médiumban szabadon felhasználható, megosztható és újraközölhető, feltéve, hogy az eredeti szerző és a közlés helye, illetve a CC License linkje és az esetlegesen végrehajtott módosítások feltüntetésre kerülnek. (SID\_1)