

Nehezen elérhető csoportok és szegregáció vizsgálata a hálózati méretgenerátor módszerével: egy magyarországi vizsgálat tanulságai¹

Estimating the size of hard-to-reach subpopulations and segregation with the network scale-up method: lessons from a Hungarian survey

Kisfalusi Dorottya² – Susánszky Pál³ – Susánszky Éva⁴

<https://doi.org/10.51624/SzocSzemle.2022.4.4>

Beérkezés: 2022. 06. 29.

Átdolgozott változat beérkezése: 2022. 12. 04.

Elfogadás: 2022. 12. 09.

Összefoglaló: Jelen tanulmányban empirikus, leíró elemzését adjuk a magyar társadalom kapcsolathálózati jellemzőinek, valamint módszertani szempontból értékeli a méretgenerátoros technikával adott becslések pontosságát a társadalmi csoportok méretére és szegregáltságára vonatkozóan. Egy 7000 fős magyarországi reprezentatív kérdőíves felmérés (HS2021) adatai alapján megbecsüljük a válaszadók kapcsolathálójának átlagos méretét. Ezt követően vizsgáljuk, hogy a méretgenerátoros módszer milyen becslést ad különböző társadalmi csoportok (például munkanélküliek, romák, Covid-fertőzöttek, tünetesen részt vevők) létszámára és társadalmon belüli szegregáltságára. A létszámbecsléseket összehasonlítjuk más adatforrásokból (népszámlálási adatok, adminisztratív adatok, kérdőíves adatfelvételek) származó létszámadatokkal. Eredményeink azt mutatják, hogy a méretgenerátoros módszer a tartós és jól látható tulajdonságok esetében képes nagyságrendileg jól becsülni a tényleges csoportlétszámokat (például romák, hajléktalanok), míg az alacsony láthatóságú csoportokét alulbecsli. A társadalmi törésvonalak a hajléktalanok, a tünetesen részt vevők és a civil szervezeti tagok ismertségét tekintve a legnagyobbak. Tanulmányunk végén ismertetjük a méretgenerátoros technikán belül a pontosabb becslések érdekében megalkotott módszertani újításokat, és javaslatokat teszünk arra, hogy a módszert használó jövőbeli kutatások hogyan tervezzék meg a kutatási dizájnt, hogy pontosabb becsléseket érjenek el.

1 A „Hungarostudy 2021” kutatás a Kopp Mária Intézet a Népesedésért és a Családokért és a Semmelweis Egyetem Magatartástudományi Intézet együttműködésében valósult meg 2021-ben, N = 7000. Kutatásvezetők: Purebl György, Engler Ágnes. A tanulmány a Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal (NKFIH) FK 135274 számú projektje keretében készült. Köszönjük a bírálóknak a kézirat korábbi változatához fűzött értékes hozzászólásait.

2 Társadalomtudományi Kutatóközpont, email: kisfalusi.dorottya@tk.hu

3 Társadalomtudományi Kutatóközpont, email: susanszky.pal@tk.hu

4 Semmelweis Egyetem, Magatartástudományi Intézet, email: susanszky.eva@med.semmelweis-univ.hu

Kulcsszavak: kapcsolatháló, méretgenerátor, társadalmi struktúra

Abstract: In this paper, we provide an empirical, descriptive analysis of the social networks of Hungarian society and illustrate how the network scale-up method estimates the size of hard-to-reach subpopulations and segregation of social groups. Based on a representative survey of 7000 respondents from Hungary (HS2021), we first estimate the average size of the respondents' personal networks. Then, we examine the social fault lines along various social groups and how accurately the network scale-up method estimates the size of these groups (e.g., unemployed, protesters, the Roma, Covid-infected). These estimates are then compared with data from other sources (census data, administrative data, surveys). Our results show that the network scale-up method estimates the size of visible social groups (e.g., the Roma, homeless people) quite well. The visibility of other social groups appears to be much lower. Social fault lines are greatest in the case of homeless people, protesters, and members of NGOs. Finally, we describe recent methodological advancements and summarize our suggestions for future research using this method.

Keywords: network scale-up method, social network analysis, social structure

Bevezető

A szociológiai vizsgálatok számára kiemelten fontosak az egyén társadalmi pozícióját leíró kutatások, hiszen a társadalmi státusz szorosan összefügg az egyének életmódjával, viselkedésével, értékrendjével. A társadalomban elfoglalt hely vizsgálatakor gyakran a szocioökonomiai státusz (Huszár 2015) mellett a társas kapcsolathálózatokba való beágyazottságot is figyelembe veszik (Albert–Dávid 2012), ugyanis a személyes kapcsolatok alapvetően határozzák meg az egyének társadalomba való integrálódását (Granovetter 1983, Coleman 1988, Lin 1999, Kmetty–Koltai 2016, Dávid–Albert–Husztai 2020).

Tanulmányunkkal a magyarországi survey adatokon alapuló hálózatkutatáshoz kapcsolódunk. Célunk a magyar társadalom néhány strukturális jellemzőjének leírása a méretgenerátor használatával, valamint a módszer bizonyos tulajdonságainak értékelése. Két témára fókuszálunk tehát: egyrészt empirikus, de főleg leíró elemzését kívánjuk adni a magyar társadalom kapcsolathálózati jellemzőinek. Másrészt módszertani szempontból kívánjuk értékelni a méretgenerátor technikával a társadalmi csoportok méretére és szegregáltságára adott becslések pontosságát.

A hálózati méretgenerátoros vagy összegző módszer szociológiai alkalmazása alapvetően három különböző kutatási területre terjed ki (Kmetty–Koltai 2015, Laga–Bao–Niu 2021, McCormick 2021): 1) az egóhálózatok méretének becslésére, 2) a nehezen elérhető vagy rejtett csoportok méretének becslésére, valamint 3) a társadalmon belüli szegregációs minták feltárására.

A fenti hálózatkutatási technikát több magyarországi kutatásban is alkalmazták (Kmetty 2015, Kmetty–Koltai 2016, 2018). Az eddigi vizsgálatokból kimaradt, vagy kisebb hangsúlyt kapott három fontos téma: 1) a módszer alkalmazhatóságának értékelése a nehezen elérhető csoportok méretének becslésére, 2) a társadalmi csoportok szélesebb körére vonatkozó szegregációs minták leírása, valamint 3) az előző két szempont összekapcsolása: a csoportok létszámbecslésének pontossága és azok szegregáltságának empirikus összevetése.

A tanulmány első felében bemutatjuk az összegző módszer használatát, majd összefoglaljuk a társadalmi csoportok méretének becslésével és a szegregációs mintákkal kapcsolatos legfontosabb kutatási eredményeket. A cikk második felében egy 7000 fős magyarországi reprezentatív kérdőíves felmérés adatain illusztráljuk a módszer használatát. Először megbecsljük a válaszadók kapcsolathálójának átlagos méretét. Ezt követően rátérünk az egyes csoportok mentén tapasztalható társadalmi törésvonalak vizsgálatára. Ezután megvizsgáljuk, hogy a méretgenerátoros módszer leggyakrabban alkalmazott technikája (Killworth–McCarty et al. 1998) milyen becslést ad különböző társadalmi csoportok (például munkanélküliek, tüntetesen részt vevők, romák, Covid-fertőzöttek) méretére. Ezeket a becsléseket aztán összehasonlítjuk más adatforrásokból (népszámlálási adatok, adminisztratív adatok, kérdőíves adatfelvételek) származó létszámadatokkal és becslésekkel. A csoportméret becslésének pontosságát/pontatlanságát összevetjük a csoport társadalmon belüli szegregáltságával. Végül ismertetjük az érvényes eredmények útjában álló módszertani akadályokat, illetve az ezek orvoslására felmerült technikai megoldásokat.

A méretgenerátoros vagy összegző technika

A méretgenerátoros vagy másként összegző módszer alkalmazásakor kérdőíves adatfelvételnél megkérdezzük a válaszadókat arról, hány embert ismernek bizonyos társadalmi csoportokból. A kérdezett csoportok nagyobb részének ismert az összlétszáma a populáción belül (például nyugdíjasok vagy Sándor nevű emberek). Ezeket a csoportokat használjuk fel arra, hogy a válaszadók kapcsolathálózatának méretét megbecsljük, illetve hogy kiszámoljuk a populációban az egyének kapcsolathálózatának átlagos méretét. A kérdezett csoportok egy kisebb részének nem ismert az összlétszáma a populáción belül (például szerhasználók, hajléktalanok, szexmunkások). Ezek azok a nehezen elérhető csoportok, amelyeknek a létszámát becsülni kívánjuk. A becsléshez felhasználjuk egyrészt azt az információt, hogy a válaszadók saját bevallásuk szerint hány embert ismernek az adott csoportból, másrészt hogy a többi kérdésre adott válaszuk alapján mekkora a kapcsolathálójuk becsült mérete.

A méretgenerátoros technika alapfeltevése, hogy egy alpopuláció létszáma úgy aránylik a teljes populáció létszámához, ahogyan az alpopulációból ismert emberek száma a válaszadók teljes kapcsolathálózához (Killworth–McCarty et al. 1998, Laga–Bao–Niu 2021). Azaz:

$$\frac{y_{ik}}{d_i} = \frac{N_k}{N} \quad 1. \text{ egyenlet,}$$

ahol N a teljes populáció létszáma, N_k a k alpopuláció létszáma, y_{ik} az i egyén által ismert emberek száma k alpopulációból, és d_i az i egyén kapcsolathálójának mérete. A teljes populáció létszáma általában ismert a népszámlálási és egyéb adminisztratív adatokból. Az y_{ik} létszámot a válaszadók rögzítik a kérdőívben. Ahhoz, hogy az alpopuláció méretét, N_k -t becsülni tudjuk, egyrészt arra van szükség, hogy a válaszadók y_{ik} -ra vo-

natkozóan minél pontosabb számot mondjanak a kérdőíves adatfelvétel során, másrészt arra, hogy az egyén kapcsolathálójának méretét, d_i -t valahogyan megbecsüljük.

A következő fejezetben áttekintjük, milyen torzításokhoz vezethet, ha az adatfelvétel során a megkérdezettek az egyes társadalmi csoportokhoz tartozó ismerőseik létszámára nem pontos számmal válaszolnak, illetve hogy miért nem igaz minden egyes egyén esetében, hogy a társadalmi arányuknak megfelelően ismernek egyes társadalmi csoportokból embereket.

A társadalmi csoportok méretének becslése és a becslési torzítás lehetséges forrásai

Az összegző módszer első alkalmazásai az 1980-as évek végéről, illetve az 1990-es évek elejéről származnak. Az első tanulmányokban a nehezen elérhető populációk méretének lehetséges alsó és felső határaitra vonatkozóan adtak becsléseket (Bernard et al. 1991). Ezt az eljárást fejlesztették tovább Killworth és munkatársai (Killworth–Johnsen et al. 1998, Killworth–McCarty et al. 1998), akik az általuk kidolgozott modell segítségével már sokkal pontosabb becslést tudtak adni mind a személyes kapcsolatháló nagyságára, mind a nehezen elérhető társadalmi csoportok létszámára. A módszert használták például földrengés áldozatainak (Bernard et al. 1991), szerhasználók létszámának (Salganik et al. 2011), vagy abortuszok számának (Sully–Giorgio–Anjur–Dietrich 2020) becslésére is.⁵

A Killworth és munkatársai (1998) által megalkotott módszer máig a leggyakrabban alkalmazott méretgenerátoros technika a nehezen elérhető csoportok létszámának becslésére (Laga–Bao–Niu 2021), ebben a cikkben is erre támaszkodunk, ezért ezt részletesebben is ismertetjük. A szerzők az Egyesült Államokban végeztek kutatást, melyben egy országos reprezentatív minta válaszadóit kérdezték meg arról, hány embert ismernek 32 alpopulációból. Ebből 29 csoport létszáma ismert volt a kutatók számára (például 12 keresztnév tulajdonosai, postai alkalmazottak, az elmúlt 12 hónapban motorbalesetben elhunytak), míg három alpopuláció létszámát nem ismerték (HIV-fertőzöttek, hajléktalanok, nemi erőszak áldozatául esett nők), ezeket a válaszok alapján becsülték. A módszer első részében az ismert populációkra adott válaszok összegzése alapján megbecsülték a válaszadók személyes kapcsolathálóját az alábbi egyenlet segítségével:

$$\hat{d}_i = \frac{N * \sum_{k=1}^L y_{ik}}{\sum_{k=1}^L N_k} \quad \text{2. egyenlet,}$$

ahol \hat{d}_i az i egyén kapcsolathálójának becsült mérete, N a teljes populáció létszáma, N_k a k alpopuláció létszáma, y_{ik} az i egyén által ismert emberek száma k alpopulációból, és L az ismert létszámú alpopulációk száma.

5 A módszerről két áttekintő cikk is megjelent a közelmúltban. Míg McCormick (2021) összefoglaló írása elsősorban a kérdőíves adatfelvétel megtervezésére és elemzésére, valamint a módszer korábbi alkalmazásainak bemutatására koncentrált, addig Laga és munkatársai (2021) a módszer keretében létrejött különböző statisztikai modelleket ismertetik.

Ezek után a becsült kapcsolathálózati nagyságot felhasználva, maximum likelihood becslés (MLE) segítségével megbecsülték az ismeretlen létszámú alpopulációk méretét:

$$\hat{N}_u^{MLE} = \frac{N * \sum_{i=1}^n y_{iu}}{\sum_{i=1}^n d_i} \quad 3. \text{ egyenlet,}$$

ahol u az ismeretlen, becsülni kívánt alpopulációt jelöli.

Ez a méretbecslés azonban gyakran lehet pontatlan. A pontatlanságot nevezzük torzításnak, vagyis amikor a hálózati tulajdonságok alapján becsült csoportnagyság és a tényleges csoportnagyság között eltérés van.

A méretgenerátoros kérdések során a torzítás négy főbb fajtája ismert (Killworth et al. 2006, McCormick–Salganik–Zheng 2010, Laga–Bao–Niu 2021). A transzmissziós hiba (*transmission error*) akkor jelentkezik, ha a válaszadók nincsenek tudatában annak, hogy egy ismerősük egy bizonyos társadalmi csoportba tartozik. Például van olyan ismerősük, aki éppen munkanélküli, de ezt az információt nem tudják róla, és ezért az adott ismerőst nem veszik figyelembe a válaszadás során.

A korláthatás (*barrier effect*) abból adódik, hogy a válaszadók eltérő valószínűséggel ismernek különböző társadalmi csoportokba tartozó embereket. A társadalmi kapcsolathálóik jellemzője az erős homofília (McPherson–Smith-Lovin–Cook 2001), vagyis az emberek nagyobb valószínűséggel ismernek olyanokat, akik különböző tulajdonságaikban hozzájuk hasonlóak. Ha például megkérdezzük a válaszadókat arról, hogy hány olyan nőt ismernek, aki nemrég szült, akkor a harmincas éveikben járó női válaszadók jóval több embert fognak megnevezni, mint a nyolcvanas éveikben járó férfi válaszadók (McCormick–Salganik–Zheng 2010).

A felidézési hibáról (*recall error*) akkor beszélhetünk, ha a válaszadók nem tudják pontosan felidézni, hány embert ismernek az adott kategóriából. Például ismernek ugyan Milán nevű embert, de a kérdőíves adatfelvétel során ez nem jut eszükbe. Korábbi kutatások alapján a válaszadók hajlamosak a kérdőívben túlbecsülni azon ismerőseik számát, akik kis létszámú populációkba tartoznak, és alulbecsülni azon ismerőseik számát, akik nagy létszámú populációkba tartoznak (Zheng–Salganik–Gelman 2006, Laga–Bao–Niu 2021).

A válaszadási torzítás (*response bias*) abból adódik, hogy a válaszadók hajlamosak lehetnek szándékosan eltitkolni, hogy ismernek bizonyos társadalmi csoportokba tartozó embereket. Ismernek például olyan embert, aki használ drogokat, de a kérdőívben ezt az információt elhallgatják.

A szegregáció vizsgálata az összegző módszerrel

Az összegző módszer használható a társadalom szegregáltságának vizsgálatára is. Zheng és munkatársai (2006) egy olyan modellt alkottak, melyben a korláthatásból és az eltérő kapcsolathálózati méretekből fakadó túlszórást negatív binomiális modellek becslésével korigálták annak érdekében, hogy a nehezen elérhető alpopulációk létszámára pontosabb becslést adjanak. A modellen belül a túlszórás paramétere információt nyújt arról, hogy az egyes csoportok ismertségének mekkora a szórása a tár-

sadalmon belül, emiatt a társadalmi törésvonalak vizsgálatára is lehetőséget ad. Erről a módszerről magyarul részletes leírást nyújt Kmetty és Koltai (2015).

DiPrete és munkatársai (2011) úttörő cikkükben az Egyesült Államok társadalmának struktúráját elemezték ezzel a módszerrel. Arra voltak kíváncsiak, hogy az egyének társas kapcsolataiból kiindulva milyen következtetések vonhatóak le különböző társadalmi csoportok közötti viszonyokról. Mennyire jellemző például, hogy a vallásukat rendszeresen gyakorlók és az ateisták között személyes kapcsolatok alakulnak ki. Tanulmányukban tehát bizonyos társadalmi csoportok szegregáltságát ragadták meg.

Szegregáció alatt a szociológiai irodalomban azt értjük, amikor egyének bizonyos tulajdonságaik vagy magatartásmódjuk alapján csoportba rendezve egymással gyakrabban kerülnek interakcióba, a csoporton kívüli egyénnel pedig ritkábban kerülnek interakcióba, mint ahogyan azt a csoportnagyságaik alapján véletlenül várnánk (Duncan–Duncan 1955).

DiPrete és szerzőtársai elméleti kiindulópontja a társadalmi kapcsolatokra jellemző homofília jelensége volt. Eszerint a társas kapcsolatok nem véletlenszerűen alakulnak ki a populáció tagjai között, mivel különböző pszichológiai okok és a társadalmi körülmények hatására az egyének hajlamosabbak a hozzájuk hasonló személyekkel kapcsolatokat kialakítani, mint az eltérő tulajdonságú személyekkel (McPherson–Smith-Lovin–Cook 2001). Emiatt az egyének eltérő valószínűséggel ismernek különböző csoportokhoz tartozó embereket. A társadalmi szegregáció ezek alapján nemcsak elkülönülésként fogható fel, hanem az adott társadalmi csoportra jellemző homofil kapcsolatok arányaként is.

DiPrete és munkatársai (2011) vizsgálata szerint az amerikai társadalomban az etnikai elkülönülés mértéke hasonló a vallási, munkaerőpiaci státusz szerinti, és politikai ideológia szerinti elkülönülés mértékéhez. A melegek ennél jóval kevésbé elkülönülő csoportot alkotnak. Ugyanakkor az erősen liberális és erősen konzervatív beállítódásúak azonos mértékben szegregált csoportok. Kmetty (2015) magyar adatokon végzett elemzése nagyon hasonló eredményeket mutat. 2014-ben a jobb- vagy baloldali, illetve liberális vagy konzervatív orientációjú személyek közötti törésvonalak azonos mértékűek voltak.

Adatfelvétel

A következőkben bemutatásra kerülő elemzésekhez a Hungarostudy 2021 felmérés⁶ adatait használtuk (Susánszky–Székely–Székely 2022). A Hungarostudy kutatásorozat a lakosság egészségi állapotát vizsgálja, és olyan témákat mér fel, mint a lakosság fizikai és mentális egészségi állapota, valamint az egészség- és rizikómagatartás különböző formáinak elterjedtsége. Az adatfelvétel 2021 nyarán készült, és 7000 főt kérdeztek meg személyes formában, kérdezőbiztosok közreműködésével. A minta

6 A „Hungarostudy 2021” kutatás a Kopp Mária Intézet a Népesedésért és a Családokért és a Semmelweis Egyetem Magatartástudományi Intézet együttműködésében valósult meg 2021-ben, N = 7000. Kutatásvezetők: Purebl György, Engler Agnes.

nem, kor és lakóhely dimenziók mentén reprezentálja a Magyarországon élő, magyar állampolgárságú, 18 éves és idősebb népességet. Az általunk használt kérdés a társadalmi beágyazottság mérésére kialakított kérdésblokk része.

Elemzési stratégia

A hálózati méretgenerátorra vonatkozó kérdést az 1. táblázat szemlélteti. A transzmissziós hiba csökkentésének érdekében a válaszadók kapcsolathálózatának becslésére csak a nevekre adott válaszokat használjuk fel (McCormick–Salganik–Zheng 2010). A neveket Kmetty (2015) alapján választottuk. A válaszadók kapcsolathálóját a korábban ismertetett 2. egyenlet segítségével becsljük meg. Korábbi kutatásokhoz hasonlóan azokat a válaszokat, ahol harmincnál nagyobb létszámot adtak meg a válaszadók egy-egy név esetében, harmincra kódoljuk (Zheng–Salganik–Gelman 2006, McCormick–Salganik–Zheng 2010).⁷

1. táblázat: A hálózati méretgenerátorra vonatkozó kérdés a kérdőívben

A következő kérdések olyan személyekkel lesznek kapcsolatosak, akiknek ismeri a nevét, és ha valahol összefutnának, megállna velük beszélgetni. Kérem, mondja meg, hogy hány olyan embert ismer...

1. akit Juditnak hívnak.
2. akit Milánnak hívnak.
3. akit Krisztiánnak hívnak.
4. akit Barbarának hívnak.
5. akit Júliának hívnak.
6. akit Sándornak hívnak.
7. akiről tudja, hogy munkanélküli.
8. aki az elmúlt években részt vett tüntetésen.
9. akinek van nyaralója.
10. aki cigány származású.
11. aki rendszeresen jár istentiszteletre.
12. aki meleg vagy lesbikus.
13. aki átesett Covid-fertőzésen.
14. aki hajléktalan.
15. aki öngyilkosságot kísérelt meg.
16. aki tagja bármilyen típusú civil szervezetnek.
17. akivel, politikai, közéleti kérdéseket megbeszélne.

Megjegyzés: a 17. kérdés nem egy társadalmi csoportra vonatkozik, hanem annak felmérésére, hogy az egyének kapcsolathálóján belül mekkora azok aránya, akikkel politikai, közéleti kérdéseket megbeszélnének.

Elemzésünkben ezután különböző társadalmi csoportok – például a munkanélküliek, tüntetésen részt vevők, romák, Covid-fertőzöttek – esetében negatív binomiális modellek segítségével megbecsljük a túlszórás mértékét. Majd a méretgenerátoros módszer segítségével megbecsljük a csoport méretét, és összehasonlítjuk ezeket más adatforrásokban (népszámlálási adatok, adminisztratív adatok, kérdőíves adatfelvé-

⁷ A becsléni kívánt csoportokra adott válaszokat ugyanakkor nem korlátozzuk harmincra, előfordulhat ugyanis, hogy például egy roma válaszadó ismerőseinek a többsége szintén roma.

telek) található adatokkal. Ezek alapján az információk alapján következtetünk a csoportok társadalmon belüli láthatóságára, és a társadalmi törésvonalakra.

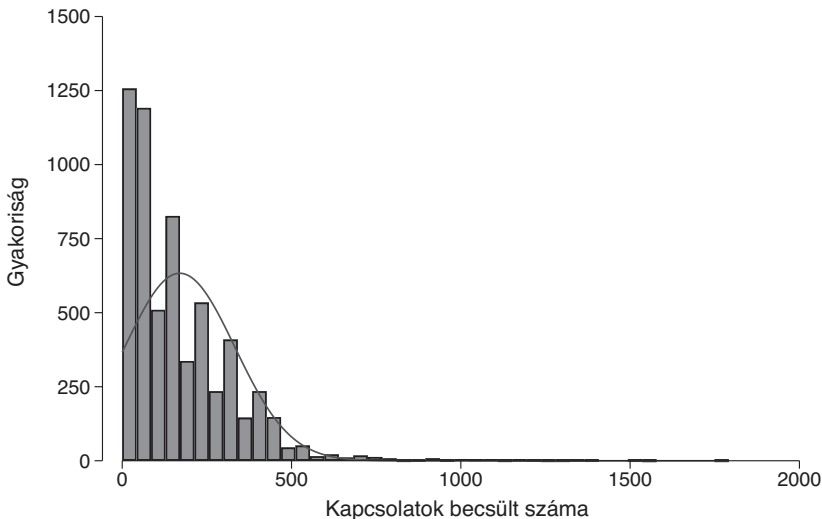
A negatív binomiális modellek becslésére a Stata programot használjuk. A csoportok méretének becslését a korábban bemutatott 3. egyenlet segítségével végezzük el, a networkreporting R csomagot (Feehan–Salganik 2016b) használva.

Eredmények

A becsült kapcsolatok száma

A válaszadók átlagos kapcsolathálójának mérete a mintánkban 170 (szórás: 163, medián: 138). Ez némileg alacsonyabb, mint a Kmetty (2015) által becsült átlagos kapcsolatháló-méret egy korábbi, 2014-es magyarországi adatfelvétel alapján (átlag = 195, szórás = 146), valamint a Kmetty és Koltai (2016) által becsült méret egy 2015-ös adatfelvétel alapján (átlag = 184). Ebben az alacsonyabb kapcsolatszámban könnyen lehet, hogy a Covid-járvány hatása mutatkozik meg. A személyes kapcsolatok változását vizsgáló kutatások szerint ugyanis a baráti kapcsolatok csökkenése (Dávid et al. 2021), valamint a bizalmi kapcsolatok átrendeződése figyelhető meg ebben az időszakban (Albert–Dávid 2022). Természetesen az eltérő becslések adódhatnak a minták eltéréseiből is, bár mindegyik kutatás országos reprezentatív mintát használt. Az összes eddigi magyarországi adatfelvétel becslése jelentős mértékben elmarad az amerikai becslésektől, ahol a kapcsolatszám mediánértékét 290 és 610 közötti értékekre becsülték attól függően, hogy a felidézési hibára korrigáltak-e a becslésben, vagy sem (McCarty et al. 2001, Zheng–Salganik–Gelman 2006, DiPrete et al. 2011).

1. ábra: A kapcsolatok becsült számának megoszlása a válaszadók között



Az 1. ábrán jól látszik, hogy a kapcsolatok száma nem normál eloszlást követ a társadalomban. Az eloszlás balra ferde, ami azt jelenti, hogy viszonylag sok embernek van kevés kapcsolata, és kevés embernek van sok vagy nagyon sok kapcsolata.

A kapcsolatok száma és a főbb demográfiai változók összefüggését a 2. táblázat szemlélteti. Két regressziós modellt becsültünk: az első modellben a kapcsolatok száma a függő változó, a második modellben a balra ferde eloszlás miatt egy transzformált függő változót használtunk: a kapcsolatok számának természetes alapú logaritmusát. Míg a férfiak és nők között nem találunk statisztikailag szignifikáns különbséget kontrollálva a többi demográfiai változóra, a kapcsolathálózatok mérete egyértelmű különbségeket mutat az iskolai végzettség, az életkor és a lakóhely településtípusa szerint. Az iskolai végzettség növekedésével az átlagos kapcsolatháló-méret növekszik. A középkorúak (30–59 évesek) között magasabb a kapcsolatháló átlagos mérete, mint a fiatalabbak és idősebbek között. Budapesthez képest mind a nagy- és kisvárosokban, mind a községekben magasabb a kapcsolatháló átlagos mérete.

Ezek az összefüggések jól illeszkednek a korábbi kutatások eredményeihez. Az iskolai végzettséget tekintve Kmetty és Koltai (2016) hasonló összefüggést talált a kapcsolati miliókat vizsgálva egy korábbi kutatásuk során. Továbbá a jelenlegi eredményekhez hasonlóan azt találták, hogy Budapesten alacsonyabb a gazdag, sokszínű kapcsolathálóval rendelkezők aránya, mint a községekben és más városokban, ugyanakkor Budapesten a kapcsolat nélküliek aránya is alacsonyabb volt.

2. táblázat: A kapcsolatok számának regressziós modellje a főbb demográfiai változók szerint

	(1) Kapcsolatok száma	(2) Kapcsolatok számának logaritmus
Nem (ref.kat.: férfi)		
Nő	-4,181 (4,473)	-0,043 (0,047)
Iskolai végzettség (ref.kat.: alapfok)		
Középfok	19,037*** (4,788)	0,187*** (0,053)
Felsőfok	41,944*** (7,250)	0,503*** (0,068)
Életkor (ref.kat.: 18–29 éves)		
30–39 éves	20,836** (7,627)	0,179* (0,077)
40–49 éves	25,977*** (7,548)	0,249** (0,077)
50–59 éves	31,136*** (7,804)	0,307*** (0,075)
60 év fölött	-0,576 (6,674)	-0,051 (0,074)
Településtípus (ref.kat.: Budapest)		
Megyeszékhely, megyei jogú város	44,324*** (7,419)	-0,032 (0,084)
Egyéb város	49,128*** (5,430)	0,317*** (0,069)
Község	68,644*** (5,947)	0,570*** (0,070)
Konstans	99,659*** (7,297)	3,875*** (0,090)
N	6063	6063
R ²	0,031	0,033

Megjegyzés: a kapcsolatok számának logaritmusát becsülő modell esetében a 0-ra becsült kapcsolatháló értékekhez 0,001-et hozzáadtunk, hogy a 0 becsült kapcsolattal rendelkező válaszadók esetében is tudjuk alkalmazni a transformációt. Zárójelben a standard hibák. *** p<0,001, ** p<0,01, * p<0,05, + p<0,1

Szegregációs mintázatok

Ahogy korábban ismertettük, a Zheng és munkatársai (2006) által használt negatív binomiális modellek túlszórás-paramétere információt nyújt arról, hogy az egyes csoportok ismertségének mekkora a szórása a társadalmon belül, emiatt a társadalmi törésvonalak vizsgálatára is lehetőséget ad. A paraméter értéke azt mutatja meg, hogy az egyes társadalmi csoportok ismertsége mennyire tér el a véletlentől, mennyire szegregált az adott társadalmi csoport a sokaságon belül (Moksony 2006, Zheng–Salganik–Gelman 2006, DiPrete et al. 2011, Kmetty–Koltai 2015). Minél kisebb a túlszórás értéke, az egyes csoportok ismertsége annál kiegyenlítettebb a lakosságon belül. Lehetőség van továbbá arra is, hogy megvizsgáljuk, egyes demográfiai változók bevonásának hatására hogyan változik a túlszórás-paraméter értéke, vagyis ezek a változók mennyiben magyarázzák a társadalmi törésvonalakat (Kmetty 2015).

A 3. táblázat szemlélteti a negatív binomiális modellben becsült túlszórás mértékét. Az első modell egy üres modell, melyben csak az egyének kapcsolatszámára kontrolláltunk. A további modellekben külön-külön a nemzet, iskolai végzettséget, életkort és településtípust vontuk be az elemzésbe, majd a végső modellben ezeket a kontrollváltozókat egyszerre szerepeltettük.

Lássuk először az üres modelleket. A Covid-fertőzésen átesettek, a nyaralótulajdonosok és a munkanélküliek esetében nem találunk nagy szórást a lakosságon belül a tekintetben, hogy milyen valószínűséggel ismernek ilyen csoportokba tartozó embereket a válaszadók. Ezeknek a csoportoknak az ismertsége viszonylag kiegyensúlyozott a magyar társadalmon belül. Az etnikai hovatartozás, a templomba járás, a homoszexualitás, illetve az öngyilkosságot megkísérlők tekintetében azonban nagyobb ez a szórás, ezen jellemzők mentén szegregáltabb a magyar társadalom. Különösen nagy a szórás a hajléktalanok, a tüntetésen részt vevők, és a civil szervezeti tagok ismertségét illetően.

Összehasonlításképpen, Kmetty (2015) a politikai beállítódások szerint vizsgálta a társadalmi törésvonalakat a méretgenerátoros módszerrel. Míg a bal- és jobboldali kapcsolatok esetében nem talált jelentős törésvonalakat (a túlszórás paraméter értéke 1,8 és 2,0), addig a liberális, konzervatív és radikális ideológiák alapján 5 és 6 közötti paraméterértéket becsült. Mintánkban a hajléktalanok, a tüntetésen részt vevők és a civil szervezeti tagok esetében is meghaladja a túlszórás a 6-os értéket. Vagyis ezen csoportok ismertsége a társadalmon belül jelentősen kiegyenlítetlen.

Az egyes demográfiai változókat vizsgálva azt látjuk, hogy bizonyos társadalmi csoportok esetében iskolai végzettség (tüntetésen részt vevők, romák, melegek és leszbikusok, civil szervezeti tagok), életkor (melegek és leszbikusok) és településtípus (tüntetésen részt vevők, romák, melegek és leszbikusok, hajléktalanok) szerint látunk egyfajta szegregációt, ezen változók bevonásával jelentősebben csökken a túlszórás mértéke. A magasabb iskolai végzettségűek például több tüntetésen részt vevőt, melegeket és leszbikusokat, valamint civil szervezeti tagot, de kevesebb romát ismernek, mint az alacsony iskolai végzettségűek. A Budapesten lakók is több tünte-

tésen részt vevőt, melegét és leszbikusot, valamint hajléktalant, de kevesebb romát ismernek, mint az egyéb városokban és községekben lakók. A melegek és leszbikusok ismertsége a magasabb korcsoportokban alacsonyabb.

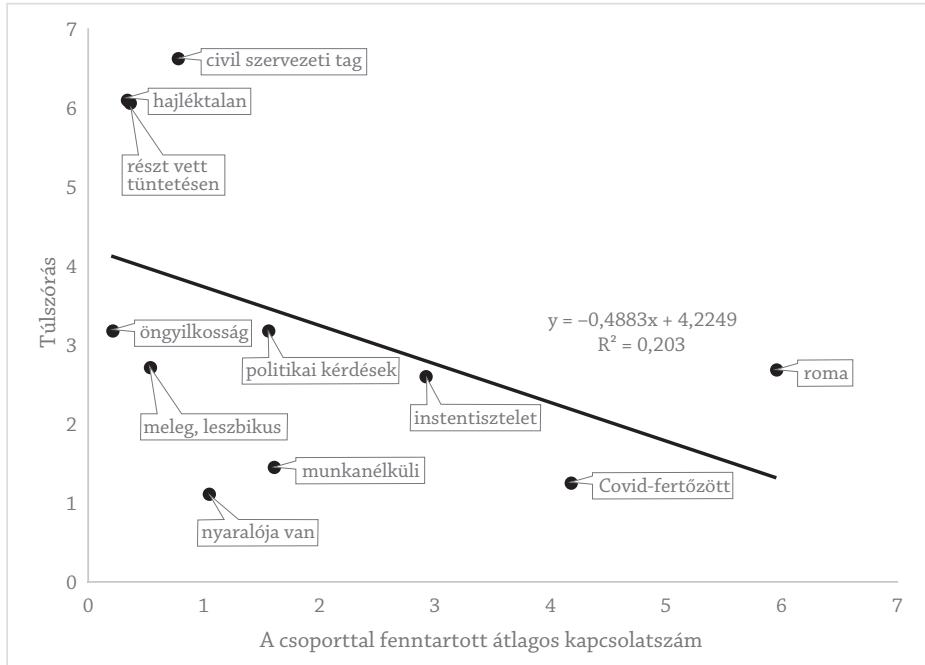
3. táblázat: A túlszórás-paraméter értéke a negatív binomiális modellekben

	(1) Üres modell	(2) Nem	(3) Iskolai végzettség	(4) Kor	(5) Település-típus	(6) Minden kontroll
Munkanélküli	1,4	1,4	1,4	1,4	1,4	1,3
Részt vett tüntetésen	6,1	6,1	5,3	5,9	5,2	4,7
Nyeralója van	1,1	1,1	0,9	1,1	1,0	0,9
Roma	2,7	2,7	2,4	2,7	2,3	2,2
Rendszeresen jár istentiszteletre	2,6	2,5	2,6	2,5	2,4	2,3
Meleg, leszbikus	2,7	2,7	2,5	2,4	2,2	1,9
Covid-fertőzött	1,2	1,2	1,2	1,2	1,2	1,2
Hajléktalan	6,1	6,1	6,1	6,1	5,3	5,2
Öngyilkosságot kísérelt meg	3,2	3,2	3,2	3,1	3,1	3,0
Civil szervezeti tag	6,6	6,6	6,2	6,5	6,6	6,0
Politikai kérdéseket megvitatna vele	3,2	3,2	3,2	3,1	3,1	3,0

Megjegyzés: minden modellben kontrollálunk az egyének kapcsolatszámára. Az egyes oszlopokban számolt túlszórás-paraméterek az adott kontrollváltozók bevonása melletti paraméterértéket mutatják.

A 2. ábra az adott társadalmi csoportokkal fenntartott átlagos kapcsolatszám és a túlszórás közötti összefüggést mutatja. Kmetty (2015) Blau (1994) alapján amellel érvel, hogy egy adott csoport nagysága összefügg a túlszórás mértékével: a nagyobb csoportok esetében kisebbek a törésvonalak, míg a kisebb csoportok esetében nagyobbak. Az ábrán a regressziós vonal felett helyezkednek el azok a csoportok, amelyeknél a mérhetőből adódó törésvonalakon felül is van további elválasztó erő. Ez alapján a már korábban is említett hajléktalanok, tüntetésen részt vevők és civil szervezeti tagok, valamint a romák csoportja különösen szegregált a társadalomban.

2. ábra: A csoporttal fenntartott átlagos kapcsolatszám és a túlszórás közötti összefüggés



E csoportok magasabb szegregáltságára többféle magyarázat is adható. Egyrészt állhatnak a szegregáltság mögött tényleges interakciós korlátok: ahogy a 3. táblázatban szereplő regressziós modellek is mutatják, az iskolai végzettség és a lakóhely településtípusa szerepet játszik abban, hogy a válaszadók találkoznak-e ezekben a csoportokba tartozó emberekkel, és ha találkoznak is, interakcióba lépnek-e velük. A csoportok szegregáltságának mértékét tehát befolyásolja a tényleges térbeni és társadalmi elkülönülés.

Másrészt azonban a túlszórás-paraméter értékét, tehát a becsült szegregáció mértékét befolyásolhatják a méretgenerátoros technikából adódó torzítások is. A transzmissziós hiba akkor jelentkezik, ha a válaszadók nincsenek tisztában ismerőseik bizonyos tulajdonságaival. Éppen ezért, amikor azt kérdezzük tőlük például, hogy hány civil szervezeti tagot ismernek, akkor ezen ismerőseik számát alulbecslik, mert ezt az információt nem tudják a kapcsolathálójukhoz tartozókról. A transzmissziós hiba akkor jelentős, ha az adott csoporthoz tartozó személyek (például civil szervezeti tagok) inkább tisztában vannak azzal, hogy ismerőseik közül kik a civil szervezeti tagok, mint a csoporton kívüliek (a nem civil szervezeti tagok).

Ez a fajta torzítás jelen lehet a civil szervezeti tagok és a tüntetési résztvevők szegregációjában. A civil és politikai aktivitás olyan tulajdonság lehet, amit az emberek inkább tudnak egymásról akkor, ha közös civil vagy politikai aktivitásban vesznek részt, vagy ha általában a közéleti aktivitás fontos számukra. A hajlékta-

lanok, valamint a romák esetében ez a magyarázat kevésbé állja meg a helyét, ezek könnyebben észlelhető tulajdonságok. Fontos megjegyezni, hogy a transzmissziós hiba statisztikai jelentőségén túl szociológiai szempontból is érdekes jelenséget ragad meg. Ugyanis a transzmissziós hiba egy tulajdonság, vagy magatartás gyengébb társadalmon belüli láthatóságát mutatja, vagy másként fogalmazva arra utal, hogy a társadalom tagjai nem tartják számon egymásról az adott ismérvet. Az eredményeink fényében azt gondoljuk, hogy a civil és politikai aktivitás olyan magatartásformák, amelyeket a magyar társadalom tagjai nem tartanak számon.

A transzmissziós hibához hasonlóan, a felidézési hiba (*recall error*) is befolyásolhatja a túlszórás mértékét, ha a csoporthoz tartozók könnyebben idézik fel a válaszadás során, hogy kik tartoznak az adott csoporthoz, mint a csoporthoz nem tartozó személyek.

Fontos azonban megjegyezni, hogy a 2. ábra esetében a csoport nagyságát a csoporttal fenntartott átlagos kapcsolatszámmal mértük. Korábban bemutattuk, hogy többféle torzítás is vezethet ahhoz, hogy a kérdőívben bevallott, a csoporttal fenntartott kapcsolatszám alacsonyabb lehet a ténylegesnél, és ezáltal a csoport nagyságát a méretgenerátoros technikával alulbecsüljük. A következő fejezetben rátérünk a csoportméretbecslések pontosságának vizsgálatára.

Nehezen elérhető csoportok becsült mérete

A 4. táblázatban egyrészt a korábban bemutatott 3. egyenlet segítségével megbecsüljük az adott társadalmi csoport nagyságát, másrészt összevetjük ezeket a becsléseket más adatforrásokban (adminisztratív adatok, kérdőívek) szereplő létszámbecslésekkel.

4. táblázat: A különböző csoportok becsült mérete, a csoportok mérete más adatforrások alapján, valamint a két becslés aránya

	Becsült létszám	Külső adatforrások alapján becsült létszám	A becsült létszám és a külső létszámforrás aránya ¹¹
Munkanélküli	97 002	189 000 ¹	0,51
Hajléktalan	20 258	12 347 ²	1,64
Roma	354 540	315 583 ³ – 600 000 ⁴	0,59
Rendszeresen jár istentiszteletre	165 701	kb. 820 000 ⁵	0,20
Meleg, leszbikus	32 354	kb. 280 000 – 450 000 ⁶	0,07
Covid-fertőzött	241 290	809 646 ⁷	0,30
Öngyilkosságot kísérelt meg	12 458	16 282 ⁸	0,77
Részt vett tüntetésen	20 336	kb. 240 000 ⁹	0,08
Civil szervezeti tag	42 341	kb. 1 600 000 ¹⁰	0,03

1 Forrás: a KSH nyilvántartásában szereplő érték az adatfelvétel idején, 2021 júliusában <https://www.ksh.hu/docs/hun/xftp/gyor/mun/mun2107.html>

2 A 2011-es népszámlálás adatai alapján, lásd Györi (2020).

3 A 2011-es népszámláláskor magukat romának vallók aránya, forrás: KSH, https://www.ksh.hu/docs/hun/xftp/idoszaki/nepsz2011/nepsz_09_2011.pdf

4 Szociológiai kutatások a roma lakosság arányát a magyarországi népesség 5-6%-ára becslik (Kemény–Janky 2006).

5 Az ESS 2018-as adatai és a KSH adatai (https://www.ksh.hu/stadat_files/nep/hu/nep0003.html) alapján számított érték: legalább hetente jár istentiszteletre.

6 Alsó becslés: a Williams Institute 11 kérdőíves kutatás adatai alapján úgy becsli, hogy az amerikai népességen belül 3,5% a meleg vagy leszbikus emberek aránya. A magyarországi becsléshez a KSH adatait használtuk még (https://www.ksh.hu/stadat_files/nep/hu/nep0003.html). (How Many People are Lesbian, Gay, Bisexual, and Transgender? – Williams Institute (ucla.edu)). Felső becslés: a Gallup 2020-ban készült kérdőíves kutatása alapján úgy becsli, hogy az amerikai népességen belül 5,6% a meleg vagy leszbikus emberek aránya. A magyarországi becsléshez a KSH adatait használtuk még (https://www.ksh.hu/stadat_files/nep/hu/nep0003.html). (LGBT Identification Rises to 5.6% in Latest U.S. Estimate [gallup.com].)

7 A regisztrált fertőzöttek száma 2021. augusztus 2-án, forrás: <https://koronavirus.gov.hu>

8 Forrás: Országos Mentőszolgálat, az öngyilkossági kísérletek száma 2020. szeptember 1. és 2021. augusztus 31. között.

9 Forrás: Szabó–Gerő 2022 és a KSH adatai (https://www.ksh.hu/stadat_files/nep/hu/nep0003.html) alapján számított érték.

10 Forrás: Szabó–Gerő 2022 és a KSH adatai (https://www.ksh.hu/stadat_files/nep/hu/nep0003.html) alapján számított érték.

11 A becsült létszám és a külső létszámforrás arányának számításakor a külső források esetén a magasabb becsléseket használtuk.

A csoportok méretgenerátoron alapuló becsült létszáma érdekes mintázatot mutat. A romák esetében például a becsült létszám (354 540) jól közelíti azt a létszámot, amelyet a 2011-es népszámlálásban látunk (315 583), de jól közelíti a Hungarostudy 2021 felmérés adataiból becsült létszámot is (409 228). Itt fontos azonban megjegyezni, hogy feltehetően a népszámlálás is alulbecsli a romák tényleges létszámát. Egyrészt az etnikai-nemzetiségi kérdésekre nem kötelező válaszolni, másrészt a válaszadók között is lehetnek, akik inkább nem vallják be, hogy romaként identifikálják magukat. Más kutatások a roma lakosság arányát a magyarországi népesség 5-6 százalékára becslik (Kemény–Janky 2006). Ennek nagyjából fele az általunk kapott becslés. Ehhez hasonlóan a munkanélküliek méretgenerátorral becsült létszáma is (97 002) körülbelül fele a 2021. júliusi regisztrált munkanélküliek létszámának (189 000).

A hajléktalanok számát a módszer valamivel felülbecsli a 2011-es népszámlálás során becsült hajléktalanlétszámhoz képest (20 258 vs. 12 347), de itt lehetséges, hogy a népszámláláskor sem sikerült minden hajléktalant felkutatni. A hajléktala-

nok felülbecslése mögött továbbá az a hatás is állhat, hogy a válaszadók hajlamosak túlbecsülni azon ismerőseik számát, akik kis létszámú populációkba tartoznak (Zheng–Salganik–Gelman 2006, Laga–Bao–Niu 2021). A vizsgált csoportok közül a hajléktalanoké volt a legkisebb csoport.

A Covid-fertőzésen átesettek esetében már jóval nagyobb eltérést látunk a becsült (241 290) és az adminisztratív adatokból származó létszám között (809 646, 2021. augusztus 2-án), még úgy is, hogy az adminisztratív adatok is csak a regisztrált fertőzöttek létszámáról adnak tájékoztatást, a tényleges fertőzöttségről – ami valószínűleg jóval magasabb volt – nem. Hasonlóan nagy az eltérés a rendszeresen istentiszteletre járó emberek méretgenerátoros módszerrel becsült (165 701) és egyéb források alapján számított létszámadatai között (kb. 820 000).

A méretgenerátoros módszer valószínűleg alulbecsli az öngyilkosságot megkísérlők számát is. Az Országos Mentőszolgálat adatai szerint 16 282 ember kísérelt meg az adatfelvételt megelőző egy évben öngyilkosságot. Ezzel szemben a modell becsülése alapján a becsült létszám ennél valamivel alacsonyabb, 12 458 fő. Itt meg kell jegyeznünk, hogy a kérdőívben nemcsak azokra kérdeztünk rá, akik az elmúlt egy évben kíséreltek meg öngyilkosságot, hanem akik valaha így tettek.

A legnagyobb eltérések a méretgenerátorral becsült és az alternatív források alapján becsült csoportlétszámok között a tüntetésen részt vevők, a civil szervezeti tagok és a melegek, leszbikusok csoportja esetében tapasztalhatóak. A méretgenerátorral valamivel több mint 20 ezer tüntetőt becslünk, a kérdőíves kutatások alapján ez nagyjából ennek tizenkétszeresére (240 000) tehető.⁸

A civil szervezeti tagság mérése sokkal nehezebb feladat. Míg a válaszadók nagyjából tudják, milyen eseményeket kell a tüntetés, demonstráció alatt gondolniuk, a civil szervezeti tagságról ez nem mondható el. A Szabó és Gerő (2022) által végzett kutatásban a válaszadóknak összesen tízféle civil szervezetről kellett megmondaniuk, hogy kötődnek-e hozzájuk vagy sem. Így a civil szervezeti részvétel 20 százalékos aktivitási szintet mutat⁹ (több mint 1,6 millió fő). A civil szervezeti tagok méretgenerátorral becsült száma 42 ezer fő, ami két nagyságrenddel kisebb, mint azt más adatforrások alapján a téma kutatói becslik.

A melegek és leszbikusok arányát az amerikai populáción belül 3,5-5,6 százalékra becslik. Saját becsléseinkhez ezeket az adatokat használtuk, mivel a magyar társadalomra vonatkozó pontosabb becslések nem léteznek. Ez alapján kb. 280-450 ezer homoszexuális ember élhet Magyarországon, a méretgenerátoros becslésünk ennél jóval kevesebb embert mutat. Mindössze 32 ezer főt, tehát ez a becslés lényegesen alatta marad az alternatív forrásokból származó csoportnagyságoknak.¹⁰

8 Szabó és Gerő (2022) a tüntetésen való részvételt a standardnak tekinthető „előfordult-e már Önnel az elmúlt évben, hogy részt vett demonstráción (tüntetésen)?” kérdéssel mérték, ami lényegében megegyezik a European Social Survey (ESS) kutatási programban használt kérdéssel. Az ESS 2002 óta végzett kérdőíves felméréseiben rendre 3 százalék körüli részvételi arányt mérünk, ami megfelel a körülbelül 240 000 fő tüntetőnek.

9 Ifjúsági oktatási, egészségügyi, valamint a környezetvédelemmel, szociális ügyekkel foglalkozó szervezetek.

10 Itt szeretnénk megköszönni Takács Juditnak a segítségét az adatok és irodalmak megtalálásában.

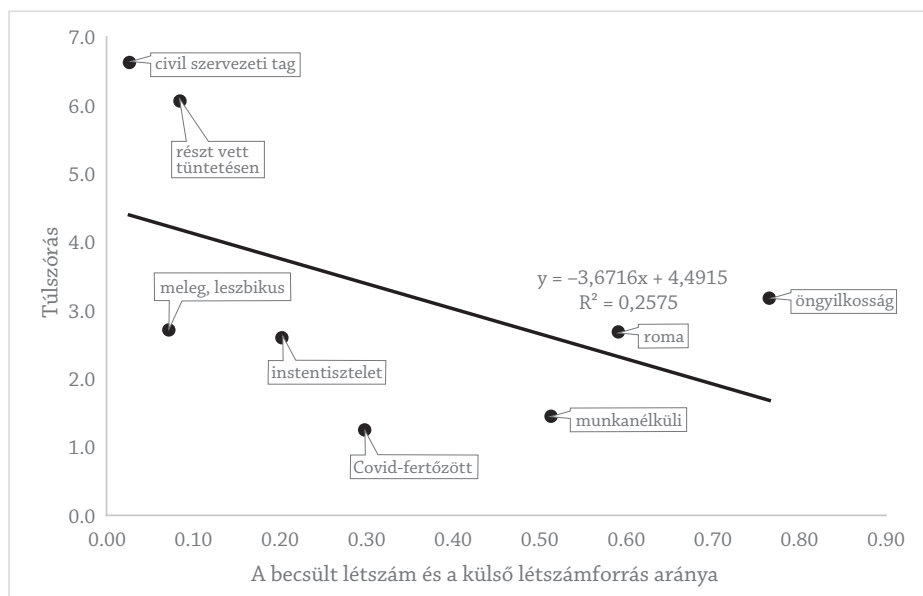
Az adatok azt mutatják, hogy a méretgenerátoros módszerrel becsült csoportlétszám elsősorban a tartós és jól látható tulajdonságok esetében képes nagyságrendileg jól közelíteni a tényleges létszámhoz (romák, hajléktalanok). A többi csoportra vonatkozó alacsonyabb létszámbecslés mögött többféle magyarázat is állhat. Az egyik ok lehet az alacsony láthatóság: a Covid-fertőzésen való átesettség esetében például plazibilisnek tűnik, hogy a válaszadók nem tudják minden ismerősükről, hogy a lekérdezés előtti másfél évben átestek-e a fertőzésen, hiszen az ismerőseik nagy részével ezen időszakban valószínűleg nem találkoztak vagy még csak nem is kommunikáltak. Szintén a láthatósághoz kapcsolódik, hogy bizonyos, kevésbé látható tulajdonságokat (például melegek és leszbikusok, öngyilkosságot megkísérlők) az érintettek nem fednek fel ismerőseik előtt, vagy csak bizonyos ismerőseik előtt fedik fel őket (Park 2021). Egy másik ok lehet a válaszkonformitás, tehát ha a válaszadók nem akarják a kérdőívben felfedni, hogy ismernek bizonyos csoportokhoz tartozó egyéneket.

Kmetty és Koltai (2015) a kapcsolathálózati mérőeszközökről szóló elméleti és metodológiai cikkükben felhívják a figyelmet arra, hogy – mint azt fentebb mi is bemutatunk – a méretgenerátoros módszer rendszerint alulbecsli a csoportok méretét. A látencia felelős lehet a nagyobb fokú szegregáltságért is, hiszen, ha egy csoport tagjai nehezen azonosíthatóak, akkor valószínűbb, hogy a csoporttagok a csoporton kívüliek számára kevésbé láthatóak, mint a csoporton belüliek számára, és így a túlszórás is magasabb. Ez alapján tehát negatív összefüggés feltételezhető a becslés pontossága (alacsony látencia / magas láthatóság) és a szegregáltság között: minél kevésbé tudunk pontos létszámbecslést adni a csoport méretére, annál szegregáltabbnak tűnhet a csoport.

Ennek vizsgálatára a 3. ábrán felrajzoltuk a csoportok létszámbecslésének pontossága és a túlszórás paramétere közötti kapcsolatot. A csoportok létszámbecslésének pontosságát a méretgenerátoros módszerrel becsült létszám és a külső források alapján becsült létszám arányával mérjük. A mutató értéke 1, ha a becsült érték és a külső forrásból származó adat pontosan megegyezik. Minél kisebb a hányados, annál inkább alulbecsülünk a csoport méretét.

Ha az outlier értékek tekinthető hajléktalanokat kihagyjuk, a pontosság és a túlszórás között valóban negatív kapcsolatot figyelhetünk meg. A méretgenerátoros módszer során fellépő torzítások tehát mind a nehezen elérhető társadalmi csoportok létszámának becslésekor, mind az adott csoport szegregáltságának vizsgálatakor problémát okozhatnak. A következő fejezetben javaslatokat fogalmazunk meg a további kutatásokhoz arra, hogyan lehet ezeket a torzításokat csökkenteni az adatfelvétel vagy az elemzés során.

3. ábra: A becslült létszám és a külső létszámforrás aránya, valamint a túlszórás közti kapcsolat



Megjegyzés: az ábrán nem szerepel az outlier értékkel rendelkező hajléktalan csoport.

A méretgenerátoros módszer során fellépő torzítások csökkentésének lehetőségei

Tanulmányunkban a társadalmi csoportok becslését a méretgenerátoros technikán belül leggyakrabban alkalmazott Killworth-féle módszerrel (Killworth–Johnsen et al. 1998, Killworth–McCarty et al. 1998) illusztráltuk. Az elmúlt években többféle előrelépés is történt az alcsoportok becslésének pontosítására, melyek, mint fentebb bemutattuk, a csoportok szegregáltságának pontosabb becslését is segítik.

A különböző torzításokat a méretgenerátoros módszer különböző irányzatai eltérő módon próbálják kezelni. Míg egyes irányzatok a kutatási dizájn megtervezésével kapcsolatban fogalmaznak meg olyan javaslatokat, melyek az egyéni kapcsolatháló, valamint az egyes társadalmi csoportok méretének pontosabb becsléséhez vezetnek (például McCormick–Salganik–Zheng 2010, Feehan–Salganik 2016a), addig más, főleg bayesi becsléseket kidolgozó irányzatok a statisztikai modellen belül próbálják ezeket kezelni (például Maltiel et al. 2015, Teo et al. 2019).

A torzításokat a statisztikai modellen belül kezelő bayesi becslési eljárások jóval idő- és számításgényesebbek, mint a hagyományos statisztikai módszerek. Nagy létszámú adatfelvételek esetén tapasztalataink alapján ez problémát jelenthet.¹¹ Ezért érdemes ezeket a torzításokat már a kutatási dizájn megtervezésekor figyelembe venni.

¹¹ Maltiel és munkatársai (2015) nyomán például megpróbáltunk a transzmissziós hibára és a korláthatásra korrigált bayesi becslést is adni a csoportméretekre. Ehhez az NSUM R csomagot (Maltiel–Baraff 2015) használtuk. Mivel azonban ez a módszer minden egyén kapcsolathálóját is random hatásként becsli MCMC eljárással, a 7000 válaszadó tartalmazó adatbázisunk esetében a kapott adatfájl 17 GB volt egyetlen csoport nagyságának becslésére és csak egy Markov-láncot becsülve. Amikor a konvergenciakritériumok ellenőrzéséhez négy párhuzamos láncot próbáltunk becsülni, ez meghaladta egy jobb minőségű laptop számítási kapacitását (abban az esetben is, ha az elemzést egy 1000 fős almintán végeztük el).

Az egyéni kapcsolathálók pontosabb becslésére McCormick és munkatársai (2010) fogalmaztak meg javaslatokat. A szerzők azt tanácsolják, hogy az ismert létszámú társadalmi csoportok közül csak keresztneveket használjunk, mert a nevek használata önmagában kiküszöböli a transzmissziós hibából fakadó torzítást. Emiatt becsültük az egyéni kapcsolathálót mi is csak a nevek alapján. További javaslatuk, hogy a keresztneveket a populációs előfordulásuk szerinti gyakoriságnak megfelelően szerepeltessük a kérdőívben. Ha például a teljes népesség 20 százaléka harminc év alatti nő, akkor a kérdőívben szereplő nevek 20 százaléka legyen olyan női keresztnév, ami jellemző, gyakori a harminc év alatti nők körében. Javasolják továbbá, hogy olyan neveket válasszunk, amik a populáció 0,1-0,2 százalékát teszik ki, valamint hogy kerüljük azokat a neveket, amelyeknél gyakori a becenevek használata, ami megnehezítheti az adott név felidézését. Érdeemes lehet kerülni azokat a neveket is, amelyeknek egymáshoz hasonló, de különböző változatai vannak (például Vera és Veronika, vagy hasonló okok miatt az általunk, illetve korábban Kmetty [2015] által használt Júlia név helyett is érdemes lehet másik nevet választani a Julianna névváltozat miatt). Ha a kérdőíves adatfelvétel megengedi, akkor érdemes továbbá az általunk használt hat névnél több névre is rákérdezni, ahogyan azt az amerikai adatfelvételekben tették (például DiPrete et al. 2011).

McCormick és munkatársai (2010) a korláthatásból fakadó torzítást úgy küszöbölik ki, hogy egy olyan negatív binomiális modellt használnak, ami expliciten megbecsüli, hogy egy válaszadó az e egőcsoportból milyen valószínűséggel ismer valakit az a altercsoportból (ez az ún. *látens, nem random keveredési modell*). Ehhez az szükséges, hogy a populációt kölcsönösen kizáró csoportokra osszuk (például nem és korcsoportok alapján: 0–20 éves férfiak, 0–20 éves nők, 21–40 éves férfiak, 21–40 éves nők), és ismerjük az egyes nevek előfordulási valószínűségét az egyes csoportokban (McCormick–Salganik–Zheng 2010). Esetünkben csak a teljes magyar lakosságra vonatkozóan álltak rendelkezésre a nevek gyakoriságára vonatkozó adatok, a korcsoportos gyakoriságokat az egyes években adott keresztnemek összegzésével lehetne kiszámolni.

Ha a kutatás célja a nehezen elérhető csoportok létszámának vagy a szegregáció mértékének becslése, akkor javasoljuk az *általánosított méretgenerátor* (Feehan–Salganik 2016a) használatát. Ez egy olyan becslőfüggvény, ami jelentős mértékben csökkenti a korláthatásból és a transzmissziós hibából fakadó torzításokat. Ehhez arra van szükség, hogy a kérdezést kiegészítsük egy, az ismeretlen alpopulációból vett mintával is. Az alpopuláció tagjaitól információt gyűjtünk arról, hogy ismerőseik mekkora hányada tudja róluk azt, hogy ők ebbe az alpopulációba tartoznak. Ez alapján becsülni tudjuk az alpopuláció láthatóságát, és ezzel az információval korrigálni lehet az alpopuláció méretére vonatkozó becslést. Az általunk alkalmazott adatfelvétel során sajnos nem volt lehetőség egy ilyen bővített méretgenerátoros blokk lekérdezésére.

Összegzés

A méretgenerátoros technikát alkalmazó korábbi magyarországi vizsgálatok fókuszából eddig hiányzott, vagy kisebb hangsúlyt kapott az egyes társadalmi csoportméretek becslésének értékelése, a társadalmi csoportok szélesebb körére vonatkozó szegregációs minták leírása, valamint az előző két szempont összekapcsolása, vagyis a csoportok létszámbebecslésének pontossága és azok szegregáltóságának empirikus összevetése.

Vizsgálatunkat a Hungarostudy 2021 kutatás 7000 fős személyes megkérdezésen alapuló reprezentatív adatain végeztük. Elemzésünk első részében megállapítottuk, hogy a válaszadók átlagos kapcsolathálójának becsült mérete 170 fő (medián: 138 fő), ami valamivel alacsonyabb, mint amit korábbi (2014-es és 2015-ös) kutatások találtak (Kmetty 2015, Kmetty–Koltai 2016). Erre a változásra plauzibilis magyarázatot adhat, hogy a Covid-járvánnyal összefüggésben többen megállapították a személyes kapcsolatok csökkenését (Dávid et al. 2021, Albert–Dávid 2022). Az egyéni kapcsolatháló nagyságát magyarázó regressziós modellek eredményei szerint a magas iskolai végzettségűekre, a középkorúakra (30–59 évesek), valamint a kisebb városokban és községekben élőkre jellemző a magasabb kapcsolatszám.

Elemzésünk második részében bemutattuk a vizsgált csoportok magyar társadalmon belüli szegregáltóságának mértékét a méretgenerátoros módszerrel becsülve. A szegregáció megragadására negatív binomiális regressziós modelleket becsültünk, és ezek túlszórás-paraméterét használtuk (DiPrete et al. 2011, Kmetty–Koltai 2015). Eredményeink szerint a legmagasabb szegregáltósági mutatóval rendelkező csoportok a kérdőívben vizsgáltak közül a hajléktalanok, a civil szervezeti tagok, valamint a tüntetésen részt vevők. Később pedig megmutattuk, hogy ezen csoportok esetében a szegregáltóság mértéke lényegesen meghaladja a csoportok méretéből fakadó szegregáltóságot.

A következő lépésben a modellekbe társadalmi háttérváltozókat vontunk be, hogy lássuk, a szegregáltóságért mely társadalmi dimenziók felelhetnek. Elmondható, hogy a tüntetésen részt vevők, a romák, a melegek és leszbikusok, valamint a civil szervezeti tagok szegregációja részben az iskolai végzettség és a lakóhely településtípusa mentén alakult ki. A magasabb iskolai végzettségűek több tüntetőt, meleget és leszbikusot, valamint civil szervezeti tagot, de kevesebb romát ismernek, mint az alacsony iskolai végzettségűek. A Budapesten lakók is több tüntetésen részt vevőt, meleget és leszbikusot, valamint hajléktalant, de kevesebb romát ismernek, mint az egyéb városokban és községekben lakók. A melegek és leszbikusok ismertsége az idősebb korcsoportokban alacsonyabb. Ezek alapján tehát azt látjuk, hogy a vizsgált közéleti aktivitási formák és a kisebbségi csoportok elkülönülése elsősorban a budapesti / nem budapesti lakóhely, az iskolai végzettség, valamint az életkor mentén történik.

Ezt követően a nehezen elérhető csoportok méretét becsültük meg az összegző módszerrel. A kapott létszámbebecsléseket összevetettük különböző külső adatforrásokból származó adatokkal, illetve korábbi, elfogadott bebecslésekkel. Megállapítottuk, hogy a külső forrásokhoz leginkább illeszkedő bebecslést a méretgenerátoros módszer elsősorban a tartós, jól látható tulajdonságokkal rendelkező társadalmi csoportok

esetén ad (például hajléktalanok, romák). A legpontosabb becslést pedig a civil és politikai aktivitást leíró csoportoknál (civil szervezeti tagok, tüntetésen részt vevők) kaptuk. Bemutattuk, hogy a méretgenerátoros technika során fellépő torzítások hogyan befolyásolhatják egyrészt a nehezen elérhető társadalmi csoportok becslését, másrészt a csoportok társadalmon belüli szegregáltságának becsült mértékét. Végül ismertettük e torzítások kezelésére megalkotott módszertani újításokat, és javaslatokat tettünk arra, hogy az összegző módszert használó jövőbeli kutatások hogyan tervezzék meg a kutatási dizájnt, hogy pontosabb becsléseket érjenek el.

A bemutatott eredmények alapján úgy véljük, hogy a méretgenerátoros, összegző módszer számos új eredményt hozhat a társadalom szegregáltságának/integráltságának vizsgálatában. Továbbá a kapott eredményeink alapján úgy tűnik, hogy a jövőbeli kutatásokban érdemes volna mélyebben is megvizsgálni a szegregáció politikai dimenzióit, valamint a politikai részvétel és polarizáció mögött meghúzódó kapcsolathálózati mechanizmusokat.

Irodalom

- Albert, F. – Dávid, B. (2012): Az interperszonális kapcsolathálózati struktúra átrendeződése Magyarországon. In Kovách, I. – Dupcsik, C. – P. Tóth, T. – Takács, J. (szerk.): *Társadalmi integráció a jelenkori Magyarországon*. Budapest: Argumentum Kiadó, 343–356.
- Albert, F. – Dávid, B. (2022): A bizalmas kapcsolathálózatok változása a pandémia idején. In Purebl, G. – Susánszky, A. – Susánszky, P. (szerk.): *Csak az egészségg meglegyen: Tanulmányok a mentális egészségről és egészségmagatartásról*. Budapest: Semmelweis Kiadó, 7–19.
- Bernard, H. R. – Johnsen, E. C. – Killworth, P. D. – Robinson, S. (1991): Estimating the size of an average personal network and of an event subpopulation: Some empirical results. *Social Science Research*, 20(2): 109–121. [https://doi.org/10.1016/0049-089X\(91\)90012-R](https://doi.org/10.1016/0049-089X(91)90012-R)
- Blau, P. M. (1994): *Structural Contexts of Opportunities*. University of Chicago Press.
- Coleman, J. S. (1988): Social capital in the creation of human capital. *American Journal of Sociology*, 94 95–121.
- Dávid, B. – Albert, F. – Huszti, É. (2020): Kapcsolati tőke és földrajzi mobilitás. In Kovách, I. (szerk.): *Mobilitás és integráció a magyar társadalomban*. Budapest: Argumentum Kiadó, 163–192.
- Dávid, B. – Herke, B. – Huszti, É. – Tóth, G. – Túry-Angyal, E. – Albert, F. (2021): A pandémia hatása a személyes kapcsolathálózati struktúrára. Otthoni munkavégzés: védelem a vírustól vagy társas izoláltság? *Scientia et Securitas*, 2(3): 350–357. <https://doi.org/10.1556/112.2021.00056>
- DiPrete, T. A. – Gelman, A. – McCormick, T. – Teitler, J. – Zheng, T. (2011): Segregation in Social Networks Based on Acquaintanceship and Trust. *American Journal of Sociology*, 116(4): 1234–83. <https://doi.org/10.1086/659100>

- Duncan, O. D. – Duncan, B. (1955): A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20(2): 210–217. <https://doi.org/10.2307/2088328>
- Feehan, D. M. – Salganik, M. J. (2016a): Generalizing the Network Scale-up Method: A New Estimator for the Size of Hidden Populations. *Sociological Methodology*, 46(1): 153–186. <https://doi.org/10.1177/0081175016665425>
- Feehan, D. M. – Salganik, M. J. (2016b): Package ‘networkreporting’. Tools for using Network Reporting Estimators. <http://cran.r-project.org/package=networkreporting> (elérés: 2022. március 15.)
- Granovetter, M. (1983): The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 1 201–233. <https://doi.org/10.2307/202051>
- Győri, P. (2020): Idősorok a hazai hajléktalanságról. In Kolosi T. – Széleányi I. – Tóth I. Gy. (szerk.): *Társadalmi riport 2020*. Budapest: TÁRKI, 332–357.
- Huszár, Á. (szerk.). (2015): *A társadalom rétegződése*. Budapest: KSH.
- Kemény, I. – Janky, B. (2006): Roma Population of Hungary 1971–2003. In Kemény, I. (szerk.): *Roma of Hungary*. New York: Columbia University Press, 70–225.
- Killworth, P. D. – Johnsen, E. C. – McCarty, C. – Shelley, G. A. – Bernard, H. R. (1998): A social network approach to estimating seroprevalence in the United States. *Social Networks*, 20(1): 23–50. [https://doi.org/10.1016/S0378-8733\(96\)00305-X](https://doi.org/10.1016/S0378-8733(96)00305-X)
- Killworth, P. D. – McCarty, C. – Bernard, H. R. – Shelley, G. A. – Johnsen, E. C. (1998): Estimation of Seroprevalence, Rape, and Homelessness in the United States Using a Social Network Approach. *Evaluation Review*, 22(2): 289–308. <https://doi.org/10.1177/0193841X9802200205>
- Killworth, P. D. – McCarty, C. – Johnsen, E. C. – Bernard, H. R. – Shelley, G. A. (2006): Investigating the Variation of Personal Network Size Under Unknown Error Conditions. *Sociological Methods & Research*, 35(1): 84–112. <https://doi.org/10.1177/0049124106289160>
- Kmetty, Z. (2015): Ideológiai és kapcsolathálózati törésvonalak a társadalmi-politikai térben a 2014-es országgyűlési választások előtt. *Studies in Political Science*, (1): 8–34.
- Kmetty, Z. – Koltai, J. (2015): Kapcsolathálózatok mérése – elméleti és gyakorlati dilemmák, lehetőségek. *Socio.hu*, 2015(4): 34–49. <http://dx.doi.org/10.18030/socio.hu.2015.4.3>
- Kmetty, Z. – Koltai, J. (2016): Státuszélerés, társas támogatás, társadalmi törésvonalak. A kapcsolathálózati integráció aspektusai. *Socio.hu*, 6(3): 1–21. DOI: 10.18030/socio.hu.2016.3.1
- Kmetty, Z. – Koltai, J. (2018): Társadalmi mobilitás – kapcsolathálózati megközelítésben. In Kolosi, T. – Tóth, I. G. (szerk.): *Társadalmi riport 2018*. Budapest: TÁRKI, 47–63.
- Laga, I. – Bao, L. – Niu, X. (2021): Thirty Years of The Network Scale-up Method. *Journal of the American Statistical Association*, 116(535): 1548–1559. <https://doi.org/10.1080/01621459.2021.1935267>
- Lin, N. (1999): Social Networks and Status Attainment. *Annual Review of Sociology*, 25(1): 467–487. <https://doi.org/10.1146/annurev.soc.25.1.467>

- Maltiel, R. – Baraff, A. J. (2015): Package 'NSUM'. Network Scale Up Method. <https://CRAN.R-project.org/package=NSUM> (elérés: 2022. március 15.)
- Maltiel, R. – Raftery, A. E. – McCormick, T. H. – Baraff, A. J. (2015): Estimating Population Size Using the Network Scale Up Method. *The annals of applied statistics*, 9(3): 1247–1277. <https://doi.org/10.1214/15-AOAS827>
- McCarty, C. – Killworth, P. D. – Bernard, H. R. – Johnsen, E. C. – Shelley, G. A. (2001): Comparing Two Methods for Estimating Network Size. *Human Organization*, 60(1): 28–39. <https://doi.org/10.17730/humo.60.1.efx5t9gjtgmga73y>
- McCormick, T. H. (2021): The Network Scale-Up Method. In Light, R. – Moody, J. (szerk.): *The Oxford Handbook of Social Networks*. Oxford University Press, 152–169.
- McCormick, T. H. – Salganik, M. J. – Zheng, T. (2010): How Many People Do You Know?: Efficiently Estimating Personal Network Size. *Journal of the American Statistical Association*, 105(489): 59–70. <https://doi.org/10.1198/jasa.2009.ap08518>
- McPherson, M. – Smith-Lovin, L. – Cook, J. M. (2001): Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Moksony, F. (2006): A Poisson-regresszió alkalmazása a szociológiai és demográfiai kutatásban. *Demográfia*, 49 121–144.
- Park, B. (2021): Segregated in Social Space: The Spatial Structure of Acquaintanceship Networks. *Sociological Science*, 8 397–428. <https://doi.org/10.15195/v8.a20>
- Salganik, M. J. – Fazito, D. – Bertoni, N. – Abdo, A. H. – Mello, M. B. – Bastos, F. I. (2011): Assessing Network Scale-up Estimates for Groups Most at Risk of HIV/AIDS: Evidence From a Multiple-Method Study of Heavy Drug Users in Curitiba, Brazil. *American Journal of Epidemiology*, 174(10): 1190–1196. <https://doi.org/10.1093/aje/kwr246>
- Sully, E. – Giorgio, M. – Anjur-Dietrich, S. (2020): Estimating abortion incidence using the network scale-up method. *Demographic Research*, 43(56): 1651–1684. <https://doi.org/10.4054/DemRes.2020.43.56>
- Susánszky, É. – Székely, B. A. – Székely, A. (2022): A Hungarostudy 2021 felmérés célkitűzései és módszertana. In Engler, Á. – Purebl, G. – Susánszky, É. – Székely, A. (szerk.): *Magyar lelkiállapot 2021*. Megjelenés alatt.
- Szabó, A. – Gerő, M. (2022): *A magyar társadalom politikai integrációjának dimenziói a COVID-19 árnyékában*. Budapest: Társadalomtudományi Kutatóközpont.
- Teo, A. K. J. – Prem, K. – Chen, M. I. C. – Roellin, A. – Wong, M. L. – La, H. H. – Cook, A. R. (2019): Estimating the size of key populations for HIV in Singapore using the network scale-up method. *Sexually Transmitted Infections*, 95(8): 602–607. <https://doi.org/10.1136/sextrans-2018-053747>
- Zheng, T. – Salganik, M. J. – Gelman, A. (2006): How Many People Do You Know in Prison? *Journal of the American Statistical Association*, 101(474): 409–423. <https://doi.org/10.1198/016214505000001168>