

<https://helda.helsinki.fi>

---

## Sotsiaalmeedia andmete sotsiaal-ruumiline analüüs

Järv, Olle

TLÜ Kirjastus, Tallinn University Press  
2020

---

Järv , O & Müürisepp , K 2020 , Sotsiaalmeedia andmete sotsiaal-ruumiline analüüs . in A Masso , K Tiidenberg & A Siibak (eds) , Kuidas mõista andmestunud maailma? : Metodoloogiline teejuht . Gigantum Humeris , TLÜ Kirjastus, Tallinn University Press , Tallinn, Estonia , pp. 597-621 . < <http://www.digar.ee/id/nlib-digar:448917> >

---

<http://hdl.handle.net/10138/339482>

---

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

## 5.3. SOTSIAALMEEDIA ANDMETE SOTSIAAL-RUUMILINE ANALÜÜS

Olle Järv, Kerli Müürisepp<sup>1</sup>

### LUGEMISSOOVITUSED

- Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage 2014.
- Shih-Lung Shaw, Daniel Sui (eds.), *Human Dynamics Research in Smart and Connected Communities*. Springer 2018.
- Luke Sloan, Anabel Quan-Haase (eds.), *The Sage Handbook of Social Media Research Methods*. London: Sage 2017.

### 5.3.1. Sissejuhatus

Paljud sotsiaalteoreetikud jagavad üldist arusaama, et ühiskonnanähtustel ja -protsessidel on nii sotsiaalne, ruumiline kui ka ajaline mõõde, mis on omavahel tihedalt seotud ja pidevas vastastikmõjus (Bourdieu 2013; Giddens 1985; Lefebvre 1991; Soja 2013; Urry 2000). Teisisonu on inimeste sotsiaalsed praktikad ja ühiskondlikud protsessid mõjutatud füüsilisest keskkonnast, kus need aset leiavad, ning samaaegselt osalevad need praktikad ja protsessid ise ümbritseva ruumi loomises. Toimub sotsiaal-ruumiline dialektika ehk koostoime (Soja 2013).

Kui seni on ühiskonna sotsiaal-ruumilisi nähtusi ja protsesse uuritud valdavalt kas ainult ruumilisest (nt ruumiline segregatsioon) või ainult sotsiaalsest (nt integratsioon) vaatenurgast, siis tänapäeva globaliseerunud ja võrgustunud ühiskonnas, kus inimesed, asjad ja informatsioon on pidevas liikumises (Urry 2000), muutub sotsiaal-ruumilise koostoime ja selle dünaamika mõistmine üha olulisemaks (Kwan 2013). Näiteks mõjutavad liikumine, kokkupuude ümbritseva keskkonnaga ja

---

<sup>1</sup> Autorid tänavad kolleege Helsingi Ülikooli tööruhmast Digital Geography Lab sisukate arutelude ja koostöö eest sotsiaalmeedia analüütika valdkonnas. Peatüki valmimist toetas Kone Fond ning Helsingi Ülikooli matemaatika- ja loodusteaduste teaduskond.

erinevates kohtades viibimine inimese identiteedi, aga ka tema kuuluvustunde, eelarvamuste ja hinnangute kujunemist ja muutumist (Valentine, Sadgrove 2012). Samuti on inimese füüsiline liikuvus tihedalt seotud tema subjektiivse heaolu (De Vos *et al.* 2013) ja sotsiaalse liikuvusega (mobiilsusega) (Savage 1988). Seega aitab sotsiaalse ja ruumilise mõõtme kombineeritud uurimine selgitada näiteks inimeste identiteedi, sotsiaalse tõrjutuse ja võõraviha kujunemist või seda, miks teatud linnaosad segregeeruvad.

Sotsiaal-ruumilise koostoime uurimisel on olnud kaua piiranguks kulutõhusate andmekogumismeetodite ja sobivate andmeallikate puudumine. Viimase kümne aasta jooksul on aga virtuaalsete võrgustike üha laialdasem kasutuselevõtt ning selle tulemusena tekkinud uued andmekogumismeetodid ja -allikad – suurandmed – toonud teadlaste jaoks kaasa uusi võimalusi ühiskonnaprotsesside uurimiseks ja selgitamiseks (Kitchin 2014). Suurandmetest on ühiskonnateadlastele ühed huvitavamad sotsiaalmeedia andmed, mis sisaldavad nii aegruumilist kui ka rikkalikku sotsiaalset informatsiooni, andes nii pretsedenditu võimaluse samaaegselt uurida üksikisikute sotsiaalseid ja ruumilisi praktikaid ning seda võrdlemisi detailselt. Lisaks võimaldavad need selgitada seoseid virtuaalse ja füüsilise keskkonna vahel (Croitoru *et al.* 2015).

Sotsiaalmeedia andmetel (nt Facebook, Flickr, Instagram, Twitter) põhinev teadus on viimase kümnendi jooksul eksponentsiaalselt kasvanud, kusjuures veelgi enam on andmeid rakendanud erasektor nii ärilistes kui ka poliitilistes huvides. Teadusvaldkondadest on sotsiaalmeedia andmeid kasutatud laialdaselt sotsioloogias (Golder, Macy 2014), demograafias (Longley *et al.* 2015), psühholoogias (Shaughnessy *et al.* 2018), majandusteaduses (Lloyd, Cheshire 2017), keskkonnateaduses (Toivonen *et al.* 2019), aga ka turismi (Hawelka *et al.* 2014), transpordi ja inimeste reisikäitumise (Rashidi *et al.* 2017) uurimiseks. Sotsiaal-ruumilise koostoime uurimiseks on sotsiaalmeedia andmeid kasutatud ennekõike linnauuringute raames (Poorthuis, Zook 2017). Näiteks on uuritud sotsiaalmeedia andmete põhjal linna rohealade mõju inimese emotsioonidele (Roberts *et al.* 2018), kohatunnetust ja õnnetunnet (Mitchell *et al.* 2013), sotsiaal-ruumiliste struktuuride mõju naabruskondade tekkele (Poorthuis 2018), avaliku ruumi kasutust (Salas-Olmedo, Rojas Quezada 2017), sotsiaal-ruumilist ebavõrdsust ja segregatsiooni (Shelton

et al. 2015; Wang et al. 2018) ning linnade valitsemist ja planeerimist (Zook 2017).

Kahes järgmises alapeatükis anname täpsema ülevaate sotsiaalmeedia andmetest ja nende kvantitatiivsetest analüüsivõimalustest, mille abil saab selgitada sotsiaalse ja ruumilise mõõtme koostoimet. Seega keskendume ennekõike mobiilselt kasutatavatele sotsiaalmeedia platvormidele, mille andmed sisaldavad nii postituse sotsiaalset sisu kui ka asukohateavet. Alapeatükis 5.3.4 toome esile sotsiaalmeedia andmete kasutamise tugevused, väljakutsed ja eetilised aspektid ning lõpetuseks arutleme sotsiaalmeedia andmete kasutamise väljavaadete üle tuleviku teadustöös.

## 5.3.2. Andmed

### 5.3.2.1. SOTSIAALMEEDIA PLATVORMID

2019. aasta seisuga kasutab sotsiaalmeedia eri platvorme 45% kogu maailma rahvastikust. Nendest omakorda umbes 93% kasutab sotsiaalmeediat mobiilselt, st nutitefonis või tahvelarvutis (Kemp 2019). Sotsiaalmeedia platvormid võimaldavad kasutajatel teha teksti, pilte ja videot (sh heli) sisaldavaid postitusi. Lisaks saavad kasutajad märkida üksteise postitustele meeldimisi ning neid jagada ja kommenteerida. Seeläbi moodustuvad sotsiaalmeedia kasutajate ja sisu vahel võrgustikud. Samuti on kasutajatel võimalik oma postitusi asukohamärgistada ehk lisada neile geomärgis (*geotag*), mis võimaldab sotsiaalmeedia andmeid käsitleda analoogselt teiste ruumiandmetega (Sui, Goodchild 2011). Nii saab siduda sotsiaalmeedia kasutajad ja nende virtuaalse suhtluse füüsilise ruumiga.

Selles peatükis keskendume eelkõige sotsiaalmeedia platvormidele, mis võimaldavad uurida ühiskonna sotsiaal-ruumilisi nähtusi ja protsesse – need on platvormid, kus sotsiaalset suhtlust ja postituste tegemist saab siduda füüsilise asukohaga. Sellised on ennekõike mobiilselt, peamiselt nutitefonis kasutatavad platvormid, nagu Twitter, Instagram ja Hiinas populaarne Weibo. Kindlasti pakuvad sotsiaalmeedia teadlaste jaoks mitmekülgset teavet ka üleilmselt populaarseim platvorm Facebook, venekeelsetes riikides aga Vkontakte ja Odnoklassniki ning Hiinas QZone. Teatud teemade uurimiseks huvitavat teavet sisaldavad ka

sellised meedia jagamise platvormid nagu Flickr, Panoramio, TripAdvisor ja Foursquare.

Kui kasutada sotsiaalmeedia andmeid ühiskonnaähtuste uurimiseks, tuleb arvestada platvormidevaheliste erinevustega, mis mõjutavad nii seda, milliseid analüüsimeetodeid on sobiv rakendada, kui ka analüüsitulemusi ja nende põhjal järelduste tegemist (Tenkanen *et al.* 2017). Platvormidevahelised erinevused saab jagada kolme tüüpi (vt ka ptk 5.4):

- 1) **platvormil edastatava sisu vorm** – kas postituste sisuks on tekst, pilt, video või kombinatsioon nendest? Näiteks sisaldavad üldised sotsiaalvõrgustiku platvormid (nt Facebook) eri tüüpi sisu, samas kui Twitteris postitatakse peamiselt lühikest tekstilist sisu, lisades vahel pilte ja viiteid muule veebisivule, ning Instagramis ennekõike visuaalset sisu, kombineerides seda tavaliselt lühikese selgitava tekstiga;
- 2) **platvormil edastatava sisu tüüp** – kas postituste sisu on tööalane, meelelahutuslik või isikliku eluga seotud? Kui Twitter sisaldab peamiselt info edastamist ja arutelusid päevakajaliste ja tööalaste teemade ümber, siis Instagrami visuaalne sisu keskendub pigem inimeste eraeluliste tegevuste ja emotsioonide edastamisele. Vaatamata sisu rõhuasetuse erinevustele ei ole platvormidevaheline jaotus selgepiiriline, näiteks paljud Twitteri geomärgisega postitused viitavad Instagrami piltidele (Toivonen *et al.* 2019);
- 3) **platvormi kasutajaskonna profiil** – kes on sotsiaalmeedia platvormi kasutajad? Oluline on teada, keda sotsiaalmeedia platvormi andmete analüüsimisel täpsemalt uuritakse (Ruths, Pfeffer 2014). Platvormide kasutamine varieerub riikide, kultuuride ja rahvastikurühmade kaupa (Jiang *et al.* 2019; Kemp 2019). Seetõttu on platvormi kasutajaskonna profiili väljaselgitamine ja uuringutes sellega arvestamine üks kriitilisemaid lähtekohti teadustöö usaldusvääruse tagamisel (Longley *et al.* 2015; Sloan *et al.* 2015). Kuna täpne kasutajaskonna statistika on platvormide ärisaladus, lähtutakse uuringute põhjal saadud hinnangutest (vt tekstikast 5.3.1).

### TEKSTIKAST 5.3.1. ERINEVUSED SOTSIAALMEEDIA PLATVORMIDE KASUTAJASKONNAS

Platvormide kasutajaskonna profiilidel on selged soolised ja vanuselised erinevused, millega teadustööd tehes tuleb arvestada (Kemp 2019). Üleilmselt on kasutajad pigem noored, näiteks Facebooki ja Twitteri kasutajatest 65% on kuni 34-aastased. Sooline jaotus aga sõltub platvormist: kui Instagrami kasutajate seas on mehi ja naisi võrdselt, siis naised moodustavad Facebooki kasutajatest 43% ja Twitteri kasutajatest kõigest 34%.

Lisaks tuleb arvesse võtta, et kasutajaskonna profiilid erinevad riigiti. Näiteks kasutab Eestis aktiivselt sotsiaalmeediat 55% rahvastikust ning suurema osa kasutajatest moodustavad naised – Facebooki kasutajatest 54% ja Instagrami kasutajatest 59%. Eestis on Facebooki kasutajatest 52% kuni 34-aastased (*ibid.*).

#### 5.3.2.2. SOTSIAALMEEDIA POSTITUSE ELEMENDID

Üldistatult saab sotsiaalmeedia postituse teabelise sisu (*information content*) jagada viieks elemendiks (Toivonen *et al.* 2019): 1) kasutaja isiku-profiil, 2) postituse sisu (tekst, pilt, video, heli), 3) ruumiline asukoht, 4) postituse tegemise aeg ja 5) sotsiaalne võrgustik (kommentaarid, meeldimised, profiili jälgijad) (joonis 5.3.1). Neid elemente saab analüüsida nii eraldi kui ka kombineerituna. Kõik viis elementi on olemas näiteks Twitteri ja Instagrami postitustel.

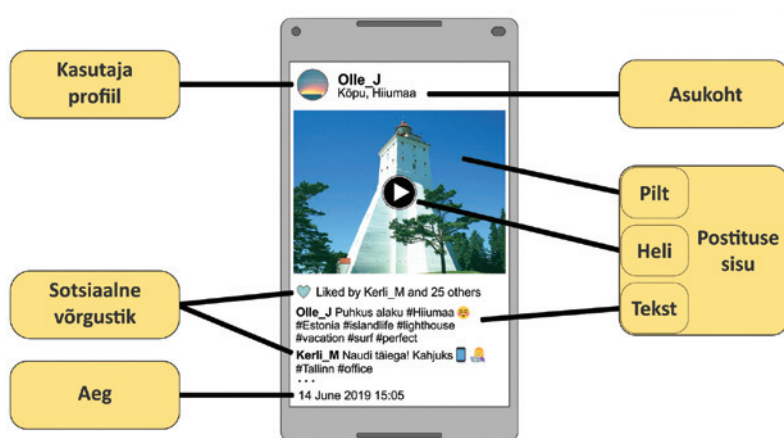
**Kasutaja isikuprofiil** annab teavet sotsiaalmeedia konto omaniku kohta. Kasutajanimi, profiilipilt ja lühike kirjeldus (*bio*) võivad sisaldada infot kasutaja taustatunnuste ja elukoha kohta. Samuti on võimalik avatud profiilide puhul näha kasutaja sotsiaalvõrgustikku ehk jälgijate või sõprade arvu ja nende profiile (nt Twitter, Facebook). Selline kasutajaprofiili (meta)andmete analüüs kombineerituna postituse teiste elementidega võimaldab välja selgitada kasutaja tõenäolise sotsiaal-demograafilise tausta (nt vanus, rahvus, leibkonna tüüp ja päritolu; Longley *et al.* 2015; Sloan *et al.* 2015).

**Postituse sisu** jaguneb tekstiliseks, visuaalseks ja heliliseks. Platvormide erisustest hoolimata koosneb tekstiline sisu enamasti lühemast sõnumist ja märkest (*caption*), nagu teemaviide (*hashtag*), emotikon (*emoji*) või interaktiivne link välisele veebisivule (pilt, video, veebileht). Näiteks on Twitteri postitus ehk säuts praegu piiratud 280 tähemärgiga,

kuid enne novembrit 2017 oli lubatud kõigest 140 tähemärki. Suurt hulka tekstilist sisu saab analüüsida automatiseeritud ja masinõppe meetoditega, kuid seda raskendab sage kõnekeele, lühendite, emotikonide ja teema- viidete kasutus, aga ka postituste sage sarkastilisus ja kontekstitundlikkus. Keeruliseks muudab selle ka erinevate keelte kasutamine, sh võib üks postitus sisaldada teksti mitmes keeles (Hiippala *et al.* 2018; Väisänen 2018). Visuaalne sisu koosneb enamasti kasutaja tehtud fotodest, aga ka meemidest, infograafikast ja teist tüüpi visuaalidest. Video koosneb sisuliselt piltide jadast, millele võib olla taustaks lisatud heli. Vähemal määral võib postituse sisu koosneda ka ainult helist (nt muusika, kõne).

**Ruumilise asukoha** teave võimaldab sotsiaalmeedia postituse siduda konkreetse geograafilise asukohaga, kust postitus tehti ja/või millega postituse sisu on seotud. Postituse geomärgise täpsus oleneb sellest, kas platvorm fikseerib koha geograafiliste koordinaatide või platvormi poolt eelnevalt fikseeritud tegevuskohtade (nt Viru keskus) kujul (Toivonen *et al.* 2019). Seetõttu tuleb postituste ruumilise analüüsi puhul andmete ruumilist täpsust kriitiliselt hinnata. Arvestada tuleb ka sellega, et vaid väike osa kõigist postitustest on geomärgisega. Näiteks on Twitteris vaid 3% postitustest geomärgisega, kuid seda kompenseerib postituste suur hulk (Poorthuis, Zook 2017). Ruumilist teavet sisaldavad aga ka ilma geomärgiseta postitused, näiteks tekstis mainitud kohanimed ja teema- viited (nt „Unistuste päikeseloojang #Hiiumaa #Estonia“), visuaalsest sisust äratuntavad kohad või tuntud maamärgid (nt Kõpu tuletorn või Egiptuse püramiidid).

**Ajamärgis** sotsiaalmeedia postituse juures kajastab kuupäeva ja kella- laaega, millal kasutaja postituse veebis avaldas ja teatud juhtudel täien- davalt, millal postitatud foto on tehtud. Ajamärgise abil saab sotsiaal- meedia postituse siduda konkreetse ajahetkega. See võimaldab uurida inimtegevuse ja ühiskonnaähtuste ajalist dünaamikat ja trende, näiteks analüüsida ööpäevaseid, nädalapäevaseid ja sesooneid mustreid, aga ka pikemaajalisi trende aastate kaupa (Tenkanen *et al.* 2017). Ka ajamärgi- seid analüüsides tuleb olla kriitiline, kuna postitus võib olla avaldatud postituse sisu loomise hetkest hiljem. Näiteks ei pruugi sotsiaalmeedias postitatud foto olla tehtud selle avaldamise ajal. Samuti võib postituse sisu olla seotud ajaloolise sündmusega (nt #ajastagasi, #nooruses) või rääkida tulevikuplaanidest (nt „järgmisel aastal tööle #Austraalia“).



**Joonis 5.3.1.** Sotsiaalmeedia andmete viis elementi Instagrami postituse näitel; elemente saab analüüsida nii eraldi kui ka kombineerituna. Allikas: Toivonen *et al.* 2019; autorite mugandatud

Sotsiaalmeedia platvormi kasutajate **sotsiaalseid võrgustikke** võimaldavad uurida postituse juures olevad kommentaarid, meeldimised ja jagamised ning profiilidevahelised jälgimised. Näiteks salvestub Twitteris iga postituse juurde sellele reageerinud kasutajate unikaalsed koodid ja postitusele tehtud kommentaaride sisu. Sellised kasutajatevahelised seosed võimaldavad tuvastada kasutajate sotsiaalvõrgustikke ja avastada temaatilisi kogukondi (vt täpsemalt ptk 2.2).

### 5.3.2.3. ANDMETE HANKIMINE JA TÖÖVOOG

Sotsiaalmeedia andmete hankimiseks on erinevaid meetodeid, alates manuaalsetest otsingutest ja lõpetades andmetele automatiseeritud programmilise juurdepääsu loomisega (Batrinca, Treleaven 2015). Andmete hankimise viis sõltub nii teadlase oskustest kui ka uurimisküsimusest tulenevast andmevajadusest. Sotsiaalmeedia postituste ja grupisisese suhtluse manuaalne läbisirvimine (nt digitaalne etnograafia) on ajamahukas ja piirab analüüsitavat andmemahtu, kuid ei nõua see-eest erilisi arvutialaseid oskusi ja võimaldab andmete kontekst tundlikku analüüsi. Andmete hankimine automatiseeritud meetoditega võimaldab koguda ja analüüsida oluliselt suuremaid andmemahte,



kuid nõuab tehnilisi teadmisi ja programmeerimisoskust (Sloan, Quan-Haase 2017).

Automatiseeritud meetoditest on andmetele kõige lihtsam juurdepääsu saada andmeanalüüsi tarkvarade abil. Need võivad olla maksulised (suletud lähtekood) või tasuta (avatud lähtekood) ning pakuvad erinevaid valmisprogrammeeritud funktsioone sotsiaalmeedia andmete allalaadimiseks ja analüüsimiseks platvormi rakendusliidese (*application programming interface*, API) kaudu. Samas võimaldavad rakendusliideseid programmeerimisoskusega teadlastel suhelda otse platvormiga ning programmeerida oma uuringuks vajalik ja tõhus automaatne andmekogumine (Lomborg, Bechmann 2014). Oluline on arvestada, et rakendusliideste kaudu saab ligipääsu vaid avalikele sotsiaalmeedia kontodele, st privaatsete kontode andmeid ei edastata.

Rakendusliideseid on üldiselt kahte tüüpi: voogedastusega liides (*streaming API*) edastab jooksvalt uusi postitusi ja hajustöötlusega liides (*representational state transfer; REST API*) võimaldab spetsiifilistele andmetele ja paindlikumate päringute tegemist ka tagasivaatavalt (Poorthuis, Zook 2017). Tuleb muidugi arvestada, et platvormi on rakendusliidesele ligipääs ja selle kasutusvõimalused erinevad ning platvormi omanikul on õigus liidese kasutusreeglid igal ajal muuta (boyd, Crawford 2012; Freelon 2018). Näiteks suleti Instagrami rakendusliides, mis varem oli avatud, jaanuaris 2018, kui Facebook platvormi omandamise järel liidese kasutustingimusi muutis. Täpsem ülevaade võimalustest kasutada Twitteri avatud andmeid on antud tekstikastis 5.3.2.

Lisaks saab sotsiaalmeedia andmeid osta volitatud andmete müüjatelt ning kasutada veebilehtedelt andmete allalaadimiseks veebiroomamist (*web crawling*) ja veebikraapimist (*web scraping*). See on siiski enamasti vastuolus sotsiaalmeedia platvormide kasutustingimustega (Freelon 2018). Olenemata andmete hankimise viisist on teadlane alati vastutav, et andmete hankimine, hoidmine, analüüsimine ja tulemuste avaldamine oleks kooskõlas nii õigusaktide, kasutustingimuste kui ka teaduseetika põhimõtetega (Zook *et al.* 2017).

Sotsiaalmeedia andmetest tähendusliku teabeni jõudmine on mitmetapiline tsükliline protsess, mis järgib tüüpilist andmekaeve töövoogu (Han *et al.* 2011). Arvestada tuleb etappidega alates andmete tõhusast ja turvalisest hankimisest ja hoidmisest kuni andmete puhastamise,

### TEKSTIKAST 5.3.2. TWITTER – AVATUD LÄHTEKOODIGA SOTSIAALMEEDIA PLATVORM

Praeguse seisuga on Twitter üks vähestest avatud lähtekoodiga sotsiaalmeedia platvormidest, mille andmeid saab kasutada kvantitatiivse sotsiaal-ruumilise analüüsi tegemiseks. Twitteri andmetele on mitu juurdepääsuviimast:

- rakendusliides: <https://developer.twitter.com>;
- tarkvarad, nt COSMOS: <http://socialdatalab.net/COSMOS>;
- akadeemilised andmebaasid, nt DOLLY Project: [www.floatingsheep.org](http://www.floatingsheep.org).

Twitteri kasutajakontosid on maailmas praegu 326 miljonit, millest 75% on aktiivsed (Kemp 2019). Teadustööd tehes tuleb arvestada riikidevahelisi erinevusi nii Twitteri kasutajate soolises ja vanuselises jaotuses kui ka platvormi üldises populaarsuses. Kui Jaapanis kasutab Twitterit 34% rahvastikust, siis Soomes 11%, Eestis 8%, Saksamaal 5% ja Hiinas 0,1% rahvastikust. Soomes on 64% Twitteri kasutajatest mehed, Eestis on see näitaja 67% ja Lätis koguni 72% (*ibid.*).

filtrerimise, töötlemise, rikastamise, analüüsi ja järelduste tegemiseni (Toivonen *et al.* 2019).

Enne reaalselt andmeanalüüsi on tähtis andmete puhastamine ja filtreerimine, näiteks robotprogrammide (*bot*) poolt automaatselt loodavate andmete välistamiseks (vt ka ptk 3.3). Andmete töötlemine ja rikastamine toimub tsükliliselt: uute teadmiseni jõutakse sageli suurandmete töötlemisel väikeandmeteks (*small data*), mida analüüsitakse ja mille põhjal tehakse järeldused (Poorthuis, Zook 2017). Tähtis on tulemusi alati kriitiliselt hinnata, võttes arvesse platvormi kasutajaskonna profiili ja andmete iseloomu, sh sisemist varieeruvust, kallutatust ja ajas muutuvust (Kitchin 2014; Jiang *et al.* 2019).

### 5.3.3. Analüüsivõimalused

Sotsiaalmeedia andmete analüüsivõimaluste tutvustamisel keskendume eelkõige mobiilsetele platvormidele (Twitter, Instagram, Weibo), kus postitused sisaldavad ajamärgist ja geograafilist teavet. Geograafilisi andmeid ehk ruumiandmeid analüüsitakse peamiselt geograafias ja geoinformaatikas mitmesuguste ruumianalüüsi ja statistiliste meetoditega. Näiteks tehakse Twitteri andmete töötlust ja analüüsi sarnaselt mobiilpositsioneerimise andmetega (vt ka ptk 5.4). Sotsiaal-ruumilise analüüsi

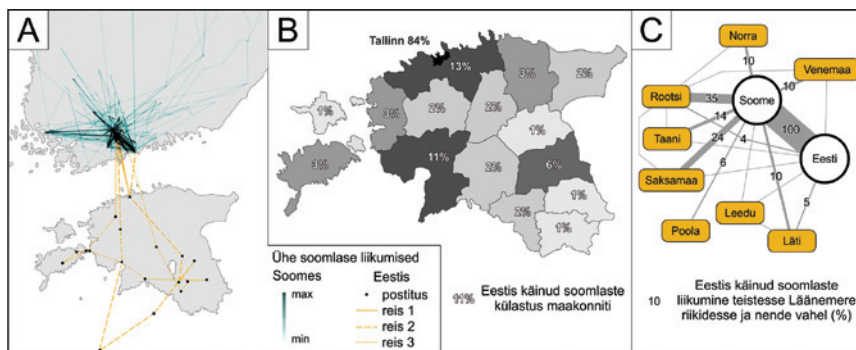
puhul saab välja tuua kolm peamist käsitlusviisi vastavalt uuritavale objektile:

- inimesed (*people*);
- kohad ja ruumid (*places, spaces*);
- vood ja võrgustikud (*flows, networks*).

### 5.3.3.1. AEGRUUMILISE ANALÜÜSI VIISID

**Indiviidipõhise käsitluse** korral tehakse analüüs üksikisiku ehk sotsiaalmeedia platvormi kasutaja tasandil. Näiteks on Twitteri andmetest võimalik tagasiulatuvalt analüüsida kasutaja kuni 3200 viimast postitust (Hasnat, Hasan 2018). Postituste tegemise aja ja asukohtade alusel saab selgitada välja üksikisiku külastatud kohad ja liikumistrajektorid ehk tema ruumikasutuse (joonis 5.3.2A). Ennekõike on oluline tuvastada indiviidi igapäevased tegevuskohad (elukoht ja töökoht/kool; Osorio-Arjona, García-Palomares 2019; Järv *et al.* 2015), mis võimaldab omakorda eristada, kes on teatud piirkonnas kohalikud ja kes külastajad (Hasnat, Hasan 2018). Lisaks on indiviidi aegruumilise käitumise ja teiste täiendavate andmete (nt maakasutus – elumumaa, tööstusmaa, ärimaa, roheala) alusel võimalik tuletada tema teisi tegevuskohti, mis on seotud näiteks hobide, vaba aja veetmise või majapidamise ja perekonnaga (García-Palomares *et al.* 2018). See võimaldab omakorda uurida indiviidi tegevuskohtadevahelist reisikäitumist, sh igapäevaseid kodu ja töö vahelisi ning rahvusvahelisi liikumisi (Rashidi *et al.* 2017). Üksikisikute ruumikasutuse ulatust saab statistiliselt hinnata näiteks tegevusruumide arvutamise ja seejärel on võimalik analüüsida seda mõjutavaid tegureid, nagu isiku taustatunnused (nt sugu, vanus, rahvus) ning füüsilise ja sotsiaalse väliskeskonna omadused (Järv *et al.* 2015). Lisaks on kvantitatiivset aegruumilist analüüsi võimalik kombineerida postituste sisuanalüüsiga, mis võimaldab saada veelgi täpsemat teavet inimese ruumikasutuse, tema elustiili, aga ka ruumikogemuse ning kohtade eelistamise ja vältimise kohta (vt alaptk 5.3.3.2).

**Kohapõhine käsitlus** vaatleb uurimisobjektina geograafiliselt määratletud kohta või piirkonda – ruumi, kus sotsiaalmeediat kasutatakse. Geograafilisi kohti ja piirkondi saab sotsiaalmeedia postituste põhjal analüüsida kahel viisil. Kõige tavapärasem viis on määratleda



**Joonis 5.3.2.** Erinevad võimalused Soomes elavate ja Eestit külastanud Twitteri-kasutajate ( $n = 1379$ ) aegruumilise liikumise uurimiseks. A. Indiviidipõhine käsitlus võimaldab postituste ajaloo põhjal uurida indiviidi liikumisi Soomes ja tema reise Eestisse. B. Kohapõhise käsitlusega saab teada, millistes Eesti maakondades kui palju soomlasti on käinud. C. Voogude käsitlus võimaldab uurida, kui palju Eestit külastanud soomlastest on käinud erinevates Läänemere riikides. Allikas: autorite arvutused ja kujundus

konkreetne uurimisala ning koguda ja analüüsida selle alaga seotud postitusi. Näiteks saab uurida Twitteri postitusi, mille geomärgis asub Eestis või mille tekst sisaldab teavet, mis viitab Eestile (joonis 5.3.2B). Nii saab näiteks analüüsida rahvastiku paiknemist linnas ja selle ajalist dünaamikat (Li *et al.* 2013). Sidudes postitused täiendavate andmeallikatega, näiteks ruumiliste registriandmetega, saab luua seoseid postituste ruumilise paiknemise ja piirkonna elanike sotsiaalmajandusliku tausta vahel (*ibid.*), mida on rakendatud näiteks etnilise segregatsiooni uurimisel (Wang *et al.* 2018). Koos maakasutuse andmekihiga saab uurida linna rohealade kasutust (Heikinheimo *et al.* 2020). Samas ei analüüsita kirjeldatud käsitlusviisi puhul täpsemalt konkreetseid postitajaid ja nende tausta ehk seda, kas nad on kohalikud või külastajad või mis vanusest, soost või rahvusest nad on. Kui analüüsitava ruumiüksuses ei ole postitajate profiil teada, tuleb üldistuste tegemisega olla ettevaatlik.

Teine ja põhjalikum viis kohapõhise analüüsi teostamiseks on kombineerida seda indiviidipõhise käsitlusega. Sel juhul tuleb esimeses analüüsietapis välja selgitada sotsiaalmeedia platvormi kasutajad, kes on uuritava ruumiüksuses postitanud, ja teha nende kasutajate kohta indiviidipõhine analüüs. Teises etapis tuleb kohapõhist käsitlust rakendades analüüsida täpsemalt nende kasutajate postitusi uuritava ruumiüksuses.

Selline kombineeritud meetod nõuab rohkem andmetöötlust ja aega, kuid pakub täpsemat teavet postitajate tausta kohta ja seeläbi usaldusväärsemaid analüüsitulemusi. Lisaks võimaldab see täpsemalt analüüsida sotsiaal-ruumilist dialektikat eri tegevuskohtades ja piirkondades, näiteks eristada kohalikud ja külastajad ning selgitada nende erinevat kohatunnetust (vt alaptk 5.3.3.2). Samuti lubab postitajate taustatunnuste teadmine analüüsida, mis kohtades eri rahvastikugrupid käivad (Järv *et al.* 2015) ja millist sotsiaal-ruumilist ebavõrdsust see linnaruumis tekitab (Shelton *et al.* 2015).

**Voogude ja võrgustike käsitlus** vaatleb inimeste kohtadevahelisi liikumisi rahvastikurühmade vaatenurgast. Uurimisobjektiks on siin liikumine, st eesmärk on tuvastada üksikisikute liikumistrajektooid tegevuskohtade vahel ning agregeerida liikumised kohtade- või piirkondadevahelisteks liikumisvoogudeks, näiteks on nii uuritud kodu ja töö vahelist pendelrännet (Osorio-Arjona, García-Palomares 2019) ning hinnatud riikidevahelisi turismivooge (Hawelka *et al.* 2014). Samuti saab üksikisikute tegevuskohtadevaheliste liikumiste alusel luua kohtadevahelised liikumisvõrgustikud kogu rahvastiku või mingi rahvastikurühma kohta (joonis 5.3.2C). Linnasisese liikumisvõrgustiku põhjal saab tuvastada näiteks funktsionaalseid naabruskondi ehk kogukondade ühiseid toimepiirkondi, kus tehakse igapäevaseid toimetusi (Poorthuis 2018). Ka siin võib postituste sisuanalüüs (vt alaptk 5.3.3.2) anda täiendavat teavet liikumise põhjuste kohta ning võimaldada uurida näiteks vaba ajaga seotud liikumisvõrgustikke linnas ja nende muutumist ajas.

### 5.3.3.2. POSTITUSTE SISU SIDUMINE AEGRUUMIGA

Sotsiaalmeedia postituste aegruumilise analüüsi kombineerimine postituste sisuanalüüsiga annab täiendava vaatenurga sotsiaal-ruumiliste nähtuste ja protsesside uurimiseks. Postituste manuaalne sisuanalüüs on aja- ja tömahukas ning võimaldab uurida vaid väikest osa kogu andmehulgast. Nii jääb ära kasutamata üks sotsiaalmeedia tugevustest: suur andmehulk ja andmete väga kiire juurdekasv. Tänu masinõppe meetodite hulka kuuluva süvaõppe (*deep learning*) kiirele arengule on saanud aga võimalikuks postituste sisu automatiseeritud analüüsimine (Goodfellow *et al.* 2016). Süvaõpe võimaldab programmil õppida analüüsi teostama

eelnevalt koostatud treeningandmestiku alusel. Järjest paremaid tulemusi saadakse ka neurovõrkudele tugineva *word embeddings* -tehnika rakendamisel (Bojanowski *et al.* 2017, vt ka ptk 2.4).

Parim viis tekstilise sisu automatiseeritud töötlemiseks on kasutada loomuliku keeletöötlemise (*natural language processing*) meetodit, milles nähakse suurt potentsiaali tekstilisest sisust uurijate jaoks kasuliku informatsiooni tuvastamisel (Becken *et al.* 2017, vt ka ptk 3.1). Lihtsusatult öeldes uuritakse loomulikke keeli (nt inglise või eesti keel) arvutusliku töötlemise (*computational processing*) kaudu. Kõigepealt tuvastatakse automaatselt postitustes kasutatud keel(ed), mille järel on võimalik edasine postituste kvantitatiivne sisuanalüüs (Toivonen *et al.* 2019, vt ka tekstikast 5.3.3).

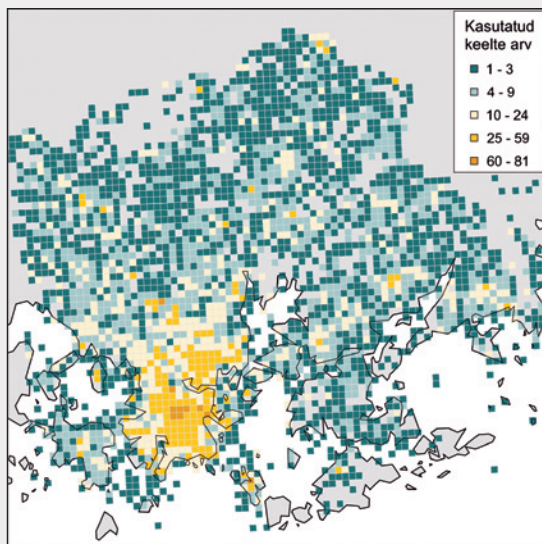
Parim viis visuaalse sisu automatiseeritud töötlemiseks on rakendada masinnägemise (*computer vision*) meetodikat. Tegemist on laia ja interdistsiplinaarse meetodite kogumiga, mis tegeleb piltide automaatse töötlemise ja analüüsimisega. Süvaõppega on saavutatud järjest täpsemaid tulemusi keeruliste masinnägemise ülesannete teostamisel, näiteks piltidelt objektide leidmisel ja nende piirjoonte määratlemisel (He *et al.* 2017), kirjelduste loomisel tervete piltide või nende osade kohta (Karpathy, Fei-Fei 2015) ning lõpuks fotode sisu klassifitseerimisel (Rawat, Wang 2017). Samas sisaldavad postitused sageli nii visuaalset kui ka tekstilist sisu (nt Instagrami pilt koos selgitava tekstiga). Sellisel juhul on kõige kasulikum rakendada multimodaalset meetodit (*multimodality*), mis võtab töötlusel samaaegselt arvesse nii visuaalset kui ka tekstilist sisu (Bateman *et al.* 2017).

Järgmise sammuna pärast andmetöötlust kasutatakse postituse sisu analüüsiks eri meetodeid vastavalt uurimuse eesmärgile: teemade modelleerimist, diskursusanalüüsi (vt ptk 4.3) ja meelestatuse analüüsi (vt ptk 3.3). Samas tuleb automatiseeritud analüütikat kasutades olla alati kriitiline ning arvestada meetodite puudujääke, esinduslikkust ja võimalikku kallutatust (vt alaptk 5.3.4).

Teadustöid, mis kombineerivad sisuanalüüsi aegruumilise analüüsiga, ei ole veel palju, kuid seni teostatud uuringud on paljulubavad. Meetod võimaldab uurida inimeste poolt kohtadele ja piirkondadele omistatud tunnetust, hoiakuid, emotsioone ja väärtusi (Hu 2018). Näiteks on analüüsitud inimeste väärtushinnanguid seoses maastikega (vt

### TEKSTIKAST 5.3.3. KEELEKASUTUSE RUUMILINE ANALÜÜS

Huvitavaid võimalusi pakub postituste keelekasutuse analüüs. Üksikisiku kohta on võimalik välja selgitada tema tõenäoline emakeel (kombineerituna postituste ruumilise teabega) kui oluline rahvuse ja kultuurilise tausta näitaja, aga ka üldine keelteoskus, mis on sotsiaalse suhtluse võimaldajana oluline identiteedi kujunemisel (Noels *et al.* 1996). Twitteri-postituste alusel on uuritud üleilmselt suurlinnade keelemaastikke ja hinnatud eri kultuurirühmade ühiskonda integreerumist keelekasutuse ruumilise segunemise alusel (Lamanna *et al.* 2018). Samuti saab uurida keelekasutust konkreetses kohas või piirkonnas ning kirjeldada sealset keelemaastikku, st siduda keelekasutus füüsilise ruumiga (Väisänen 2018; Hliippala *et al.* 2020). Helsingis eristuvad näiteks selgelt rahvusvaheline kesklinn ja linna äärealad (joonis 5.3.3). Esile tulevad ka huvitavad keelte koondumiskohad, mille tekkepõhjusti ja tagamaid saab täpsemalt selgitada postituste sisuanalüüsi ja täiendavate andmete kasutamisel näiteks maakasutuse ja turismiobjektide kohta.



**Joonis 5.3.3.** Üle 150 000 Instagrami kasutaja poolt tehtud ligi 850 000 postituse põhjal kaardistatud keelemaastik Helsingis ajaperioodil 2015–2016 paljastab linnaruumi sotsiaalse mitmekesisuse ja selle erisused linnaosati. Allikas: Väisänen 2018; autorite mugandatud

Toivonen *et al.* 2019), inimeste õnnetaset USA osariikides (Mitchell *et al.* 2013) ja linnaruumi piirkondi, mille suhtes on mingitel rahvastikurühmadel negatiivsed hoiakud (Shelton *et al.* 2015). Alkoholiteemaliste postituste alusel saab uurida näiteks alkoholi tarbimist ning selle piirkondlikke erisusi ja koondumiskohti (Zook, Poorthuis 2014).

Palju arenguruumi on individipõhise käsitlemise rakendamisel, kus sisuanalüüs seotakse aegruumilise analüüsiga (vt alaptk 5.3.3.1), ning

pikaajaliste longituud-uuringute teostamisel. Üksikisiku tegevusruumi ja tema taustatunnuste (nt vanus, sugu, rahvus, sissetulek, haridus, elustiil, identiteet) teadmine võimaldaks täpsemalt uurida sellist sotsiaalruumilist nähtust nagu segregatsioon, näiteks seda, kuidas erineb rahvastikurühmade ruumikasutus Tallinnas tervikuna või mingis piirkonnas. Samuti saaks uurida, kuidas eri rahvastikurühmad tunnetavad erinevaid kohti ja piirkondi. Millised hoiakud on näiteks Tallinna vanalinna suhtes kohalikel noortel, eakatel välisturistidel ja Kalamajas elavatel tööelistel?

Lisaks saab huvipakkuvates teemades (nt vähemusrahvused, alkoholi tarbimine, hirm) tuvastada, kellele teema on oluline, millistes kohtades ja piirkondades sel teemal räägitakse ning milline on postitajate meelsus teema suhtes. Nii on võimalik välja selgitada, millistesse kohtadesse mingid hoiakud teatud teema suhtes koonduvad ning millised on võimalikud konfliktipiirkonnad, kus esineb vastakaid hoiakuid. Perspektiivseks uurimisvaldkonnaks on ka pikemaajalised muutused kohtadega seotud teemades, tunnetustes ja hoiakutes (vt Hawelka *et al.* 2014). Näiteks saab uurida, kuidas linnaosa gentrifitseerumise tulemusena muutub koos elanikkonna muutumisega aja jooksul ka asumi identiteet ning millised tegurid seda potentsiaalselt mõjutavad.

### 5.3.3.3. SOTSIAALVÕRGUSTIKE SIDUMINE AEGRUUMIGA

Sotsiaalmeedia postituste aegruumilist analüüsi (vt alaptk 5.3.3.1) on võimalik kombineerida ka kasutajate- ja postitustevaheliste seoste kaardistamise ja mõõtmisega ehk sotsiaalvõrgustike uurimisega (Scott, Carrington 2011). Lihtsustatult öeldes koosneb sotsiaalvõrgustik sõlmede (*nodes*) ja neid ühendavate seoste (*links*) kogumist, kusjuures sõlmedeks võivad olla inividid, postitused või teemad (vt ka ptk 2.2). Sotsiaalmeedia postituste meeldimiste, kommentaaride ja jagamiste ning profiili jälgijate põhjal saab tuvastada üksteisest eristuvaid sotsiaalseid kogukondi ja neid geograafilise ruumiga siduda (vt ka tekstikast 5.3.4).

Teave sotsiaalvõrgustike kohta rikastab sotsiaalruumiliste nähtuste ja protsesside uurimist mitmel viisil, kuid lähtekohaks on virtuaalse suhtluse sidumine füüsilise keskkonnaga. Kogukondi saab eristada selle põhjal, milline on nende meelestatus arutelu all olevate teemade suhtes, kombineerides loomuliku keele töötluse tehnikaid ja



kogukonnatuvastuse algoritme. Näiteks tuleb esmalt määratleda uuritav ühiskondlik nähtus (nt liikumine Okupeerige Wall Street) ning seejärel saab tuvastada, millised virtuaalsed ja geograafilised kogukonnad sellest räägivad, kuidas teema kajastamine ajas ja geograafilises ruumis levib ning kuidas suhtlusvõrgustikud muutuvad (Croitoru *et al.* 2015). 2016. aasta Zika viiruspuhangu näitel on uuritud, kuidas viirusega seotud suhtlus sotsiaalmeedias ajaliselt ja geograafiliselt arenes ning reaalse epideemia levikuga ühtis ning kuidas epideemiasse suhtumine ajas muutus ja millised toimijad suhtlust mõjutasid (Stefanidis *et al.* 2017).

Sarnaselt saaks määratleda huvipakkuva piirkonna Eestis, näiteks Ida-Virumaa, tuvastada kohalikud inimesed sotsiaalmeedias ja uurida seejärel nende virtuaalmaailma sotsiaalseid võrgustikke. Nii on võimalik uurida füüsilise ruumi mõju virtuaalsete kogukondade tekkele, aga ka hinnata ruumiliselt määratletud kogukondade sisemist suhtlust. Tallinlaste ja Tallinnas viibivate inimeste näitel saaks tuvastada nende meelsuse avalikus linnaruumis oleva tänavakunsti suhtes ja välja selgitada: 1) kus elavad teatud meelsusega inimesed, 2) millistes kohtades tänavakunstist räägitakse ja millistes piirkondades on erimeelsusi, 3) millised sotsiaal-ruumilised tunnused kirjeldavad neid kogukondi ja kuidas need ajas muutuvad. Samuti saaks uurida üksikisiku sotsiaalvõrgustiku ajalist muutumist. Näiteks on sisserändaja kohanemine ja kontaktide loomine uues ühiskonnas pikaajaline protsess, mida saaks uurida tema sotsiaalvõrgustikus toimivate muutuste kaudu: aja jooksul kohalike elanike lisandumine sotsiaalvõrgustikku ja nendega sotsiaalmeedias suhtlemine võiks olla üks näitaja sisserändaja ühiskonda integreerumise hindamisel.

#### **TEKSTIKAST 5.3.4. ISIKU TAUSTATUNNUSTE TULETAMINE SOTSIAALVÕRGUSTIKU ALUSEL**

Sotsiaalvõrgustiku analüüsimise põhjal saab uuritavatele tuletada täiendavaid taustatunnuseid. Näiteks saab sotsiaalvõrgustiku alusel hinnata isiksuse tüüpi (Hughes *et al.* 2012). Kui uuritava elukoht pole teada või see on ebaselge, saab selle tuletada sotsiaalvõrgustiku ruumilise paiknemise alusel (Pontes *et al.* 2012). Lisaks peegeldavad uuritavaga sotsiaalvõrgustikus seotute teemapostitused, hoiakud ja väärtushinnangud tõenäoliselt ka uuritava maailmavaadet, arvestades inimeste kalduvust suhelda endasarnastega (McPherson *et al.* 2001).

### 5.3.4. Tugevused ja väljakutsed

Eespool tutvustasime sotsiaalmeedia andmete sotsiaal-ruumilist analüütikat, toetudes peamiselt meetodi tugevustele. Kokkuvõttes on sotsiaalmeedia andmete oluliseks tugevuseks võimalus andmete eri elemente (vt alaptk 5.3.2.2) kombineerides samaaegselt uurida ühiskonnanähtuste ja -protsesside aegruumilist ja sotsiaalset mõõdet ning seejuures ulatuslikumalt ja täpsemalt kui teiste andmekogumismeetoditega. Teiseks tugevuseks on hoomamatu andmemahd, mille juurdekasv on väga kiire ja mis on reaalselt kättesaadav rakendusliidest kaudu (Kitchin 2014). See võimaldab omakorda konkreetse ühiskonnanähtuse uurimiseks muuta suurandmed väikeandmeteks (Poorthuis, Zook 2017). Kolmas oluline tugevus põhineb oletusel, et üha enam digitaliseerivas ühiskonnas hakatakse sotsiaalmeediat veelgi laialdasemalt kasutama ja see muutub igapäevaseks suhtlusviisiks. Sotsiaalmeedia kasutus levib ka vanemaalaste sekka ning praegu laialdasemat kasutamist piiravad tehnoloogilised lahendused arenevad kiiresti ja muudavad digitaalse suhtlemise lähiaastatel veelgi mugavamaks (nt hääljuhtimine ja virtuaalne klaviatuur).

Samas on selge, et sotsiaalmeedia andmete kasutamisel inimeste ja ühiskonna uurimiseks on mitmeid väljakutseid ja piiranguid, millega tuleb analüüse teostades kindlasti arvestada ning saadud tulemuste usaldusväarsust kriitiliselt hinnata. Toome esile neli olulisemat väljakutset ja piirangut. **Esimene suurem sotsiaalmeedia andmete väljakutse tuleneb andmete iseloomust.** Sotsiaalmeedia platvormid ja seal talletatavad andmed varieeruvad sisemiselt ning muutuvad ajas; see mõjutab nende kättesaadavust, usaldusväarsust ja kvaliteeti (Kitchin 2014). Sotsiaalmeedia andmeid omavad ettevõtted, kelle peamine eesmärk on kasumi teenimine. Sellest tulenevalt võidakse muuta nii platvormi, andmestruktuuri kui ka andmete kättesaadavust teadlastele (nt Instagram API juhtum). Seega võib analüüsi teostades tekkida raskuseid seoses ajas muutuvate andmetega, andmestruktuuride ühilduvusega ja analüüside korratavusega, mis on takistuseks ennekõike pikaajalisi longituuduringuid tehes (Lomborg, Bechmann 2014).

**Teine väljakutse on andmete kallutatus.** Keda ja mida tegelikult sotsiaalmeedia andmetega uuritakse? Esiteks on sotsiaalmeedia platvormid sageli suunatud ja/või sobilikud teatud tüüpi suhtluseks. Näiteks

on Twitter teabe ja seisukohtade ning Instagram pigem tegevuste ja emotsioonide edastamiseks. Seega esindab üks sotsiaalmeedia platvorm vaid teatud osa digitaalsest suhtlusest. Teiseks tuleb arvestada nii uuritava platvormi kasutajaskonna profiiliga (vt alaptk 5.3.2.1) kui ka selle demograafiliste ja sotsiaal-majanduslike eripäradega geograafiliselt: platvormi kasutajaskonna profiil erineb ka riigiti ja maakonniti (Jiang *et al.* 2019). Rahvusvaheliste uuringute puhul tuleb arvestada, et kasutajaskonna profile eri riikides mõjutavad kultuurilised ja poliitilised eripärad: Hiinas on Twitter ja Facebook keelatud, Venemaal eelistatakse venekeelseid platvorme jne. Sotsiaalmeedia platvormide kasutust mõjutab ka innovatsiooni aegruumiline levik ehk see, milline platvorm kus, millal ja kelle seas kasutusele võetakse. Samas saab kallutatust vähendada andmete kombineerimise ja valimi modelleerimisega (Tenkanen *et al.* 2017; Jiang *et al.* 2019).

**Kolmas väljakutse on piirangud analüüsi tehnilises teostuses,** kuigi meetodite areng on muljetavaldavalt kiire (Bojanowski *et al.* 2017). Näiteks on süvaõppe rakendamisel mudelite väljakoolitamiseks vajalik suure hulga treeningandmete olemasolu, mida seni leidub veel piiratud koguses ning teatud valdkondade ja teemade kohta. Samas tõstab treeningandmete loomine sotsiaalmeedia andmete analüüsivõimet samm-sammult. Lisaks tuleb arvestada algoritmidel ja masinõppel põhinevate analüüside „musta kasti“ kriitikaga: pole võimalik täpselt tuvastada, kuidas tulemus on saadud. Samas eeldab avatud teaduse põhimõtete rakendamine üha enam lähtekoodide avaldamist ja analüüsi täpset kirjeldamist.

**Neljandaks väljakutseks on andmekaitse ja eetikaga seotud teemad** (vt alaptk 5.3.4.1).

#### 5.3.4.1. ANDMEKAITSE, PRIVAATSUS JA EETIKA

Suurandmete ülikiire tekkimine ja nende rakendamine nii teadustöös, avalikus sektoris kui ka ettevõtluses on toonud jõudsalt päevakorda nende kasutamise seotud eetilised küsimused ja inimeste privaatsuse tagamise põhimõtted (nt Markham *et al.* 2018). Seda kiirendavad avalikuks tulevad probleemid andmete rakendamisel, näiteks Cambridge Analytica skandaal (*ibid.*). Sotsiaalmeedia andmete analüütika on

privaatsuse ja eetika vaatenurgast veelgi tundlikum kui mobiiltelefonide kasutajate digitaalsete jäljeridade uurimine (vt ptk 5.4), kuna lisaks geomärgiste põhjal leitud sotsiaalmeedia kasutaja aegruumilisele jäljereale saab selle siduda postituste (delikaatse) sisuga.

Lisaks kehtivatele Euroopa Liidu ja riigisiseste isikuandmete kaitse seadustele ja regulatsioonidele peab sotsiaalmeedia andmete kasutamine olema kooskõlas ka sotsiaalmeedia platvormide tingimustega (Freelon 2018) ja teaduseetika põhimõtetega (Zook *et al.* 2017; Markham *et al.* 2018). Parim viis andmete seadusliku ja eetilise kasutamise ning uuritava privaatsuse tagamiseks on uurija otsene koostöö platvormi haldava ettevõttega. Alternatiivne võimalus uurija jaoks on luua andmete kogumiseks oma programm, mis avatud rakendusliidese kaudu kogutavad avalikud sotsiaalmeedia andmed automaatselt anonümiseerib, tagades nii, et andmed ei sisalda postituse teinud kasutaja isikuandmeid.

Üks keerulisemaid ülesandeid sotsiaalmeedia andmete analüütikas on tagada andmete kogumine seadusi ja eetilisi põhimõtteid järgides ning isikuandmete kaitsmine ehk privaatsuse tagamine andmeid hoides, analüüsides ja tulemusi esitades. Lisaks on seoses kasvava suurandmete automatiseeritud ja ennustava analüütika rakendamisega tekkinud oht inimeste ekslikuks profileerimiseks, mis võib endaga kaasa tuua süveneva diskrimineerimise ja ebavõrdsuse kasvu ühiskonnas (Kitchin 2014). Sellised ohud võivad tekkida näiteks sotsiaal-ruumiliste suurandmete rakendamisel politsei järelevalves (Brayne 2017).

### 5.3.5. Kokkuvõte

Sotsiaalteadustes laiemalt on alles nüüd hakatud mõistma, milliseid võimalusi sotsiaalmeedia andmed inimeste ja ühiskonna uurimiseks pakuvad. Paranemas on ka teadlaste võimekus suurandmeid analüüsida ja tulemusi tõlgendada (Kitchin 2014). Ennekõike annavad sotsiaalmeedia andmed suurepärase võimaluse teostada sotsiaal-ruumilist analüüsi, võimaldades korraga arvesse võtta ühiskonnanähtuste ja -protsesside sotsiaalset, ruumilist ja ajalist mõõdet.

Sotsiaal-ruumilise analüüsiga on seni vähe tegeldud (nt Shelton *et al.* 2015). Suurem osa uuringuid on keskendunud kas ainult aegruumilisele või ainult sotsiaalsele mõõtmele. Võib siiski ennustada, et lähitulevikus

hakatakse neid kahte enam kombineerima ja sotsiaal-ruumiline analüüs leiab laiemat rakendust (Poorthuis, Zook 2017). Selle eelduseks on sotsiaalteadustes keerukate analüüsimeetodite rakendamine, mida võimaldab interdistsiplinaarne teaduskoostöö sotsiaal- ja arvutiteadlaste vahel, ning sotsiaalmeedia postituste eri elementide (vt alaptk 5.3.2.2) kombineerimine, sh sotsiaalmeedia kasutajatele täiendavate taustatunnuste tuletamine. Kiirelt areneb meetodika sotsiaalmeedia andmete kallutuse paremaks arvestamiseks ja sellest tulenevate vigade vähendamiseks järelduste tegemisel. Sellele aitab kaasa andmete kaudne kombineerimine teiste täiendavate andmeallikatega (sh registrid, küsitlusuuringud) ja tulemuste kriitiline valideerimine (sh kontrolluuringud).

Suurem areng leiab lähiaastatel aset sisuanalüüsi valdkonnas. Järjest paremad (süvaõppe) meetodid tekstilise ja visuaalse sisu automaatseks tuvastamiseks ja analüüsiks ning nende kombineerimine võimaldavad saada täpsemaid tulemusi ja teha usaldusväärsemaid järeldusi. Näiteks on hoiakute, hinnangute ja meelsuse automatiseeritud tuvastamine praegu veel keeruline postitustes kasutatavate eri keelte, slängi, lühendite, emotikonide ja sarkasmi tõttu. Oodata on ka videote ning kombineerituna hääle-pildi-teksti sisuanalüüsi analüüsimeetodeid (Baveye *et al.* 2015), kuivõrd videote osakaal kasvab sotsiaalmeedias kiiresti seoses ülikiire 5G-mobiilsidevõrgu üleilmse kasutuselevõtuga.

Kokkuvõttes peab tõdema, et sotsiaalmeedia platvormid ja nende andmed erinevad suuresti ja muutuvad ajas: platvormide luuakse, muudetakse ja pannakse kinni, nende andmestruktuurid muutuvad vastavalt ettevõtete vajadustele. Lisaks muutuvad andmete kasutustingimused, täieneb andmete teadusuuringutes kasutamise seotud seadusandlus ning paraneb teadlikkus ja muutuvad arusaamad üksikisiku privaatusest. Need on ülesanded meetodite ja tarkavarade arendamisele. Sotsiaalmeedia kui digitaalse suhtlusviisi populaarsus ühiskonnas aga aina kasvab ning praegustest raskustest hoolimata on sotsiaalmeedia andmed tulevikus üks olulisemaid andmeallikaid ühiskondlike nähtuste ja protsesside uurimisel.

## VIIDATUD KIRJANDUS

- Bateman, J. A.; Wildfeuer, J.; Hiippala, T. 2017. *Multimodality: Foundations, Research and Analysis a Problem-Oriented Introduction*. Boston: De Gruyter, Mouton.
- Batrinca, B.; Treleaven, P. C. 2015. *Social Media Analytics: A Survey of Techniques, Tools and Platforms*. – *AI and Society* 30, 1, 89–116.
- Baveye, Y.; Dellandrea, E.; Chamaret, C.; Chen, L. 2015. *LIRIS-ACCEDE: A Video Database for Affective Content Analysis*. – *IEEE Transactions on Affective Computing* 6, 1, 43–55.
- Becken, S.; Stantic, B.; Chen, J.; Alaei, A. R.; Connolly, R. M. 2017. *Monitoring the Environment and Human Sentiment on the Great Barrier Reef: Assessing the Potential of Collective Sensing*. – *Journal of Environmental Management* 203, 87–97.
- Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. 2017. *Enriching Word Vectors with Subword Information*. – *Transactions of the Association for Computational Linguistics* 5, 135–146.
- Bourdieu, P. 2013. *Distinction: A Social Critique of the Judgement of Taste*. Routledge.
- boyd, d.; Crawford, K. 2012. *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*. – *Information, Communication and Society* 15, 5, 662–679.
- Brayne, S. 2017. *Big Data Surveillance: The Case of Policing*. – *American Sociological Review* 82, 5, 977–1008.
- Croitoru, A.; Wayant, N. M.; Crooks, A.; Radzikowski, J. R.; Stefanidis, A. 2015. *Linking Cyber and Physical Spaces through Community Detection and Clustering in Social Media Feeds*. – *Computers, Environment and Urban Systems*, 53, 47–64.
- De Vos, J.; Schwanen, T.; Van Acker, V.; Witlox, F. 2013. *Travel and Subjective Well-Being: A Focus on Findings, Methods and Future Research Needs*. – *Transport Reviews*, 33, 4, 421–442.
- Freelon, D. 2018. *Computational Research in the Post-API Age*. – *Political Communication* 35, 4, 665–668.
- García-Palomares, J. C.; Salas-Olmedo, M. H.; Moya-Gómez, B.; Condeço-Melhorado, A.; Gutiérrez, J. 2018. *City Dynamics through Twitter: Relationships between Land Use and Spatiotemporal Demographics*. – *Cities* 72, 310–319.
- Giddens, A. 1985. *Time, Space and Regionalisation*. – *Social Relations and Spatial Structures*. Springer, 265–295.

- Golder, S. A.; Macy, M. W. 2014. Digital Footprints: Opportunities and Challenges for Online Social Research. – *Annual Review of Sociology* 40, 1, 129–152.
- Goodfellow, I.; Bengio, Y.; Courville, A. 2016. *Deep Learning. Adaptive Computation and Machine Learning*. Cambridge MA: MIT Press.
- Han, J.; Pei, J.; Kamber, M. 2011. *Data Mining: Concepts and Techniques*. Elsevier.
- Hasnat, M. M.; Hasan, S. 2018. Identifying Tourists and Analyzing Spatial Patterns of Their Destinations from Location-Based Social Media Data. – *Transportation Research Part C: Emerging Technologies* 96, 38–54.
- Hawelka, B.; Sitko, I.; Beinat, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. 2014. Geo-Located Twitter as Proxy for Global Mobility Patterns. – *Cartography and Geographic Information Science* 41, 3, 260–271.
- He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. 2017. Mask R-CNN. – 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2980–2988. doi:10.1109/ICCV.2017.322.
- Heikinheimo, V.; Tenkanen, H.; Bergroth, C.; Järv, O.; Hiippala, T.; Toivonen, T. 2020. Understanding the use of urban green spaces from user-generated geographic information. – *Landscape and Urban Planning* 201, 103845.
- Hiippala, T.; Hausmann, A.; Tenkanen, H.; Toivonen, T. 2019. Exploring the Linguistic Landscape of Geotagged Social Media Content in Urban Environments. – *Digital Scholarship in the Humanities* 34, 2, 290–309.
- Hiippala, T.; Väisänen, T.; Toivonen, T.; Järv, O. 2020. Mapping the Languages of Twitter in Finland: Richness and Diversity in Space and Time. – *Neuphilologische Mitteilungen*, 1–42.
- Hu, Y. 2018. Geo-Text Data and Data-Driven Geospatial Semantics. – *Geography Compass* 12, 11, e12404.
- Hughes, D. J.; Rowe, M.; Batey, M.; Lee, A. 2012. A Tale of Two Sites: Twitter vs. Facebook and the Personality Predictors of Social Media Usage. – *Computers in Human Behavior* 28, 2, 561–569.
- Jiang, Y.; Li, Z.; Ye, X. 2019. Understanding Demographic and Socioeconomic Biases of Geotagged Twitter Users at the County Level. – *Cartography and Geographic Information Science* 46, 3, 228–242.
- Järv, O.; Müürisepp, K.; Ahas, R.; Derudder, B.; Witlox, F. 2015. Ethnic Differences in Activity Spaces as a Characteristic of Segregation: A Study Based on Mobile Phone Usage in Tallinn, Estonia. – *Urban Studies* 52, 14, 2680–2698.
- Karpathy, A.; Fei-Fei, L. 2015. Deep Visual-Semantic Alignments for Generating Image Descriptions. – *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3128–3137.

- Kemp, S. 2019. Digital 2019: Global Digital Report. Kepios. <https://datareportal.com/>.
- Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kwan, M.-P. 2013. Beyond Space (As We Knew It): Toward Temporally Integrated Geographies of Segregation, Health, and Accessibility. – *Annals of the Association of American Geographers* 103, 5, 1078–1086.
- Lamanna, F.; Lenormand, M.; Salas-Olmedo, M. H.; Romanillos, G.; Gonçalves, B.; Ramasco, J. J. 2018. Immigrant Community Integration in World Cities. – *Plos One* 13, 3, e0191612.
- Lefebvre, H. 1991. *The Production of Space*. Malden: Blackwell Publishing.
- Li, L.; Goodchild, M. F.; Xu, B. 2013. Spatial, Temporal, and Socioeconomic Patterns in the Use of Twitter and Flickr. – *Cartography and Geographic Information Science* 40, 2, 61–77.
- Lloyd, A.; Cheshire, J. 2017. Deriving Retail Centre Locations and Catchments from Geo-Tagged Twitter Data. – *Computers, Environment and Urban Systems* 61, 108–118.
- Lomborg, S.; Bechmann, A. 2014. Using APIs for Data Collection on Social Media. – *The Information Society* 30, 4, 256–265.
- Longley, P. A.; Adnan, M.; Lansley, G. 2015. The Geotemporal Demographics of Twitter Usage. – *Environment and Planning A* 47, 2, 465–484.
- Markham, A. N.; Tiidenberg, K.; Herman, A. 2018. Ethics as Methods: Doing Ethics in the Era of Big Data Research: Introduction. – *Social Media + Society* 4, 3, 205630511878450.
- McPherson, M.; Smith-Lovin, L. Cook, J. M. 2001. Birds of a Feather: Homophily in Social Networks. – *Annual Review of Sociology* 27, 1, 415–444.
- Mitchell, L.; Frank, M. R.; Harris, K. D.; Dodds, P. S.; Danforth, C. r M. 2013. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. – *Plos One* 8, 5, e64417.
- Noels, K. A.; Pon, G.; Clement, R. 1996. Language, Identity, and Adjustment: The Role of Linguistic Self-Confidence in the Acculturation Process. – *Journal of Language and Social Psychology* 15, 3, 246–264.
- Osorio-Arjona, J.; García-Palomares, J. C. 2019. Social Media and Urban Mobility: Using Twitter to Calculate Home-Work Travel Matrices. – *Cities* 89, 268–280.
- Pontes, T.; Magno, G.; Vasconcelos, M.; Gupta, A.; Almeida, J.; Kumaraguru, P.; Almeida, V. 2012. Beware of What You Share: Inferring Home Location in Social Networks. – 2012 IEEE 12th International Conference on Data Mining Workshops, 571–578. Brussels: IEEE.



- Poorthuis, A. 2018. How to Draw a Neighborhood? The Potential of Big Data, Regionalization, and Community Detection for Understanding the Heterogeneous Nature of Urban Neighborhoods. – *Geographical Analysis* 50, 2, 182–203.
- Poorthuis, A.; Zook, M. 2017. Making Big Data Small: Strategies to Expand Urban and Geographical Research Using Social Media. – *Journal of Urban Technology* 24, 4, 115–135.
- Rashidi, T. H.; Abbasi, A.; Maghrebi, M.; Hasan, S.; Waller, T. S. 2017. Exploring the Capacity of Social Media Data for Modelling Travel Behaviour: Opportunities and Challenges. – *Transportation Research Part C: Emerging Technologies* 75, 197–211.
- Rawat, W.; Wang, Z. 2017. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. – *Neural Computation* 29, 9, 2352–2449.
- Roberts, H.; Resch, B.; Sadler, J.; Chapman, L.; Petutschnig, A.; Zimmer, S. 2018. Investigating the Emotional Responses of Individuals to Urban Green Space Using Twitter Data: A Critical Comparison of Three Different Methods of Sentiment Analysis. – *Urban Planning* 3, 1, 21.
- Ruths, D.; Pfeffer, J. 2014. Social Media for Large Studies of Behavior. – *Science* 346 (6213), 1063–1064.
- Salas-Olmedo, M. H.; Rojas Quezada, C. 2017. The Use of Public Spaces in a Medium-Sized City: From Twitter Data to Mobility Patterns. – *Journal of Maps* 13, 1, 40–45.
- Savage, M. 1988. The Missing Link? The Relationship between Spatial Mobility and Social Mobility. – *The British Journal of Sociology* 39, 4, 554–577.
- Scott, J.; Carrington, P. J. (eds.) 2011. *The Sage Handbook of Social Network Analysis*. London: Sage.
- Shaughnessy, K.; Reyes, R.; Shankardass, K.; Sykora, M.; Feick, R.; Lawrence, H.; Robertson, C. 2018. Using Geolocated Social Media for Ecological Momentary Assessments of Emotion: Innovative Opportunities in Psychology Science and Practice. – *Canadian Psychology* 59, 1, 47–53.
- Shelton, T.; Poorthuis, A.; Zook, M. 2015. Social Media and the City: Rethinking Urban Socio-Spatial Inequality Using User-Generated Geographic Information. – *Landscape and Urban Planning* 142, 198–211.
- Sloan, L.; Morgan, J.; Burnap, P.; Williams, M. 2015. Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. – *Plos One* 10, 3, e0115545.
- Sloan, L.; Quan-Haase, A. (eds.) 2017. *The Sage Handbook of Social Media Research Methods*. London: Sage.

- Soja, E. W. 2013. *Seeking Spatial Justice*, vol. 16. University of Minnesota Press.
- Stefanidis, A.; Vraga, E.; Lamprianidis, G.; Radzikowski, J.; Delamater, P. L.; Jacobsen, K. H.; Pfoser, D.; Croitoru, A.; Crooks, A. 2017. Zika in Twitter: Temporal Variations of Locations, Actors, and Concepts. – *JMIR Public Health and Surveillance* 3, 2, e22.
- Sui, D.; Goodchild, M. 2011. The Convergence of GIS and Social Media: Challenges for GIScience. – *International Journal of Geographical Information Science* 25, 11, 1737–1748.
- Zook, M. 2017. Crowd-Sourcing the Smart City: Using Big Geosocial Media Metrics in Urban Governance. – *Big Data & Society* 4, 1, 205395171769438.
- Zook, M.; Poorthuis, A. 2014. Offline Brews and Online Views: Exploring the Geography of Beer Tweets. – Mark Patterson, Nancy Hoalst-Pullen (eds.), *The Geography of Beer*. Dordrecht: Springer, 201–209.
- Zook, M.; Barocas, S.; boyd, d.; Crawford, K.; Keller, E.; Gangadharan, S. P.; Goodman, A. *et al.* 2017. Ten Simple Rules for Responsible Big Data Research. – *Plos Computational Biology* 13, 3, e1005399.
- Tenkanen, H.; Di Minin, E.; Heikinheimo, V.; Hausmann, A.; Herbst, M.; Kajala, L.; Toivonen, T. 2017. Instagram, Flickr, or Twitter: Assessing the Usability of Social Media Data for Visitor Monitoring in Protected Areas. – *Scientific Reports*, 7, 1–11.
- Toivonen, T.; Heikinheimo, V.; Fink, C.; Hausmann, A.; Hiippala, T.; Järv, O.; Tenkanen, H.; Di Minin, E. 2019. Social Media Data for Conservation Science: A Methodological Overview. – *Biological Conservation* 233, 298–315.
- Urry, J. 2000. *Sociology beyond Societies: Mobilities for the Twenty-First Century*. International Library of Sociology. London: Routledge.
- Valentine, G.; Sadgrove, J. 2012. Lived Difference: A Narrative Account of Spatiotemporal Processes of Social Differentiation. – *Environment and Planning A* 44, 9, 2049–2063.
- Väisänen, T. 2018. Suomen- ja englanninkieliset digitaaliset kaupunkilat Helsingissä? – Case Instagram. Pro Gradu -tutkielma. Helsingin yliopisto.
- Wang, Q.; Phillips, N. E.; Small, M. L.; Sampson, R. J. 2018. Urban Mobility and Neighborhood Isolation in America's 50 Largest Cities. – *Proceedings of the National Academy of Sciences* 115, 30, 7735–7740.