

Poliitikkojen ja politiikan toimittajien (näkyvät) verkostot Twitterissä.

Aineistonäkökulma
Paavo Arvola (informaatiotutkimus)



Tutkimuskohde ja tutkijat



Jussi Halla-aho
@Halla_aho

Kylläpä puhelin soi. Perussuomalaiset kuuluvat mellakoivan Yhdysvalloissa.

10.29 ap. · 7. tammik. 2021 · Twitter Web App

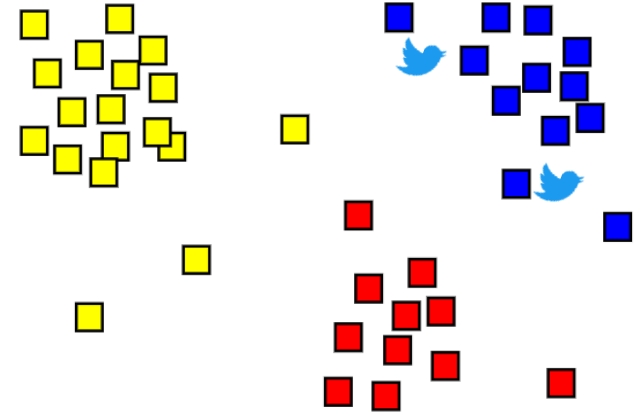
188 uudelleentwiittausta 74 Twiitin lainaukset 3 005 tykkäystä

- Twitter tilit ja tviitit
 - Kansanedustajat (176) ja politiikan toimittajat (164)
- Miten politiikan toimittajat ja kansanedustajat ovat verkostoituneet Twitterissä?
 - Kuka seuraa ketä? Ketkä käyvät keskusteluita?
 - Case: Capitol Hillin valtauksen jälkeiset päivät (n. 4500 tviittiä)
- Tiedonlähteet
 - Twitterin julkinen rajapinta, eduskunnan avoin data, digi- ja väestötietoviraston avoin data, semanttinen web, Wikipedia, internet...
- Tutkijat: Paavo Arvola (informaatiotutkimus), Jaana Kuusipalo (sukupuolen ja politiikan tutkimus) ja Iiris Ruoho (mediatutkimus).
 - Monimenetelmäistä tutkimusta

Mitä tietokoneella ei voi
laskea, sen
IHMINEN RATKAISEE

Laskennalliset menetelmät Twitterin tutkimuksessa

- Tviittejä voidaan käsitellä automaattisesti
 - Ryhmitellä
 - Luokitella
 - Hakea
- Perustuvat tviittien tekstisisältöön
 - Joskus verkostoihin
- Mikä on menetelmän saanti ja tarkkuus?
 - Menetelmiä arvioidaan vertaamalla niitä ihmisten tekemiin arvioihin.
- Hyvä renki, mutta huono isäntä
 - Erityisesti Twitterissä



”Algoritmien” ongelmia Twitterissä



Pia Kauma
@PiaKauma



Eilisten tapahtumien jälkeen tuntuu siltä, että demokratian vaaliminen ja kehittäminen on entistäkin tärkeämpi asia juuri nyt. Ja myös henkilökohtainen missiini. "Kauma tarkkailemaan Kazakstanin vaaleja" eduskunta.fi/FI/tiedotteet/...

12.40 ip. · 7. tammik. 2021 · Twitter for Android

- Tviitit ovat erittäin lyhyitä dokumentteja
 - Vähän tekstiä tai ei ollenkaan tekstiä (1-280 merkkiä)
 - Multimediaa, linkkejä jne.
- Tviittien merkitys ja tulkinta riippuvat *kontekstista*
 - Keskustelukonteksti tai muu
 - Niiden aihetta EI aina voi päätellä edes teoriassa pelkän sisällön perusteella.
- Luonnollisen kielen aiheuttamat haasteet
 - Varsinkin suomen kielessä
- Käsitteisiin voidaan viitata lukemattomin erilaisin ilmaisuin
 - Ilmiön alkaessa terminologia on vakiintumatonta

Capitol Hill Twitter keskustelu 6.1. – 14.1.

- Kansanedustajat ja politiikan toimittajat
 - 176 + 160 twittertiliä, 4387 tviittiä
 - Twitter REST API
- Luokittelu teemoihin
 - "Käsityönä"
- Relevantit, 1815 tviittiä
 - "Demokratia", 285 tviittiä
 - "Media ja sananvapaus" 567 tviittiä
 - "Populismi ja poliittinen väkivalta" 924 tviittiä
 - "Muu", 39 tviittiä
- Ei relevantit 2572 tviittiä



By Tyler Merbler from USA - DSC09156, CC BY 2.0,
<https://commons.wikimedia.org/w/index.php?curid=98780310>

Information Retrieval Virtual Lab

- Tviittien manuaaliseen arvioimiseen ja/tai luokitteluun
 - Soveltuu myös muille dokumenteille
- Aineiston selaaminen ja siihen tehtävät haut
 - ”Haku-Lulu”
 - Laadullisen analyysin tueksi
- Toimii virtuaalisesti ja verkon yli millä tahansa selaimella
- Käytettävissä muuhunkin tutkimukseen
 - Ota yhteyttä (Paavo Arvola), jos tarvetta



Kysymyksiä: Kerääminen – tallentaminen – analysointi (aineiston käsittely)

- *Kiinnostaisi kuulla konkreettisia esimerkkejä siitä, miten aineistoja on kerätty, tallennettu ja analysoitu. Esim. mitä ohjelmia tai alustoja käytössä.*
- *Miten aineiston keruu kannattaa käytännössä toteuttaa? Mitä someaineiston rajaamisessa kannattaa ottaa huomioon.*
- *Miten someaineistoja voi tallentaa tutkimusta varten? Myös tekniset vinkit kiinnostavat.*

Aineiston keruu ja analyysi



Keruun työkaluja

- Twitter API, Twarc



Aineistojen jalostamisen ja analyysin työkaluja

- Gephi (verkostojen analyysi ja ladontaohjelmisto)
- MS PowerBI
- Taulukkolaskenta ja tilasto-ohjelmistot (SPSS, MSEXcel)
- Information Retrieval Virtual Lab (Python, PHP)

Verkostoanalyysiprosessi (“algoritmi”)

1. Ilmiön tulkinta verkostoksi
 - Mitä ovat solmut ja yhteydet tässä ilmiössä?
2. Datan kerääminen
3. Datan korjaaminen, muokkaus ja jalostaminen
4. Datan muuntaminen verkostoksi
5. Verkoston sijoittaminen kaksiulotteiseen tasoon (ladonta)
 - Visualisointi käyttökelpoiseksi esitykseksi
6. Havaintojen tekeminen
 - Havainnoista keskusteleminen
 - 2-5 voidaan yl. automatisoida esim. Twitter-tutkimuksessa
 - Toistetaan vaiheita (tarvittaessa)



Verkostoanalyysi Twitterissä

Seuraajaverkostot (User's network)

- Tutkitaan yhden tai useamman tietyn profiilin verkostoitumista

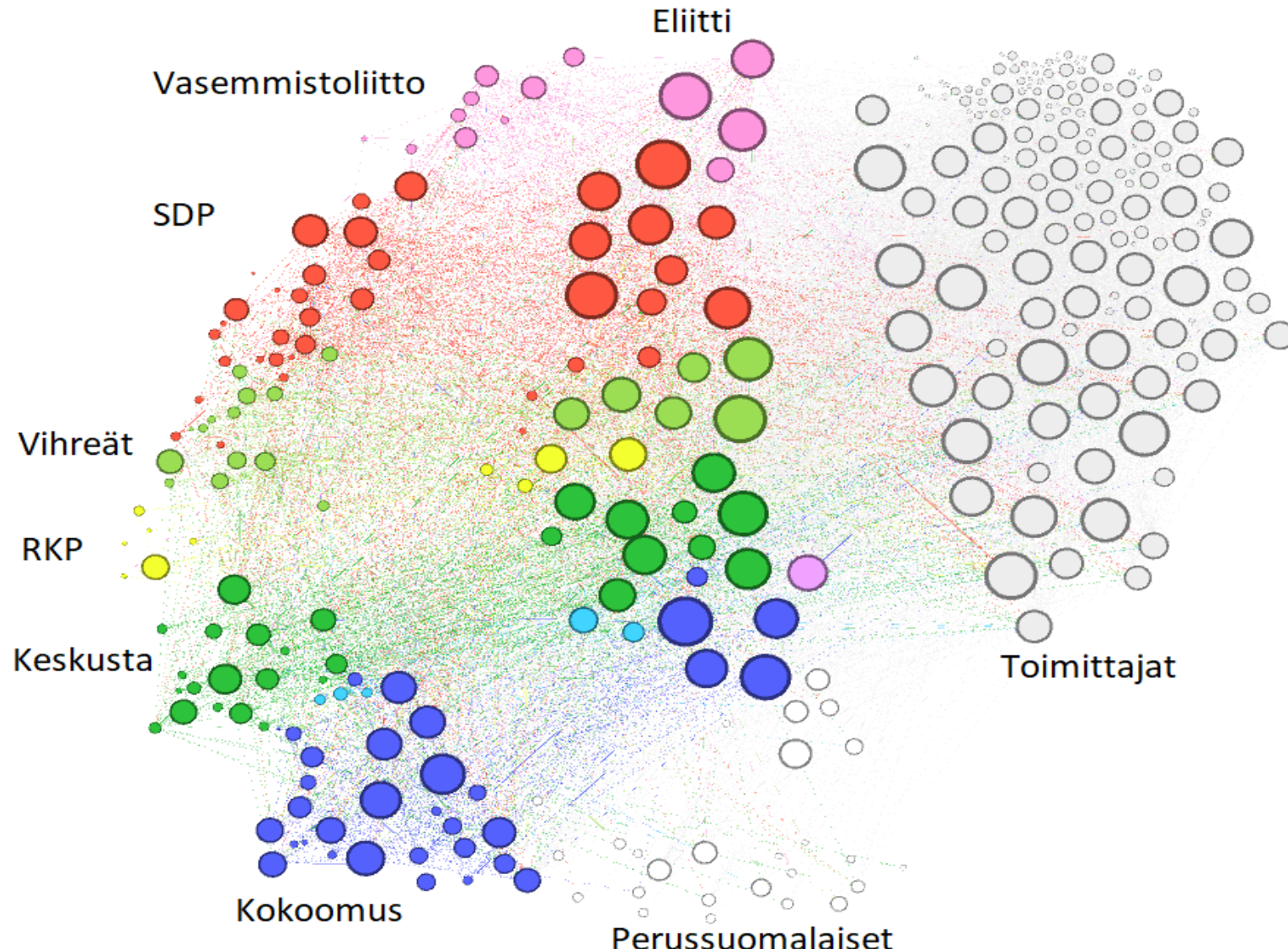
Keskusteluverkostot

- Tutkitaan tietyn teeman ympärillä tapahtuvaa verkostoitumista
- Verkoston virittävät kaikki ne tviitit, jotka täsmäävät hakulausekkeeseen.

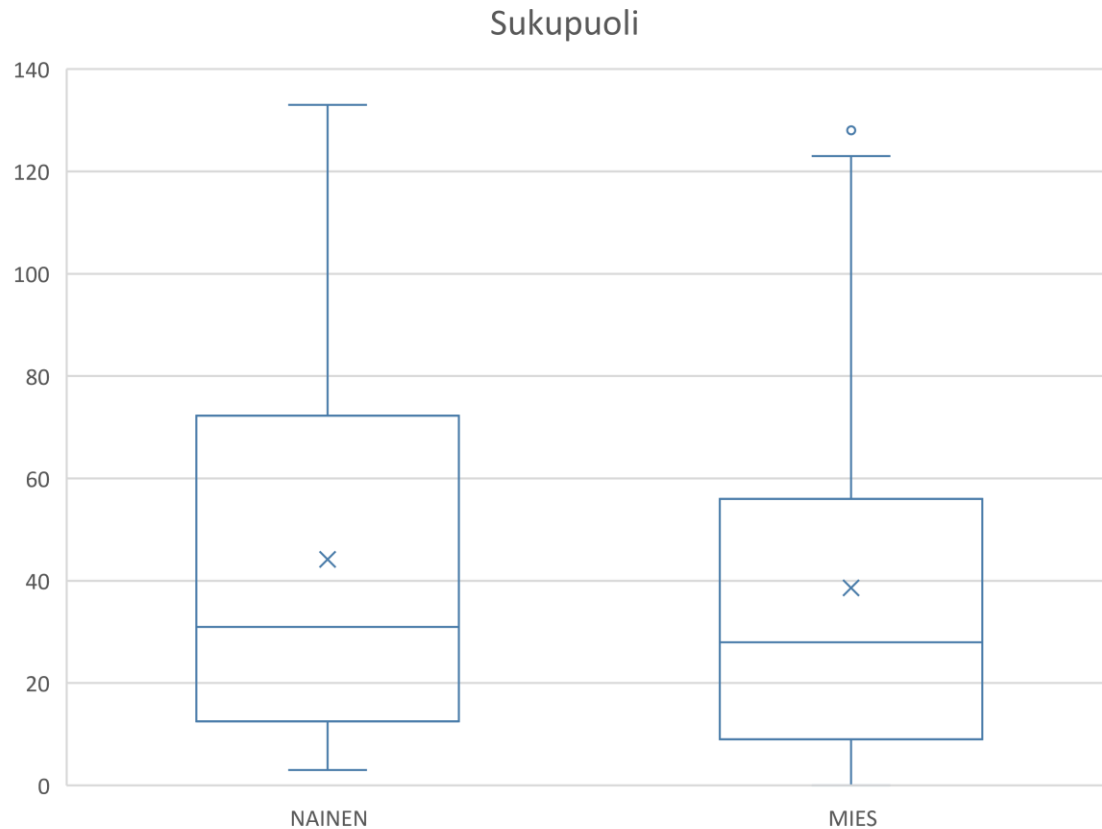
Seuraajaverkostot

- Tilit seuraavat toisiaan Twitterissä
 - Informaation saaminen seuratuilta tileiltä omaan uutisvirtaan
 - Muodostaa eksplisiittisen seuraajaverkoston
- Seuraaminen on myös sosiaalista
 - Seurattu tili on englanniksi "friend".
 - Molemminpuolinen seuranta => "sosiaalinen" suhde
- Määrällinen tutkimus

Seuraajaverkosto



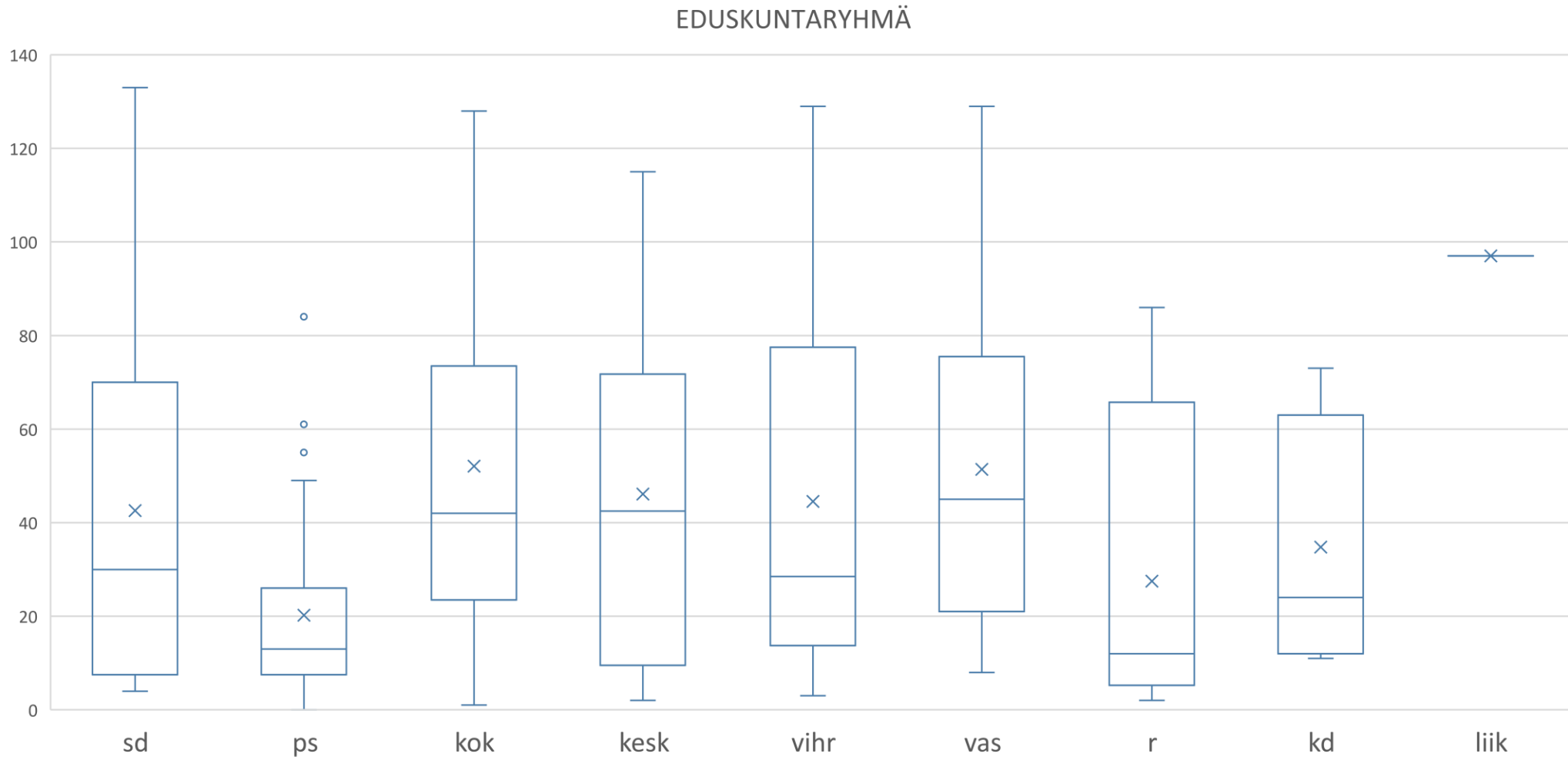
Toimittajien seuraamisten jakautuminen 1



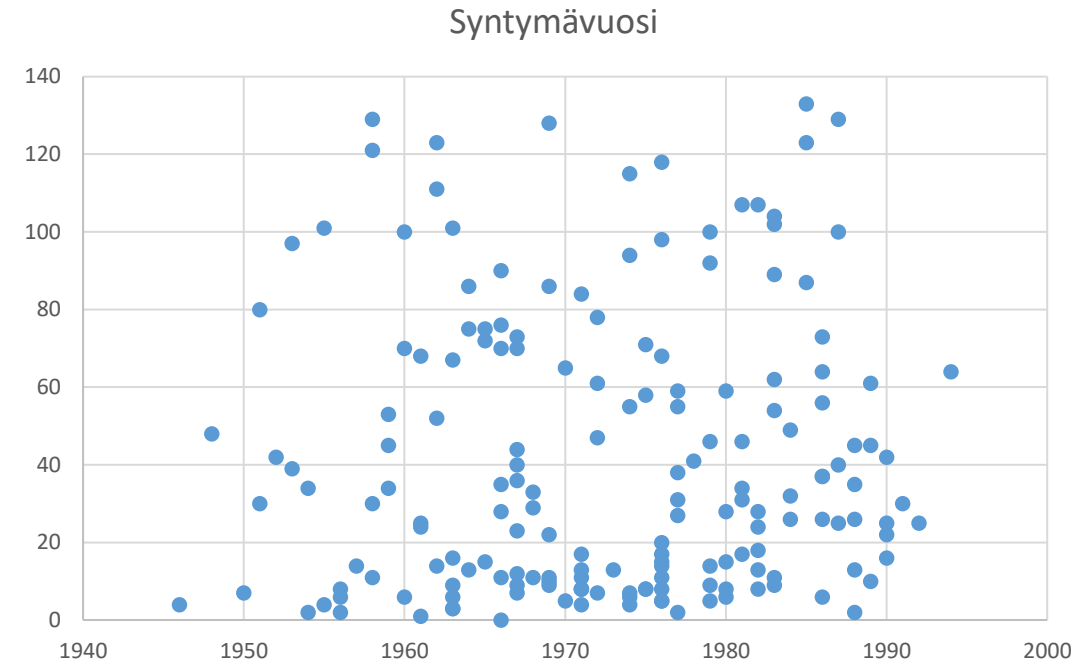
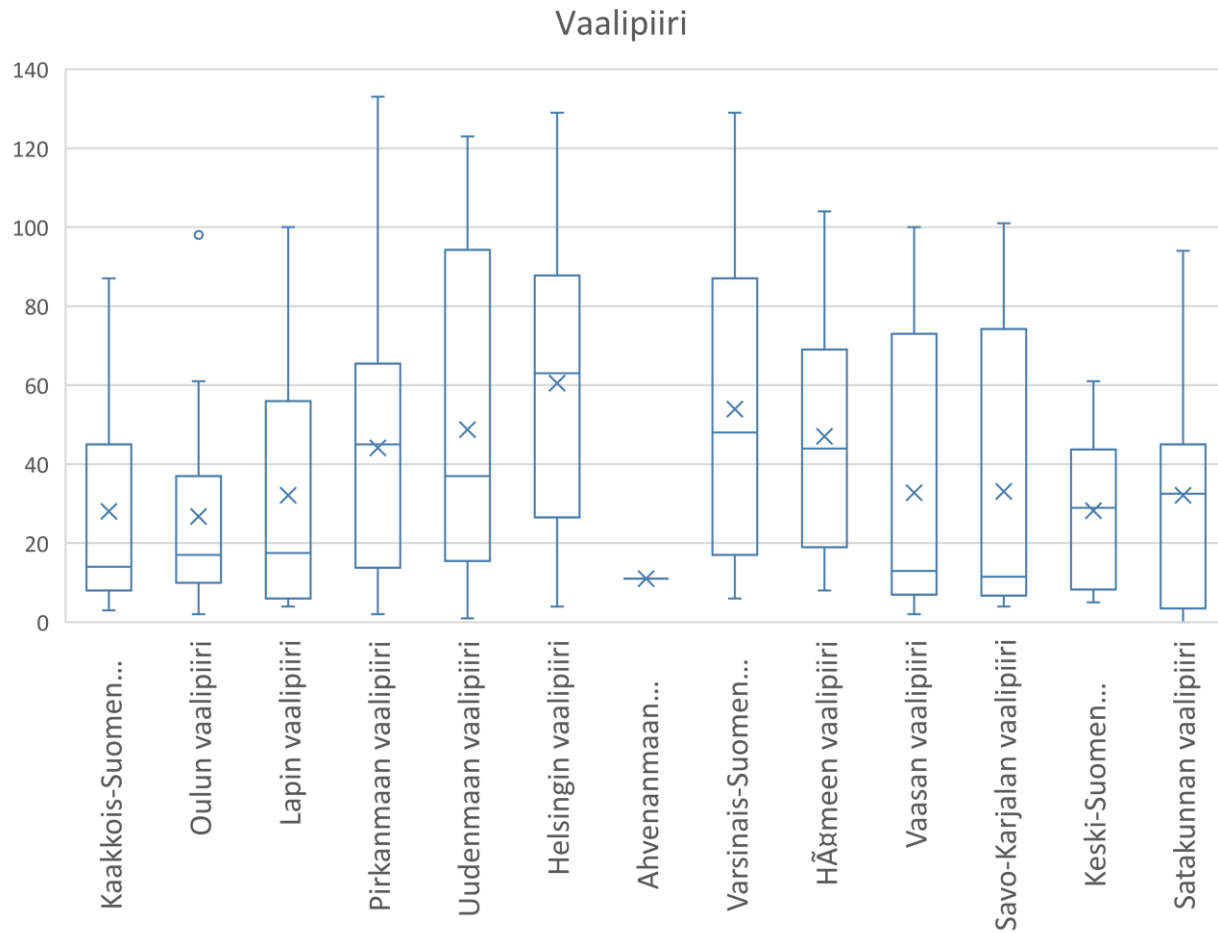
Kansanedustajien itsensä eniten seuraamat:
Petteri Orpo, Sanna Marin, Antti Rinne, Antti Kaikkonen, Pekka Haavisto, Li Andersson, Juhana Vartiainen, Katri Kulmuni, Antti Lindtman, Paula Risikko.

Aineiston eniten seuraamat koko Twitterissä:
Sanna Marin, Petteri Orpo, Sauli Niinistö, Ville Niinistö, Valtioneuvosto, Alexander Stubb, Pekka Haavisto, Merja Ylä-Anttila, Li Andersson, Antti Rinne

Toimittajien seuraamisten jakautuminen 2



Toimittajien seuraamisten jakautuminen 3



Analyysimenetelmiä

- Määrällinen tutkimusote
 - Luokittelu
 - Ryhmittely
 - Verkostoanalyysi
 - Kuvailevat menetelmät
 - Tilastotestit (Kruskal-Wallis, t-testi)
- Laadullinen tutkimusote
 - Verkostoanalyysi
 - Sisällönanalyysi

Kiitos!

Lisälukemista: <https://research.tuni.fi/comet/yleinen/valtatwitterissa/>