



Turun yliopisto
University of Turku

INFORMAATION USKOTTAVUUDEN ME- KANISMIT SOSIAALISESSA MEDIASSA

**Sosiaalinen media sijoittajamentimentin lähteenä ja mielipidevaikut-
tamisen verkostona**

Taloussosiologian pro gradu -tutkielma
Master's Thesis
in Economic Sociology

Laatija/Author:
Aki Raittila

Ohjaajat/Supervisors:
VTT Kim Holmberg
VTT Erica Åberg

22.2.2024
Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkas-
tettu Turnitin OriginalityCheck -järjestelmällä.



<input type="checkbox"/>	Kandidaatintutkielma
<input checked="" type="checkbox"/>	Pro gradu -tutkielma
<input type="checkbox"/>	Lisensiaatintutkielma
<input type="checkbox"/>	Väitöskirja

Oppiaine	Taloussosiologia	Päivämäärä	22.2.2024
Tekijä(t)	Aki Raittila	Sivumäärä	64 + 13 liitesivua
Otsikko	Informaation uskottavuuden mekanismit sosiaalisessa mediassa – Sosiaalinen media sijoittajacentimentin lähteenä ja mielipidevaikuttamisen verkostona.		
Ohjaaja(t)	VTT Kim Holmberg & VTT Erica Åberg		

Tässä Pro gradu tutkielmassa tutkitaan informaation uskottavuuden muodostumista, sekä mielipidevaikutajuuksia sosiaalisen median verkostoissa. Tutkimuksessa selvitetään käyttäjien aktiivisuuden ja huomion jakautumista, sekä niihin vaikuttavia mekanismeja Redditiin wallstreetbets kanavalla. Tutkimuksen tavoitteena on selvittää erilaisten auktoriteettisignaalien, kuten käyttäjän karman ja palkintojen vaikutusta julkaisuiden uskottavuuteen. Lisäksi tutkimuksessa pyritään paikantamaan kanavalta mielipidevaikuttajia etsimällä verkostosta keskeisiä ja aktiivisia käyttäjiä. Tutkimusaineistona käytetään Redditiä kerättyä käyttäjädataa, jota analysoidaan muun muassa regressioanalyysillä ja verkostanalyysillä.

Tutkimuksen tulosten mukaan sekä käyttäjien aktiivisuus, että julkaisuiden saama huomio jakautuvat pääosin pienelle joukolle kanavalla. Jakaumien, sekä verkostanalyysin avulla saatiin alustavia tuloksia mielipidevaikuttajien määrästä ja aktiivisuudesta kanavalla. Regressioanalyysissä aineistosta löydettiin säännönmukaisia yhteyksiä esimerkiksi käyttäjän karman ja julkaisun saaman huomion välillä. Tulokset antoivat myös viitteitä siitä, minkälaista sisältöä kanavalla arvostetaan.

Tämä tutkimus toteutetaan osana *Developing an early signal system to identify shifts in global stock market trends* -tutkimushanketta, jonka tavoite on kehittää pörssikurssien heilahtelua verkkokeskustelun avulla ennustava signaalijärjestelmä. Tämän tutkimuksen tuloksia voidaan hyödyntää signaalijärjestelmässä yksittäisten viestien vaikuttavuuden arvioimiseksi, antamalla viesteille julkaisijan verkosto- ja auktoriteettiasemaan perustuvia painokertoimia.

Asiasanat	Mielipidevaikuttaja, luottamus, uskottavuus, informaatio, sosiaalinen media, sijoittaminen, Reddit
Muita tietoja	

Sisällysluettelo

1	JOHDANTO	1
2	SIJOITUSTOIMINTA JA PÖRSSIKURSSIEN ENNUSTAMINEN	8
3	INFORMAATION ARVIOINTI.....	14
3.1	Kognitiivisen auktoriteetin teoria.....	16
3.2	Informaation uskottavuus sosiaalisessa mediassa	17
3.3	Mielipidevaikuttajuus sosiaalisen median verkostoissa.....	22
4	TUTKIMUSASETELMA	25
4.1	Tutkimuskysymykset	26
4.2	Aineisto ja menetelmät.....	27
5	TULOKSET.....	36
5.1	Julkaisuiden ja kommenttien jakautuminen	36
5.2	Mielipidevaikuttajien paikantaminen.....	38
5.3	Auktoriteettisignaalit.....	43
6	JOHTOPÄÄTÖKSET	59
	LÄHDELUETTELO.....	65
7	LIITTEET	72

Lista kuvaajista

Kuva 1. Julkaisun luotettavuuden vaikutussuhteet	25
Kuva 2. Datan siistimisprosessi	30
Kuva 3. Mielipidevaikuttajien asema huomion ja aktiivisuuden suhteen.....	39
Kuva 4. Verkostokartta kanavan keskeisimmistä käyttäjistä.....	40
Kuva 5. Kommenttien ja karman hajontakuviot	51
Kuva 6. Kommenttien saamien äänten ja karman hajontakuviot.....	55

Kuva 7. Upvote ration ja kommenttien määrän hajontakuvio	57
--	----

Lista taulukoista

Taulukko 1. Aktiivisuuden ja huomion jakautuminen r/wallstreetbets -kanavalla....	37
Taulukko 2. Muuttujien väliset korrelaatiot.....	43
Taulukko 3. Flairien yhteys huomioon	45
Taulukko 4. Julkaisuiden saamien kommenttien regressioanalyysi (Malli 1.).....	48
Taulukko 5. Julkaisun saamien äänten regressioanalyysi (Malli 2.).....	52
Taulukko 6. Kommenttien saamien äänten regressioanalyysi (Malli 3.).....	54

Lista liitteistä

Liite 1. Julkaisuiden jakauma (logaritmuunnettu).	72
Liite 2. Kommenttien jakauma (logaritmuunnettu).	72
Liite 3. Julkaisuiden saamat kommentit (logaritmuunnettu).	73
Liite 4. Julkaisuiden saamat äänet (logaritmuunnettu).	73
Liite 5. Julkaisun saamien kommenttien keskiarvo per käyttäjä (logaritmuunnettu).74	
Liite 6. Kommenttien saamat äänet (logaritmuunnettu)	74
Liite 7. Kommenttien saamien äänten keskiarvo per käyttäjä (logaritmuunnettu) 75	
Liite 8. Julkaisuiden saamien äänten keskiarvo per käyttäjä (logaritmuunnettu).. 75	
Liite 9. Mieli-pidevaikuttajat julkaisemisen osalta	76
Liite 10. Mieli-pidevaikuttajat kommentoinnin osalta	77
Liite 11. Hajontakuvio keskeisyysaste vs. julkaisun saamat kommentit (logaritmuunnettu).....	78

Liite 12. Hajontakuvio keskeisyysaste vs. julkaisun saamat äänet (logaritmimuunnettu).....	78
Liite 13. Hajontakuvio karma vs. julkaisun saamat kommentit (logaritmimuunnettu)	79
Liite 14. Hajontakuvio karma vs. julkaisun saamat äänet (logaritmimuunnettu)	79
Liite 15. Hajontakuvio trophyt vs. julkaisun saamat kommentit (logaritmimuunnettu)	80
Liite 16. Hajontakuvio trophyt vs. julkaisun saamat äänet (logaritmimuunnettu).....	80
Liite 17. Eri flairien saamien äänten määrän visualisointi.....	81
Liite 18. Julkaisuiden saamien kommenttien regressioanalyysi, residuals vs. fitted - kuvaaja ja Q-Q kaavio.	81
Liite 19. Julkaisun saamien äänten regressioanalyysi, residuals vs. fitted -kuvaaja ja Q-Q kaavio.	82
Liite 20. Hajontakuvio keskeisyys vs. kommenttien saamat äänet (logaritmimuunnettu).....	82
Liite 21. Hajontakuvio karma vs. kommenttien saamat äänet (logaritmimuunnettu)	83
Liite 22. Hajontakuvio trophyt vs. kommenttien saama äänet.....	83
Liite 23. Regressioanalyysi kommenttien saamat äänet, residuals vs. fitted kuvaaja ja Q-Q kaavio.	84

1 JOHDANTO

Tässä pro-gradu tutkielmassa tutkitaan informaation uskottavuuden muodostumiseen vaikuttavia mekanismeja sosiaalisen median verkostoissa. Tutkimuksen tavoitteena on selvittää, mitkä tekijät vaikuttavat luottamukseen informaatiota kohtaan Redditiin Wallstreetbets -kanavalla. Taustoitetaan aluksi tutkimuksen innoitusta ja Redditiin toimintaperiaatteita, ja siirrytään sitten syvemmälle teoriataustaan.

Vuoden 2021 tammikuussa Yhdysvaltain pörssissä tapahtui jotain ennenkuulumatonta, kun videopelien jälleenmyyntiin erikoistuneen yrityksen GameStopin osakekurssi lähti odottamattomasti jyrkkään nousuun. Yksi keskeisin nousun taustalla vaikuttanut tekijä oli Redditiin wallstreetbets -nimisellä sijoituskanavalla syntynyt liike, jonka myötä suuri joukko piensijoittajia ryhtyi tarkoituksellisesti ajamaan osakkeen hintaa ylös.

Reddit on vuonna 2005 perustettu sosiaalisen median alusta, jossa käyttäjät voivat keskustella toistensa kanssa erilaisista aiheista. Redditillä on nykyisin yli 70 miljoonaa päivittäistä käyttäjää ympäri maailmaa, ja yli sata tuhatta aktiivista eri aiheille omistettua kanavaa (Reddit 2023)¹. Redditiä voidaan pitää puolianonyyminä alustana, sillä se ei vaadi käyttäjiä kirjautumaan omalla nimellään ja tiedoillaan, vaan Redditiin kirjaututaan vapaasti keksityllä nimimerkillä, joka ei yleensä ole käyttäjän todellinen identiteetti. Nimimerkki ja profiili kuitenkin näkyvät muille käyttäjille. (Kilgo ym. 2016, van der Nagel & Firth 2015.)

Reddit rakentuu eri aihekanavien varaan, jotka ovat eri aiheille omistettuja keskustelufoorumeja. Käyttäjät liittyvät itseään kiinnostaville kanaville, jonka jälkeen he voivat seurata kanavalla käytävää keskustelua. Käyttäjät voivat julkaista kanavalla keskustelunavauksia, jonka jälkeen muut käyttäjät voivat kommentoida julkaisua ja käydä siitä keskustelua. (Reddit 101 2017.)² Julkaisut ja keskustelut voivat olla enemmän tai vähemmän informatiivisia ja asiallisia. Ne voivat sisältää esimerkiksi kuvia, videoita tai linkkejä ulkopuolisille sivuille, tai olla puhtaasti julkaisijan omaa tekstiä. Monesti käyttäjät käyvät

¹ <https://www.redditinc.com/>

² https://www.reddit.com/wiki/reddit_101/

esimerkiksi kysymässä apua tai mielipiteitä kanavan muilta käyttäjiltä. Redditissä on käytössä äänijärjestelmä, joka muistuttaa esimerkiksi Facebookin tykkäyksiä. Käyttäjillä on mahdollisuus antaa julkaisuille ja kommentteille positiivisia tai negatiivisia ääniä julkaisuun kiinnitetyillä nuolinäppäimillä (Reddit Help 2023a)³.

Aktiivinen osallistuminen ja sisällön tuottaminen kerryttää Reddit käyttäjille niin kutsutua karmaa. Karma -pisteet ovat käyttäjille kertyviä pisteitä, jotka heijastavat käyttäjän panoksen merkitystä Reddit yhteisölle. Karma näkyy julkisesti käyttäjän profiilissa myös muille käyttäjille. Käyttäjän julkaisuiden ja kommenttien saamat positiiviset äänet kerryttävät karmaa ja vastaavasti negatiiviset äänet vähentävät sitä. (Reddit Help 2023b.)⁴ Tämä pistejärjestelmä kannustaa julkaisemaan laadukasta sisältöä ja antaa yhteisölle myös mahdollisuuden arvioida muita käyttäjiä. Käyttäjiä, joilla on korkea karma, arvostetaan Redditissä. (Anderson 2015.)

Julkaisuiden sisältö riippuu paljon kanavan aiheesta, jotka vaihtelevat vakavammista vähemmän vakaviin. Kanavia on omistettu esimerkiksi uutisille, politiikalle, erilaisille harrastuksille, kulttuurille ja huumorille. Redditin eri kanavat ovat käyttäjien perustamia ja käyttäjät voivat perustaa uusia kanavia, mikäli vastaavanlaista kanavaa ei ole jo olemassa. Perustaja määrittelee kanavan aiheen ja säännöt, ylläpitää kanavaa ja voivat nimittää sille uusia ylläpitäjiä. Ylläpitäjä on vapaaehtoinen henkilö, jonka tehtävänä on seurata, että julkaisut noudattavat kanavan sääntöjä. Ylläpitäjien tehtävä on esimerkiksi poistaa epäasiallista tai sääntöjä rikkovaa sisältöä ja roskapostia. (Reddit Help 2023c)⁵

r/Wallstreetbets⁶ on sijoittamiselle omistettu Reddit kanava, jolla oli joulukuussa 2020 alle kaksi miljoonaa käyttäjää. GameStop -tapauksen herättämän huomion myötä käyttäjämäärä kasvoi nopeasti yli viiteen miljoonaan (Business Insider 28.1.2021). Nykyisin kanavalla on yli 14 miljoonaa käyttäjää. Ensisilmäyksellä r/wallstreetbets ei välttämättä luo kuvaa vakavasta sijoitusfoorumista räväkän kulttuurinsa takia, jossa sijoitusdiskurssi

³ <https://support.reddithelp.com/hc/en-us/articles/7419626610708-How-does-voting-work-on-Reddit->

⁴ <https://support.reddithelp.com/hc/en-us/articles/204511829-What-is-karma->

⁵ <https://support.reddithelp.com/hc/en-us/articles/204533569-What-are-communities-or-subreddits->

⁶ r”kanavan nimi” on Redditin tapa ilmaista kanavan nimi. Merkkaustapa on johdettu verkkosivun URL merkkijonosta. Esimerkiksi www.reddit.com/r/wallstreetbets/

yhdistyy meemikulttuuriin. Kanavalla julkaistaan vakavien sijoittamiseen liittyvien keskusteluiden lisäksi myös sijoittamiseen liittyviä meemejä, ja usein vakavatkin keskustelut ovat meemikulttuurin värittämiä. Keskusteluita voi olla vaikea seurata, koska niissä käytetään termejä kuten ”YOLO” (you only live once) ja ”diamond hands”, joka viittaa siihen, että käsissäsi on timantti, eli hyvä osake.

Kanavan kulttuuriin kuuluu myös vahva itseironia. Käyttäjät voivat esimerkiksi viitata itseensä tyhminä apinoina ”we just dumb apes”, jotka tekevät irrationaalisia ja holtittomia päätöksiä. Kanavalla käytetty kieli on usein tunnepohjaista, spekulatiivista, viljeä, ja sen kulttuuri ihannoit riskinottoa. Esimerkiksi joku saattaa julkaista kuvan tosittesta, josta ilmenee, että hän on juuri sijoittanut suuren summan korkean riskin kohteeseen. Tämänlainen julkaisu voitaisiin julkaista tunnisteella ”YOLO”. Käyttäjiä kannustetaan julkaisemaan sekä suuret voittonsa että suuret tappionsa, ja niiden julkaiseminen tuottaa usein julkaisijalle paljon karmaa. (Quinn 2021.)

Kuitenkin tarkastelemalla r/wallstreetbets kanavaa pintaa syvemältä, voidaan huomata, ettei kyse ole vain meemejä pursuavasta huumorikanavasta. Kanava on hyvin ylläpidetty ja kanavalla on tarkat säännöt sille, millaiset julkaisut ovat hyväksyttäviä. Keskusteluiden tulisi pysyä puhtaasti taloudessa ja esimerkiksi poliittiset julkaisut on kielletty. Kanavan säännöissä painotetaan vahvasti sitä, että käyttäjien ei tulisi koskaan puhua asioista, joista he eivät tiedä. Sijoitusneuvot tulee antaa vastuullisesti ja julkaisuiden tulee olla informatiivisia. Säännöt kieltävät kaikenlaisen harhaanjohtamisen ja markkinamanipulaation. (r/wallstreetbets content guide 2023.)⁷

Kanavan ylläpitäjät ilmoittavat poistavansa kaikki sääntöjä rikkovat julkaisut ja ylläpitäjien lisäksi muut käyttäjät voivat äänestää asiattoman tai tarpeettoman sisällön poistamisesta. Kanavalta löytyy myös suuri määrä linkkejä erilaisiin sijoitusoppaisiin. Meemit ovat yksinkertaisesti osa kanavan kommunikaatiota. Niiden avulla voidaan välittää tunteita ja luoda merkityksiä (Quinn 2021). Seuraavaksi esittelen lyhyesti vuoden 2021 tammikuussa tapahtunutta GameStop -tapausta.

⁷ <https://www.reddit.com/r/wallstreetbets/wiki/contentguide/>

Vuoden 2021 tammikuussa GameStopin osake lähti jyrkkään nousuun. Tapaukseen vaikuttivat piensijoittajien lisäksi niin kutsutut hedge-rahastot, jotka ovat yksityisiä, usein veroparatiiseissa sijaitsevia sijoitusrahastoja. Hedge rahastoilla on perinteisiä rahastoja laajempi keinovalikoima. Perinteisten rahastojen toiminta on tarkoin säänneltyä, eivätkä ne saa sijoittaa kaikkein riskialttiimpiin sijoitustuotteisiin. Hedge-rahastoja eivät tällaiset sääntelyt koske, vaan ne voivat sijoittaa varallisuuttaan vapaasti ja ottaa suurempia riskejä. Hedge-rahastot ovat tarkoitettu suurille pääomille, eikä tavallisilla sijoittajilla ole usein mahdollisuutta sijoittaa niiden kautta. (Eichengreen & Mathieson 1999.)

Yksi hedge-rahastojen keinovalikoimaan kuuluvista sijoitusinstrumenteista on niin kutsuttu lyhyeksi myyminen, joka on keino tehdä rahaa yhtiön osakekurssin laskusta (Eichengreen & Mathieson 1999). Kun sijoittaja, tai tässä tapauksessa sijoitusrahasto olettaa tietyn yhtiön osakekurssin laskevan tulevaisuudessa, voidaan yhtiön osakkeita lainata joltain, joka on valmis lainaamaan niitä pientä korvausta vastaan. Tämän jälkeen lainatut osakkeet myydään. Kun osakkeen hinta on odotusten mukaisesti pudonnut, osakkeet ostetaan takaisin ja palautetaan alkuperäiselle omistajalleen. Näin tehdään voittoa myyntihinnan ja takaisinostohinnan erotuksen verran. (Kumar 2015 s.209–219)

Lyhyeksi myymiseen sisältyy kuitenkin ankara riski. Mikäli lyhyeksi myydyin osakkeen hinta alkaa vastoin odotuksia nousemaan, joutuu lyhyeksi myynyt taho ostamaan osakkeet takaisin kalliimmalla välttääkseen tappioiden kasvamisen. Tämä ajaa osakkeen hintaa edelleen ylöspäin. Mikäli useat toimijat ovat myyneet kyseistä osaketta paljon lyhyeksi, on reaktio voimakas. Jo suhteellisen pieni yllättävä nousu voi laukaista todella jyrkän kurssinousun. (Kumar 2015 s. 196–197.)

Tämä riski toteutui GameStopin tapauksessa. Useat hedge-rahastot, suurimpana niistä Melvin Capital Management olivat myyneet lyhyeksi vaikeuksissa olevan GameStopin osaketta raskaalla kädellä. Osakkeen hinnan odotettiin laskevan muun muassa siksi, että videopelien myynti on siirtynyt pitkälti verkkoon, eikä fyysisiä kopioita enää osteta. Lyhyeksi myytyjen osakkeiden määrä oli poikkeuksellisen suuri. Osakkeita oli lainattu eteenpäin useaan kertaan, ja niitä oli myyty lyhyeksi enemmän, kuin niitä oli markkinoilla (Reuters 18.2.2021). Tämä loi erittäin räjähdysherkän tilanteen markkinoille.

Wallstreetbets kanavan sijoittajat huomasivat hedge-rahastojen strategian ja alkoivat joukolla ostamaan GameStopia ajaakseen sen hintaa ylöspäin. Yksi vaikuttavimmista toimijoista liikkeen takana oli Redditissä nimimerkillä DeepFuckingValue toiminut pörssimeklari ja analyytikko Keith Gill. Hän näki tilaisuuden jo kesällä 2020, ja julkaisi oman arvionsa GameStopin arvosta ja sen laajasta lyhyeksi myymisestä. Hänen mukaansa lyhyeksi myyneet tahot joutuvat vetäytymään ja GameStopin osake nousee ”kuuhun asti” (to the moon), kunhan r/wallstreetbetsin sijoittajat jatkavat osakkeen ostamista. (Quinn 2021.)

Kanavalle muodostui organisoitu sosiaalinen liike, joka pyrki voittojen tekemisen lisäksi osoittamaan massojen voiman⁸ ja rankaisemaan Wall Streetin isoja toimijoita (The Guardian 28.1.2021; Umar 2021). Asetelmaa on verrattu kertomukseen Daavidista ja Goljattista (abc News 13.2.2021). Liike otti nopeasti tuulta alleen ja GameStopin osakekurssi lähti räjähtävään nousuun. Osakkeen hinta nousi vajaassa kuukaudessa noin 16 dollarista noin 347 dollariin (Umar 2021). Arvioiden mukaan hedge-rahastot menettivät tämän seurauksena noin 12.5 miljardia dollaria (Reuters 18.2.2021). Wallstreetbetsin piensijoittajista jotkut tekivät satumaisia voittoja, toiset taas murskaavia tappioita.

28.1.2021 sijoitusalueelta RobinHood, joka on yksi Yhdysvaltojen käytetyimmistä osakkeiden oston ja myyntiin käytetyistä palveluista, päätti estää GameStopin osakkeen ostaminen alustallaan. Päätös johti osakekurssin jyrkkään laskuun ja sen voidaan katsoa olleen tämän tapauksen loppu. RobinHoodia vastaan nostettiin joukkokanne kaupankäynnin estämisen vuoksi. Myös Keith Gilliä vastaan nostettiin syyte markkinamanipulaatiosta. Kumpikaan oikeustapauksista ei johtanut tuomioon. (Reuters 2.2.2022; Reuters 27.1.2022.)

Tapaus on harvinainen – kenties jopa ainutlaatuinen, sillä sen toteutumiseen taustalla vaikuttivat sosiaalisen liikkeen lisäksi myös edellä mainitut lyhyeksi myymiseen liittyvät mekanismit (Umar ym. 2021). Myös koronapandemialla on voinut olla vaikutusta tapaukseen. Pandemian takia ihmisillä oli paljon vapaa-aikaa ja kenties myös ylimääräistä rahaa,

⁸ "Apes together strong", kuten r/wallstreetbets kanavalla sanotaan. (viittaus Apinoiden planeetta -elokuvaan)

kun palkkoja ei päästy kuluttamaan. Nämä tekijät saattoivat lisätä kiinnostusta sijoitus-toimintaa kohtaan. Joka tapauksessa tämä tapaus on esimerkki siitä, kuinka verkossa käytävä keskustelu voi vaikuttaa osakemarkkinoihin. Tapaus on myös erityisesti innoittanut tutkijoita tutkimaan verkossa käytävää keskustelua sijoittamisesta ja sen vaikutusta pörs-sikursseihin.

Sosiaalinen media on avannut uudenlaisen ikkunan ihmisten ajatusmaailmaan ja sosiaa-liseen toimintaan. Ihmiset jakavat verkossa näkemyksiään ja käyvät keskustelua monen-laisista aiheista. Näitä keskusteluja analysoimalla voidaan tarkastella ihmisten ajatuksia ja näkemyksiä eri aiheista, kuten sijoittamisesta. Tutkimalla sosiaalisessa mediassa käy-tävää keskustelua sijoittamisesta ja taloudesta voidaan kerätä tietoa esimerkiksi ihmisten luottamuksesta tiettyä yritystä kohtaan ja näkemyksistä sen osakkeen hintaan.

Ihmiset voivat vaikuttaa toistensa näkemyksiin ja olla vaikuttamisen kohteena sosiaali- sessa mediassa, joko tietoisesti tai tietämättään. Ajatukset voivat syntyä keskustelufooru-meilla ja levitä sieltä markkinoille, kuten esimerkiksi GameStopin tapauksessa oli käynyt. Sijoittamista koskevien verkkokeskusteluiden lukemisen ja niihin osallistumisen voidaan olettaa edeltävän osto- tai myyntipäätöstä, joten tutkimalla näitä keskusteluja voidaan saada perinteisiä analyysikeinoja aikaisempia signaaleja kurssikehityksestä.

Tämä pro gradu -tutkielma toteutetaan osana Turun yliopiston taloussosiologian laitok-sella käynnissä olevaan *Developing an early signal system to identify shifts in global stock market trends* -tutkimushanketta. Tutkimushankkeen tarkoituksena on kehittää signaali-järjestelmä, joka pystyy tunnistamaan Redditistä keskustelut, jotka voivat mahdollisesti vaikuttaa osakekurssien kehitykseen. Järjestelmä laskee osakkeiden mainintoja kanavalla ja huomioi keskusteluun osallistuvien henkilöiden aseman verkostossa. Tulevaisuudessa järjestelmän on tarkoitus huomioida myös keskustelun konteksti ja sävy. Lopullinen jär-jestelmä tulee keräämään ja analysoimaan dataa lähes reaaliaikaisesti r/wallstreetbets ka-navalta, ja tunnistamaan automaattisesti keskustelussa nousevia trendejä.

Tämä tutkimus tulee etenemään seuraavanlaisesti: ensin esitellään sijoittamista toimintana, finanssimarkkinoiden toimintaperiaatteita ja ihmisten markkinakäyttäytymistä. Tämän jälkeen siirrytään käsittelemään informaation uskottavuuden ja luottamuksen muodostumista sosiaalisen median verkostoissa. Luottamus on keskeisessä roolissa sijoittamiseen liittyvän tiedon etsinnässä, sillä se mitä omaksumme muilta, riippuu vahvasti siitä, kuinka luotettavana pidämme lähdettä. (ks. Chaiken ym. 1980; Wilson 1983 s.13–16; Lin 2016) Teoriaosuuden lopuksi tarkastellaan vielä mielipidevaikuttamista verkostoteoreettisesta näkökulmasta.

Analyysiosiossa analysoidaan r/wallstreetbets -kanavalta tutkimushanketta varten kerättyä dataa. Tätä dataa analysoimalla pyritään selvittämään, kuinka julkaisujen saama huomio jakautuu kanavalla ja vaikuttavatko esimerkiksi julkaisijan karma julkaisun saamien äänten ja kommenttien määrään. Aineistosta pyritään myös löytämään keskeisiä mielipidevaikuttajia kanavalla. Tavoitteena on valottaa sitä, mitkä tekijät vaikuttavat julkaisun näkyvyyteen ja uskottavuuteen. Tutkimuksen tuloksia tullaan hyödyntämään signaalijärjestelmän kehittämisessä.

2 SIJOITUSTOIMINTA JA PÖRSSIKURSSIEN ENNUSTAMINEN

Rahoitusjärjestelmä muodostuu rahoitusmarkkinoista, sekä vaihtoa ja sopimuksia hoitavista instituutioista, kuten pankeista, pörssistä ja finanssivalvonnasta. Rahoitusjärjestelmän tarkoituksena on ohjata pääomaa sieltä missä sitä on, sinne missä sitä tarvitaan – esimerkiksi sijoittajilta yrityksille. Tämä tapahtuu joko suoraan osakemarkkinoiden kautta, tai välillisesti luottolaitoksen kautta. Toisin sanoen yritykset voivat hankkia tarvitsemaansa pääomaa myymällä osakkeita, tai ottamalla lainaa. Tämän tutkimuksen kannalta keskeinen osa rahoitusjärjestelmää ovat osakemarkkinat. Osake on osuus yhtiöstä. Omistamalla yhtiön osakkeita omistaa siis osan yhtiöstä ja on oikeutettu osuuteen sen voitosta, mikäli yhtiön hallitus päättää jakaa voittoja osinkoina. Yhtiön omistajat myös valitsevat yhtiön hallituksen. (Bodie & Merton 2000 s.2, 7, 22–23.)

Sijoittamisessa on kyse taloudellisten kustannusten ja hyötyjen hajauttamisesta ajan yli. Sijoittaja pyrkii arvioimaan, voiko hän saada tulevaisuudessa enemmän rahaa sijoittamalla rahaa tässä hetkessä. Oma varallisuus pyritään valjastamaan tuottamaan lisäarvoa. Rahan aika-arvolla, tarkoitetaan sitä, että jokin summa rahaa nyt on arvokkaampi, kuin sama summa rahaa tulevaisuudessa. Tämä johtuu siitä, että sijoittamalla rahalle voidaan saada korkoa, ja toiseksi rahan arvo laskee jatkuvasti inflaation takia. (Bodie & Merton 2000 s.101–102.)

Ennen sijoituspäätöksen tekemistä sijoittajan on järkevää tutkia potentiaalista sijoituskohdetta sekä yleistä markkinatilannetta, ja pyrkiä valitsemaan kannattavat sijoituskohdet. Tietoa markkinatilanteesta saa esimerkiksi uutisista. Myös yrityksen kurssin kehitys kertoo markkinasta ja tulevaa kehitystä voidaan pyrkiä ennustamaan teknisen analyysin avulla esimerkiksi kurssin trendistä ja vaihdon määrästä (Lim 2015 s.8). Yritysten toimintaa ja taloudellista tilannetta koskevaa tietoa taas saa esimerkiksi yritysten vuosikertomuksista, tilinpäätöksistä ja taseesta. Näitä tulkitsemalla pyritään ennustamaan yrityksen arvoa tulevaisuudessa. Tätä kutsutaan fundamenttianalyysiksi. Yrityksen taloutta ja suorituskykyä voidaan analysoida esimerkiksi myynnin ja pääoman tuottoa, osakkeen hinnan ja voiton välistä suhdetta, sekä velkaantumistasetta kuvaavilla tunnusluvuilla. (Bodie & Merton 2000 s.64, 77–79.)

Tehokkaiden markkinoiden hypoteesin (efficient market hypothesis) mukaan osakkeiden hinnat refleктоivat suoraan markkinoilla kaiken saatavilla olevaa yritystä ja sen tulevaisuutta koskevaa tietoa. Eli sijoittajat muodostavat mielipiteensä rationaalisesti kaiken saatavilla olevan tiedon perusteella, eikä mitään saatavilla olevaa tietoa jätetä käyttämättä tai huomioimatta. Ainoastaan uusi tieto voi vaikuttaa osakkeen hintaan. Esimerkiksi hyvät uutiset yrityksestä voivat nostaa osakkeen hintaa, ja huonot uutiset puolestaan voivat heikentää uskoa yhtiön tulevaisuuteen ja ajaa osakkeen hintaa alaspäin. (Bodie & Merton 2000 s. 206–207.) Lisäksi uusklassinen taloustiede olettaa markkinoiden olevan täydelliset siten, ettei mikään vääristä hintoja, eikä yksittäisillä toimijoilla ole mahdollisuutta vaikuttaa hintoihin (Forbes 2009 s.9–10).

Vaikka on totta, että saatavilla oleva tieto vaikuttaa sijoittajien toimintaan markkinoilla, ovat uusklassiselle taloustieteelle ominaiset oletukset täydellisistä markkinoista ja täysin rationaalisista toimijoista karkeita yksinkertaistuksia, jotka eivät pysty selittämään koko todellisuutta. Ihmiset eivät ole täysin rationaalisia ja informoituja toimijoita, jotka tekevät optimaalisia valintoja. Todellisuus on monimutkainen ja täynnä epävarmuutta, jota ei voi ennustaa matemaattisesti. (ks. Granovetter 1985; Boudon 1998.)

Ihmisten käyttäytymistä ja valintoja markkinoilla ohjaavat esimerkiksi tunteet, intuitio ja erilaiset kognitiiviset harhat (Baker ym. 2014 s.45). Sijoittaminen on sosiaalinen prosessi, joka on altis muoti-ilmioille ja villityksille (Forbes 2009 s.9–11, 133). Myös ajan ja kykyjen puute rajoittavat rationaalista valintaa, vaikka yksilö lähtökohtaisesti haluaisikin tehdä rationaalisia valintoja (Baker ym. 2014 s.31–32). Behavioraalinen taloustiede kuitenkin tunnustaa ihmiselle tyypilliset rajoitukset.

Esimerkkejä irrationaalisista valinnoista sijoitusmarkkinoilla on useita. Esimerkiksi se, kuinka asiat esitetään tai kehystetään, vaikuttaa sijoittajien päätöksiin. Näkökulma, joka korostaa riskejä näyttäytyy usein vähemmän houkuttelevana kuin hyötyjä korostava näkökulma, vaikka todellisuudessa lopputulos olisi sama. Tämä johtuu siitä, että ihmiset pyrkivät välttämään riskejä. Lisäksi ihmiset pyrkivät välttämään tappioita ja ovat usein valmiita uhraamaan voittoja välttääkseen tappioita. Tästä syystä on tyypillistä, että voitokkaat osakkeet myydään liian nopeasti, ja tappiollisia osakkeita pidetään liian pitkään. (Baker ym. 2014 s. 35, 52.)

Ihmiset ovat myös usein turhan itsevarmoja omista tiedoistaan ja taidoistaan. Yli-itsevarmuus onkin herättänyt käyttäytymistaloustieteilijöiden huomion, ja sillä on vaikutusta ihmisten tekemiin sijoituspäätöksiin. Liika itsevarmuus näyttäytyy esimerkiksi liiallisena riskin ottamisena, optimismina ja oman informaation tarkkuuden yliarviointina. Tämä johtaa tarpeettomaan kaupan käymiseen ja huonoihin tuottoihin. (Baker ym. s.36, 53)

Yhtiön markkinahinta osakemarkkinoilla ja yhtiön arvo sen kirjanpidon mukaan poikkeavat usein toisistaan. Osakkeen hinta määräytyy osakemarkkinoilla kysynnän ja tarjonnan mukaan. Osakkeen hinta ei ole suoranaisesti yhtiön todellinen arvo, vaan yhtiöt voivat olla yli- tai aliarvostettuja osakemarkkinoilla, jos sijoittajat eivät ole onnistuneet arvioimaan yhtiön todellista arvoa. Osakkeen hinta kertoo ainoastaan sen, kuinka paljon sijoittajat ovat valmiita maksamaan osakkeesta, tai minkä arvoiseksi sijoittajat arvioivat yhtiön. Yhtiön todellisen arvon määrittäminen on monimutkaista, ja se vaatii esimerkiksi yhtiön pääoman ja velan määrän huomioimisen. Osakkeen hinnanmuodostukseen markkinoilla vaikuttavat esimerkiksi yhtiön maine ja brändi, jotka eivät näy arvona paperilla. Yhtiön hintaan vaikuttaa myös sen kyky tuottaa voittoa. Jos sijoittajat odottavat yhtiön kasvavan ja tuottavan yhä suurempaa voittoa tulevaisuudessa, sen hintaan on ikään kuin ennakoivasti laskettu tulevaisuuden tuottoja. Yhtiön markkinahinta on tällöin yhtiön kirjanpidollista arvoa korkeampi. (Bodie & Merton 2000 s.73–75, 194, 203.)

Osakkeiden hintoihin voivat vaikuttaa myös ulkoiset tekijät, kuten inflaatio ja ihmisten ostovoima, sekä erilaiset markkinatrendit. Hintoihin voi vaikuttaa myös sijoittajien senttimentti. Senttimentti on subjektiivinen psykologinen tekijä, joka kuvaa sijoittajan tunteita markkinatilanteesta. Korkean senttimentin aikaan sijoittajat ovat toiveikkaita tulevaisuuden suhteen ja osakkeiden hinnat nousevat. Matala senttimentti taas näkyy pelkona taantumasta tai romahduksesta, jolloin hinnat laskevat. Senttimenttiin voivat vaikuttaa esimerkiksi uutiset maailman tilanteesta ja politiikasta tai yrityksen tekemästä uudesta innovaatiosta. (Harper 2022.)

Pörssikurssien kehityksen ja sijoittajien sentimentin välistä yhteyttä on alettu tutkia sosiaalisessa mediassa. Esimerkiksi analysoimalla Twitter⁹-keskusteluja sentimenttianalyysin avulla on onnistuttu ennustamaan osakekurssien kehitystä (ks. Pagolu ym. 2016; Huang 2021). Myös käyttäjien välistä luottamusta on mallinnettu ja sen avulla on onnistuttu parantamaan kurssien kehitystä ennustavia malleja (Ruan ym. 2018). Juuri tämänkaltaisen järjestelmän kehittämiseen pyritään myös tässä tutkimusprojektissa.

Ihmisten vaikutukset toisiin ihmisiin ovat erityisen tärkeä tutkimusteema globalisoituneessa maailmassa. Eri markkinat ja toimijat ovat yhteydessä toisiinsa ja esimerkiksi osakkeiden välittäminen sekä maksut tapahtuvat virtuaalisesti. Globaalissa rahoitusjärjestelmässä pääomat voivat virrata vaivattomasti ympäri maailman, ja uutiset yhdessä paikassa tai yhdellä alalla voivat vaikuttaa sijoittajiin myös muualla. (Bodie & Merton 2000 s.22.) Uutiset leviävät ennennäkemättömän nopeasti sosiaalisen median verkostoissa. Sijoittajat ympäri maailman voivat vaihtaa tietoa ja keskustella keskenään erilaisilla sijoituskanavilla vaikuttaen täten jatkuvasti toistensa näkemyksiin markkinoista.

Osakekaupan käyminen on lisääntynyt länsimaissa jatkuvasti, erityisen jyrkästi -90 luvun jälkeen. Yksi syy tähän on se, että kulutusluoton ottaminen on yleistynyt. Ihmiset voivat rahoittaa tarpeitaan, kuten asuntoja, autoja ja jopa elektroniikkaa velalla ja rahaa jää sijoitettavaksi (Forbes 2009 s.119). Lisäksi Teknologia on tehnyt sijoittamisesta todella helppoa. Nykyisin ei tarvitse kauppaa käydessä soittaa pankkiin salkunhoitajalle toimeksiantojen tekemiseksi, vaan osakekauppaa voidaan käydä vaivattomasti puhelimella. Myös kaupankäyntikulut ovat nykyään alhaisia (Quinn 2021.)

Taloustieteen klassinen teoria ei pysty selittämään suurta vaihdon määrää. Jos kaikki sijoittajat ovat informoituja ja rationaalisia, yhtiön hinnan ja arvon tulisi olla tasapainossa. Kauppaa tulisi käydä ainoastaan uuden tiedon valossa, mutta todellisuudessa kauppaa käydään jatkuvasti. Käyttäytymistaloustiede selittää tätä sillä, että on olemassa suuri joukko tietotaidottomia sijoittajia, joiden sijoituspäätökset eivät perustu yhtiön todelliseen arvoon tai uuteen tietoon. Nämä ei-informoidut sijoittajat käyvät kauppaa tunteiden

⁹ Toimii nykyisin nimellä X. Tässä tutkimuksessa kuitenkin puhutaan Twitteristä, kun viitataan tutkimuksiin, jotka on julkaistu ennen nimen vaihtamista.

ja trendien ohjaamana, ja tekevät huonoja hinta-arvioita. Tämä aiheuttaa hintojen heilahtelua markkinoilla. (Forbes 2009 s.119–122)

Jotkut sijoittajat pyrkivät hyötymään kurssien nopeasta heilahtelusta käymällä lyhyen aikavälin kauppaa huippujen ja pohjien välillä (Quinn 2021). Vaikka teknologinen kehitys onkin tehnyt tällaisesta keinottelusta helpompaa ja täten oletetusti yleisempää, kyse ei kuitenkaan ole uudesta ilmiöstä. Esimerkiksi jo Max Weber (1896) puhuu Pörssi -kirjassaan pienkeinottelijasta, joka *”yrittää ansaita pienillä hintaeroilla, tekee pörssistä paikan, jossa hän metsästää omaisuutta, jota hänellä ei ole”*. Tällaista toimintaa kutsutaan nykyisin ”treidaamiseksi” (day trading).

Modernien sijoitusmarkkinoiden on esitetty jopa pelillistyneen. Konkreettisen toiminnan ja rahoitusmarkkinoiden välinen yhteys on muuttunut hyvin abstraktiksi. Sijoittamisesta on tehty helppoa ja perinteisten osakkeiden ja rahastojen lisäksi markkinoilla on lukuisia uusia ja yhä monimutkaisempia sijoitusinstrumentteja. Voitot ja tappiot eivät tunnu todellisilta, ennen kuin sijoitukset realisoidaan, eli myydään. (Quinn 2021.) Yhdysvaltojen eläkejärjestelmä mahdollistaa esimerkiksi kertyvien eläkkeiden sijoittamisen. Kertyneet varat voidaan kuitenkin käyttää vasta eläkkeellä. (Fernando 2023.) Tämä saa voitot ja tappiot tuntumaan entistäkin hypoteettisemmilta.

Myös teknologian kehityksellä voi olla rooli sijoittamisen pelillistymisessä. Kauppaa voidaan käydä helposti puhelimella ja sijoittamiseen tarkoitetut sovellukset tarjoavat värikkäitä kuvaajia ja analyyssejä omasta menestyksestä. Nettifoorumeilla käydään vertailemassa muiden sijoittajien kanssa strategioita ja ”pisteitä”. (Quinn 2021.) Tästä näkökulmasta tarkasteltuna on helppo samaistua r/wallstreetbets -kanavan sijoittajaan vuoden 2021 GameStop tapauksen aikana ja kenties muutenkin modernina aikana.

Tähän mennessä tämä työ on avannut ja taustoittanut tutkimusprojektia ja sen innoitusta, kuvaillut Redditiä foorumina, sekä käsitellyt finanssijärjestelmän perusteita ja ihmisten toimintaa sijoitusmarkkinoilla. Seuraavaksi tarkastellaan informaation uskottavuutta sekä luottamuksen muodostumista sosiaalisessa mediassa liikkuvaan informaatioon. Oletuksena on, että verkossa ja sosiaalisessa mediassa liikkuva informaatio ja keskustelu vaikuttavat sijoittajien sijoituspäätöksiin. Sijoittajan on kuitenkin itse tehtävä päätelmänsä siitä, kuinka luotettavaa informaatio on. Informaation uskottavuus ja sijoittajan luottamus

siihen toimivat siis väliin tulevina tekijöinä sijoituspäätöksessä. Seuraavassa luvussa tarkastellaan niitä mekanismeja, jotka vaikuttavat informaation uskottavuuteen, luottamuksen muodostumiseen ja informaation näkyvyyteen sosiaalisessa mediassa.

3 INFORMAATION ARVIOINTI

Internet on nykyisin tärkeä tiedonhaun kanava ihmisille. Verkossa on saatavilla valtavia määriä informaatiota, ja erilaisilla sosiaalisen median kanavilla sekä keskustelufoorumilla käydään lukemattomia keskusteluita erilaisista aiheista ja jaetaan esimerkiksi uutisia. Meistä jokainen etsii tai kohtaa informaatiota verkossa päivittäin, ja moni meistä myös jakaa informaatiota verkossa säännöllisesti. Esimerkiksi Yhdysvalloissa noin puolet väestöstä lukee uutisensa sosiaalisesta mediasta joskus tai usein (Pew Research Center 2022). Euroopassa vastaava luku on 45 % (Eurobarometer Media & News survey 2022). Sosiaalisen median sisällöstä suuri osa on ihmisten itse tuottamaa, ja käytännössä kuka tahansa voi saada verkossa paljon huomiota asialleen.

Viime vuosina on keskusteltu paljon verkosta saatavan informaation luotettavuudesta. On esitetty huolta misinformaatiosta ja disinformaatiosta, sekä painotettu medialukutaidon tärkeyttä. Vallitsevan tietotulvan keskellä ihmisten on jotenkin pystyttävä erottamaan verkkosisällöstä luotettava informaatio epäluotettavasta. Akateeminen mielenkiinto informaation luotettavuuden muodostumista kohtaan ei kuitenkaan koske ainoastaan nykyistä mediaympäristöä, vaan aiheita on tutkittu jo perinteisen median kontekstissa. Ihmisten luottamusta informaatioon on pyritty selittämään muun muassa sosiodemografisilla tekijöillä, poliittisilla kannoilla ja uskonnollisuudella (Golan 2010). Tarkastellaan ensin informaation luotettavuuden teoriaperinnettä ja perinteistä mediaa koskevaa tutkimusta. Edetään sitten kohti uudempaa tutkimusta ja uutta mediaympäristöä.

Ihmiset arvioivat informaation luotettavuutta joko systemaattisesti tai heuristisesti (Chaiken ym. 1980). Systemaattisessa tarkastelussa informaation vastaanottaja pyrkii rationaalisesti tarkastelemaan ja ymmärtämään viestin sisältöä ja argumentteja, sekä arvioimaan niiden validiteettia. Heuristisessa tulkinnassa taas ei nähdä juurikaan vaivaa viestin arviointiin, vaan arvio tehdään intuitiivisesti esimerkiksi lähteen, tai muun sisällöllisesti merkityksettömän tekijän perusteella. Esimerkiksi auktoriteettiheuristiikka tarkoittaa sitä, että kun informaation vastaanottaja tunnistaa lähteen olevan auktoriteetti, kuten valtion virasto, tietoon luotetaan helposti (Lin 2016). On myös yleistä, että ihmiset arvioivat in-

formaatiota sen perusteella, mitä mieltä he ovat sen toimittajasta ja tämän uskottavuudesta. Myös se, pitääkö viestin vastaanottaja henkilökohtaisesti viestin toimittajasta voi vaikuttaa arvioon informaation uskottavuudesta. (Chaiken ym. 1980.)

Kokemus tiedon lähteen luotettavuudesta voi vaihdella myös kontekstin mukaan. Eri asioissa luotetaan eri medioihin. Esimerkiksi paikallisuutisia saatetaan pitää luotettavampana, kun ne luetaan paikallislehdestä ja kansallisia uutisia puolestaan pidetään luotettavampana televisiosta katsottuna. (Gaziano & McGrath 1986.) Vastaava mekanismi esiintyy myös esimerkiksi eri alojen välillä. Johonkin tiettyyn alaan erikoistunutta lähdettä, ja siitä saatavaa informaatiota, pidetään automaattisesti luotettavampana omalla alallaan, kuin yleistä lähdettä (Wilson 1983 s.168–169).

Systemaattisen ja heuristisen arvioinnin käyttäminen riippuu jossain määrin tilanteesta. Arviointiin vaikuttavat esimerkiksi tarjotun tiedon määrä ja aiheen tärkeys kuulijalle. Jos aihe on tärkeä ja jos tietoa on saatavilla paljon, systemaattisen arvioinnin todennäköisyys kasvaa. Jos taas tietoa on niukasti, tai jos aihe ei ole kuulijalle tärkeä, kasvaa heuristisen tulkinnan todennäköisyys. (Chaiken ym. 1980.)

Tässä tutkimuksessa keskitytään pääasiassa heuristiseen arviointiin ja sen mekanismeihin. Sosiaalisen median julkaisuiden sisällöllä on varmasti merkittävä vaikutus niiden uskottavuuden arvioinnissa, joten sisällönanalyysi tai sentimenttianalyysi sijoitusfoorumien keskustelulle olisi varmasti hyvä työkalu pörssikurssien kehityksen ennustamiselle. Heuristiikkojen vaikutus julkaisuiden uskottavuudelle on kuitenkin vähemmän tutkittu, ja vähemmän ilmiselvä aihe, johon tässä tutkimuksessa halutaan pureutua. Tarkoituksena on selvittää muiden tekijöiden, kuin julkaisun sisällön vaikutusta luottamukseen julkaisua kohtaan.

3.1 Kognitiivisen auktoriteetin teoria

Tiedon arviointia voidaan lähestyä kognitiivisen auktoriteetin teorian kautta. Teorian mukaan on olemassa kahdenlaista informaatiota: oman kokemuksen kautta opittua ja muilta opittua. Kaikki muilta saatava tieto, eli kuulopuhe, ei kuitenkaan ole yhtä uskottavaa tai luotettavaa. Eri ihmiset tietävät eri asioita, ja jokaisen on pääteltävä, kuka on luotettava lähde, eli kenellä on ”auktoriteetti” minkäkin asian suhteen. Luottamus on tiedon vastaanottajan subjektiivinen kokemus, joka muodostuu lähteen koetusta asiantuntemuksesta, uskottavuudesta ja hyväntahtoisuudesta. (Wilson 1983 s.13–16.)

Näiden ominaisuuksien perusteella lähteen uskottavuus tai luotettavuus on kuulijan näkemys siitä, tietääkö lähde mistä puhuu, kertooko hän totuuden ja ajaako hän kuulijan etua (Lin ym. 2016). Kognitiivinen auktoriteetti selittää informaation vaikutusta ihmisten ajatuksiin. Auktoriteettina pidetyn henkilön sanomisilla on vaikutusvaltaa ihmisiin, koska ihmiset pitävät sitä uskomisen arvoisena.

Ihmiset päättävät kuinka korkeassa arvossa tai auktoriteettiasemassa he pitävät lähdetä erilaisten auktoriteettisignaalin tai ”vihjeiden” perusteella. Kognitiivinen auktoriteetti voidaan luokitella neljään luokkaan auktoriteettisignaalin perusteella. Ensimmäinen on henkilökohtainen auktoriteetti, joka perustuu informaation vastaanottajan kokemukseen sen julkaisijan luotettavuudesta. Esimerkiksi luottamus tekstin kirjoittajaa kohtaan nostaa luottamusta myös julkaisun sisältöön. Toinen luokka on institutionaalinen auktoriteetti, joka syntyy luottamuksesta tiedon julkaisijaan, kuten kustannusyhtiöön, lehteen tai muuhun kyseessä olevaan mediaan. Kolmas luokka liittyy tekstin tyyppiin. Esimerkiksi tietokirjoilla on itsessään jonkinasteinen auktoriteetti. Neljäs luokka on lukijan kokemus tekstin varsinaisen sisällön uskottavuudesta. (Wilson 1983 s.166–169.) Tämänlaista jakoa on käytetty useissa tutkimuksissa myöhemmin sekä perinteisen, että sosiaalisen median kontekstissa (ks. Golan 2010, Sterret ym. 2019, Fritch & Cromwell 2001).

Perinteisessä mediassa informaation luotettavuutta voidaan arvioida esimerkiksi uutistoinniston maineen avulla. Myös digitaalisen median kohdalla tämän voidaan olettaa toimivan, ainakin jos kyse on uutisista. Luotettavan lähteen lisäksi informaation tulee itsessään olla tarkkaa, objektiivista, ajantasaista ja kattavaa (Fritch & Cromwell 2001). Verkossa

liikkuvan informaation ja sosiaalisen median tapauksessa luotettavuuden arviointi muuttuu monimutkaisemmaksi. Wilsonin teoria perustuu perinteisen median toimintatapaan, mutta internet on luonut uudenlaisen toimintaympäristön informaation julkaisemiselle.

Verkossa on esimerkiksi tyypillistä, että kirjoittajan identiteetti ei ole tiedossa lainkaan. Toisaalta jos identiteetti on tiedossa, myös kirjoittajan kompetenssin selvittäminen on helpompaa kuin perinteisesti. Verkosta hakemalla voidaan esimerkiksi taustoittaa kirjoittajan koulutusta ja kokemusta. Verkossa liikkuvan informaation validiteettia tarkasteltaessa on syytä huomioida myös informaation ajantasaisuus ja lähteet. Myös julkaisun tuotanteen organisaation uskottavuutta voidaan arvioida sen verkkosivuja ja kaupallisia yhteyksiä tarkastelemalla. (Fritch & Cromwell 2001.)

3.2 Informaation uskottavuus sosiaalisessa mediassa

Toisin kuin perinteisessä mediassa, verkossa ei välttämättä ole minkäänlaista laatutakuuta. Sosiaalinen media on kaikille avoin ja kynnyksensä sisällön tuottamiseen on matala, eikä julkaisuiden portinvartijana toimi esimerkiksi editoria niin kuin perinteisessä mediassa. Tämän takia verkossa jaetaan myös paljon virheellistä informaatiota – tarkoituksella ja tahattomasti. Kuka tahansa voi jakaa kutakuinkin minkäläistä sisältöä tahansa ja perinteiset auktoriteettisignaalit, eli vihjeet tiedon luotettavuudesta saattavat puuttua kokonaan. Sosiaalisessa mediassa kirjoittaja saattaa olla hyvin tuntematon, tai toimia nimimerkin takaa. Lisäksi tarjolla oleva tietomäärä on niin valtava, että sen käsitteleminen systemaattisesti on mahdotonta, jolloin informaatiota arvioidaan luottamalla heuristiikkoihin. (Lin ym. 2016.) Loppupeleissä vastuu informaation arvioinnista on pitkälti kuluttajalla.

Informaation luotettavuuden arviointi sosiaalisessa mediassa on tutkimusaiheena melko tuore. Sosiaalinen media on ollut jatkuvassa muutoksessa yleistymisestään asti, ja eri kanavat toimivat eri tavalla. Kenties näistä syistä selkeää konsensusta ei olla saavutettu, vaan tutkimusten tulokset vaihtelevat. Kuitenkin tulosten välillä on selkeitä yhtäläisyyksiä. Siinä missä yksi tutkimus ei ole antanut tukea jollekin hypoteesilleen, usein jokin toinen tutkimus on.

Sosiaalisen median kohdalla vaikuttavat hyvin pitkälti samat mekanismit ja heuristiikat kuin perinteisessä mediassa. Esimerkiksi luottamus tiedon jakajaan ja julkaisun sisällön johdonmukaisuus vaikuttavat kokemukseen tiedon luotettavuudesta myös sosiaalisessa mediassa. (Fritch & Cromwell 2001, Rieh 2002, Lin 2016.) Seuraavaksi tarkastelemme keskeisimpiä teorioita ja tutkimustuloksia tiedon luotettavuuden arvioinnista ja siihen vaikuttavista tekijöistä sosiaalisessa mediassa.

Yksi tiedon arviointiin vaikuttavista mekanismeista tai auktoriteettisignaaleista on identiteettivihje, eli se kuka tietoa jakaa. Esimerkiksi Twitterissä Twiitin lähettäjän henkilökohtainen uskottavuus vaikuttaa siihen, kuinka uskottavana Twiitin sisältöä pidetään. Tämä ei kuitenkaan tarkoita sitä, että ainoastaan asiantuntijoita pidettäisiin luotettavina, vaikka auktoriteettiheuristiikan mukaan näin voisi päätellä. Ihmiset luottavat uutiseen enemmän myös silloin, kun sen jakaa sosiaalisessa mediassa julkisuuden henkilö, jota he pitävät luotettavana. (Sterret ym. 2019) Myös luotettavina pidettyjen kavereiden jakamia uutisia pidetään luotettavampina, kuin tuntemattomien jakamia uutisia (Turcotte ym. 2015). Jo se, että lukija tunnistaa tiedon jakajan voi lisätä luottamusta tietoon (Metzger ym. 2010). Ihmisten luottamus sosiaalisessa mediassa jaettuihin uutisiin on siis yhteydessä siihen, kuka uutisen jakaa.

Myös muiden ihmisten mielipide julkaisusta voi vaikuttaa sen luotettavuuden arviointiin. Tätä kutsutaan kannatusheuristiikaksi (endorsement heuristic). Ihmiset luottavat lähteeseen suuremmalla todennäköisyydellä, jos muutkin luottavat siihen. Jos muut taas eivät luota lähteeseen, on ohjaava vaikutus vielä vahvempi. (Metzger & Flanagin 2013; Metzger ym. 2010.) Sosiaalisen median vuorovaikutteisessa ympäristössä nämä mekanismit tulevat uudella tavalla näkyviksi.

Sosiaalisessa mediassa jaot, tykkäämiset ja kommentit voivat laukaista kannatusheuristiikan ja vaikuttaa julkaisun luotettavuuden arviointiin (Xiao & Borah 2018). Tutkimustulokset tästä ovat kuitenkin jossain määrin ristiriitaisia. Esimerkiksi Facebookia koskevassa tutkimuksessaan Luo ym. (2022) osoittivat yhteyden julkaisuiden saaman tykkäysmäärän ja koetun luotettavuuden välillä. Toisaalta Lin ym. (2016) Twitteriä koskevassa tutkimuksessa uudelleentwiittausten määrä ei lisännyt merkittävästi julkaisun koettua luotettavuutta. On kuitenkin syytä huomata, että eri heuristiikat voivat toimia eri tavoin

eri konteksteissa. Esimerkiksi vertaisten mielipiteillä voi olla suurempi vaikutus ostopäätöksiä tehdessä, mutta auktoriteetilla voi olla suurempi vaikutus, kun kyse on terveyteen liittyvästä tiedosta. Lisäksi eri mekanismit voivat kumuloitua ja vaikuttaa luottamukseen yhdessä (Lin ym. 2016).

Myös viestin esitystapa tai kehystäminen vaikuttaa ihmisten kokemuksiin sen luotettavuudesta. Positiivisessa viitekehyksessä esitettyjä väitteitä uskotaan helpommin kuin negatiivisessa viitekehyksessä esitettyjä. Ihmisillä on taipumus luottaa esimerkiksi hyötyjä korostavaan näkökulmaan helpommin, kuin haittoja korostavaan näkökulmaan. Negatiivisia asioita ei haluta uskoa, koska ne koetaan uhkana vapaudelle. (Xiao ja Borah 2018.)

Redditin tapauksessa informaation luotettavuuden arvioinnin kannalta merkityksellistä lienee erityisesti julkaisun sisältö. Esimerkiksi kommenttien sentimentti ja relevanssi ovat yhteydessä kommentin saamaan äänimäärään, jota voidaan pitää luottamuksen mittana. Uutta tietoa tuovat kommentit keräävät usein paljon ääniä. Sen sijaan viestin sentimentin ja tunteellisuuden yhteys viestin saamiin ääniin vaihtelee. Joillain kanavilla arvostetaan enemmän tunnepitoista sisältöä, kun taas toisilla kanavilla arvostetaan objektiivista ja analyttistä otetta. (Horne ym. 2017.) Redditiä tutkiessa on aina syytä huomioida kanavien väliset eroavaisuudet. Yleistysten tekemistä eri kanavien välillä on syytä harkita tarkasti.

Redditin julkaisuiden ja kommenttien äänestysjärjestelmä on tärkeässä roolissa luottamuksen vuorovaikutteisessa muodostumisessa. Julkaisun saama äänimäärä viestii muiden näkemyksestä julkaisua kohtaan ja positiivisia ääniä voidaan pitää hyväksynnän merkinä. Paljon positiivisia ääniä saaneet julkaisut nousevat syötteessä ylöspäin ja saavat tätä myötä myös lisää näkyvyyttä (Reddit Help 2023a)¹⁰. Äänestäminen antaa käyttäjille mahdollisuuden nostaa näkyville kiinnostavimmat ja relevanteimmat julkaisut ja kommentit. Näin luottamus rakentuu vuorovaikutuksen kautta.

Myös kannatusheuristiikan voidaan olettaa toimivan Redditissä vastaavasti kuin muillakin sosiaalisen median alustoilla, eli julkaisun saamien kommenttien ja äänten määrä voi vaikuttaa julkaisun uskottavuuteen. (Horne ym. 2017.) Kognitiivisen auktoriteetin teorian

¹⁰ <https://support.reddithelp.com/hc/en-us/articles/7419626610708-How-does-voting-work-on-Reddit->

valossa voidaan argumentoida äänten ja kommenttien määrän toimivan auktoriteettisignaalinä ja viestivän muille käyttäjille julkaisun luotettavuudesta. Näin tarkasteltuna kognitiivisen auktoriteetin teorian auktoriteettisignaalit ovat hyvin lähellä kannatusheuristiikkaa.

Tekstin kirjoittajan uskottavuuden vaikutus Redditissä on monimutkaisempi kysymys. Reddit on puolianonyymi alusta, joten lähteen identiteetti ja sen mukanaan tuoma uskottavuus ei ole yksiselitteistä. Intuitiivisesti voisi olettaa lähteen identiteetin vaikutuksen uskottavuuteen olevan tällaisessa tapauksessa vähäinen. Toisaalta jotkut käyttäjät voivat olla hyvinkin tunnettuja kanavalla. Esimerkiksi GameStopin tapauksessa nimimerkillä DeepFuckingValue toiminut käyttäjä, joka oli keskeisessä roolissa liikkeen agitoimisessa, tuli tapauksen myötä hyvinkin tunnetuksi alustalla (Hartzell ym. 2021).

Kognitiivisen auktoriteetin teoriaa soveltaen voidaan päätellä, että äänten ja kommenttien lisäksi käyttäjien karma -pisteet, flairit (flair), ja palkinnot (trophy) voivat toimia auktoriteettisignaaleina. Flairit ovat eräänlaisia kanavakohtaisia tunnisteita julkaisun sisällölle. Ne näkyvät lukijalle julkaisun otsikon vieressä ja ne ovat usein kategorioita, kuten ”keskustelu” tai ”uutiset”. Flairien tarkoitus on antaa lukijalle käsitys siitä, millaisesta julkaisusta on kyse (Reddit Help 2024)¹¹. Trophyt ovat Redditiin käyttäjilleen myöntämiä palkintoja, jotka näkyvät käyttäjän profiilissa (Reddit Wiki 2022)¹². Karman ja trophyjen voidaan ajatella viestivän julkaisijan henkilökohtaisesta auktoriteetista (ks. Wilson 1983 s.166–169).

Flaireilla on vaikutus julkaisun saamien äänten ja kommenttien määrään ainakin joillain kanavilla. Erityisesti julkaisut, jotka on merkattu auktoriteettia ilmentävällä flairilla, kuten ”expert”, saavat paljon ääniä. Tämä voi tietysti johtua viestin laadukkaasta sisällöstä ja luultavasti selittyikin suurelta osin sillä. (Horne ym. 2017.) Kuitenkin asiantuntijaläheteitä pidetään lähtökohtaisesti luotettavampana, joten kyseinen flair voi jo itsessään herättää luottamuksen ja johtaa suureen äänimäärään (Wilson 1983 s.168–169). Flairit voivat siis toimia auktoriteettisignaaleina.

¹¹ <https://support.reddithelp.com/hc/en-us/articles/15484503095060-User-Flair>

¹² <https://www.reddit.com/wiki/trophies/>

Myös julkaisemisen ajoitus on yhteydessä julkaisun saamaan äänimäärään Redditissä. Julkaisuun ensimmäisenä tulevat kommentit saavat ensimmäisenä ääniä ja näkyvät tämän takia myös jatkossa ylempänä syötteellä. Tähänkin liittyy kuitenkin kanavien välisiä eroavaisuuksia. Joillain kanavilla julkaisutahti voi olla niin hidas, että kommentit saavat ääniä useissa ryppäissä. Myös kellonaika, jolloin julkaisu on tehty voi vaikuttaa äänien jakautumiseen ja koko julkaisun näkyvyyteen. (Horne ym. 2017.) Julkaisun kellonaika voisi olla mielenkiintoinen muuttuja pörssikurssien ennustamisessa, sillä pörssit sulkeutuvat viideltä. On syytä miettiä vaikuttaako julkaisun ajankohta ostopäätöksiin.

On syytä muistaa, että Reddit -kanavien välillä voi olla suuriakin eroja siinä, mikä tekee julkaisusta uskottavan. Uskottavuuteen voivat vaikuttaa esimerkiksi kanavan ylläpitoperiaatteet, kanavalle ominainen kulttuuri ja käyttäjien määrä (Hartzell ym. 2021). Esimerkiksi ylläpidon taso voi vaikuttaa kanavan uskottavuuteen. Hyvin ylläpidetyllä kanavalla keskustelu on kontrolloitua. Huonosti ylläpidetyn kanavan syöte taas saattaa täyttyä ros-kapostista ja asiattomista kommenteista.

Myös kanavan kulttuuri voi vaikuttaa julkaisuiden uskottavuuteen. Alakanavien kulttuuriin ja kieleen kuuluvat usein sisäpiirin vitsit ja meemit, joilla on oma historiansa ja jotka noudattavat omanlaistaan logiikkaa. Sisäpiirin meemit luovat yhteenkuuluvuuden tunnetta ja kollektiivista identiteettiä (Quinn 2021). Lukijan on tunnettava kanavalle ominaista kieltä pystyäkseen seuraamaan keskustelua ja luomaan kanavalle uskottavaa sisältöä. Toisaalta kanavalle ominaisen kielen tai kulttuurin ilmentäminen ei takaa julkaisulle suurta äänimäärää (Horne ym. 2017). Tämä voi johtua esimerkiksi siitä, että kanavan kulttuurille ominainen sisältö on usein meemejä, jotka eivät sinällään tarjoa sisällöllisesti mitään uutta.

Ylipäätään on syytä huomioda, että tekijät, jotka ennustavat suurta äänimäärää yhdellä alustalla tai kanavalla, eivät välttämättä ennusta sitä toisella (Horne ym. 2017). Tästä syystä kanavien välisten yleistysten tekeminen on hankalaa ja jokaista kanavaa tulee tutkia erikseen. Sosiaalisen median tutkimuskenttä on vielä melko nuori. Tutkimustulokset poikkeavat jossain määrin toisistaan, eikä absoluuttisia totuuksia tai oikeita lähestymistapoja ole löydetty. Informaation luotettavuuden arvioinnin subjektiivisen luonteen takia absoluuttisten totuuksien etsiminen ei liene mielekäästä lainkaan. Arviointiin vaikuttavat

monet yksilökohtaiset tekijät, ja tutkimuksella voidaan korkeintaan arvioida sitä, mitä vaikuttavia tekijöitä on, ja kuinka suuri vaikutus niillä kutakuinkin on.

3.3 Mieli-pidevaikuttajuus sosiaalisen median verkostoissa

Tässä tutkimuksessa ollaan ennen kaikkea kiinnostuneita siitä, kuinka sosiaalisessa mediassa liikkuva informaatio vaikuttaa ihmisten mielipiteisiin. Tätä kysymystä voidaan lähestyä informaation uskottavuuden ja luotettavuuden arvioinnin lisäksi myös verkostoteoreettisesta näkökulmasta. Verkostoteoria korostaa sosiaalisten siteiden ja niiden välisten suhteiden merkitystä ihmisten välisen toiminnan tutkimisessa. Verkostoteoria ei ole niinkään kiinnostunut toimijoiden yksilöllisistä ominaisuuksista, vaan toimijoiden välisistä suhteista. Toimijat ovat riippuvaisia toisistaan, ja verkosto voi vaikuttaa toimijoihin suoraan tai epäsuorasti. Toimijat vaikuttavat esimerkiksi toistensa mielipiteisiin vaihtamalla informaatiota keskenään. (Ackland 2013 s. 55–56, 104.)

Toimijoiden välistä sosiaalisten suhteiden verkostoa voidaan ajatella sosiaalisena järjestelmänä, jossa asenteet, tieto ja käyttäytyminen rakentuvat. Verkosto mahdollistaa, ohjaa, tai rajoittaa informaation kulkemista ja sen vaikutusta toimijoihin verkostossa. Informaation ja sen lähteiden uskottavuuden lisäksi verkostolla on keskeinen rooli ihmisten mielipiteiden ja käytöksen muodostumisessa. Ihmiset omaksuvat näkemyksiä, asenteita ja tapoja omalta verkostoltaan kommunikaation kautta (Konstantinou ym. 2021). Toisiinsa yhteydessä olevat henkilöt jopa arvioivat riskejä samalla tavalla, joka on tämän tutkimuksen kontekstissa erityisen mielenkiintoista, sillä riskien arviointi on keskeinen osa sijoittamista (Scherer & Cho 2003).

Informaation kulkemiseen mediassa vaikuttavat monet, esimerkiksi kanavasta riippuvat tekijät, ja sitä on pyritty selittämään erilaisilla malleilla (Thorson&Wells 2016). Informaation on esimerkiksi esitetty virtaavan verkostossa kahden askeleen kautta (two step flow of communication). Tämän teorian mukaan mielipidevaikuttajat ensin keräävät informaatiota eri kanavista, ja sitten levittävät sitä omalle verkostolleen, joka reagoi informaatioon. Mieli-pidevaikuttajien julkaisut saavat paljon näkyvyyttä ja niitä myös jaetaan usein eteenpäin. (Choi 2015.) Informaatio siis virtaa ja valikoituu mielipidevaikuttajien

kautta. Tämä malli pätee nykyisin erityisesti sosiaalisessa mediassa (Thorson&Wells 2016).

Verkoston rakenne voi tarjota kiinnostavaa tietoa informaation kulkemisesta sosiaalisessa mediassa. Verkoston rakennetta analysoimalla voidaan tehdä erilaisia päätelmiä informaation virtaamisesta ja keskusteluiden kehittymisestä kanavalla. Tiiviissä verkostoissa tieto kulkee hyvin, ja toisaalta rakenteesta voi löytyä myös aukkoja, jotka hidastavat tiedon kulkua tai katkaisevat sen kokonaan (Mason, Conrey & Smith 2007). Verkostoanalyysin avulla voidaan vertailla erilaisia sosiaalisia verkostoja määrällisesti (Ackland 2013 s.58).

Verkoston rakennetta tarkastelemalla voidaan havaita myös erilaisia rooleja kanavalla. Jotkut käyttäjät voivat esimerkiksi osallistua kanavalla lähinnä vastaamalla muiden käyttäjien kysymyksiin, kun taas toiset käyttäjät käyvät vastavuoroisempaa keskustelua (Buntain & Goldbeck 2014). Lisäksi verkostossa mahdollisesti olevien klusterien välillä voi olla laadullisia eroja esimerkiksi mielipiteissä tai keskustelunaiheissa. Klustereiden välisten laadullisten erojen tarkastelu vaatisi kuitenkin sisällönanalyysia, joka on rajattu tämän tutkimuksen ulkopuolelle. Verkostoanalyysi voi kuitenkin antaa viitteitä sen tarpeellisuudesta.

Vaikutusvalta sosiaalisessa mediassa perustuu suurilta osin näkyvyyteen eli saadun huomion määrään. Vaikutusvallan osoittaminen sosiaalisen median ja todellisen maailman välillä on kuitenkin vaikeaa. Emme pysty mittaamaan sitä, kuinka paljon suoraa vaikutusta yhden käyttäjän viesteillä on toisen käyttäjän mielipiteeseen. Näkyvyyttä alustalla voidaan kuitenkin pitää mittarina vaikutuksen suuruusluokasta. Suurempi yleisö tuo mukanaan suuremman vaikutusvallan. (Ackland 2013 s.106–107.) On myös syytä huomata, että sen lisäksi että ideat leviävät sosiaalisessa mediassa ja vaikuttavat markkinoihin, sosiaalisen median voidaan myös ajatella olevan ikkuna sijoittajien mielipiteisiin, josta voidaan tutkia yleistä ilmapiiriä markkinoilla. Täten keskustelun analysointi on jo itsessään mielekästä.

Verkossa eri sivustojen käyttäjämäärät jakautuvat niin kutsutun potenssilain (power law) mukaisesti. Tämä tarkoittaa sitä, että vain pieni määrä verkkosivuista saavuttaa suuria käyttäjämääriä, kun taas suurimmalla osalla verkkosivuista on vain vähän käyttäjiä.

(Holmberg 2009 s.27–28; Mahanti ym. 2013.) Myös verkossa käytävään keskusteluun osallistuminen noudattaa vastaavaa jakaumaa. Usein hyvin pieni osa kanavan käyttäjistä julkaisee suurimman osan kanavan julkaisuista (Carron-Arthur 2014; Kilgo 2016). Lisäksi julkaisuiden saama huomio jakautuu samalla tavalla epätasaisesti. Pieni joukko aktiivisia käyttäjiä kerää suurimman määrän huomiosta, joka näkyy esimerkiksi vastausten määränä. Monet käyttäjistä vain seuraavat keskustelua, eivätkä julkaise tai kommentoi lainkaan, ja useimpien käyttäjien julkaisut saavat vain vähän huomiota. (Choi 2015.)

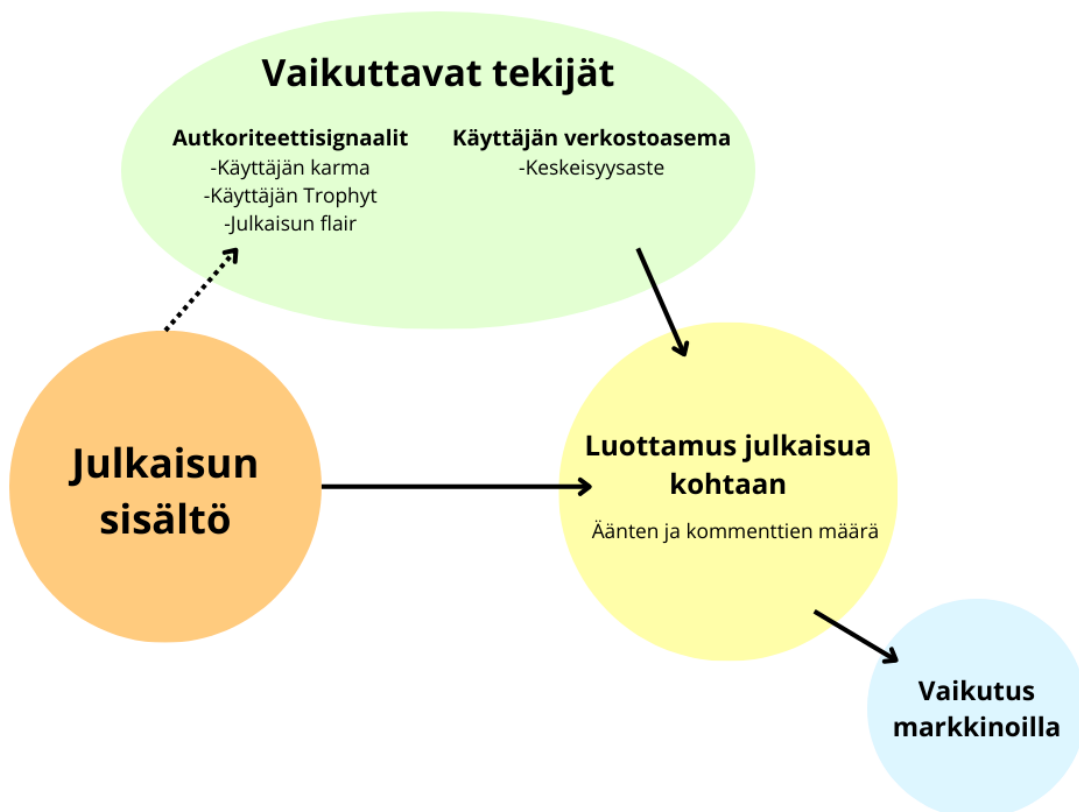
Käyttäjät, jotka ovat muita huomattavasti aktiivisempia, julkaisevat enemmän ja saavat enemmän huomiota, voidaan määritellä mielipidevaikuttajiksi (Choi 2015, Kilgo ym. 2016). Mielipidevaikuttajilla on tiivis verkosto ja he ovat keskeisessä asemassa verkostossaan, joten heillä on mahdollisuus vaikuttaa suureen joukkoon käyttäjiä nopeasti (Rohde ym. 2023). Redditissä julkaisun saamien kommenttien määrä on hyvä tapa mitata mielipidevaikuttajuutta, ja myös korkea karma, sekä äänen määrä voivat viestiä mielipidevaikuttaja-asemasta (Kilgo 2016).

Sosiaalisen vaikuttamisen näkökulmasta tarkasteltuna voidaan olettaa mielipidevaikuttajien olevan hyvin keskeisessä roolissa ohjaamassa ihmisten mielipiteitä verkostossa. Kanavalla aktiivisesti toimivat mielipidevaikuttajat esimerkiksi määrittelevät hyvin pitkälti sen, mistä aiheista kanavalla keskustellaan ja mikä on keskustelun näkökulma. Mielipidevaikuttajat ovat oletettavasti myös toimineet kanavalla pitkään ja täten tuntevat kanavan kulttuurin. Tämä antaa valmiuksia tuottaa kanavalle sisältöä, jota siellä arvostetaan, johtaen aseman vahvistumiseen entisestään.

Tässä tutkimuksessa pyritään paikantamaan mielipidevaikuttajia r/wallstreetbets -kanavalta luokittelemalla käyttäjät aktiivisuuden ja käyttäjien saaman huomion mukaan. Aktiivisuuden mittareina käytetään käyttäjän kirjoittamia julkaisuja sekä kommentteja. Käyttäjän saamaa huomiota puolestaan mitataan käyttäjän julkaisuiden saamalla äänillä ja kommentteilla. Tarkemmat määrittelyt ja rajaukset esitellään seuraavaksi menetelmäosiossa.

4 TUTKIMUSASETELMA

Tässä tutkimuksessa tutkitaan informaation luotettavuuteen vaikuttavia mekanismeja sosiaalisen median verkostoissa. Tutkimuksen tavoitteena on selvittää aktiivisuuden ja huomion jakautumista r/wallstreetbets kanavalla, sekä julkaisuiden luotettavuuteen vaikuttavia mekanismeja. Luottamusta julkaisua kohtaan mitataan tässä tutkimuksessa julkaisuiden saamalla ääni- ja kommenttimäärillä. Luottamuksen lisäksi nämä kertovat julkaisun näkyvyydestä. Toiseksi tutkimuksen tavoite on paikantaa kanavalta mielipidevaikuttajia. Mielipidevaikuttajat paikannetaan etsimällä verkostosta keskeisiä ja aktiivisia käyttäjiä, joiden julkaisut saavat paljon huomiota. Lisäksi mielipidevaikuttajien asemaa tarkastellaan verkostanalyysin avulla.



Kuva 1. Julkaisun luotettavuuden vaikutussuhteet

Julkaisun luotettavuuden vaikutussuhteita ja luottamusta väliin tulevana tekijänä ennen markkinoita on havainnollistettu kuvassa 1. Julkaisun sisältö on luultavasti keskeinen julkaisua kohtaan koettua luottamusta määrittelevä tekijä. Hyvin argumentoitu, informatiivinen sisältö herättää enemmän luottamusta kuin huonosti argumentoitu sisältö. Muita vaikuttavia tekijöitä ovat muun muassa heuristiseen päättelyyn perustuvat auktoriteettisignaalit, sekä käyttäjän verkostoasema, joiden vaikutuksesta julkaisua kohtaan koettuun luottamukseen tässä tutkimuksessa ollaan kiinnostuneita. On syytä myös huomata, että julkaisun sisältö voi osittain vaikuttaa myös vaikuttaviin tekijöihin. Julkaisija esimerkiksi valitsee flairin julkaisun sisällön perusteella ja käyttäjän karma kertyy käyttäjän julkaisuiden saadessa ääniä. Kausaliteettisuhteet ovat siis monimutkaiset.

4.1 Tutkimuskysymykset

Tällä tutkimuksella on kaksi päätavoitetta. Ensinnäkin selvittää mahdollisten auktoriteettisignaalien vaikutusta julkaisuiden uskottavuuteen kanavalla. Potentiaalisina auktoriteettisignaaleina pidetään käyttäjän karma -pisteitä, käyttäjien saamia trophyja ja julkaisun flairia, sillä nämä voivat viestiä julkaisijan kokemuksesta ja julkaisun sisällöstä. Toiseksi pyrin tunnistamaan kanavalta mielipidevaikuttajia analysoimalla käyttäjien aktiivisuutta ja saaman huomion määrää, sekä käyttäjien verkostorakennetta. Tämän tutkimuksen tutkimuskysymykset ovat seuraavat:

1. Minkälainen on julkaisuiden, kommenttien ja niiden saaman huomion jakauma kanavalla?
2. Saavatko paljon julkaisevat henkilöt enemmän huomiota, kuin vain vähän julkaisevat?
3. Minkälainen on kanavan käyttäjien välinen verkostorakenne?
4. Ovatko karma, flairit tai trophyt yhteydessä käyttäjän saamaan huomioon?

Tämä tutkimus on merkityksellinen ensisijaisesti *Developing an early signal system to identify shifts in global stock market trends* -tutkimushankkeen kannalta. Tutkimusprojektin tavoite on kehittää signaalijärjestelmä, joka ennakoii pörssikurssien heilahduksia r/wallstreetbets kanavalla käytävän keskustelun perusteella. Tämä tapahtuu käytännössä laskemalla osakkeen mainintamääriä kanavalla, ja mallintamalla pörssikurssin heilahtelua mainintojen, niiden saaman huomion ja niiden kirjoittajan auktoriteettiaseman funktiona. Tuloksia mielipidevaikuttajien verkostoasemasta ja auktoriteettisignaaleista voidaan hyödyntää signaalijärjestelmässä yksittäisten viestien vaikuttavuuden arvioimiseksi, antamalla viesteille verkosto- ja auktoriteettiasemaan perustuvia painokertoimia.

Tutkimuksen aihe on merkityksellinen myös laajemmassa kontekstissa. Sosiaalisessa mediassa liikkuva informaatio on muodostunut yhä tärkeämmäksi lähteeksi ihmisten jokapäiväisessä elämässä. Tästä syystä sosiaalisen median dynamiikkaa, kuten auktoriteetin ja luottamuksen muodostumista, on syytä tutkia. Informaation uskottavuuden ja auktoriteetin muodostuminen verkossa on monimutkainen prosessi, johon vaikuttavat useat tekijät, kuten viestin sisältö, kirjoittajan uskottavuus ja muiden käyttäjien näkemykset informaation uskottavuudesta. Myös kanavien oma kulttuuri ja käytännöt voivat vaikuttaa tulkintoihin uskottavuudesta.

4.2 Aineisto ja menetelmät

Tutkimusaineistona käytetään Redditistä kerättyä käyttäjätietoa, joka on muodostunut käyttäjien välisessä vuorovaikutuksessa. Menetelmällisesti tämä tutkimus kuuluu laskennallisen sosiaalitieteen (computational social science) alaan. Seuraavaksi esittelen lyhyesti laskennallisen sosiaalitieteen alaa, sen oletuksia ja ongelmia. Tämän jälkeen kuvailen aineistoani, sekä esittelen tässä tutkimuksessa käytettävät menetelmät. Lopuksi pohdin vielä tutkimuksen etiikkaa.

Tietotekniikan kehitys ja sosiaalisen median yleistymisen ovat tuoneet erilaiset data-analyysin menetelmät osaksi yhteiskunta- ja käyttäytymistieteitä. Laskennallinen sosiaalitiede on yleistynyt sosiaalitieteiden suuntaus, joka hyödyntää tietojenkäsittelymenetelmiä

ja verkossa syntyvää massadataa. Kyse on poikkitieteellisestä suuntauksesta, joka toimii tietojenkäsittelytieteen ja sosiaalitieteiden rajalla. (Hox 2017.)

Modernit informaatioteknologiat tuottavat valtavia määriä käyttäjätietoa, joka antaa tutkijoille mahdollisuuden tutkia ihmisten toimintaa täysin uudella tavalla ja havainnoida suuria ihmisjoukkoja pitkän ajanjakson yli. Ihmisten toiminta internetissä jättää jälkeensä digitaalisia jälkiä, joita analysoimalla voidaan tehdä päätelmiä ja testata teorioita. Tätä käyttäjätietoa kutsutaan usein nimellä ”big data”. Työlään aineistonkeräysprosessin sijaan, tämä data voidaan yksinkertaisesti ladata. (Hox 2017.)

Riippuen tutkimusasetelmasta, käyttäjätieto voi mahdollistaa jopa kokonaisotannon tutkimuksen kohteista. Käyttäjätiedon avulla voidaan esimerkiksi analysoida käyttäjien toimintaa sosiaalisessa mediassa tilastollisesti ja tehdä verkostanalyseja. Datalähtöisen analysoinnin lisäksi tietotekniikan avulla voidaan myös simuloida erilaisia malleja esimerkiksi teorioiden tai hypoteesien testaamista varten, tai toteuttaa erilaisia kokeellisia tutkimuksia. (Bravo & Farjam 2017.)

Vaikka sosiaalisesta mediasta kerätty data tuottaakin paljon mahdollisuuksia, liittyy siihen myös monenlaisia ongelmia ja epävarmuustekijöitä. Tällainen data ei ole lähtökohdaisesti tuotettu tutkimuskysymyksiä silmällä pitäen varten, vaan se on algoritmien ohjaama, sosiaalisessa mediassa tapahtuvan toiminnan sivutuote. Suurta verkkosyntyistä dataa analysoitaessa haasteita tuottavat usein datan massiivisuus, laajuus, sotkuisuus ja epävarmuus. (Kitchin 2014.) Analyysissa on myös syytä huomioda, että yhdeltä alustalta kerätty data ei kuvasta koko internetissä liikkuvaa tietoa ja sen dynamiikkaa. Yleistyksien tekeminen hankalaa. Yhdellä kanavalla liikkuvaa tietoa ei toisaalta voida myöskään rajata kyseiselle kanavalle, vaan tieto liikkuu kanavien välillä. (Tufekci 2014.)

Data, jota alustan käyttämisestä syntyy, on lähtökohdaisesti alustan teknisten toimintaperiaatteiden määrittelemää ja voi olla vinoutunutta (Lazer ym. 2020). Lisäksi alustan määrittelemät API (Application Programming Interface) oikeudet ja datan keräämisen toimintaperiaatteet vaikuttavat siihen, mitä dataa on saatavilla ja kuinka se on järjestetty (Snee ym. s.37–38). Datan saatavuus voi olla myös ongelma. Suuri osa datasta on yritysten tai valtioiden hallussa, eli tutkijoiden saavuttamattomissa. Alustat myyvät käyttäjätietoa kaupallisille toimijoille, joten suuri osa datasta ei ole ilmaiseksi saatavilla API:n

kautta. API on rajapinta, jonka kautta voidaan hakea alustalta julkista tietoa esimerkiksi alustan käyttäjistä ja julkaisuista (Snee ym. 2016 s.19). Tutkijat ovat myös esittäneet huolta siitä, että tullaanko API oikeuksia rajoittamaan jatkossa yhä enemmän. (Snee ym. s.224–225.)

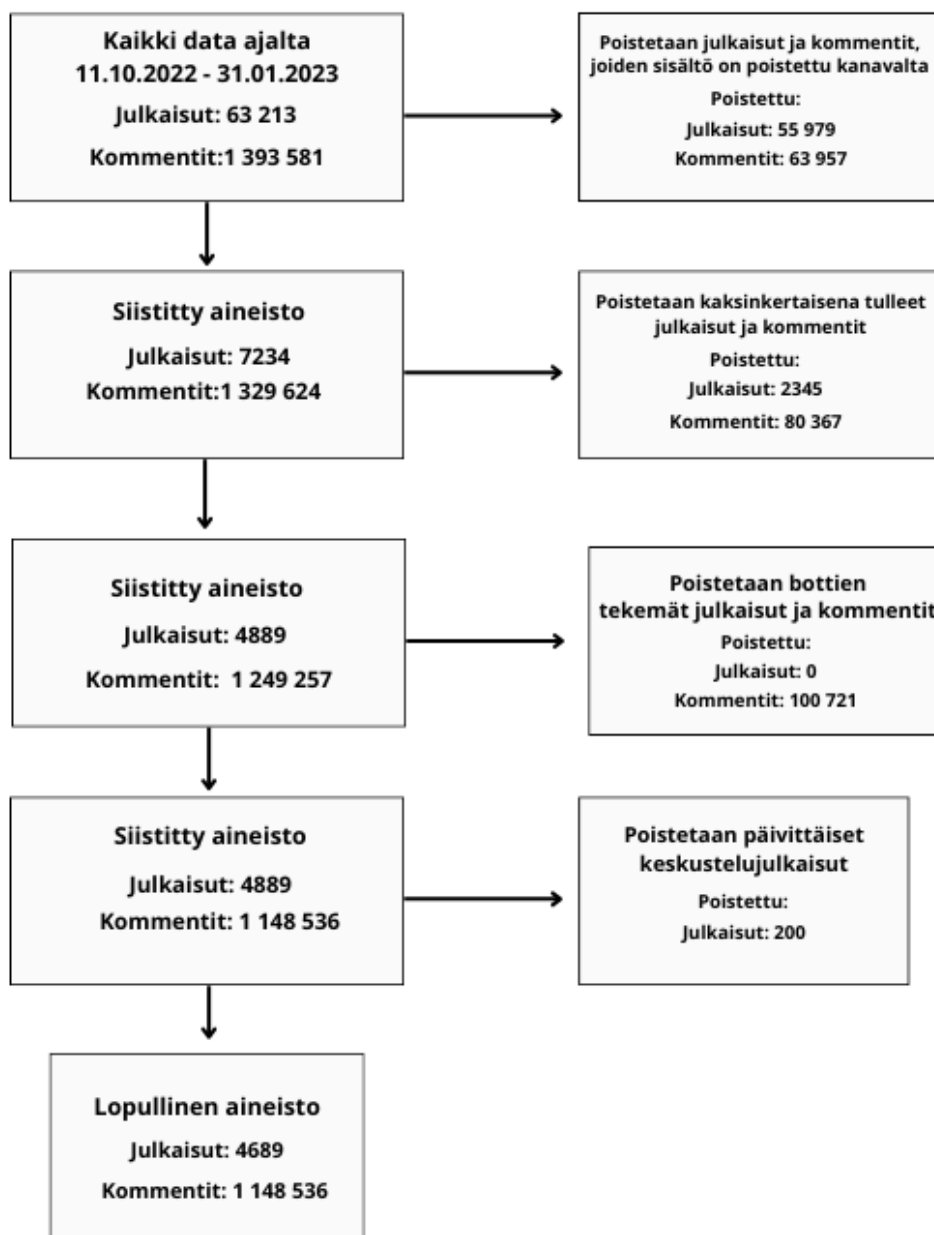
Tämän tutkimuksen aineisto on kerätty Redditin API:n kautta. Käytössämme oleva data sisältää kaikki julkaisut ja ensimmäisen tason kommentit ajalta 11.10.2022 – 31.1.2023. Ensimmäisen tason kommentti tarkoittaa julkaisuihin tulleita kommentteja, mutta ei kommentteihin tulleita kommentteja. Lisäksi data sisältää tiedot muun muassa julkaisuiden ja kommenttien saamista äänistä, julkaisun flairista, sekä tiedot käyttäjien karmasta ja pal-kinnoista.

Datassa on paljon puutteellisia havaintoja, joista puuttuu esimerkiksi kirjoittajan karma, julkaisun sisältö, tai julkaisun äänimäärä. Osa näistä on julkaisuja, jotka on poistettu joko käyttäjän tai ylläpitäjän toimesta. Syitä poistamiselle voi olla monia, mutta esimerkiksi r/wallstreetbetsin tiukkojen sääntöjen takia ylläpitäjät poistavat luultavasti paljon julkaisuja. Myös käyttäjien poistuminen kanavalta voi aiheuttaa julkaisujen poistumista. Poistetut julkaisut ovat kuitenkin jääneet Redditin API:in, ja päätyneet täten dataan. Tämä on käyttäjädatalle tyypillistä sotkuisuutta (ks. Kitchin 2014). Poistetut julkaisut ja kommentit poistettiin datasta kokonaan analyysiä varten.

Lisäksi dataa tarkasteltaessa huomattiin, että kanavalla julkaistut kuvat ovat tulleet dataan kaksinkertaisena, joka nostaa havaintojen määrää runsaasti. Pelkät julkaisuista irrotetut kuvat eivät kuitenkaan sisällä juuri muita tietoja, eivätkä tuo lisäarvoa analyysille, joten ne poistettiin datasta. Datan siistimisprosessi on havainnollistettu kuvassa 2.

Muut puutteellisia tietoja sisältävät havaintoyksiköt on otettu mukaan analyysiin mahdollisuuksien mukaan. Esimerkiksi puuttuva tieto karmasta ei tuota ongelmaa, kun analysoidaan ääniä, mutta analyyseissa, joissa tutkitaan karmaa, puutteelliset havaintoyksiköt on jätettävä huomiotta. Havaintojen määrä eri analyysien välillä voi siis vaihdella. Lisäksi datasta poistettiin kanavalla toimivien automaattisten moderointibottien tekemät julkaisut ja kommentit, sillä nämä eivät ole todellisia käyttäjiä, eivätkä siten mielekkäitä havaintoja.

Datasta poistettiin myös erään kanavan ylläpitäjän julkaisemat päivittäiset keskusteluketjut. Nämä julkaisut päätettiin poistaa niiden jakaumia vääristävän luonteen vuoksi, sillä ne keräävät yleisen luonteensa vuoksi todella paljon vastauksia. Analyysissä on tarkoitus selvittää tavallisten käyttäjien aktiivisuutta ja heidän saamaansa huomiota. Ylläpitäjän päivittäiset yleiseen keskusteluun tarkoitetut julkaisut eivät toimi samalla logiikalla kuin muu keskustelu kanavalla. Tästä syystä nämä julkaisut on jätetty tämän analyysin ulkopuolelle, ja näiden keskusteluketjujen sisältöä tulisi tutkia erikseen. Lopullinen aineisto sisältää 4689 julkaisua ja 1148536 kommenttia.



Kuva 2. Datan siistimisprosessi

Analyysi tulee etenemään seuraavanlaisesti: ensin tarkastellaan keskustelun jakautumista kanavalla. Oletuksena on, että kanavalla on pieni määrä aktiivisia julkaisijoita ja suuri määrä hiljaisia seuraajia (Carron-Arthur 2014; Kilgo 2016). Tämän tarkoituksena on muodostaa yleinen käsitys aktiivisuuden jakautumisesta kanavalla. Tämän jälkeen tarkastellaan julkaisuiden ja kommenttien saaman huomion määrän jakautumista kanavalla. Oletuksena on, että vain pieni osa julkaisuista ja kommentteista saa paljon huomiota (Mahanti 2013, Choi 2015). Aktiivisuuden ja huomion jakautumisen tarkastelun tarkoituksena on luoda yleiskuva aineistosta.

Seuraavaksi pyritään paikantamaan mielipidevaikuttajia kanavalta. Tämä tehdään huomion ja aktiivisuuden jakautumisen pohjalta luokittelemalla käyttäjiä aktiivisimpien ja eniten huomiota saavien käyttäjien ryhmiin. Aktiivisuutta mitataan käyttäjän kirjoittamien julkaisuiden ja kommenttien määrällä. Huomiota taas mitataan käyttäjän saamien äänten ja kommenttien keskiarvolla. Koska aineistossa on vain ensimmäisen tason kommentit, kommenttien saamaa huomiota voidaan tarkastella vain äänten avulla. Yhdistämällä aktiivisimmat ja eniten huomiota saavien käyttäjien ryhmät, löydetään käyttäjät, joilla on sekä korkea aktiivisuus, että korkea huomion taso. Näitä käyttäjiä voidaan todennäköisin perustein pitää mielipidevaikuttajina (Choi 2015, Kilgo ym. 2016). Näitä henkilöitä verrataan verkostanalyysin avulla verkstorakenteessa keskeisimpiin henkilöihin, joita voidaan myös pitää mielipidevaikuttajina (Rohde ym. 2023).

Verkostoanalyysin avulla lasketaan käyttäjille keskeisyysaste ja tarkastellaan verkoston rakennetta, sekä mielipidevaikuttajien asemaa verkostossa. Verkosto muodostetaan käyttäjien ja näiden toistensa julkaisuihin kirjoittamien kommenttien avulla. Verkostokartassa solmut ("node") kuvastavat käyttäjiä ja yhteydet muodostuvat käyttäjien toistensa julkaisuihin kirjoittamista kommentteista. Verkosto muodostetaan suunnattuna, eli solmun vastaanottamat ja lähettämät kommentit lasketaan erikseen. Käyttäjakohtaisia keskeisyysmittareita hyödynnetään regressiomallissa ja verkostokartan avulla pyritään tekemään päätelmiä mielipidevaikuttajien asemasta ja tiedon virtaamisesta verkostossa.

Käyttäjien luokittelu aktiivisuuden ja huomion perusteella, sekä verkostoanalyysi siis tukevat toisiaan mielipidevaikuttajien etsimisessä. Menetelmät lähestyvät samaa ongelmaa hieman eri näkökulmista ja tuovat omat vahvuutensa analyysiin. Asetelman tarkoituksena on myös vertailla menetelmien välisiä eroja, sekä näiden vahvuuksia ja heikkouksia, jotta

tulevaisuudessa järjestelmää pystytään kehittämään paremmaksi. Vertailemalla näiden menetelmien tuloksia pyritään mahdollisimman perustellusti tunnistamaan potentiaaliset mielipidevaikuttajat kanavalla. On esimerkiksi oletettavissa, että tarkastelemalla aktiivisuutta, saatua huomiota ja keskeisyyttä verkostossa, esiin nousevat osittain samat käyttäjät.

Julkaisuiden osalta mielipidevaikuttajiksi määritellään ne käyttäjät, jotka kuuluvat sekä aktiivisimpaan 10 %, että eniten huomiota saavaan 10 % julkaisijoista. Kommenttien osalta raja asetetaan korkeimpaan 1 %, sillä kommenttien määrä aineistossa on huomattavasti julkaisujen määrää suurempi. Tällä menetelmällä tunnistetut potentiaaliset mielipidevaikuttajat korostetaan verkostoanalyysillä luodussa verkostokartassa, jolloin näiden käyttäjien verkostoasemaa voidaan tarkastella visuaalisesti.

Oletuksena on, että mielipidevaikuttaja-asemassa olevien henkilöiden viestit saavat paljon huomiota ja ovat keskimääräistä uskottavampia, jolloin ne myös vaikuttavat keskimääräistä enemmän lukijoiden mielipiteisiin. On kuitenkin syytä huomata, että vaikka äänten tai kommenttien perusteella voidaan tehdä joitain päätelmiä viestin saamasta huomiosta, nämä eivät kuitenkaan kerro lainkaan sitä, kuinka moni käyttäjä on nähnyt viestin, mutta päättänyt olla reagoimatta siihen millään lailla (Tufekci 2014). Todellisen huomion määrän voidaan olettaa olevan huomattavasti äänten ja kommenttien määrää suurempi. Kanavalla on yli 14 miljoonaa rekisteröitynyttä seuraajaa ja vain harva osallistuu keskusteluun.

Viimeisenä tarkastellaan käyttäjän karman, julkaisun flairin ja käyttäjän saamien trophyjen, sekä käyttäjän verkostoaseman yhteyttä julkaisun saamaan huomioon regressioanalyysin avulla. Regressioanalyysin avulla voidaan tutkia samanaikaisesti useiden muuttujien välisiä yhteyksiä, niiden voimakkuutta ja vaikutuksia toisiinsa. Regressioanalyysi on tehokas ja monipuolinen menetelmä tällaisten vaikutussuhteiden tarkasteluun (Field s.371–374).

Muuttujien jakaumien vinouden takia regressioanalyysia varten äänten, kommenttien ja karman määrille, sekä käyttäjän keskeisyysasteelle tehtiin logaritmuunnos. Logaritmuunnos on yleisesti tilastollisissa malleissa käytetty menetelmä, jonka avulla voidaan tasoittaa muuttujien jakaumaa, varianssia ja keskiarvojen keskivirheitä (Fox & Weisberg

2011 s.126–131). Logaritminuunnosta voidaan käyttää esimerkiksi oikealle vinon datan, epätasaisten varianssien ja epälineaarisuuden käsittelyyn (Field 2018 s.269).

Tämän analyysin tavoitteena on tarkentaa näiden potentiaalisten auktoriteettisignaalien ja julkaisun saaman huomion välistä yhteyttä. Regressioanalyysin avulla voidaan tarkastella muuttujien välisten yhteyksien voimakkuutta, ja mallintaa selitettävän muuttujan arvoja eri selitettävien muuttujien arvoilla (Nummenmaa ym. 2016 s. 236–237, 242–244, 249–250). Lisäksi regressiomallin yhtenä selittäjänä muuttujana käytetään käyttäjän verkostoasemasta laskettua sisään tulevaa keskeisyysastetta (”indegree centrality”). Sisääntulo-keskeisyysaste on yleisesti verkostanalyysissä käytettävä mittari, joka muodostuu solmuun, eli tässä tapauksessa käyttäjälle sisään tulevien yhteyksien summana. (Golbeck 2013 s.25–27.) Solmusta ulos päin lähtevät yhteydet on jätetty regressioanalyysin ulkopuolelle, eli vain saatujen kommenttien määrä kasvattaa keskeisyysastetta. Näin vältetään roskapostaaajien päätyminen merkityksellisiksi tekijöiksi analyysissä.

Keskeisyysasteen käyttäminen selittäjänä muuttujana regressiomallissa julkaisun saamia kommentteja selitettäessä voi olla ongelmallista, sillä keskeisyysaste lasketaan käyttäjän julkaisuja kommentoineiden käyttäjien summana. Asetelmassa piilee kehäpäätelmän riski – käyttäjät, jotka saavat paljon kommentteja, saavat paljon kommentteja. Pidän riskiä kuitenkin melko pienenä, sillä tässä verrataan yksittäisen julkaisun saamia kommentteja käyttäjän kokonaisuudessaan saamiin vastaajiin kaikille julkaisuilleen. Ainoastaan tilanteessa, jossa käyttäjä on julkaissut vain yhden julkaisun, ja saanut sille suuren määrän kommentteja, tämä voi vinouttaa tulosta. Tällaiset tapaukset ovat kuitenkin harvinaisia aineistossamme.

Analyysit toteutetaan verkostanalyysiä lukuun ottamatta R-kielellä. R on avoimen lähdekoodin toimintaympäristö, joka on tarkoitettu pääasiassa datan manipulointiin, tilastolliseen analyysiin ja tulosten visualisointiin. R on tehokas datan käsittelyn ja laskennan työkalu, joka toimii koodikielen avulla. R ohjelmiston ominaisuuksia voidaan laajentaa erilaisten laajennuspakettien (package) avulla. (R-project.) Tätä tutkimusta varten käytettiin paketteja readr, tidyverse, powerLaw ja car. Tätä tutkimusta varten kirjoitetun R koodin tuottamisessa on käytetty apuna OpenAi:n ChatGPT tekoälyä. Kaikki tätä tutkimusta

varten kirjoitettu koodi löytyy Github¹³ -palvelusta. Verkostoanalyysi tehtiin Gephi ohjelmistolla, jolla laskettiin käyttäjille keskeisyysluvut ja piirrettiin verkostokartta.

Tutkimuksen eettisyyttä pohdittaessa ihmistieteissä otetaan yleensä huomioon tutkimuskohteiden suostumus, anonymiteetti, ja tiedon yksityisyys tai julkisuus. Verkkosyntyisen datan käyttämisen eettisyys tutkimuksessa on herättänyt keskustelua tutkijoiden keskuudessa. Onko esimerkiksi sosiaalisesta mediasta kerätty data, kuten julkaisut, vapaasti käytettävissä, vai tarvitseeko sen käyttämiseen saada jokaisen yksilön suostumus. On esitetty, että tämä riippuu esimerkiksi julkaisun sisällöstä ja kanavan luonteesta. Myös sillä, minkälaista dataa kerätään, on merkitystä. (Ackland 2013 s.43–45.)

Raja julkisen ja yksityisen tilan välillä hämärtyy verkossa ja pohdinta tiedonkeruun eettisyydestä kulminoituu debattiin siitä, mitä tulisi käsitellä yksityisenä ja mitä julkisena tietona. Toiset kanavat ovat yksityisempiä kuin toiset. Esimerkiksi sosiaalisen median julkaisut on lähtökohtaisesti tarkoitettu omalle seuraajakunnalle, kun taas esimerkiksi uutisten kommenttikentät ovat selkeämmin julkista tilaa. Selkeänä linjana yksityisen ja julkisen tilan välillä verkossa voidaan pitää sitä, onko kanava avoin vai suljettu. (Snee 2016 s.214–216.)

Myös keskusteluiden aiheella on merkitystä. Esimerkiksi vertaistuelle tarkoitetuilla foorumeilla voidaan keskustella hyvinkin yksityisistä asioista, vaikka foorumi olisi sinänsä täysin avoin. Lisäksi ihmisten oma käsitys siitä, kuinka julkista keskustelua on voi vaihdella. Ihmiset voivat ajatella käyvänsä suhteellisen yksityistä keskustelua julkisellakin kanavalla olettaen, että tietoa ei käytetä muihin tarkoituksiin. Tutkijan on hyvä myös miettiä, ovatko kanavan käyttäjät teknisesti kykeneviä ymmärtämään, mitä tietoja itsestä julkaisee verkkoon ja millaiseen käyttöön ne voivat päätyä. (Ackland 2013 s.45.) Näiden seikkojen valossa on syytä pohtia suostumuksen pyytämisen tarpeellisuutta.

Tutkimuskohteiden anonymiteetistä tulee huolehtia tutkimuksen toteutuksessa. Datasta tulee poistaa tiedot, jotka voivat paljastaa yksilöiden identiteetin ja lisäksi data tulee säilyttää tietoturvallisesti. Myös suorien lainausten esittäminen esimerkiksi sosiaalisen median julkaisuista on riski anonymiteetille, sillä niiden avulla voidaan etsiä alkuperäiset

¹³ Koodi löytyy osoitteesta <https://github.com/ARthesisdump/Memestonks-code.git>

julkaisut hakukoneella (Ackland 2013 s.46). Tutkimuskohteiden identiteetin paljastumisen riski on kuitenkin hyvin pieni tutkimuksessa, jossa käsitellään määrällisesti käyttäjätietoa (Snee ym. 2016 s.217).

Redditä pääsee selaamaan kuka tahansa ilman profiilia ja lisäksi Wallstreetbets -kanavalla on yli 14 miljoonaa rekisteröitynyttä seuraajaa. Näistä syistä kanavaa voidaan pitää julkisena tilana. Lisäksi Reddit kertoo tietoturvaselosteessaan suurimman osan datasta, kuten julkaisuiden ja käyttäjäprofiilien olevan julkista tietoa, ja että kolmansilla osapuolilla on pääsy näihin tietoihin Redditiin API:n kautta. Näin ollen suostumuksen pyytäminen tutkimuskohteilta ei ole nähdäkseni tarpeellista.

Tässä tutkimuksessa tarkastellaan kvantifioitua käyttäjätietoa, joten riski yksilöiden tunnistamisesta on pieni. Tutkimuksessa ei lähtökotaisesti olla kiinnostuneita yksilöistä, vaan kanavalla yleisesti vaikuttavista säännönmukaisuuksista. Mielipidevaikuttajien paikantamisessa toki tarkastellaan tiettyjä käyttäjiä, mutta tämä joukko on oletettavasti kymmeniä tai satoja käyttäjiä. Käyttäjien nimimerkit anonymisoidaan yksilönsuojan turvaamiseksi. Tässä tutkimuksessa ei myöskään esitetä lainauksia käyttäjien julkaisuista.

5 TULOKSET

5.1 Julkaisuiden ja kommenttien jakautuminen

Aloitetaan tarkastelemalla julkaisuiden, kommenttien ja äänten jakautumista yleisellä tasolla. Taulukossa 1. on esitetty keskeisimmät tunnusluvut kanavalle tehdyistä julkaisuista ja kommentteista, sekä niiden saamat äänet julkaisu- ja käyttäjäkohtaisesti. Näiden tietojen perusteella voidaan tehdä päätelmiä käyttäjien aktiivisuudesta ja käyttäjien saamasta huomiosta, sekä näiden jakautumisesta kanavalla.

Osassa havainnoista puuttui kirjoittajan tiedot, julkaisun sisältö, äänimäärä tai kommenttimäärä. Nämä havainnot on poistettu datasta analyysiä varten. Äänten määriä koskevien tunnuslukujen laskemiseksi on poistettu julkaisut ja kommentit, joiden nettoäänimäärä on negatiivinen. Tämä tehtiin suhdelukujen laskemisen mahdollistamiseksi, sillä suhdeasteikon käyttäminen vaatii absoluuttisen nollopisteen (Nummenmaa s.57). Nettomäärältään negatiiviset äänimäärät olivat aineistossa kuitenkin harvinaisia.

Julkaisuita tehneiden käyttäjien määrä tässä aineistossa on 2990 ja kommentoijien määrä on 173737 käyttäjää. Aineistossa eniten julkaisuita kirjoittanut käyttäjä oli tehnyt 99 julkaisua ja eniten kommentoinut käyttäjä oli kirjoittanut 3722 kommenttia. On myös syytä huomata, että nämä tiedot pitävät sisällään ainoastaan käyttäjät, jotka ovat ylipäättään kirjoittaneet julkaisuja tai kommentteja. Kanavalla on yli 14 miljoonaa seuraajaa, joten suurin osa seuraajista on passiivisia.

Taulukko 1. Aktiivisuuden ja huomion jakautuminen r/wallstreetbets -kanavalla

	n	Keskiarvo	Mediaani	90% desiili	Keskihajonta	Top 10% osuus
1. Käyttäjien tekemät julkaisut	4689 (julkaisua)	1,6	1	2	3,2	47,60 %
2. Käyttäjien kirjoittamat kommentit	1148536 (kommenttia)	6,6	1	8	42,8	76 %
3. Julkaisuiden saamat äänet	4689 (julkaisua)	254,3	5	104	1932	95,7%*
4. Kommenttien saamat äänet	1148536 (kommenttia)	8,9	2	8	138,4	78,8%*
5. Julkaisuiden saamat kommentit	4689 (julkaisua)	43,1	12	87	144	65,30 %
6. Käyttäjien saamat äänet julkaisuilleen	2990 (käyttäjä)	111,1	3	88	938,9	91,70 %
7. Käyttäjien saamat kommentit julkaisuilleen	2990 (käyttäjä)	45,4	13	93	157,9	65,40 %
8. Käyttäjien saamat äänet kommentteilleen	173737 (käyttäjä)	3,3	0	2	77	95,2%*
* negatiiviset arvot jätetty pois						

Sekä julkaisuiden, kommenttien, että äänten jakaumat ovat hyvin vinoja ja noudattavat selvästi potenssijakaumaa. (Kuvaajat liitteessä 1–8.) Pieni määrä julkaisijoista on tehnyt suuren osan julkaisuista ja kommentteista, sekä saanut suuren osan huomiosta. Eniten julkaissut 10 % on julkaissut lähes 48 % julkaisuista. Vastaava luku kommenttien osalta on 76 %. Sekä julkaisuiden, että kommenttien määrien mediaani on 1, eli puolet kirjoittajista on julkaissut vain yhden julkaisun tai kommentin. 90 % desiili kuvastaa rajaa, jonka alla 90 % käyttäjistä on. Esimerkiksi käyttäjien kirjoittamien kommenttien kohdalla tämä tarkoittaa sitä, että 90 % kommentteja kirjoittaneista käyttäjistä on kirjoittanut 8 kommenttia tai vähemmän.

Tarkastellaan seuraavaksi huomion jakautumista saaduilla äänillä ja kommentteilla mitattuna. Taulukon 1. kohtia 3–5 tarkastelemalla voidaan jälleen päätellä jakaumien olevan todella vinoja. Eniten ääniä kerännyt 10 % julkaisuista ja kommentteista sai lähes kaikki äänet. Eniten kommentteja kerännyt 10 % julkaisuista sai yli 65 % kaikista kommentteista. Mediaanit ja 90 % desiilit ovat jälleen alhaisia ja keskihajonnat korkeita. Eniten ääniä saanut julkaisu sai 41 373 ääntä ja eniten kommentteja saanut julkaisu sai 4731 kommenttia. Eniten ääniä saanut kommentti sai 24 524 ääntä.

Edeltävät analyysit koskivat siis julkaisuiden ja kommenttien saaman huomion jakaumaa. Tarkastellaan vielä seuraavaksi käyttäjien keskimäärin saamaa huomiota julkaisuilleen ja kommentteilleen. Taulukon 1. kohdissa 6–8 on esitetty vastaavat luvut käyttäjäkohtaisesti. Luvut kertovat samaa tarinaa kuin aikaisemmissakin tarkasteluissa. Suurin osa käyttäjistä saa vain vähän huomiota julkaisuilleen, kun taas pieni joukko käyttäjistä kerää paljon huomiota. Suurin osa julkaisuita tehneistä käyttäjistä on saanut 3 ääntä tai vähemmän. Julkaisuilleen eniten ääniä saanut 10 % käyttäjistä saa lähes 92 % äänistä. Käyttäjän saamien kommenttien jakauma on hieman tasaisempi, mutta ei poikkea trendistä. Eniten kommentteja saanut 10 % käyttäjistä saa runsaat 65 % kommentteista.

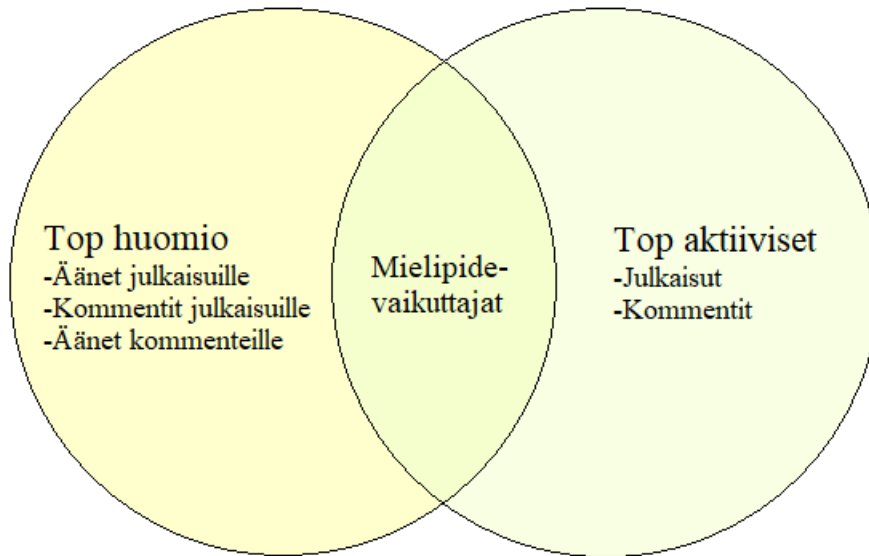
Kaiken kaikkiaan näiden tulosten perusteella voidaan todeta, että aineistossamme on suuri joukko käyttäjiä, jotka eivät ole erityisen aktiivisia julkaisemaan tai kommentoimaan, ja pieni joukko hyvin aktiivisia julkaisijoita. Vastaava jakauma on havaittavissa myös julkaisuiden ja kommenttien saaman huomion kohdalla. Jakaumat noudattavat potenssilakia ja ovat linjassa teorian kanssa. Tulosten perusteella voidaan olettaa, että kanavalla on mielipidevaikuttajia, jotka julkaisevat paljon, saavat paljon huomiota ja ovat keskeisessä roolissa luomassa kanavan keskustelua.

5.2 Mielipidevaikuttajien paikantaminen

Seuraavaksi etsitään potentiaalisia mielipidevaikuttajia aineistosta. Mielipidevaikuttajien löytämiseksi käyttäjät järjestettiin julkaisu- ja kommenttimäärien mukaan ja luotiin lista aktiivisimmasta kymmenestä prosentista käyttäjistä. Rajat valittiin taulukossa 1 esitettyjen jakaumien perusteella.

Vastaavanlainen lista tehtiin myös eniten huomiota saavista käyttäjistä. Saatua huomiota mitattiin keskimääräisenä äänten ja kommenttien määränä per käyttäjä. Listat eniten ääniä ja eniten kommentteja saaneista käyttäjistä luotiin erikseen, ja mukaan valikoitiin käyttäjät, jotka pääsivät jommallekummalle listalle. Eniten ääniä saaneiden kommenttien osalta sovellettiin tiukempaa top 1 % rajaa, sillä ylin 10 % joukko saa lähes kaikki äänet. Ver-

taamalla näitä listoja keskenään, voidaan laskea, kuinka paljon samoja henkilöitä on paljon julkaisevien, paljon kommentoivien ja paljon huomiota saavien käyttäjien joukossa. Kuva 3. havainnollistaa prosessia.



Kuva 3. Mielipidevaikuttajien asema huomion ja aktiivisuuden suhteen.

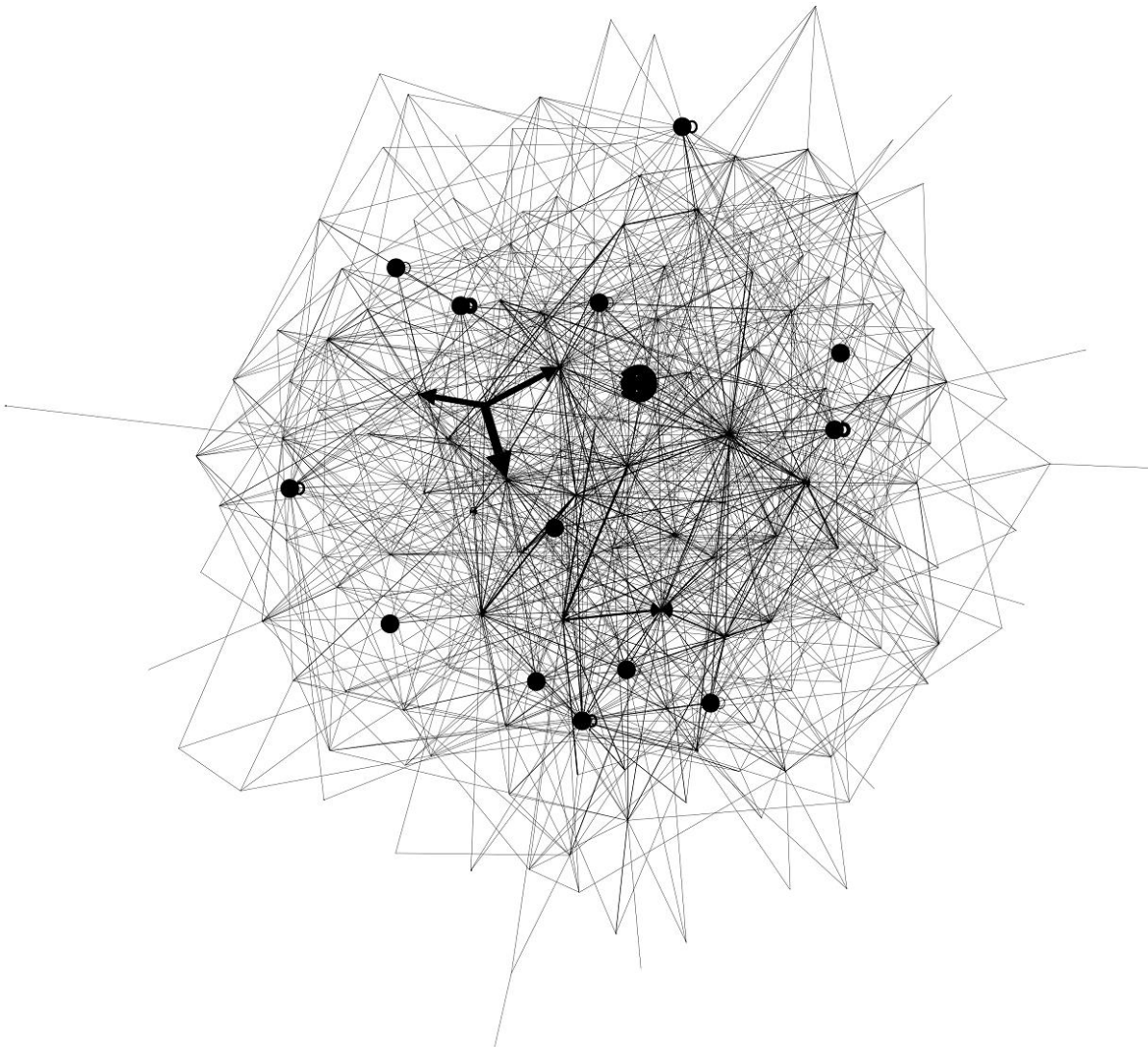
Mielipidevaikuttajien paikantamista varten datasta poistettiin flairin perusteella julkaisut, joiden flair oli ”Meme” tai ”shitpost”, sekä näihin tulleet kommentit. Tämä tehtiin siksi, että mielipidevaikuttajia etsiessä ollaan kiinnostuneita vain asiasisällöstä. ”Meme” ja ”shitpost” -flairit on tarkoitettu julkaisuille, joiden pääasiallinen sisältö on meemejä, vitsejä, satiiria, tai muuta vastaavaa¹⁴. Tämä ei kuitenkaan rajaa pois kaikkia meemijulkaisuja, sillä kaikkia julkaisuja ei ole merkattu flairilla lainkaan.

Yhdistämällä aktiivisin ja eniten huomiota saanut 10 % julkaisijoista löydettiin 55 käyttäjää. Eli aineistossa on 55 käyttäjää, jotka kuuluvat aktiivisimpaan 10 % julkaisijoista ja saavat julkaisuilleen keskimäärin enemmän ääniä tai kommentteja, kuin 90 % kaikista

¹⁴ <https://web.archive.org/web/20200313023742/https://www.reddit.com/r/wallstreetbets/wiki/linkflair>

julkaisijoista. Käyttäjien kirjoittamien julkaisuiden vaihteluväli oli 2–77 julkaisua ja huomion vaihteluväli 12–377 kommenttia ja 20–16372 ääntä. Näiden käyttäjien voidaan aktiivisuuden ja huomion puolesta sanoa olevan mielipidevaikuttajia aineistossamme. Listalla olevat käyttäjät on korostettu verkostokartassa (kuva 4.). Koko mielipidevaikuttajalista löytyy anonymisoituna liitteestä 9.

Kommenttien kohdalla sovellettiin vastaavaa menetelmää. Yhdistämällä eniten kommentteja kirjoittaneiden (10 %) lista eniten ääniä kommentteilleen saaneiden (1 %) listaan löydettiin 178 käyttäjää. Tällä listalla käyttäjien kirjoittamien kommenttien vaihteluväli oli 8–139 kommenttia ja kommenttien saamien äänten keskiarvon vaihteluväli oli 140–4337 ääntä. Näitä käyttäjiä voidaan pitää mielipidevaikuttajina kommentoinnin suhteen. Näitä ei kuitenkaan voitu korostaa verkostokartassa, sillä verkostokartta on luotu käyttäjien toistensa julkaisuihin kirjoittamien kommenttien mukaan. Anonymisoitu lista eniten kommenttoineista ja kommentteilleen huomiota saaneista käyttäjistä löytyy liitteestä 10.



Kuva 4. Verkostokartta kanavan keskeisimmistä käyttäjistä.

Top 10 % -menetelmällä luotujen mielipidevaikuttajalistojen perusteella voidaan todeta, että samat henkilöt, jotka julkaisevat paljon, ovat vain osittain samoja henkilöitä, jotka saavat paljon huomiota. Kanavalle on julkaissut otanta-aikana 2990 käyttäjää, josta 10 % on 299. Jos käyttäjät, jotka ovat aktiivisia ja saavat paljon huomiota olisivat täysin samoja, kaikki nämä 299 käyttäjää olisivat yhdistetyllä listalla. Nyt listalle pääsi noin 18 % näistä käyttäjistä. Kanavalla on siis runsaasti käyttäjiä, jotka vain julkaisevat paljon tai saavat paljon huomiota.

Kuvan 4. verkostokartta on luotu käyttäjien toistensa julkaisuihin kirjoittamien kommenttien pohjalta. Verkoston solmukohdat kuvastavat siis käyttäjiä, ja yhteydet käyttäjien toisilleen kirjoittamia kommentteja. Tämä verkostokartta on kaksisuuntainen, eli se sisältää sekä sisään, että ulospäin suuntautuvat yhteydet. Datan massiivisuuden takia verkoston kokoa jouduttiin rajoittamaan melko paljon, sillä koko aineiston kattavan verkostokartan luominen vaatisi liikaa laskentatehoa. Lisäksi, vaikka verkostokartta olisikin saatu luotua koko aineistolle, siitä tulisi luultavasti sotkuinen ja sen tulkitseminen lähes mahdotonta. Näistä syistä verkostokartta luotiin ainoastaan käyttäjistä, joiden keskeisyysaste on vähintään 500. Eli kaikki tässä kartassa näkyvät käyttäjät ovat joko kommentoineet tai saaneet kommentteja vähintään 500 eri käyttäjältä. Tällä kriteerillä verkostoon jäi 189 käyttäjää.

Keskeisyyttä verkostossa voidaan jo itsessään pitää mielipidevaikuttaja-aseman mittarina (Rohde ym. 2023). Tämän mittarin perusteella verkostokartan kaikkien käyttäjien voidaan ajatella olevan mielipidevaikuttajia kanavalla. Tämän lisäksi verkostokartassa on korostettuna aikaisemmin aktiivisuuden ja huomion perusteella paikannetut mielipidevaikuttajat. Näistä 55 käyttäjästä verkostokarttaan päätyi vain 14. Tämä johtuu siitä, että menetelmät mittaavat eri asioita. Vaikka molempien menetelmien tavoitteena on paikantaa mielipidevaikuttajia, ne tekevät sen eri tavalla.

Keskeisyysaste mittaa käyttäjän yhteyksien määrää. Täten esimerkiksi käyttäjällä, joka joko kommentoi paljon tai saa paljon kommentteja on korkea keskeisyysaste. Verkostokarttaan voi siis päästä esimerkiksi ainoastaan kommentoimalla paljon. Keskeisyysaste ei huomioi lainkaan käyttäjän tekemien julkaisuiden määrää eikä niiden saamien äänten määrää. Mielipidevaikuttajalistalle pääseminen taas vaatii sen, että käyttäjä sekä julkaisee paljon, että saa julkaisuilleen paljon kommentteja tai ääniä. Tämä ei kuitenkaan takaa

korkeaa keskeisyysastetta. Menetelmät siis mittaavat jossain määrin eri asioita, ja kumpikin mittaa mielipidevaikuttaja-asemaa omalla tavallaan.

Verkostokartassa on mustalla korostetut 14 käyttäjää ovat päässeet sekä verkostokarttaan, että mielipidevaikuttajalistaan. Tämän tutkimuksen tulosten valossa näitä käyttäjiä voidaan suurimmalla varmuudella pitää mielipidevaikuttajina kanavalla. Tarkastelemalla verkostokarttaa yleisemmällä tasolla voidaan huomata, että se on muodoltaan tiivis, eikä siinä ole toisistaan erottuvia joukkoja. Verkostokartta on keskeltä tiheämpi, ja keskialueella on havaittavissa muutama muita keskeisempi käyttäjä, jotka eivät kuitenkaan päässeet mielipidevaikuttajalistaan.

Verkoston yhteydet on esitetty nuolilla, jotka kuvaavat yhteyden suuntaa ja vahvuutta. Kartasta nousee esiin erityisesti yksi käyttäjä, jolla on kolme vahvaa ulospäin osoittavaa nuolta. Tämä käyttäjä on siis kommentoinut erityisen paljon nuolien osoittamien kolmen käyttäjän julkaisuja. Voi olla, että tämä käyttäjä kommentoi syystä tai toisesta hyvin aktiivisesti juuri näiden kolmen käyttäjän julkaisuja. Pidän kuitenkin todennäköisempänä, että kommentit ovat kertyneet pääasiassa yhdelle julkaisulle (per käyttäjä) kommenttikentässä käytävän vuoropuhelun muodossa. On hyvin mahdollista, että tämä käyttäjä on päätyntä verkostokarttaan ainoastaan korkean kommentointimäärän vuoksi.

On syytä pohtia, voidaanko kaikkia tämän verkostokartan käyttäjiä pitää lähtökohtaisesti mielipidevaikuttajina, ja toisaalta voidaanko kaikkia top mielipidevaikuttajalistan käyttäjiä pitää mielipidevaikuttajina. Menetelmät onnistuivat paikantamaan vain osittain samat henkilöt, ja molemmat menetelmät ovat luultavasti nostaneet esiin myös henkilöitä, joita ei tulisi määrittellä mielipidevaikuttajaksi. On myös mahdollista, että mielipidevaikuttajia on jäänyt näillä menetelmillä huomaamatta.

Näiden tulosten avulla pystytään kuitenkin huomioimaan mielipidevaikuttajien merkitys signaalijärjestelmää kehitettäessä. Mikäli tässä tutkimuksessa paikannettujen mielipidevaikuttajien julkaisuiden ja pörssikurssien kehityksen välillä pystytään havaitsemaan yhteys, voidaan mielipidevaikuttajien paikantamisen menetelmiä hioa tarkemmiksi. Mikäli yhteyttä ei kuitenkaan löydy, voidaan miettiä, onko sellaista lainkaan.

5.3 Auktoriteettisignaalit

Seuraavaksi tarkastellaan potentiaalisten auktoriteettisignaalien, eli käyttäjän karman, käyttäjän saamien trophyjen ja julkaisun flairin yhteyttä julkaisujen ja kommenttien saamaan huomioon. Tarkastellaan ensin muuttujien välisiä yhteyksiä korrelaatiomatriisin avulla. Tämän jälkeen tutkitaan yhteyksiä tarkemmin regressioanalyysillä. Regressiomalliin lisätään auktoriteettisignaalien lisäksi käyttäjän verkostoasemaa kuvaava keskeisyysaste, jonka avulla pyritään selvittämään käyttäjän verkostoaseman vaikutusta julkaisun saamaan huomioon.

Taulukko 2. Muuttujien väliset korrelaatiot

	Karma	Äänet	Kommentit	Trophyt	Keskeisyysaste
Karma	1				
Äänet	0,31	1			
Kommentit	0,2	0,63	1		
Trophyt	0,63	0,22	0,15	1	
Keskeisyysaste	0,47	0,2	0,18	0,27	1

Taulukossa 2. on esitetty regressioanalyysissä käytettävien muuttujien väliset korrelaatiokertoimet. (Hajontakuviot liitteessä 11–16.) Suurten hajontojen ja jakaumien vinouden takia korrelaatiot on laskettu Spearmanin järjestyskorrelaatiokertoimella, joka on perinteistä korrelaatiota vähemmän herkkä äärimmäisille havainnoille ja epälineaarisuudelle (Nummenmaa ym. 2016 s.221–224). Kaikkien muuttujien välillä vallitsee vähintään heikko korrelaatio. Äänten ja kommenttien määrän välinen korrelaatio on 0,63 joka kertoo keskivahvasta yhteydestä. Regressiomallissa kommentteja ja ääniä ei kuitenkaan käytetä toistensa selittämiseen, sillä molemmat kuvaavat julkaisun saamaa huomiota.

Karman ja trophyjen korrelaatiot julkaisun saamaan huomioon antavat syytä tarkastella niiden yhteyttä tarkemmin. Korrelaatiot eivät ole erityisen suuria, mutta ne ovat kuitenkin olemassa ja niiden perusteella voidaan olettaa, että auktoriteettisignaalien ja julkaisun

saaman huomion välillä on yhteys. Tämä yhteys voi selittyä sillä, että lukijat pitävät kirjoittajia, joilla on korkea karma ja paljon trophyjä keskimääräistä luotettavampina. On kuitenkin syytä huomioida, että nämä korrelaatiot saattavat ainakin osittain selittyä yhteisellä taustamuuttujalla, kuten julkaisijan taipumuksella tuottaa laadukasta sisältöä. Oli miten oli, näiden muuttujien avulla voidaan mahdollisesti ennustaa julkaisun saaman huomion määrää.

Sisään tulevan keskeisyysasteen korrelaatio julkaisuiden saamien äänten kanssa on 0,2 ja julkaisun saamien kommenttien kanssa 0,18. Nämä korrelaatiot ovat voimakkuudeltaan samaa luokkaa kuin muidenkin muuttujien kohdalla, ja antavat syytä tarkastella yhteyttä tarkemmin regressioanalyysissä. Tämän tuloksen perusteella voidaan olettaa, että käyttäjän keskeisyys verkostossa ennustaa tämän julkaisuiden saamaa huomiota.

On myös syytä huomioida, että trophyjen ja karman, sekä keskeisyysasteen ja karman väliset korrelaatiot ovat vahvoja ja saattavat olla ongelmallisia regressiomallin kannalta. Kahden vahvasti toistensa kanssa korreloivan selittävän muuttujan käyttäminen regressiomallissa voi johtaa siihen, että malli selittää samaa vaihtelua kahteen kertaan (Nummenmaa 2021 s.446). Tämä huomioidaan regressiomallia rakentaessa. Regressiomalliin lisätään selittäviä muuttujia yksitellen, ja tarkastellaan selityksasteen, sekä regressiokertoimien muutosta.

Regressiomalliin otetaan mukaan myös julkaisun flair luokiteltuna muuttujana. Taulukossa 3. on esitettyinä julkaisuiden saaman huomion määrä luokiteltuna kanavalla käytettyjen flairien mukaan. (Laatikko-jana-kuvio liitteessä 17.) Vertailemalla eri flaireilla merkittävien julkaisuiden saamaa huomiota huomataan, että toiset flairit keräävät enemmän huomiota kuin toiset. Julkaisuiden saaman huomion jakauman vinouden vuoksi on mielekkäintä vertailla mediaaneja. On myös syytä huomioida flairien käyttöasteen jakauma. Osaa flaireista on käytetty hyvin vähän, eikä suurinta osaa julkaisuista ole merkitty flairilla lainkaan.

Taulukko 3. Flairien yhteys huomioon

Flair	Äänet keskiarvo	Äänet mediaani	Kommentit keskiarvo	Kommentit mediaani	n
Chart	183,9	9,0	91,4	28,5	14,0
DD	114,2	25,0	59,8	30,0	89,0
Discussion	55,6	8,0	58,4	25,0	367,0
Gain	25,4	17,5	40,1	25,0	14,0
Loss	87,8	36,5	75,1	26,5	10,0
Meme	2040,3	25,0	96,2	41,0	13,0
News	109,7	29,0	60,6	24,0	59,0
Shitpost	4,5	2,5	32,8	32,5	6,0
YOLO	137,0	16,0	66,3	39,5	20,0
COSMETIC SUGGESTION	1,0	1,0	2,0	2,0	1,0
Tyhjä	274,0	3,0	40,6	10,0	4096,0
Kaikki	254,3	5,0	43,1	12,0	4688,0

Taulukosta 3. nähdään, että lähes kaikki flairilla merkatut julkaisut saavat enemmän sekä kommentteja, että ääniä, kuin julkaisut, joita ei ole merkattu flairilla. Flairien saaman huomion jakaumaa tarkastelemalla voidaan tehdä päätelmiä kanavan kulttuurista. Esimerkiksi ”Loss” flairilla merkityt julkaisut saavat muihin verrattuna paljon ääniä. Myös ”YOLO” flair kerää melko paljon huomiota, erityisesti kommentteja. Näitä flaireja käytetään, kun julkaisu käsittelee käyttäjän tekemiä suuria tappioita (vähintään 2500 \$) tai käyttäjän ottamia suuria riskejä (vähintään 10 000 \$) (r/wallstreetbets flair guidelines 2021). Näiden havaintojen voidaan ajatella kertovan kanavan riskejä ihannoivasta kulttuurista. Suurten riskien ottamisesta ja erityisesti suurten tappioiden tekemisestä palkitaan huomiolla.

Aikaisemmissa sosiaalista mediaa käsittelevissä tutkimuksissa meemijulkaisut eivät ole saaneet paljon huomiota, koska nämä eivät tuota kanavalle lisäarvoa (Horne ym. 2017). Tällä kanavalla vaikuttaisi kuitenkin siltä, että meemejäkin arvostetaan. On kuitenkin syytä huomata, että ”Meme” flair esiintyi aineistossa vain 13 julkaisussa. Käytetyimpiä

flaireja olivat ”Discussion”, ”DD”¹⁵ ja ”News”. Näiden voidaan katsoa olevan lähtökoh-
taisesti asiallisia julkaisuja, jotka tuovat kanavalle arvokasta sisältöä. Myös nämä julkai-
sut keräävät melko paljon huomiota.

Näiden tulosten valossa voidaan olettaa flairilla olevan merkitystä julkaisun saaman huo-
mion kannalta. Flairien avulla voidaan mahdollisesti ennustaa julkaisun saamaa huomiota
ja täten flair kannattaa lisätä selittäväksi muuttujaksi regressiomalliin. Aineiston kerää-
misen aikaan flairin valitseminen julkaisulle ei ollut pakollista. Nykyisin r/wallstreetbets
vaatii käyttäjiä merkkamaan kaikki julkaisut flairilla. Tämä tekee asetelmasta vielä mie-
lenkiintoisemman. On mahdollista, että flairien saaman huomion suhteet muuttuvat tai
ainakin tulevat paremmin esille, kun kaikki julkaisut merkitään flairilla.

Potentiaalisten auktoriteettisignaalien ja julkaisuiden saaman huomion yhteyden tarkem-
paa selvittämistä varten tehtiin joukko regressioanalyysijä. Julkaisuiden saamien kom-
menttien määrää, julkaisuiden saamien äänten määrää, ja kommenttien saamien äänten
määrää pyrittiin selittämään erillisillä regressiomalleilla. Lopputuloksena on kolme reg-
ressiomallia, johon kuhunkin lisättiin selittäviä muuttujia yksitellen. Tämän menetelmän
tavoitteena on tutkia sitä, kuinka yksittäiset selittävät muuttujat vaikuttavat malliin ja toi-
siinsa.

Regressioanalyysin muuttujia on logaritmuunnettu aineiston vinouden ja muuttujien
välisten yhteyksien epälineaarisuuden tasoittamiseksi. Molemmat selitettävät muuttujat,
äänten ja kommenttien määrät, on logaritmuunnettu. Selittävästä muuttujista karma ja
keskeisyysaste on logaritmuunnettu. Trophy ja flair -muuttujat on säilytetty alkuperäi-
sellä asteikollaan. Selitettävä muuttuja on siis aina logaritmuunnettu, mutta selittävästä
muuttujista vain osa. Täten regressiomalli tuottaa tuloksia kahdella asteikolla: logaritmi
– lineaarinen (esim. kommentit – trophyt) tai logaritmi – logaritmi (esim. kommentit –
karma/keskeisyys).

¹⁵ ”Due Diligence” -julkaisun tulisi sisältää käyttäjän tekemää tutkimusta yrityksestä, markkinasta tai
muusta sijoitusideasta. <https://www.reddit.com/r/wallstreetbets/wiki/linkflair/>

Regressioanalyysin tulokset ovat siis logaritmisella asteikolla, joten regressiokertoimia ei voi tulkita suoraan muuttujien alkuperäisillä yksiköillä. Tulosten tulkitseminen suoraan logaritmisella asteikolla on mahdollista, mutta ei varsinaisesti mielekäästä, tästä syystä tuloksia kannattaa tulkita prosentuaalisina muutoksina. Tulosten tulkinta riippuu siitä, ovatko molemmat muuttujat, vai vain toinen muunnettu logaritmiselle asteikolle (Benoit 2011). Aavaan seuraavaksi regressiokertoimien matemaattista tulkintaa molemmissa tapauksissa.

Jos vain selitettävä muuttuja on logaritmimuunnettu (log – lin), x yksikön lisäys selittävässä muuttujassa muuttaa selitettävää muuttujaa $e^{x\beta}$ kertoimella. Kaavassa e on Neperin luku, x on selittävän muuttujan muutos ja β on selittävän muuttujan regressiokerroin. Regressiokerrointa voidaan myös tulkita yksinkertaistetusti kertomalla se sadalla, jolloin saadaan prosentuaalinen muutos selitettävässä muuttujassa yhden yksikön muutosta kohden selittävässä muuttujassa. Suuremmilla muutoksilla tulisi kuitenkin pitäytyä matemaattisessa kaavassa. (Benoit 2011.)

Jos sekä selitettävä, että selittävä muuttuja on logaritmimuunnettu (log – log), tulosta tulkitaan odotettuna prosentuaalisena muutoksena selitettävässä muuttujassa, kun selittävässä muuttujassa tapahtuu tietyn suuruinen prosentuaalinen muutos. Tämä voidaan laskea kaavalla $e^{\beta \cdot \ln(1+p)}$, jossa e on Neperin luku, β on selittävän muuttujan regressiokerroin, ja p on haluttu selittävän muuttujan prosentuaalinen muutos desimaalilukuna (esim. 0.1 = 10 %). (Benoit 2011.)

Regressiomalleissa on mukana myös kategorinen flair -muuttuja selittävänä muuttujana. Flairien regressiokertoimia tulkitessa sovelletaan samaa kaavaa, kuin log – lin -tapauksessa (Lye & Hirschberg 2002). Regressiokerroin muutetaan selittävän muuttujan muutosta kuvaavaksi kertoimeksi kaavalla e^{β} . Julkaisuja, joilla on flair, verrataan julkaisuihin, joilla ei ole flairia ja tulosta tulkitaan kertoimena.

Taulukko 4. Julkaisuiden saamien kommenttien regressioanalyysi (Malli 1.)

	1	2	3	4	Muutos
Vakiotermi	1,67	1,61	1,62	1,5	
log Karma	0,11*** (0,008)	0,13*** (0,009)	0,12*** (0,009)	0,11*** (0,009)	0,11 %
Trophyt		-0,035*** (0,009)	-0,036*** (0,009)	-0,031*** (0,009)	-3,50 %
log Keskeisyys			0,08*** (0,017)	0,087*** (0,02)	0,09 %
Flair					
Chart				0,91* (0,386)	2,48
Cosmetic Sug- gestion				-0,45 (1,44)	
DD				0,99*** (0,16)	2,69
Discussion				0,94*** (0,08)	2,56
Gain				0,79* (0,4)	2,2
Loss				1,39** (0,46)	3,01
Meme				1,48*** (0,4)	4,39
News				0,89*** (0,19)	2,44
Shitpost				1,05 (0,59)	
YOLO				1,34*** (0,33)	3,82
n	4414	4414	4414	4414	
adj. R squared	0,049	0,05	0,056	0,1	
AIC	15996,7	15983,3	15962,6	15755,3	
BIC	16015,9	16008,9	15994,6	15851,2	

Tarkastellaan seuraavaksi julkaisuiden saamien kommenttien määrää selittävää regressiomallia (Taulukko 4.). Taulukossa on esitetty selittävien muuttujien regressiokertoimet ja niiden keskivirheet. Regressiokertoimien tilastollinen merkitsevyys on esitetty ”**” symboleilla (** = 0,01; * = 0,05) Taulukon ”Muutos” sarakkeessa on esitetty regressiokertoimesta johdettu muutos selitettävässä muuttujassa joko prosentteina tai kertoimena. Karman ja keskeisyysasteen kohdalla muutosprosentti kuvaa prosentuaalista muutosta kommenttien määrässä yhden prosenttiyksikön muutosta kohden selitettävässä muuttujassa. Trophyjen kohdalla muutosprosentti puolestaan kuvaa prosentuaalista muutosta kommenttien määrässä yhtä trophyä kohden. Flairien kohdalla muutosta kuvataan kertoimena, verrattuna julkaisuihin, joilla ei ole flairia.

Kun malliin lisätään selittäviä muuttujia, mallin selitysaste paranee jatkuvasti. Selittävien muuttujien lisääminen ei muuta aikaisempien muuttujien regressiokertoimien voimakkuutta merkittävästi. Myös AIC ja BIC arvot laskevat jatkuvasti, joka viestii mallin paranemisesta (Field s.947). Näin ollen sarakkeen 4 malli selittää julkaisuiden saamien kommenttien määrän vaihtelua parhaiten.

Mallissa kaikki selittävät muuttujat ovat tilastollisesti merkitseviä, joitain flaireja lukuun ottamatta. Mallin mukaan yhden prosentin lisäys karmassa nostaa julkaisun saamien kommenttien määrää keskimäärin 0.11 %. Keskeisyysasteen kohdalla tulos on vastaavanlainen. Yhden prosentin nousu keskeisyysasteessa lisää julkaisun saamia kommentteja keskimäärin noin 0,09 %. Nämä regressiokertoimet ovat melko pieniä. On kuitenkin syytä huomata, että erityisesti julkaisun saamien kommenttien ja karman vaihteluvälit ovat todella suuria, eikä niillä ole ylärajaa. Myös keskeisyysasteen vaihteluväli on laaja. Näin isossa mittakaavassa jo pieni prosentuaalinen muutos voi olla tärkeä. Mallin selitysaste on melko hyvä – 10 %. Tämä tarkoittaa sitä, että malli pystyy selittämään 10 % julkaisun saamien kommenttien vaihtelusta (Kutner ym. s.74–75).

Karman ja keskeisyysasteen kohdalla tulokset olivat odotetunlaisia. Trophyjen kohdalla mallin tulos oli kuitenkin yllättävä. Vaikka karman ja trophyjen määrän välillä vallitsee vahva korrelaatio, on niiden vaikutus selitettävään muuttujaan eri suuntainen. Mallin mukaan käyttäjän trophyjen ja julkaisun saamien kommenttien määrän välillä vallitsee negatiivinen yhteys. Yhden trophyyn lisäys vähentää julkaisun saamia kommentteja keskimäärin 3,5 %.

Mallin mukaan osa flaireista on yhteydessä julkaisun saamien kommenttien määrään. Tulokset mukailevat hyvin pitkälti flairien saaman huomion keskiarvovertailuja taulukossa 3. Lähes kaikki flairit ennustavat selkeää nousua julkaisun saamien kommenttien määrään, verrattuna julkaisuihin, joilla ei ole flairia. Suurin regressiokerroin on ”Meme” flairilla, joka ennustaa keskimäärin 4,39 kertaista kommenttimäärää verrattuna julkaisuihin, joilla ei ole flairia. Myös flairit ”YOLO” ja ”Loss” ennustavat yli kolminkertaista kommenttimäärää verrattuna flairittomiin julkaisuihin. Kuitenkin myös informatiiviseen sisältöön viittaavat flairit ennustavat suurta kommenttimäärää. Esimerkiksi ”DD” flair ennustaa keskimäärin 2,69 ja ”News” flair ennustaa 2,44 kertaa suurempaa kommenttimäärää kuin julkaisuilla, joilla ei ole flairia.

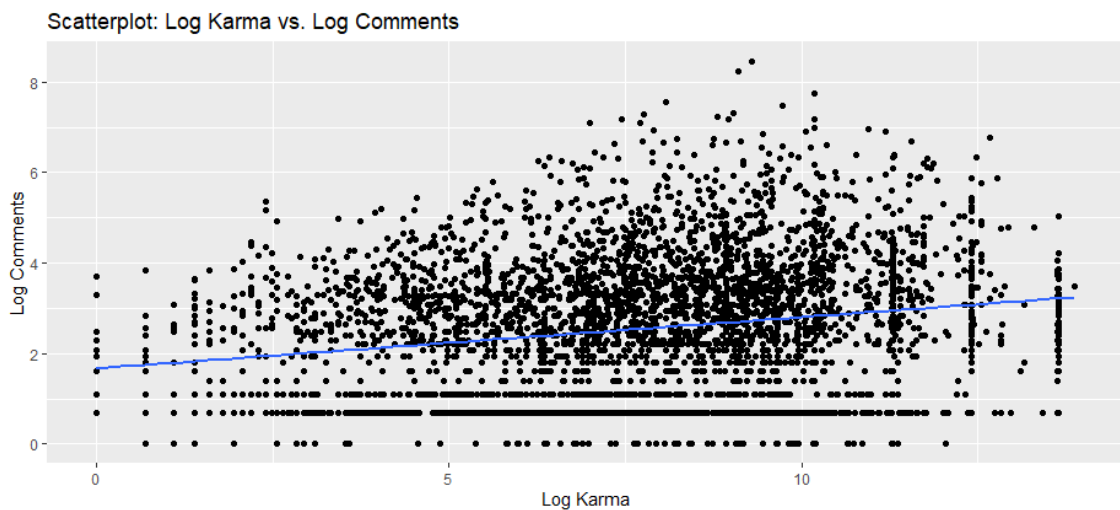
Lineaariseen regressioanalyysiin liittyy tiettyjä oletuksia, jotka takaavat mallin laadun. Ensinnäkin mallissa selittävien muuttujien on oltava lineaarisesti yhteydessä selitettävään muuttujaan. Tätä oletusta testattiin R ohjelmiston `car` paketin ”`linearHypothesis`”-funktioilla. Toiseksi regressiomallin virhetermien tulisi olla normaalisti jakautuneita ja niiden varianssin pysyä samana kaikilla selitettävän muuttujan arvoilla. Näitä oletuksia tarkasteltiin sekä visuaalisesti muuttujien välisistä hajontakuvioista, että tilastollisesti Shapiro-Wilkin ja Whiten testeillä.

Muuttujien välinen lineaarisuusolettama täyttyy ”`linearHypothesis`” testin mukaan kaikkien muuttujien kohdalla. Virhetermien normalisuutta ja variansseja koskevat tilastolliset testit eivät puolestaan menneet läpi yhdenkään muuttujan kohdalla. Virhetermien epätasaisuus, eli heteroskedastisuus on kuitenkin tyypillistä monille sosiaalitieteiden ilmiöille (Kutner ym. 2005 s.107).

Suurella otoksella pienetkin erot voivat olla tilastollisesti merkitseviä. Toisin sanoen tilastolliset testit muuttuvat herkemiksi suuremmilla otoksilla. (Field s. 89–90; 248.) Eriyisesti Shapiro Wilkin testi on hyvin herkkä vinolla aineistolla ja suurilla otoksilla (Shapiro & Wilk 1965).

Tarkastelemalla muuttujien välisiä hajontakuvioita saadaan lisätietoa virhetermien varianssien epätasaisuudesta. Kuvasta 5. nähdään, että julkaisun saamien kommenttien määrän hajonta kasvaa karman määrän kasvaessa. Tämä on merkki heteroskedastisuudesta.

Kuviota tarkastelemalla voidaan kuitenkin sanoa heteroskedastisuuden olevan näillä muuttujilla melko maltillista. Muiden muuttujien kohdalla hajontakuviosta ei löytynyt huomattavaa heteroskedastisuutta. (ks. liitteet 11, 13 ja 15.) Virhetermien normaalijakautuneisuutta voidaan tarkastella Q-Q-kuviolla (Field s.256.). Mallin Q-Q-kaaviota (liite 18) tarkastelemalla voidaan todeta, että virhetermien jakaumassa on korkeintaan pientä vinoumaa.



Kuva 5. Kommenttien ja karman hajontakuvi

Kaiken kaikkiaan mallin voidaan sanoa kuvaavan muuttujien välistä yhteyttä suuntaa antavasti. Ollakseen optimaalinen, mallin virhetermien tulisi olla tasaisia ja normaalisti jakautuneita. Näin suurella otoksella virhetermien heteroskedastisuus ja poikkeaminen normaalijakaumasta ei kuitenkaan tarkoita, että regressiokertoimet olisivat vääriä tai vinoutuneita. Ne eivät vain ole optimaalisen tarkkoja. Sen sijaan keskivirheet eivät ole virhetermien epätasaisuuden vuoksi luotettavia, joten niihin tulee suhtautua suuremmalla varauksella. (Field 2018 s.233–236, 239.) Mallin tuloksia voidaan siis pitää suuntaa antavina, mutta niitä tulee tulkita harkiten.

Tarkastellaan seuraavaksi julkaisun saamien äänten määrää selittävää regressiomallia. Mallilla pyritään selittämään äänten määrää samoilla muuttujilla, joilla edellisessä mallissa selitettiin kommenttien määrää. Taulukossa 5. on esitetty julkaisuiden saamia ääniä selittävä regressiomalli.

Taulukko 5. Julkaisun saamien äänten regressioanalyysi (Malli 2.)

	1	2	3	4	Muutos
Vakiotermi	0,37	0,61	0,62	0,56	
log Karma	0,24*** (0,009)	0,15*** (0,01)	0,14*** (0,01)	0,14 *** (0,01)	0,14 %
Trophyt		0,15*** (0,01)	0,15*** (0,01)	0,15*** (0,01)	16,20 %
log Keskeisyys			0,036 (0,02)	0,038 (0,02)	
Flair					
Chart				0,49 (0,47)	
Cosmetic Sug- gestion				-0,11 (1,77)	
DD				0,95*** (0,19)	2,59
Discussion				0,33*** (0,1)	1,39
Gain				0,89 (0,5)	
Loss				1,66** (0,56)	5,26
Meme				1,84*** (0,56)	6,3
News				1,08*** (0,24)	2,94
Shitpost				-0,76 (0,72)	
YOLO				1,19** (0,4)	3,29
n	4414	4414	4414	4414	
adj. R squared	0,13	0,16	0,16	0,18	
AIC	17834,9	17655,9	17588,4	17588,4	
BIC	17854	17681,5	17684,3	17684,3	

Julkaisun saamien äänten regressioanalyysissä ilmenee joitain poikkeamia kommenttien regressioanalyysiin verrattuna. Karman määrä säilyy yhtä voimakkaana julkaisun saamaa huomiota selittävänä muuttujana. Sen sijaan keskeisyysaste on tässä mallissa menettänyt vähäisenkin merkityksensä. Keskeisyysasteen regressiokerroin on tässä mallissa selkeästi pienempi, ja tulos on tilastollisesti merkitsevä vain 10 % tasolla. Edellisen mallin tulosten valossa tämä ei kuitenkaan tule yllätyksenä. Keskeisyysaste on muodostettu käyttäjien

toisilleen kirjoittamien kommenttien perusteella, joten oli odotettavissa, että se selittää kommenttien määrää vahvemmin, kuin äänten määrää.

Erikoisin tulos äänten regressioanalyysissä verrattuna kommenttien analyysiin ilmenee trophyjen kohdalla. Tulosten mukaan käyttäjän trophyjen määrä vaikuttaa positiivisesti julkaisun saamien äänten määrään – yksi trophy enemmän lisää julkaisun saamia ääniä keskimäärin 16,2 %. Tulos on yllättävä, sillä edellisessä mallissa trophyjen vaikutus oli negatiivinen. Muuttujien välisten korrelaatioiden perusteella olisi voinut olettaa positiivista yhteyttä molemmissa analyyseissa. Kuitenkin vaikutuksen suunnan muuttuminen on kaikkein suurin yllätys.

Flairien osalta ensimmäinen ja toinen regressiomalli ovat melko yhteneväisiä. Molemmista nousevat esiin melko pitkälti samat flairit merkityksellisinä. Myös kertoimien voimakkuudet ovat pitkälti samaa luokkaa. Joitain eroja on kuitenkin havaittavissa. Esimerkiksi ”Chart” -flair ei ole ääniä selittävässä mallissa tilastollisesti merkitsevä, ja ”Discussion” flairin vaikutus on selvästi heikompi. Tämä johtunee siitä, että ”Discussion” flairilla merkatut keskustelut lähtökohtaisesti tarkoitettu keskustelulle ja keräävät tästä syystä paljon kommentteja.

Vertaamalla ”News” flairilla merkattuja julkaisuja, jotka siis sisältävät lähtökohtaisesti kanavan aiheeseen, eli sijoittamiseen liittyviä uutisia, voimme huomata niiden keräävän lähtökohtaisesti selvästi ilman flairia olevia julkaisuja enemmän huomiota. ”News” flairilla merkatut julkaisut saavat mallien mukaan keskimäärin 2,44 kertaa enemmän kommentteja ja 2,94 kertaa enemmän ääniä, kuin ilman flairia olevat julkaisut keskimäärin. Kyseinen flair ennustaa siis hieman vahvemmin äänten määrää. Kanavalle relevanttia uutta tietoa kuten uutisia arvostetaan, joten se kerää paljon huomiota – erityisesti ääniä.

Äänten määrää selittävä regressiomalli on selitysasteeltaan hieman kommenttien mallia voimakkaampi. Mallin selitysaste on noin 18 %, eli malli pystyy selittämään 18 % äänten määrän vaihtelusta. Tällä mallilla on kuitenkin hieman enemmän heikkouksia kuin ensimmäisellä mallilla. Mallin muuttujien lineaarisuusolettama täyttyy vain karman ja trophyjen osalta. ”LinearHypothesis” testin mukaan keskeisyysasteen ja kommenttien määrän välistä yhteyttä ei voida kuvata lineaarisesti. Mikäli näin on, keskeisyysasteen tulokset tässä mallissa eivät ole luotettavia.

Mallin muuttujien välillä on heteroskedastisuutta Whiten testin mukaan, eivätkä jäännösarvot ole normaalisti jakautuneita Shapiro-Wilkin testin mukaan. Muuttujien väliset hajontakuviot (liitteet 12, 14 ja 16.) näyttävät hyvin samanlaisilta kuin kommenttienkin kohdalla. Mallin Q-Q kaaviossa (liite 19.) on hieman enemmän poikkeamaa normaalista, kuin kommenttien mallissa, mutta ei kuitenkaan hälyttävästi. Mallin 2. tulkintaa tulee lähestyä vastaavalla varauksella kuin mallin 1. tulkintaa. Tulokset ovat suuntaa antavia, mutta eivät lähtökohtaisesti vääriä tai vinoutuneita. Keskivirheet eivät kuitenkaan välttämättä ole luotettavia.

Malleissa 1 ja 2 tarkasteltiin selittävien muuttujien yhteyttä julkaisun saamiin kommentteihin ja ääniin. Tarkastellaan viimeisenä kommenttien saamien äänten määrää. Taulukossa 6. on esitetty kommenttien saamien äänten määrää selittävä regressioanalyysi. Selittävinä muuttujina on kommentin kirjoittajan karma, trophyt ja keskeisyysaste. Flairia ei käytetä tässä mallissa selittävänä muuttujana, sillä kommentteille ei voi asettaa flairia.

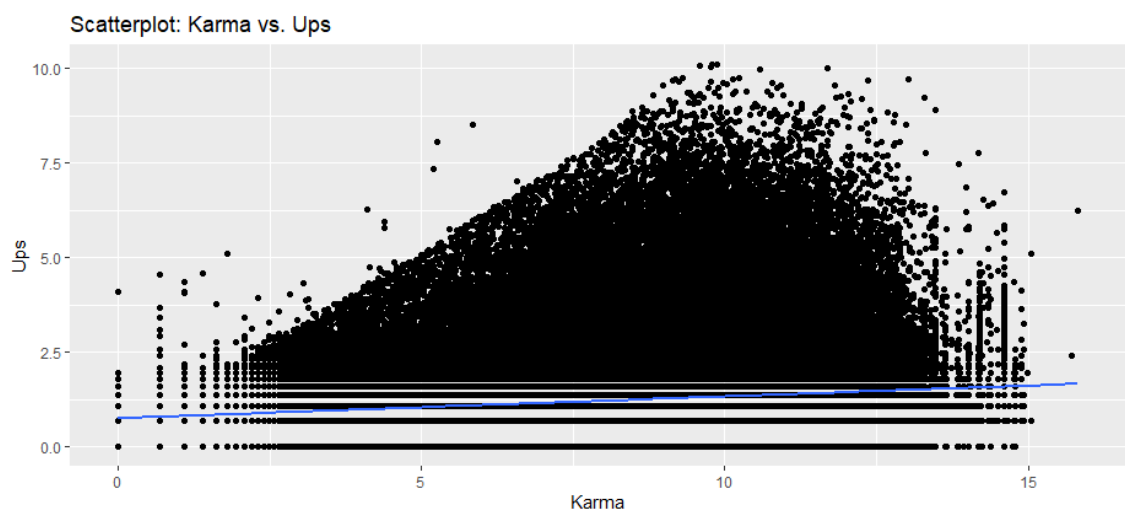
Taulukko 6. Kommenttien saamien äänten regressioanalyysi (Malli 3.)

	1	2	3	Muutos
Vakiotermi	0,76	0,77	0,78	
log Karma	0,057*** (0,000)	0,054*** (0,000)	0,054*** (0,000)	0,05 %
Trophyt		0,01*** (0,000)	0,01*** (0,000)	1,00 %
log Keskeisyys			-0,0004 (0,000)	
n	1062571	1062571	1062571	
adj. R squared	0,021	0,023	0,023	
AIC	2613529	2612724	2612726	
BIC	2613565	2612772	2612785	

Mallin 3. tuloksia tarkastelemalla havaitaan, että malli jäi selitysteeltään hyvin heikoksi. Malli pystyy selittämään vain 2,3 % kommenttien saamien äänten vaihtelusta. Yhden prosentin lisäys karmassa lisää äänten määrää keskimäärin vain noin 0,05 %. Trophyjen kohdalla yhden yksikön lisäys johtaa keskimäärin noin 1 % lisäykseen äänissä ja keskeisyysasteen merkitys on olematon.

Vaikka trophyjen regressiokerroin on mallissa korkein, on syytä huomata, että sen lisääminen malliin nostaa mallin selitystasetta vain 0,2 prosenttiyksiköllä. Lähestulkoon koko mallin selitysvaikutus tulee karmasta. Näiden tulosten perusteella voidaan sanoa kommenttien saamien äänten määrän ennustamisen olevan haastavampaa, kuin julkaisujen saaman huomion määrän ennustaminen. Jos kommenttien saamien äänten määrää kuitenkin halutaan ennustaa, tämän tiedon valossa se onnistuu parhaiten karmen avulla.

Kolmas malli kohtaa samat ongelmat, kuin toinen malli. Muuttujien lineaarisuusoletus täyttyy keskeisyysastetta lukuun ottamatta. Mallin muuttujien välillä on heteroskedastisuutta Whiten testin mukaan, eivätkä jäännösarvot ole normaalisti jakautuneita Shapiro-Wilkin testin mukaan. Muuttujien väliset hajontakuviot (liitteet 20–22) näyttävät melko samanlaisilta kuin muidenkin mallien kohdalla. Kuitenkin äänten ja karmen välinen hajontakuviokuva kertoo melko vahvasta heteroskedastisuudesta (kuva 6.). Lisäksi mallin Q-Q kaavion (liite 23) mukaan virhetermit poikkeavat normaalista. Tämä on malleista selvästi heikoin, ja sen tuloksiin tulee suhtautua suurella varauksella.



Kuva 6. Kommenttien saamien äänten ja karmen hajontakuviokuva

Kaikkien kolmen mallin eduksi voidaan todeta, että otoskoot ovat hyvin suuria, ja tämä vahvistaa mallien luotettavuutta. Esimerkiksi keskivirheet ovat monien muuttujien kohdalla lähtökohtaisesti hyvin pieniä – etenkin mallissa 3. Vaikka heteroskedastisuus ja jäännösarvojen poikkeaminen normaalijakaumasta vähentävätkin keskivirheiden luotettavuutta, monen muuttujan kohdalla keskivirheet ovat niin olemattoman pieniä, että edes niiden moninkertaistaminen ei vaikuttaisi merkittävästi malliin.

Regressioanalyysien tuloksista voidaan kaiken kaikkiaan todeta, että karman yhteys julkaisuiden ja kommenttien saamaan huomioon oli johdonmukainen kaikissa malleissa. Näiden mallien avulla karmasta voidaan muodostaa painokerroin signaalijärjestelmään. Myös flairien vaikutukset olivat johdonmukaisia ja niitäkin voidaan hyödyntää signaalijärjestelmän kehittämisessä.

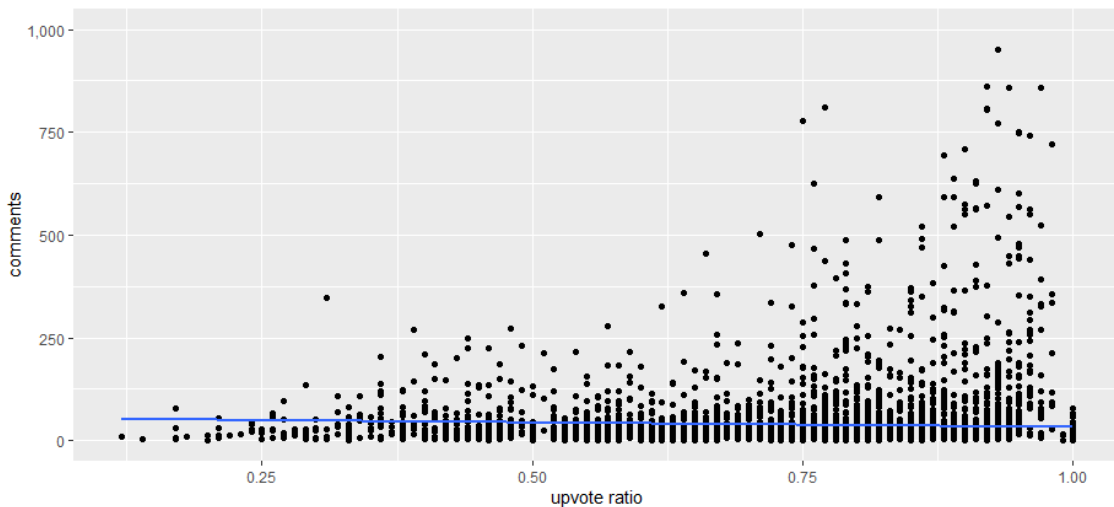
Keskeisyysasteen osalta tulokset jäävät niukoiksi. Keskeisyysasteella oli vaikutusta lähinnä kommenttien määrää tarkastelevassa mallissa. Tämä oli kuitenkin odotettavissa, sillä keskeisyysaste on muodostettu käyttäjien toisilleen kirjoittamien kommenttien pohjalta. Muissa malleissa keskeisyysaste ei selittänyt merkittävästi huomion määrää, eikä yhteyden lineaarisuusolettama mennyt läpi testissä. Näiden tulosten perusteella keskeisyysastetta ei voida hyödyntää signaalijärjestelmän kehittämisessä. Mikäli keskeisyysasteella on yhteys julkaisuiden ja kommenttien saamaan huomioon, sitä ei voida mitata lineaarisella mallilla.

Trophyjen yhteys julkaisun saamaan huomioon näyttäytyy mutkikkaana. Käyttäjän trophyjen määrä ennusti malleissa 1 ja 2 huomion määrää melko vahvasti. Kuitenkin yhteyden suunta oli mallien välillä eri. Tämän selittäminen on haasteellista. Trophyja on aineistossa havaintomäärällisesti riittävästi, joten tätä ei voida selittää puutteellisen aineiston aiheuttamalla sattumalla. Myöskään muut muuttujat eivät vaikuta trophyjen regressiokertoimiin merkittävästi. Trophyilla on positiivinen korrelaatio sekä kommenttien, että äänten kanssa, joten ristiriitaiset tulokset regressioanalyysissä yllättivät.

Yksi selitys tulokselle on se, että julkaisun kommentit ja äänet eivät mittaa täysin samaa asiaa. Positiivisten äänten voidaan olettaa lähtökohtaisesti kertovan siitä, että lukijat pitävät julkaisusta. Pidetty julkaisu kerää varmasti myös paljon kommentteja. Tämä näkyy

esimerkiksi äänten ja kommenttien välisestä korrelaatiosta ($r=0.63$). On kuitenkin mahdollista, että esimerkiksi mielipiteitä jakavat tai virheellistä informaatiota sisältävät julkaisut keräävät erityisen paljon kommentteja, kun ihmiset tulevat keskustelemaan tai korjaamaan julkaisun sisältöä sen kommentteihin. Mielipiteitä jakavat julkaisut saavat luultavasti myös paljon negatiivisia ääniä, mutta tässä tutkimuksessa käytetystä äänten määrän mittarista selviää vain nettoäänimäärä.

Aineistossa on kuitenkin saatavilla äänisuhde (upvote ratio) -mittari, joka kertoo positiivisten ja negatiivisten äänten suhteen. Ihmisten mielipiteiden jakautumista julkaisun suhteen voidaan mitata tällä mittarilla. Laskemalla julkaisun äänisuhteen ja kommenttien välinen korrelaatio huomattiin, että näiden välillä vallitsee negatiivinen yhteys ($r = -0,56$). Äänisuhteen ja kommenttien määrän välinen hajontakuvioiden esitetty kuvassa 7 (Huom. laskeva trendiviiva)



Kuva 7. Äänisuhteen ja kommenttien määrän hajontakuvioiden

Äänisuhteen ja kommenttien välisen suhteen analyysi tukee ajatusta siitä, että mielipiteitä jakavat julkaisut keräävät paljon kommentteja. Tämä voi selittää trophyjen negatiivista yhteyttä kommenttien määrään ja positiivista yhteyttä äänten määrään. Trophyja voidaan pitää osoituksena käyttäjän kokemuksesta Redditissä. Kokeneet käyttäjät luultavasti kirjoittavat laadukkaita julkaisuja, jotka harvoin jakavat mielipiteitä.

Tämä teoria on kuitenkin ristiriidassa karman ja kommenttien välisen yhteyden kanssa. Myös karmaa voidaan pitää käyttäjän kokemuksen mittarina, mutta karman yhteys kommenttien määrään on positiivinen. Tämä voi selittyä trophyjen ja karman välisillä eroavaisuuksilla. Karma muodostuu käyttäjän julkaisuilleen ja kommentteilleen saamien nettoänten summana. Trophyt taas ovat Redditin myöntämiä palkintoja ansiokkaasta toiminnasta alustalla. Esimerkiksi karman ja kommenttien määrän välisestä korrelaatiosta ($r = 0,63$) voidaan päätellä, että myös mielipiteitä jakavat julkaisut keräävät enemmän positiivisia kuin negatiivisia ääniä. Karmaa kertyy käyttäjille siis melko helposti, myös mielipiteitä jakavista julkaisuista.

Myös karman ”farmaaminen” eli kerääminen tai kartuttaminen on tunnistettu ilmiö Redditiissä. Jotkut käyttäjät keräävät tarkoituksen mukaisesti tykkäyksiä ja karmaa – internet irtopisteitä, julkaisemalla tunteisiin vetoavia tarinoita tai meemejä. (Richterich 2014.) Tällainen toiminta syö karman luotettavuutta käyttäjän kokemuksen mittarina. Trophyjen saaminen puolestaan on haastavampaa. Kenties trophyt ovatkin karmaa parempi käyttäjän kokemuksen mittari.

On myös syytä huomioida, että Redditiissä on valtava määrä eri kanavia, ja käyttäjät seuraavat yleensä useaa kanavaa. On siis mahdollista, että käyttäjät, joilla on paljon trophyja tai karmaa, ovat ansainneet ne aktiivisuudesta muilla kanavilla. Suuri trophyjen tai karman määrä ei siis automaattisesti tarkoita, että käyttäjät olisivat aktiivisia juuri r/wallstreetbets kanavalla.

6 JOHTOPÄÄTÖKSET

Tässä tutkimuksessa tutkittiin keskustelun aktiivisuutta ja huomion jakautumista sijoittamiseen liittyvällä r/wallstreetbets kanavalla. Tutkimuksen tavoitteena oli löytää muuttujia, joiden avulla voidaan ennustaa julkaisuiden saamaa huomiota. Tuloksia hyödynnetään *Developing an early signal system to identify shifts in global stock market trends* -tutkimushankkeessa kehitteillä olevan, pörssikurssien heilahtelua sosiaalisen median keskustelun avulla ennakoivan signaalijärjestelmän optimoinnissa.

Tutkimuksen teoreettinen tausta pohjautui vahvasti informaation luotettavuuden arviointiin ja siihen vaikuttaviin heuristiikkoihin. Teoriassa käsiteltiin sitä, kuinka erilaiset heuristiikat, kuten kannatusheuristiikka ja auktoriteettiheuristiikka voivat vaikuttaa informaation tulkintoihin ja päätöksentekoon. Kognitiivisen auktoriteetin teoria syvensi käsitystä auktoriteetista esittelemällä auktoriteetin muodostumisen tavat ja auktoriteettisignaalit. Nämä yksilötason tekijät sidottiin osaksi laajempaa kontekstia verkostoteoreettisella näkökulmalla, jonka avulla huomioitiin käyttäjien väliset suhteet verkostossa.

Teoreettisella taustalla pyrittiin antamaan kuva informaation luotettavuuden arvioinnin ja päätöksenteon monimutkaisesta luonteesta. Kaikkia teoriaosuudessa esitettyjä mekanismeja ei ollut tarkoituskaan tutkia tässä tutkimuksessa. Esimerkiksi kannatusheuristiikan vaikutusta ei tämän tutkimuksen asetelmalla pystytty tutkimaan. Tutkimuksen ensisijaisena tavoitteena oli etsiä muuttujia, joiden avulla pystytään ennustamaan julkaisuiden saamaa huomiota r/wallstreetbets kanavalla.

Niiden mekanismien osalta, joita tässä tutkimuksessa tutkittiin, selitysvoimaisia tuloksia saatiin ainakin karman ja joidenkin flairien osalta. Kuitenkaan kysymykseen siitä, johduko tämä selitysvoima auktoriteettisignaalien laukaisemasta heuristisesta päättelystä, julkaisun sisällön laatuun perustuvasta systemaattisesta päättelystä, vai muista mahdollisista tekijöistä, ei saatu suoria vastauksia. Tämä on kuitenkin signaalijärjestelmän kannalta yhdentekevää, sillä signaalijärjestelmän optimoinnin kannalta keskeisiä ovat säännömukaiset yhteydet.

Huomion ja aktiivisuuden jakaumien tarkastelussa osoitettiin, että sekä käyttäjien aktiivisuus, että julkaisuiden saama huomio jakautuvat pääosin pienelle joukolle. Hyvin aktiivisten ja paljon huomiota saavien käyttäjien joukot ovat osittain päällekkäisiä. Kuitenkaan samat henkilöt, jotka julkaisevat paljon, eivät suinkaan aina ole niitä, jotka saavat paljon huomiota. Kanavalla on paljon käyttäjiä, jotka ovat hyvin aktiivisia, mutta eivät saa juurikaan huomiota. Nämä käyttäjät julkaisevat innokkaasti, mutta eivät syystä tai toisesta onnistu saamaan nostetta julkaisuilleen. Tämän voidaan olettaa johtuvan heikkolaatuisesta sisällöstä. Kanavalla on myös käyttäjiä, jotka ovat saaneet paljon huomiota yksittäiselle julkaisulle, mutta eivät julkaise aktiivisesti.

Käyttäjät, jotka ovat sekä aktiivisia, että saavat säännöllisesti paljon huomiota ovat erityisessä asemassa kanavalla. Nämä käyttäjät määrittelevät pitkälti kanavalla käytävää keskustelua, sillä he ensinnäkin julkaisevat suurimman osan julkaisuista, ja toiseksi Reddit nostaa paljon huomiota saavia julkaisuja syötteellä ylöspäin. Nämä tulokset tukevat teoriaa informaation virtaamisesta sosiaalisessa mediassa mielipidevaikuttajien kautta (Choi 2015; Thorson & Wells 2016).

Tämä tutkimus antaa alustavia tuloksia mielipidevaikuttajien määrästä ja aktiivisuudesta r/wallstreetbets kanavalla. Tutkimus tuotti myös tietoa keinoista mielipidevaikuttajien tunnistamiseksi. Kahden eri menetelmän käyttäminen mielipidevaikuttajien tunnistamiseksi toi esiin menetelmien väliset eroavaisuudet, sekä niiden vahvuudet ja heikkoudet. Esimerkiksi näin vinossa aineistossa mielipidevaikuttajalistalle pääsi melko vähäisellä julkaisumäärällä. Verkostokartta ei puolestaan huomionnut käyttäjien saaman huomion määrää lainkaan.

Jatkotutkimuksessa on syytä tarkentaa sekä verkostanalyysin, että luokittelumenetelmän kriteereitä. Verkostanalyysissä voidaan esimerkiksi käyttää keskeisyysasteen lisäksi tai sijasta jotain toista mittaria. Mielipidevaikuttajien luokittelua aktiivisuuden ja huomion perusteella voitaisiin puolestaan kehittää esimerkiksi painottamalla julkaisuiden määrään merkitystä.

On myös syytä huomioida, että aineisto koostuu melko lyhyen aikavälin julkaisuista, ja kommentteista on huomioitu vain ensimmäisen tason kommentit. Luotettavimmat tulokset mielipidevaikuttajien roolista vaatisivat laajempaa aineistoa ja mahdollisesti tiettyjen toimijoiden aktiivisuuden ja näiden saaman huomion ajallista seuraamista. Myös verkostoanalyysi tarkentuisi pidemmällä tarkastelujaksolla ja syvemmästä kommenttien tasosta. Lisäksi ymmärrystä mielipidevaikuttajista olisi kiinnostavaa syventää yksilökohtaisemman verkostoanalyysin avulla. Tarkastelemalla esiin nousseiden mielipidevaikuttajien verkostoja yksitellen, voitaisiin saada arvokasta tietoa mielipidevaikuttajien roolista ja asemasta verkostossa.

Keskeisyysasteen vaikutus julkaisuiden saamaan huomioon jäi kuitenkin regressioanalyysissä niukaksi. Regressioanalyysien tulosten valossa keskeisyysastetta ei voida pitää julkaisun huomiota ennustavana tekijänä. Tämä tulos asettaa teorian mielipidevaikuttajien roolista *r/wallstreetbets* kanavalla kyseenalaiseen valoon. Kenties Redditin kaltaisella puolianonyymillä alustalla, jossa seurataan kanavia eikä käyttäjiä, käyttäjän henkilöllisyydellä ei ole yhtä suurta merkitystä, kuin käyttäjäkeskeisemmillä alustoilla, kuten X:ssä. Alustan rakenne saattaa siis ehkäistä mielipidevaikuttajien syntymistä. Tämä ei kuitenkaan suoranaisesti tarkoita, etteikö kanavalla voisi olla mielipidevaikuttajia. Tietyt nimimerkit voivat nousta tunnetuiksi kanavalla, kuten esimerkiksi GameStopin tapauksen aikana kävi.

Toisaalta ongelma voi johtua myös tämän tutkimuksen tavasta rakentaa verkosto. Verkosto rakennettiin käyttäjien toisilleen kirjoittamien kommenttien pohjalta. Mikäli paljon kommentteja keräävät julkaisut jakavat mielipiteitä, kuten analyysiosion lopussa speuloitiin, tämä heikentää keskeisyysasteen selitysvoimaa ääniin nähden. Lisäksi on syytä huomata, että lineaarinen malli ei välttämättä pysty kuvaamaan verkostoaseman ja julkaisun saaman huomion välistä yhteyttä.

Käyttäjän verkostoaseman vaikutusta julkaisuiden saamaan huomioon on kuitenkin onnistuttu mallintamaan aikaisemmissa Redditiä koskevissa tutkimuksissa (ks. Rohde ym. 2023). Tästä syystä käyttäjän verkostoaseman vaikutusta julkaisun saamaan huomioon on syytä tutkia tarkemmin jatkossa. Yhteys voidaan pyrkiä havaitsemaan esimerkiksi erilaisilla menetelmillä. Analyysimenetelmiä voidaan tarkentaa hyödyntämällä epälineaarisia

malleja, ja keskeisyysasteen lisäksi voidaan käyttää muita verkstorakenteeseen perustuvia tunnuslukuja.

Muiden muuttujien osalta regressioanalyysi tuotti sekä kelvollisia työkaluja signaalijärjestelmän kehittämiseksi, että mielenkiintoisia kysymyksiä jatkotutkimusta ajatellen. Eri-tyisesti karma tuotti johdonmukaisia tuloksia, joita voidaan hyödyntää signaalijärjestelmän kehittämisessä. Trophyjen kohdalla tulokset ovat epäselvempiä, mutta mikäli teoria mielipiteitä jakavien julkaisuiden taipumuksesta kerätä runsaasti kommentteja on oikeassa, trophyjen avulla voidaan ennustaa julkaisun saamaa äänten määrää jopa paremmin, kuin karman avulla. Tätä on kuitenkin syytä selvittää tarkemmin, ennen johtopäätösten vetämistä.

Äänisuhteen ja kommenttien välinen yhteys tuotti uutta näkemystä kommentteista luottamuksen mittarina. Pelkästään kommenttien määrä ei kerro sitä, onko huomio negatiivista vai positiivista. Mielipiteitä jakava, ristiriitainen, tai virheellinen sisältö voi herättää paljon keskustelua ja johtaa suureen kommenttien määrään. Näillä perusteilla äänten määrää voidaan pitää kommentteja parempana luottamuksen mittarina, toisin kuin Kilgo (2016) tutkimuksessaan ehdottaa. Äänten määrä ei kuitenkaan ole täydellinen luottamuksen mittari, sillä nettoäänimäärä ei huomioi negatiivisten äänten määrää. Näitä mittareita voitaisiin jatkossa tarkentaa huomioimalla keskustelun sisältö ja negatiivisten äänten määrä.

Flair analyysit tuottivat mielenkiintoisia tuloksia, joiden avulla voidaan tehdä myös päätelmiä siitä, minkälaista sisältöä kanavalla arvostetaan. Aikaisemmissa Redditiä käsittelevissä tutkimuksissa meemijulkaisut eivät ole yleensä saaneet erityisen paljon huomiota, koska nämä eivät tuota kanavalle sisällöllistä lisäarvoa (Horne ym. 2017). Tämän tutkimuksen flair analyysin tulokset ovat toisenlaisia. Esimerkiksi ”Meme”, ”Loss” ja ”YOLO” -flairien esiin nouseminen molemmissa malleissa kertoo tarinaa kanavalla vallitsevasta riskinottoa ihannoivasta kulttuurista, joka on nostettu esiin myös aikaisemmassa tutkimuksessa. Niin riskisijoitusten, kuin tappioidenkin tekemisestä palkitaan äänillä. Kanavalla vallitsee voimakas omanlaisensa kulttuuri, jossa keskustelua väritetään sisäpiirin meemeillä, ja jonka ilmentämisestä palkitaan äänillä ja kommentteilla. (Quinn 2021.)

Kuitenkin r/wallstreetbets kanavalla osataan arvostaa myös vakavamielistä sisältöä, joka huomataan esimerkiksi ”News” ja ”DD” flairilla merkattujen julkaisuiden suosiosta. Tämä tulos on linjassa aikaisemman tutkimuskirjallisuuden kanssa, jonka mukaan erityisesti julkaisut, jotka on merkattu auktoriteettia ilmentävällä flairilla, saavat paljon ääniä (Horne ym. 2017). Näitä tuloksia voidaan selittää kahdella tavalla. Flairit voivat toimia auktoriteettisignaalina ja nostaa itsessään julkaisun uskottavuutta, joka johtaa korkeampaan huomion määrään (ks. Wilson 1983 s.168–169). Voi myös olla, että käyttäjät, jotka käyttävät flaireja, julkaisevat keskimäärin parempia julkaisuja. Nämä käyttäjät voivat esimerkiksi olla kokeneempia kanavalla ja tuntea kanavan kulttuurin paremmin. Näitä tuloksia pystytään tarkentamaan tulevaisuudessa, sillä r/wallstreetbets vaatii nykyisin flairin asettamista kaikille julkaisuille.

Flairien tulokset antavat aihetta julkaisuiden sisällönanalyysille tulevaisuudessa. Laadukkaampia julkaisuja pidetään luultavasti luotettavampina ja ne keräävät enemmän huomiota. Analysoimalla kanavalla käytävää keskustelua voitaisiin kehittää signaalijärjestelmää siten, että se huomioisi myös julkaisun sisällön ja sentimentin. Flair -analyysin tulosten ja aikaisemman kirjallisuuden valossa on syytä olettaa julkaisun sisällöllä olevan vaikutusta sen uskottavuuteen (Horne ym. 2017). Uskon sisällön olevan hyvin merkityksellinen julkaisun saamaa huomiota ennustava tekijä.

Erityisesti julkaisun sentimentin huomioiminen olisi pörssikurssien ennustamisen kannalta tärkeää. Tällä hetkellä järjestelmä ei huomioi lainkaan keskustelun laatua, vain sen määrän. On kuitenkin selvää, että keskustelun laatu on keskeinen tekijä tässä yhteydessä. Keskustelun ja huomion määrä ei nimittäin kerro siitä, odottavatko kanavan sijoittajat kurssin nousua vai laskua. Sentimenttianalyysi täyttäisi tämän puutteen. Lopullista signaalijärjestelmää voidaan hienosäätää loputtomilla muuttujilla. Uskon kuitenkin julkaisun sisällön olevan erityisen tärkeä muuttuja julkaisun uskottavuuden kannalta, jonka huomioiminen on ensisijaista signaalijärjestelmän kannalta.

Tässä tutkimuksessa huomio on keskittynyt tekijöihin, jotka tekevät julkaisuista uskottavia ja herättävät ihmisissä luottamusta. Tutkimuksen ulkopuolelle on kuitenkin jäänyt keskeinen osa koko järjestelmän ajatusta – se kuinka keskustelu todellisuudessa vaikuttaa pörssimarkkinoilla. Tässä tutkimuksessa ei pystytä vastaamaan tähän kysymykseen, eikä varsinaisesti ole tarkoituskaan. Tähän kysymykseen voitaisiin jatkossa etsiä vastauksia

esimerkiksi poimimalla aineistosta pieniä tapaustutkimuksia. Aineistosta voitaisiin etsiä piikkejä keskustelumäärässä tai tietyn osakkeen maininnoissa, ja verrata niitä pörssikursien kehitykseen. Näin voitaisiin syventää tietämystä keskustelun vaikutuksista pörssikursseihin.

Sijoitusstrategioita on monenlaisia ja sijoituspäätöksiä voidaan tehdä erilaisin perustein. Wallstreetbets kanavan keskustelu on omistettu pääasiassa lyhyellä aikavälillä tapahtuvalle keinottelulle ja kanavalla on esimerkiksi linkkejä lyhyen aikavälin kaupankäyntiä käsitteleviin oppaisiin (r/wallstreetbets faq 2023).¹⁶ Tämä voi omalta osaltaan selittää riskejä ihannoivaa kulttuuria ja tämän perusteella voidaan myös olettaa, että aiheet, jotka käsittelevät tällaista treidaamista (day trading) keräävät kanavalla enemmän huomiota, kuin pitkän aikavälin strategiat.

Kanavan kiinnostusta nopeatempoiseen sijoittamiseen voidaan pitää signaalijärjestelmän kehittämisen kannalta hyvänä asiana. Pitkällä aikavälillä pörssikursseihin vaikuttavat useat tekijät, kuten markkinoiden kehitys ja erilaiset kriisit, joiden ennustaminen on vaikeaa. Sosiaalisessa mediassa käytävä keskustelu vaikuttaa tällaisen pitkän aikavälin ennusteessa melko mitättömältä. Lyhyen aikavälin heilahtelun ennustaminen sosiaalisen median keskusteluiden avulla sen sijaan tuntuu mielekkäämmältä. Lisäksi vastaavanlaiset signaalijärjestelmät ovat onnistuneet ennustamaan kurssiheilahtelua nimenomaan lyhyellä aikavälillä (Pagolu ym. 2016; Huang 2021).

¹⁶ <https://www.reddit.com/r/wallstreetbets/wiki/faq/>

LÄHDELUETTELO

abc News 13.2.2021 GameStop timeline: a closer look at the saga that upended Wall Street <<https://abcnews.go.com/Business/gamestop-timeline-closer-saga-upended-wall-street/story?id=75617315>> haettu 12.5.2023

Ackland, Robert (2013) *Web social science, Concept, data and tools for social scientists in the digital age*. SAGE Publications, London

Anderson, Katie (2015) Ask me anything: what is Reddit? *Library high tech news* Vol. 32 (5), 8-11.

Baker, H.K. – Ricciardi, V. (2014). *Investor Behavior: The Psychology of Financial Planning and Investing*. Wiley & Sons, New Jersey

Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics*. Vol. 22 (1), 23–36. London

Bodie, Z. – Merton, R.C. (2000) *Finance*. 7.p. Prentice Hall, New Jersey

Borah, P. – Xiao, X. (2018) The importance of likes: the interplay of message framing, source and social endorsement on credibility perceptions of health information in social media. *Journal of Health Communication*, Vol. 23 (4), 399– 411.

Boudon (1998) Limitations of Rational Choice Theory. *American Journal of Sociology* Vol. 104 (3), 817-828.

Buntain, C. – Goldbeck, J. (2014) Identifying social roles in reddit using network structure. Teoksessa: *Proceedings of the 23rd International Conference on World Wide Web*. 615-620. Association for Computing Machinery, New York

Business Insider 28.1.2021 The history of WallStreetBets, the Reddit group that upended the stock market with a campaign to boost GameStop. <<https://www.insider.com/wallstreetbets-reddit-history-gme-gamestop-stock-dow-futures-yolo-2021-1>>

Haettu 22.4.2023.

Chaiken, S. (1980) Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. *Journal of Personality and Social Psychology* Vol. 39 (5), 752-766.

Choi, S. (2015) The two-step flow of communication in Twitter based public forums. *Social science computer review*, Vol. 33 (6), 696-711.

Eichengreen, B. – Mathieson D. (1999) Hedge funds, what do we really know? *Economic issues*. Vol. 19, International monetary fund.

Eurobarometri (2022) Media & News survey 2022 {Raportti].
<https://europa.eu/eurobarometer/surveys/detail/2832> Haettu 3.10.2023.

Fernando, J. (2023) What is 410(k) and how does it work? *Investopedia* [Opas].
<<https://www.investopedia.com/terms/1/401kplan.asp>>, Haettu 12.5.2023.

Forbes, W. (2009) *Behavioural Finance*. John Wiley & Sons, New York.

Pew Research Center (2022) News platform fact sheet.
<<https://www.pewresearch.org/journalism/fact-sheet/news-platform-fact-sheet/>>, Haettu 19.11.2023.

Fox, J. – Weisberg, S. (2011) *An R companion to applied regression* SAGE Publications, Thousand Oaks, California.

Fritch, J. – Cromwell, R. (2001) Evaluating internet resources: indentity, affiliation and cognitive authority in a networked world. *Journal of the American Society for Information Science and Technology*, Vol. 52 (6), 499-507.

- Gaziano, C. – McGrath, K. (1986) Measuring the concept of credibility. *Journalism & Mass Communication Quarterly*, Vol 63 (3), 451-462.
- Golan, G. (2010) New perspectives on media credibility research. *American Behavioral Scientist*, Vol. 54 (1), 3-7.
- Golbeck J. (2013) *Analyzing the Social Web*. Elsevier, Waltham USA.
- Granovetter, M. (1985) Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology* Vol. 91 (3), 481-510.
- Hartzell, A. – Jackson, B. – Pinchot, J. (2021) Source credibility in social media: a case study of a Reddit community. *Issues in Information Systems*, Vol. 22 (3), 212-223.
- Ho, T. & Huang, Y. (2021) Stock price movement prediction using sentiment analysis and CandleStick chart representation. *Sensors*, Vol. 21 (23).
- Holmberg, K. (2009) *Webometric network analysis, Mapping cooperation and geopolitical connections between local government administration on the web*. Åbo Akademi University Press, Turku.
- Horne, B. – Adali, S. – Sikdar, S. (2017) Identifying social signals that drive online discussions: a case study of Reddit communities. Teoksessa: *26th International Conference on Computer Communication and Networks*, 1-9. Vancouver.
- Kilgo, D. – Yoo, J. – Sinta, V. – Geise, S. – Suran, M. – Johnson, T. (2016) Ledit on Reddit: an exploratory study examining opinion leadership on Reddit. *First Monday*, Vol. 21 (9).
- Kitchin, R. (2014) Big Data, New epistemologies and paradigm shifts. *Big Data & Society*, 1 (1).

Konstantinou, P. – Georgiou, K. – Kumar, N. – Kyprianidou, M. – Nicolaides, C. – Karakla, M. – Kassianos, A. (2021) Transmission of Vaccination Attitudes and Uptake Based on Social Contagion Theory: A Scoping Review. *Vaccines*, Vol. 9 (6), 607.

Kumar, A. (2015) *Short Selling: Finding Uncommon Short Ideas*. New York: Columbia Business School Publishing, New York.

Lim, M. (2015) *The Handbook of Technical Analysis + Test Bank: The Practitioner's Comprehensive Guide to Technical Analysis*. John Wiley & Sons, Newark.

Lin, X. – Spence, P. – Lachlan, K. (2016) Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior*, Vol. 63, 264-271.

Luo, M. – Hancock, J. – Markowitz, D. (2022) Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication research*, Vol. 49 (2), 171-195.

Lye, J. - Hirschberg, J. (2002) *Tests of inference for dummy variables in regressions with logarithmic transformed dependent variables*. Department of Economics, University of Melbourne, Australia.

Mason, W. – Conrey, F. – Smith, E. (2007) Situating Social Influence Processes: Dynamic, Multidirectional Flows of Influence Within Social Networks. *Personality and Social Psychology Review*, Vol. 11 (3), 279-300.

Metzger, M. – Flanagin, A. – Medders, R. (2010) Social and heuristic approaches to credibility evaluation online *Journal of communication* Vol. 60 (3), 413-439.

Metzger, M. – Flanagin, A. (2013) Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics* Vol. 59 part B, 210–220.

Nagel, E., van der – Firth, J. (2015) Anonymity, pseudonymity, and the agency of online identity: Examining the social practices of r/Gonewild. *First Monday* Vol. 20 (3).

Nummenmaa L. (2021) *Tilastotieteen käsikirja*. Tammi, Helsinki.

Pagolu, V. – Reddy, K. – Panda, G. – Majhi B. (2016) Sentiment analysis of Twitter data for predicting stock market movements. Teoksessa: *International Conference on Signal Processing, Communication, Power and Embedded System*. Paralakhemundi, India 2016. 1345–1350.

Quinn, M. (2021) Waiting for the squeeze to "squooze": wallstreetbets and general semantics: a review of general semantics. *et Cetera*. Vol. 78 (1), 133–150.

Reddit 101 (2017) https://www.reddit.com/wiki/reddit_101/, Haettu 10.5.2023.

Reddit (2023) <https://www.redditinc.com/>, Haettu 23.9.2023.

Reddit Help (2023a) <https://support.reddithelp.com/hc/en-us/articles/7419626610708-How-does-voting-work-on-Reddit->, Haettu 10.5.2023.

Reddit help (2023b) <https://support.reddithelp.com/hc/en-us/articles/204511829-What-is-karma->, Haettu 12.5.2023.

Reddit Help (2023c) <https://support.reddithelp.com/hc/en-us/articles/204533569-What-are-communities-or-subreddits->, Haettu 17.4.2023.

Reddit Help (2024) <https://support.reddithelp.com/hc/en-us/articles/15484503095060-User-Flair>, Haettu 15.1.2024.

Reddit Wiki (2022) <https://www.reddit.com/wiki/trophies/> Haettu 24.10.2023.

Reuters 18.2.2021 Explainer: How were more than 100% of GameStop's shares shorted? <https://www.reuters.com/article/us-retail-trading-shortselling-explainer-idUSKBN2AI2DD>, Haettu 10.5.2023.

Reuters 27.1.2022 Court dismisses claims Robinhood wrongly restricted “meme stock” trades. <https://www.reuters.com/business/court-dismisses-claims-robinhood-wrongly-restricted-meme-stock-trades-2022-01-27/>, Haettu 1.11.2023.

Reuters 2.2.2022 A year on, GameStop champion Roaring Kitty is quiet – yet much richer. Reuters <https://www.reuters.com/technology/year-gamestop-champion-roaring-kitty-is-quiet-yet-much-richer-2022-02-02/>, Haettu 10.5.2023.

Richterich, A. (2014) “Karma, Precious Karma” Karmawhoring on Reddit and the Front Page’s Econometrisation. *Journal of Peer Production*, Vol. 4 (1), 1–12.

Rieh, S. (2002) Judgement of information quality and cognitive authority in the web *Journal of the American society for information quality and technology*, Vol. 53 (2), 145–161.

Rohde, J. – Liu, J. – Rees, V. (2023) Community and opinion leadership effects on vaping discourse: a network analysis of online reddit threads. *Journal of Health Communication*, Vol. 28 (8), 487–497.

R-project: What is R? (haettu 8.1.2024) <https://www.r-project.org/about.html>

Ruan, Y. – Durrezi, A. – Alfantoukh, L. (2018) Using Twitter trust network for stock market analysis. *Knowledge-based systems*, Vol. 145, 207–218.

r/wallstreetbets content guide (2023) <https://www.reddit.com/r/wallstreetbets/wiki/contentguide/> haettu 10.4.2023.

r/wallstreetbets faq (2023) <https://www.reddit.com/r/wallstreetbets/wiki/faq/> Haettu 15.2.2024.

r/wallstreetbets flair quidelines (2021) <https://www.reddit.com/r/wallstreetbets/wiki/faq/> Haettu 31.1.2024

Scherer, C. – Scho, H. (2003) A Social Network Contagion Theory of Risk Perception. *Risk Analysis*, Vol. 23 (2), 261–267.

Shapiro, S. – Wilk, M. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, Vol. 52 (3/4), 591–611.

Snee, H. – Hine, C. – Morey, Y. – Roberts, S. – Watson, H. (2016) *Digital methods for social science. An interdisciplinary guide to research innovation*, Palgrave Macmillan, London.

Sterret, D. – Malato, D. – Benz, J. – Kantor, L. – Tompson, T. – Rosenstiel, T. – Sonderman, J. – Loker, K. (2019) Who shared it? Deciding on what news to trust on social media. *Digital Journalism*, Vol. 7 (6) 783–801.

The Guardian 28.1.2021 GameStop: how Reddit amateurs took aim at Wall Street's short-sellers. <https://www.theguardian.com/business/2021/jan/28/gamestop-how-red-dits-amateurs-tripped-wall-streets-short-sellers>, Haettu 10.5.2023.

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8 (1), 505-514.

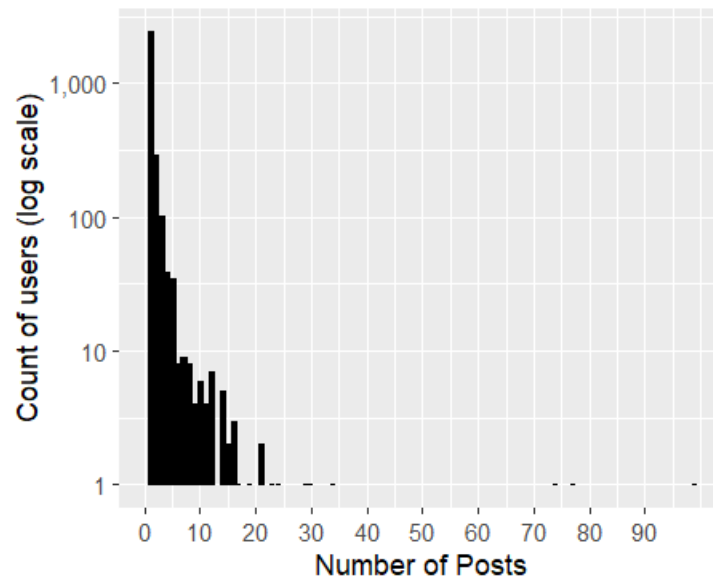
Turcotte, J. – York, C. – Irving, J. – Scholl, R. – Pingree, R. (2015) News recommendations from social media opinion leaders: Effects on media trust and information seeking. *Journal of computer-mediated communication*, Vol. 20 (5), 520-535.

Umar, Z. – Gubareva, M. – Yousaf, I. – Ali, S. (2021) A tale of company fundamentals vs. sentiment driven pricing: The case of GameStop. *Journal of Behavioral and Experimental Finance*, Vol.30.

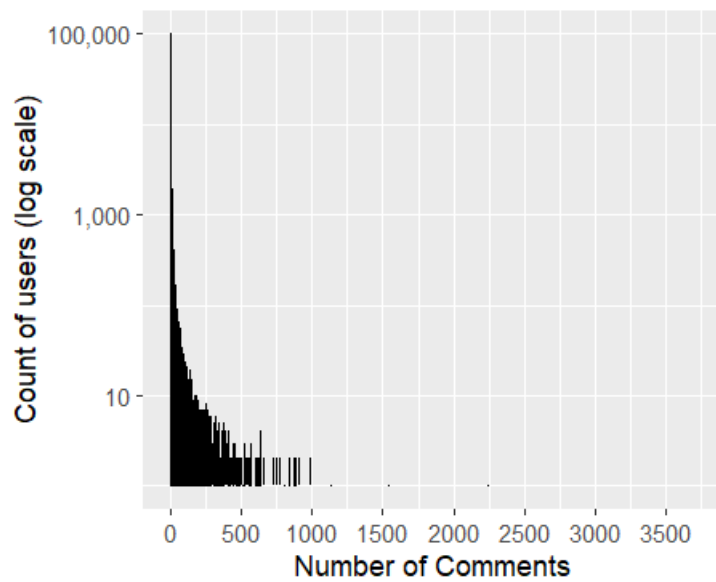
Wilson, P. (1983) *Secondhand knowledge: An inquiry to cognitive authority*. Greenwood Press, Connecticut.

R-Koodi: <https://github.com/ARthesisdump/Memestonks-code.git>

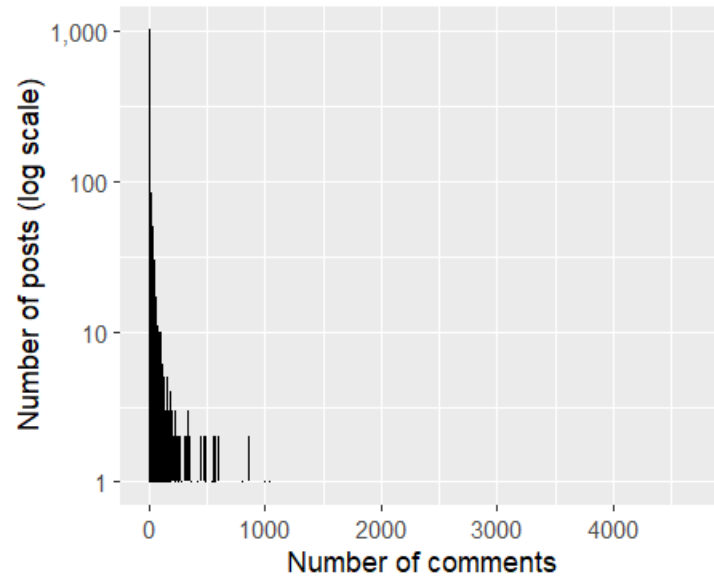
7 LIITTEET



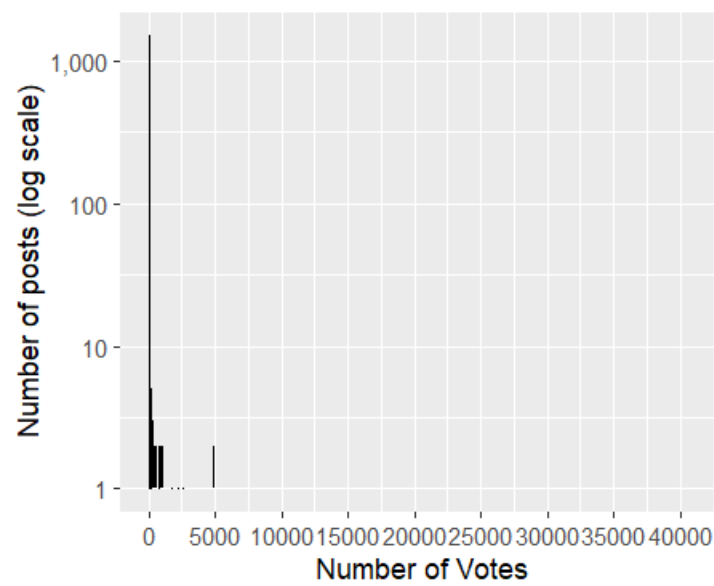
Liite 1. Julkaisuiden jakauma (logaritmuunnettu).



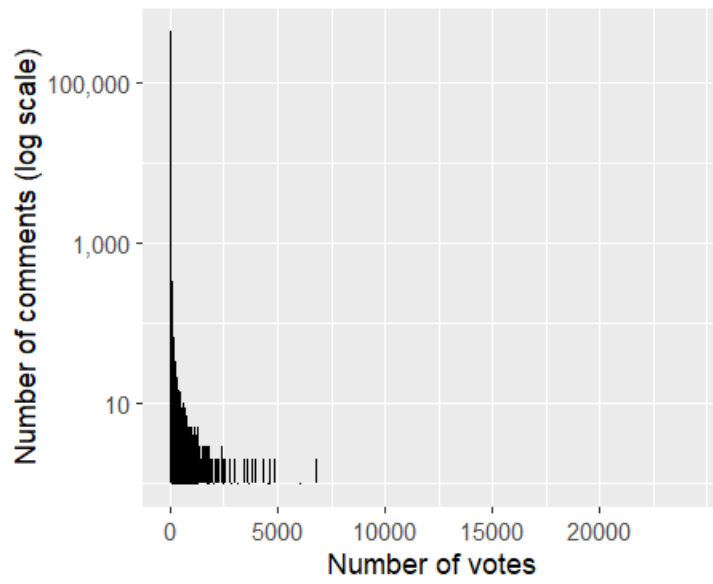
Liite 2. Kommenttien jakauma (logaritmuunnettu).



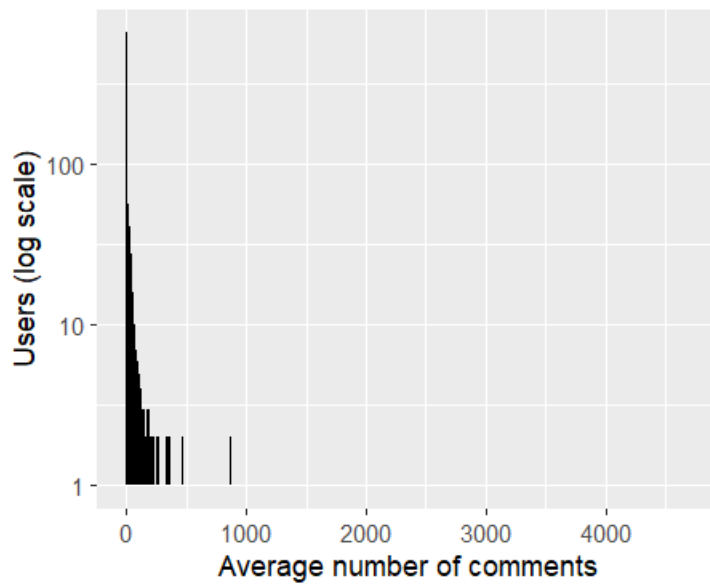
Liite 3. Julkaisuiden saamat kommentit (logaritmuunnettu).



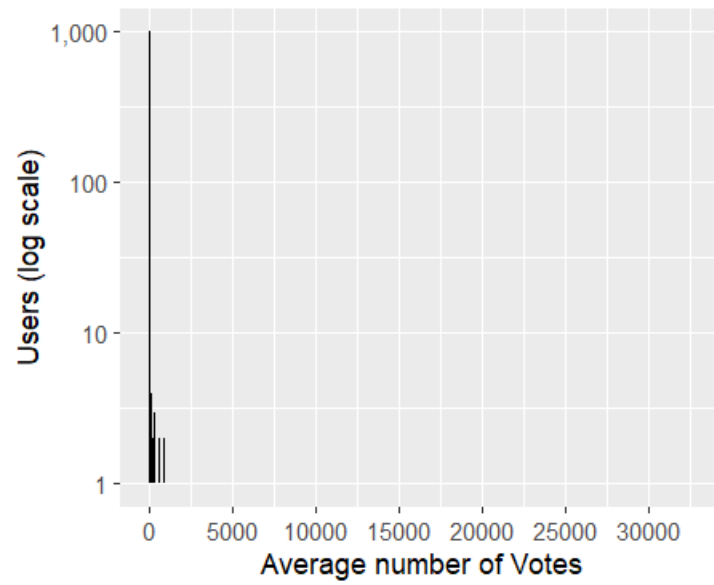
Liite 4. Julkaisuiden saamat äänet (logaritmuunnettu).



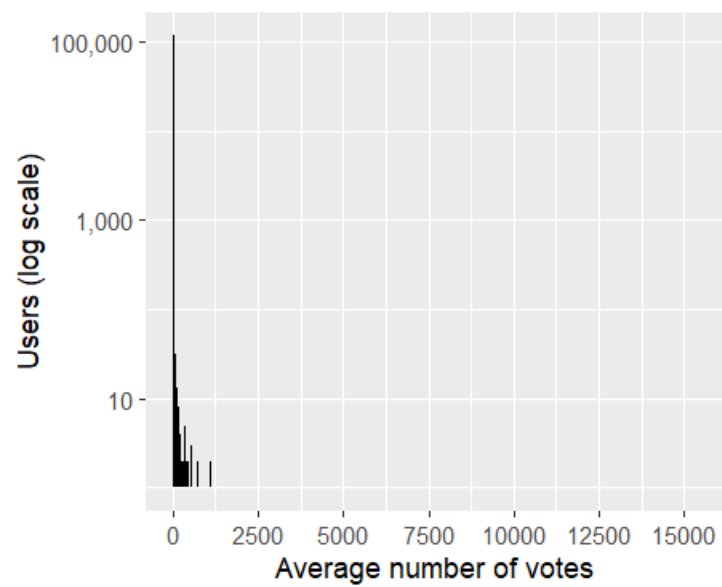
Liite 6. Kommenttien saamat äänet (logaritmuunnettu)



Liite 5. Julkaisun saamien kommenttien keskiarvo per käyttäjä (logaritmuunnettu).



Liite 8. Julkaisuiden saamien äänten keskiarvo per käyttäjä (logaritmuunnettu).



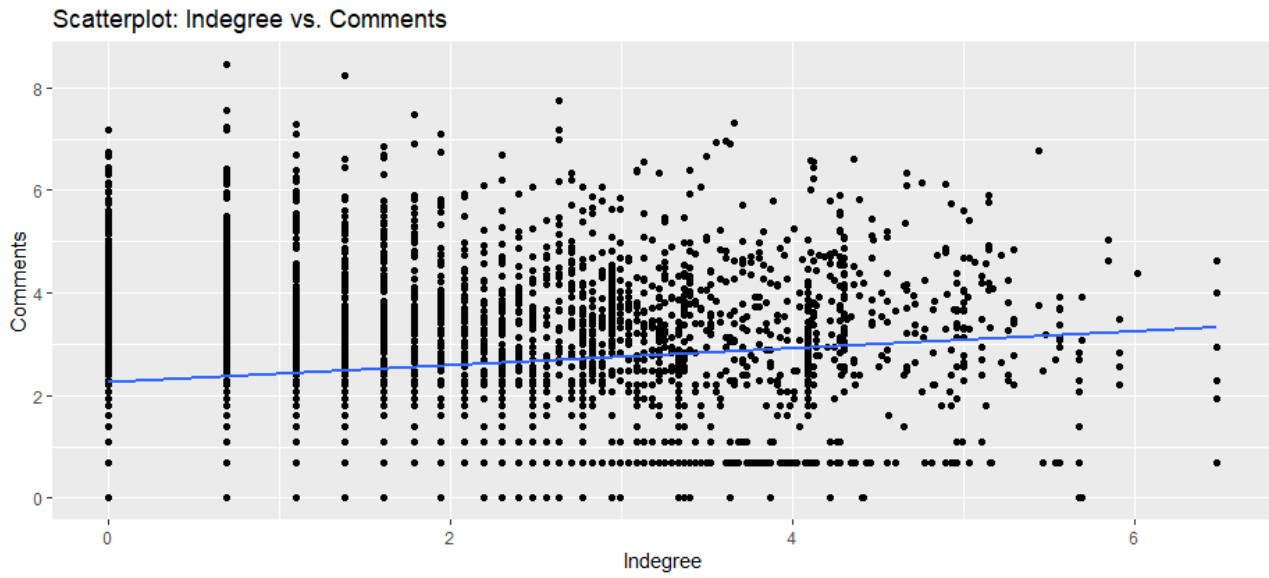
Liite 7. Kommenttien saamien äänten keskiarvo per käyttäjä (logaritmuunnettu)

id	julkaisut	kommentit	äännet
11vmg6	2	41,5	101,5
125ifc	10	77,9	112,5
129nf7	21	105,6	124,7
12rrgp	2	106,0	94,5
150mdh	3	160,3	78,3
15tehi	2	206,0	138,5
16kfrn	3	281,3	741,7
24hau26v	23	12,2	315,8
2r43g85r	7	85,6	92,3
2stk1qi5	2	159,5	209,0
3b8zco7d	2	376,5	933,0
3ca76qre	4	141,8	94,8
3dmy4	3	40,3	120,7
3m40tzn	2	242,0	246,5
3onhft77	11	119,3	33,9
3q8rx4cb	2	116,5	119,5
3uoda4sl	2	284,5	260,0
4aw9ufsy	2	215,5	170,0
4c29db0r	5	59,6	126,6
4uipp4ml	5	69,0	215,0
539j447i	3	28,7	482,0
55jujh3y	4	296,0	403,8
5b2hm5wq	5	99,2	52,6
5o77ycg9	2	23,0	121,5
5wukhd4	5	58,4	93,6
6i71rn8p	4	97,0	393,8
6ja87	4	162,3	1111,8
73sj63ly	5	140,2	103,6
77x142q5	77	14,4	6181,4
856dwna5	4	202,8	289,5
862rknk8	3	149,7	434,3
885vla44	3	286,3	1960,7
9fu8navig	3	227,7	197,3
9hmvwt6z	3	93,3	0,7
9kwzfnnu	4	84,3	175,0
a0fr0	8	83,9	146,6
a22t0bi5	9	90,2	128,8
agilmore1080	8	83,9	146,6
amboe4pe	16	19,4	16371,9
arhvk	3	35,7	7276,7
cfuqms2x	3	15,7	158,3
d0udm	4	45,0	118,5
ejyo3h15	5	36,4	162,0
p4e6n96f	8	46,8	91,3
rp6jwwox	15	401,2	1944,1
rz51bj4s	3	106,0	81,0
s3vnr6	3	165,0	695,0
s74nxv3i	7	166,4	420,4
s9lvksyl	6	122,0	29,5
sjvka6xl	4	170,3	63,8
soff1	8	72,3	167,0
t7y1kk0y	7	242,6	2396,3
uhuv68uy	5	175,8	130,2
xmsz4	3	112,7	20,7
zdt2p1m	10	74,3	106,8

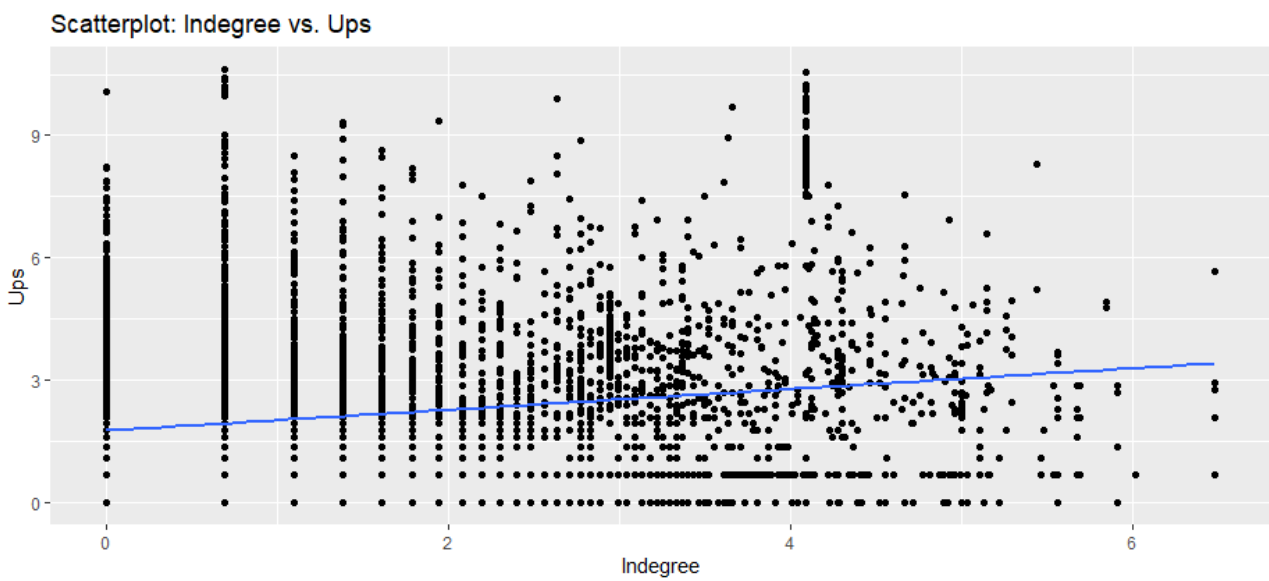
Liite 9. Mielipidevaikuttajat julkaisemisen osalta

id	comments	upvotes	id	comments	upvotes	id	comments	upvotes
11m7gy66	14	434,642857	64zrpq6h	9	299,888889	bcmql	22	373,772727
12jhcc	19	156,684211	68lez5d6	10	307,6	bek16g06	13	317
13wwin	34	429,382353	6bd6zdl2	39	157,487179	bmrhswhb	11	257,454545
167khv	27	419,481481	6by6zct7	10	284,7	br9hoqbj	49	274,857143
1ahpibjx	12	229,583333	6gmuolk7	8	203,625	bxahk	21	170,857143
1bis0pk5	8	183,125	6igz0r65	8	155,75	byfrf2yz	10	186,5
1jt5x7ob	28	234,035714	6isiupwe	19	161,578947	bztwavn2	9	1917,333333
1l7775h	11	469,909091	6ky32	9	144,111111	c2x9vm5s	23	234,043478
1mxxdxzk	8	147,75	6mbj67m7	23	585,73913	c5vm3u65	25	484,12
1nrzpnue	9	715,888889	6ylnrthh	8	636,75	c7jhh125	12	418,833333
1rnruvrzo	9	292,333333	73q3ibn8	9	1202,11111	cg8t6dd7	9	4337,11111
1sdx7y5	10	171,1	79vllawg	10	269	cj33car	68	157,323529
1t1g6zgn	17	181	7d88fjfm	12	171	d4enkp8v	13	161,692308
1v5r054	8	758	7dn5q	11	219,818182	d6qm99yc	13	422,538462
237w17q5	9	430,444444	7enfrxjd	10	420,9	d70aknz5	11	873,636364
25rdij0h	14	410,571429	7k1hq	20	165,75	d7fxpk0	26	141,538462
2ahe40fq	18	249,777778	7upibdp8	23	244,869565	d9for4ua	14	141,5
2byiwfbh	30	491,2	87t7uvr	23	180,173913	ejav07mn	19	183,578947
2eg3hfrm	11	223,363636	882jpvki	31	158,645161	fzdcxd6i	12	141,5
2fy5dg63	46	143	88xq2mzo	31	162,870968	good-crypto-d	38	147,657895
2gi79tmp	12	459,75	8b5o2pud	16	331,6875	gtfulkd3	11	487,272727
2hg9dp43	26	140,115385	8clg17hf	31	182,580645	i45hf	31	189,741935
2mnlw3b2	33	188,121212	8ijy7xc5	19	171,947368	ldidnttrythisy	12	298,333333
2vt4a4vf	9	263	8kbv7s6x	18	158,944444	idvmo	25	259,04
35ky5dgo	27	171,37037	8kucckl2	19	157,789474	iw9usbj9	12	398,833333
38b3xa1w	12	298,333333	8li864qj	16	405,1875	jf217gq	26	277,730769
3cbpe	31	173,483871	8xsice7i	44	154	ig8u2ki	11	152,636364
3gvt84ly	8	390,75	8yptksvl	13	259,538462	jinkl	45	249,177778
3i9i9om	8	207,625	91iq4pfl	28	143,5	jnhk6rug	17	212,882353
3izs0m8	12	240,083333	93mnwd8r	9	147,111111	k12pvj31	24	180,708333
3opr44h7	13	160,076923	99mmvzid	44	146,795455	k87o8epu	12	451,583333
3seeo5y	139	196,935252	9h9317gw	18	144,111111	k9ml2fa7	10	212,6
3v3586f3	12	1021,16667	9jf37reb	14	144,857143	kedop	9	177
3vf9huab	11	651,181818	9js0dd4x	50	371,84	kf15do0h	33	661,121212
3znuk6wj	9	505,666667	9p2bdv9s	10	221,6	kfln2eo5	9	144,444444
428yyhqx	16	220,1875	9qovwnqh	12	234,25	kq6q8	12	572,333333
43grc6p0	40	267,375	9u1z1lty	13	153,153846	kvdnqeo7	16	432,375
4aowp	12	1225,58333	9ydwrluu	9	176,444444	l0guufun	118	161,330508
4gpzxhs1	9	228,888889	9zns5y56	35	454,657143	l1kuf	17	536,705882
4hzb4d7o	41	382,243902	a0pvqgbp	13	319,461538	mhod2fta	35	184,742857
4wno3ci5	10	709,5	a0rybxa2	20	162,85	n0a6px9r	10	220,4
4xwhi967	11	144,272727	a1baojd8	13	364,307692	oq4hlmrh	11	348,636364
4y17br	15	145,666667	a1k90hio	43	322,837209	p7l0fk7s	10	493,3
4ya2260d	39	210,051282	a1pn6wm5	32	157,125	pobvjomc	10	384
51x96b0v	9	162,555556	a1s53t4y	22	627,181818	qf4g092v	44	242,727273
52bg2hoq	8	399,875	a1shepgg	9	345	sblrg6oh	13	300,307692
52hus32d	35	146,371429	a2l2szpp	14	242	syiitxlj	14	195,928571
55okfyrj	11	139,818182	a2r73rjh	14	207,071429	t1cago93	30	779,066667
56kpk	14	278,071429	a2y7v9ww	20	212,05	tdkmlmri	38	154,394737
5f1gs54n	12	191,916667	a33k9isb	28	237,928571	tfdnlmwd	15	151,133333
5gorkkt7	56	162,232143	a5t9oqfq	120	215,633333	tgm26ev7	11	327,363636
5jdvao0m	23	203,913043	a7f3tzt5	10	167,3	tqdnm8b7	13	192
5nf3fk8y	20	156,45	a7kivb45	9	168,222222	tvdu9z0h	87	143,126437
5xbvc	14	158,928571	ag1nfa3l	12	193,333333	txas2cme	11	295
5xkvvn0b	8	178,5	b0nrg6o6	56	147,642857	ultndjx7	10	175,8
6025641c	8	407,875	b0r8l	11	624,272727	v7y1thyq	15	223,6
61jlrnst	8	276,5	b4em2crb	23	203,826087	vg21i	26	187,961538
63lain6n	11	195,181818	b509l	10	141,6	vow523w	14	380,214286
bax33dp2	38	320,684211	bafangoolNJ	16	331,6875	Xvegantourde	15	187,666667
z2pzb	25	249,88						

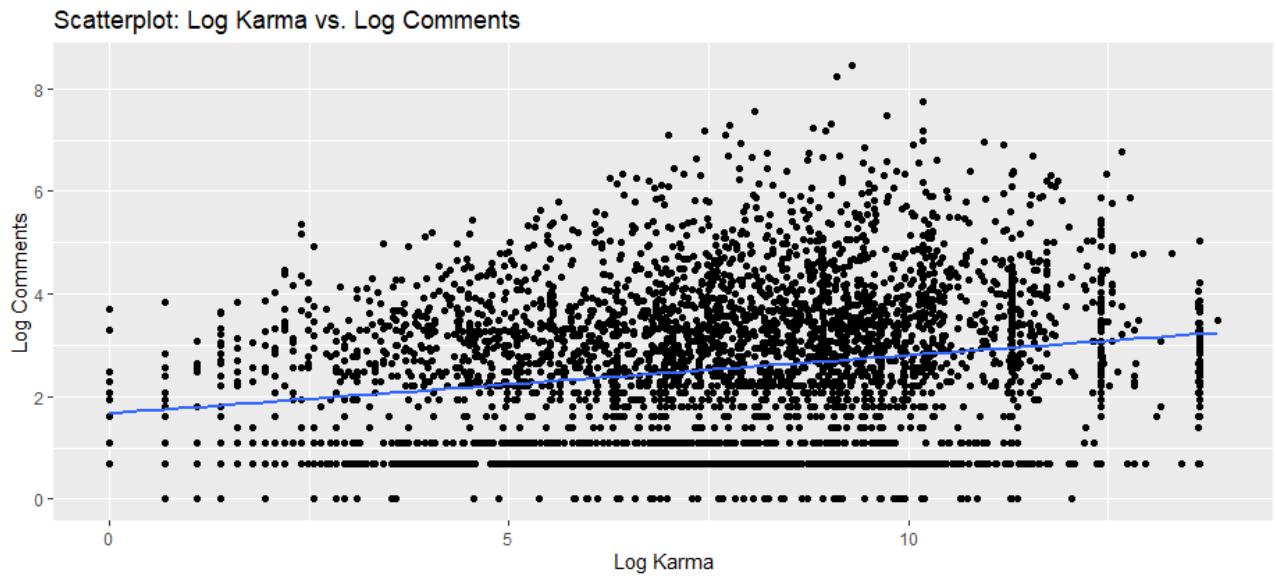
Liite 10. Mielipidevaikuttajat kommentoinnin osalta



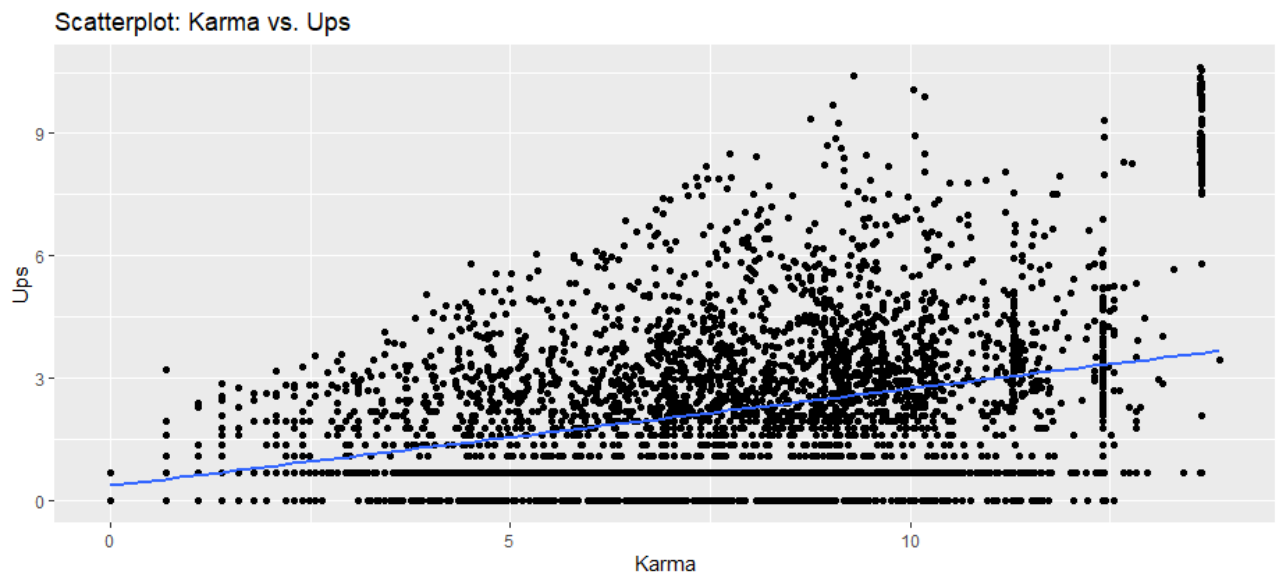
Liite 11. Hajontakuviokeskeisyysaste vs. julkaisun saamat kommentit (logaritmuunnettu)



Liite 12. Hajontakuviokeskeisyysaste vs. julkaisun saamat äänet (logaritmuunnettu)

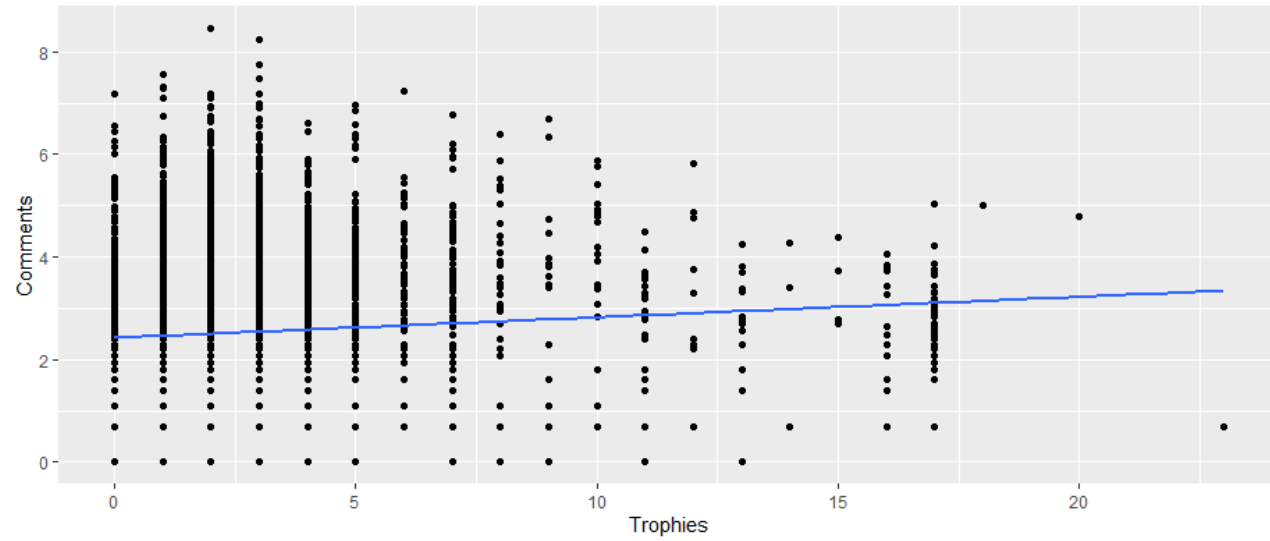


Liite 13. Hajontakuvio karma vs. julkaisun saamat kommentit (logaritmimuunnettu)

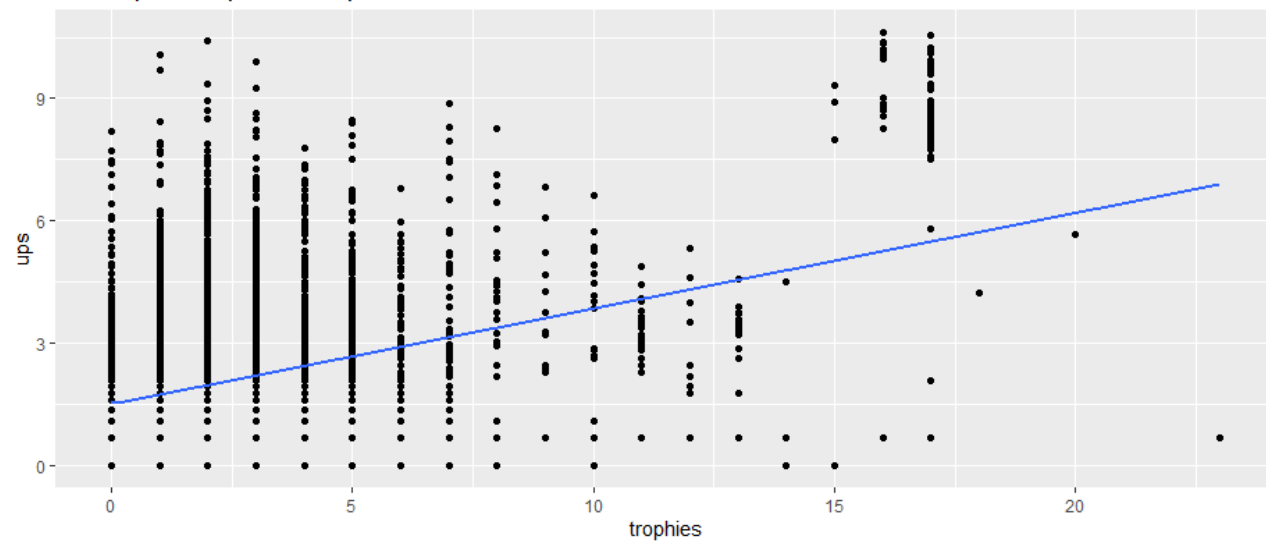


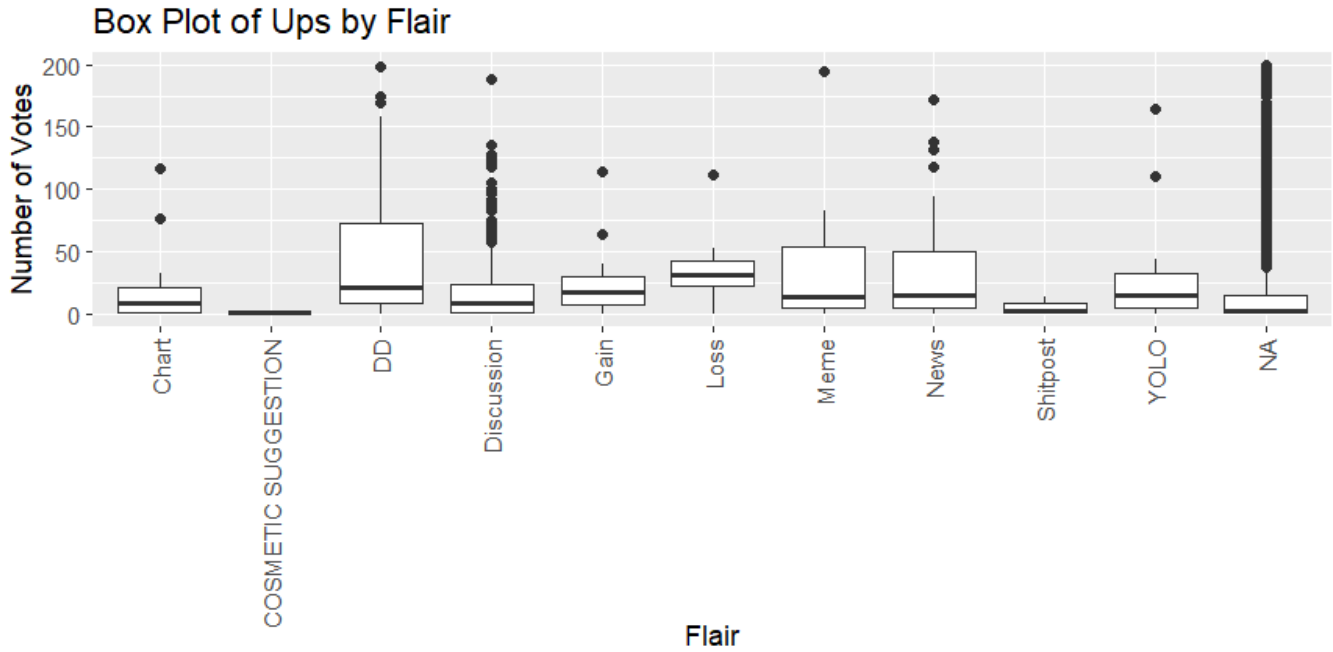
Liite 14. Hajontakuvio karma vs. julkaisun saamat äänet (logaritmimuunnettu)

Scatterplot: Trophies vs. Comments

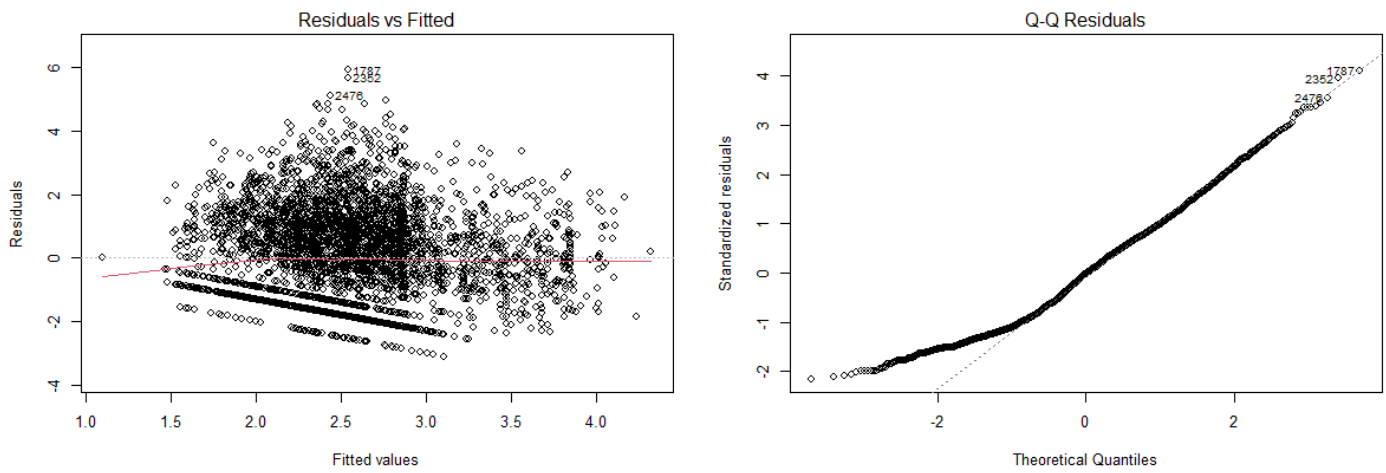
*Liite 15. Hajontakuvio trophyt vs. julkaisun saamat kommentit (logaritmimuunnettu)*

Scatterplot: Trophies vs. ups

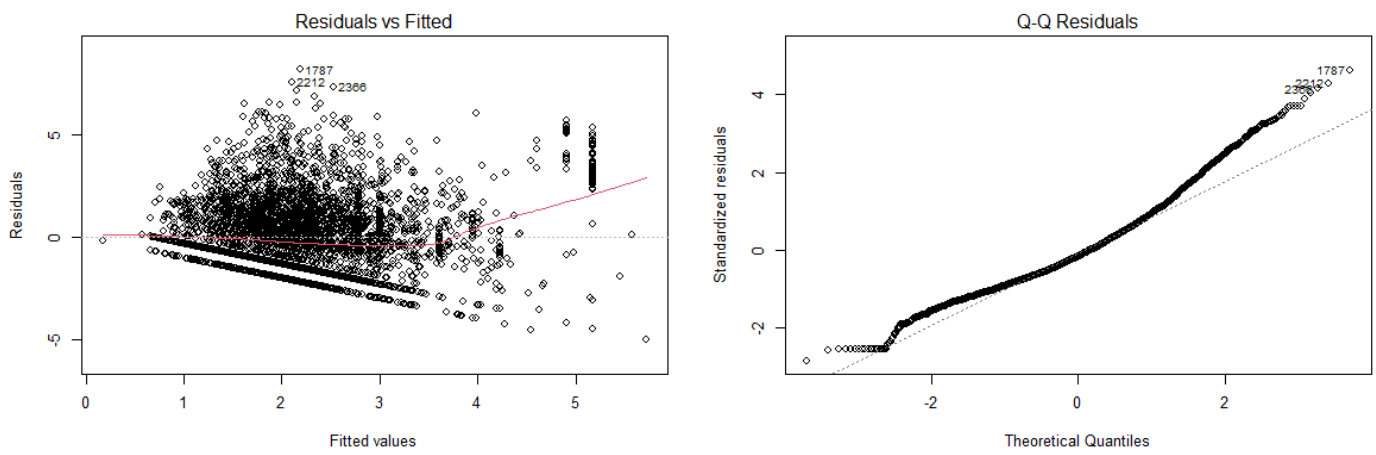
*Liite 16. Hajontakuvio trophyt vs. julkaisun saamat äänet (logaritmimuunnettu)*



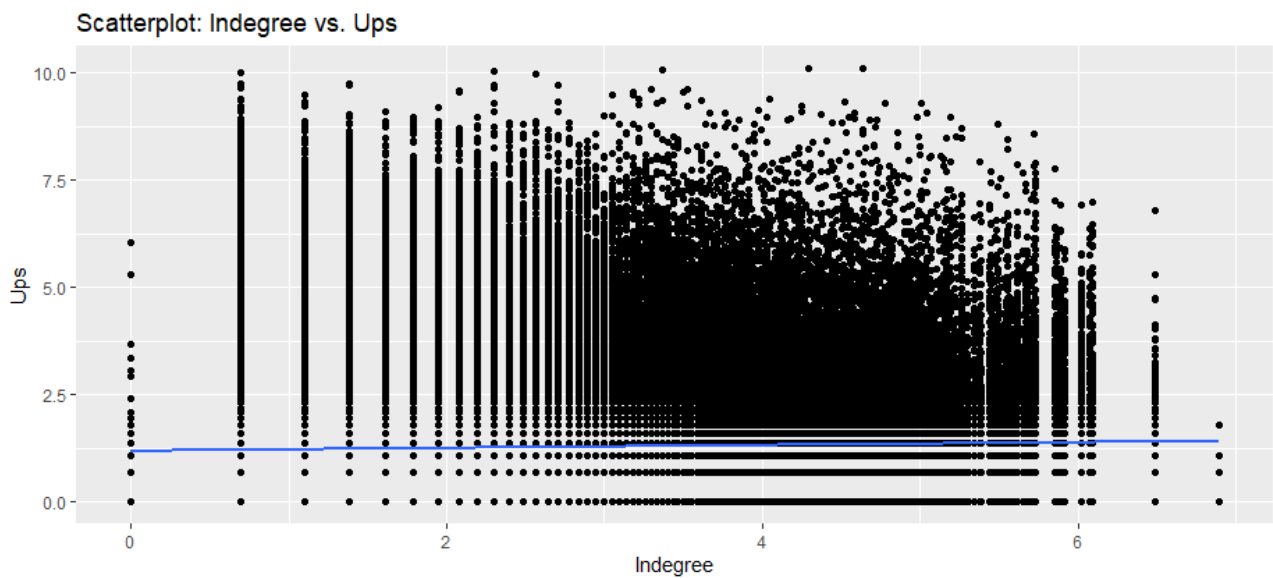
Liite 17. Eri flairien saamien äänten määrän visualisointi



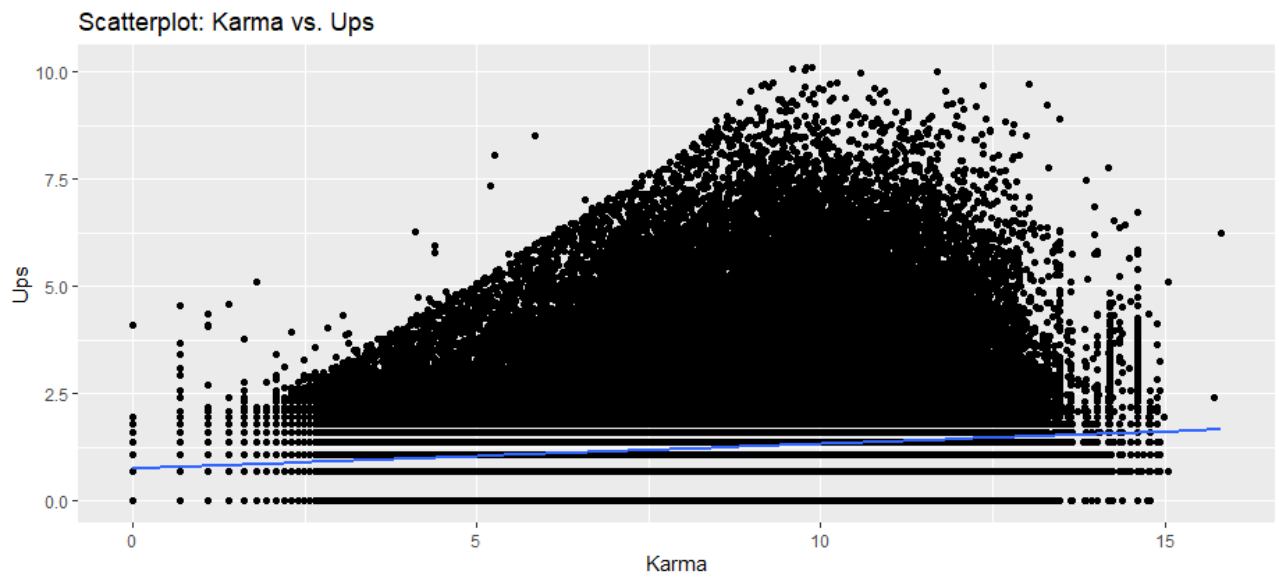
Liite 18. Julkaisuiden saamien kommenttien regressioanalyysi, residuals vs. fitted -kuvaaja ja Q-Q kaavio.



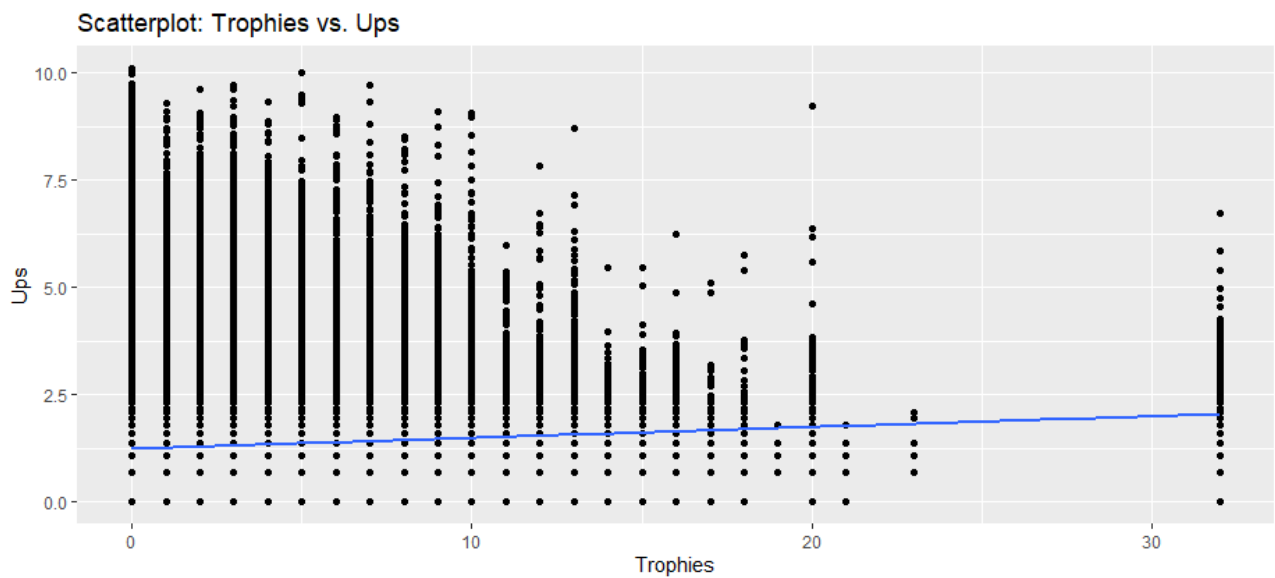
Liite 19. Julkaisun saamien äänten regressioanalyysi, residuals vs. fitted -kuvaaja ja Q-Q kaavio.



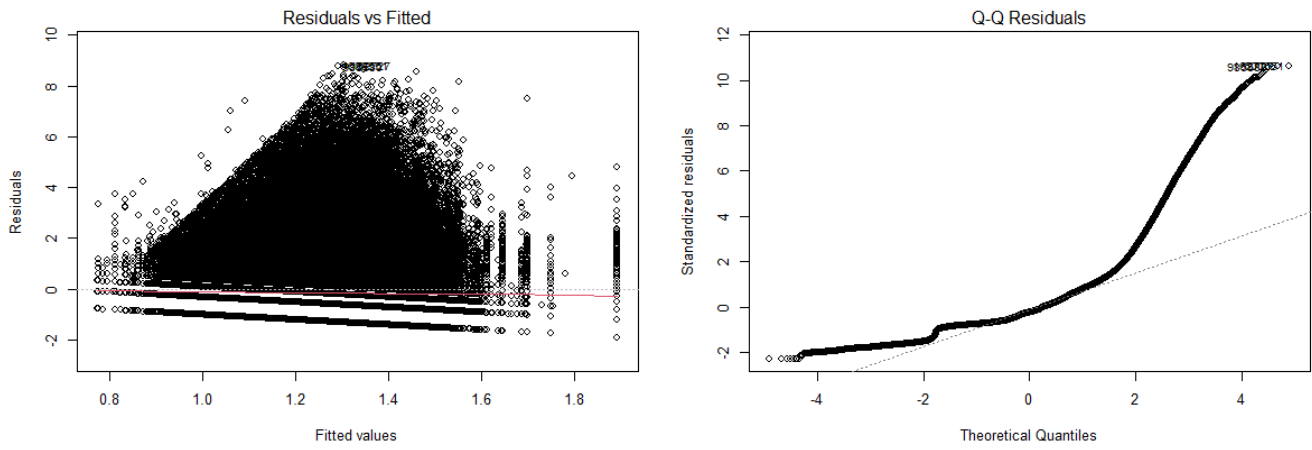
Liite 20. Hajontakuvio keskeisyys vs. kommenttien saamat äänet (logaritmuunnettu)



Liite 21. Hajontakuvio karma vs. kommenttien saamat äänet (logaritmimuunnettu)



Liite 22. Hajontakuvio trophyt vs. kommenttien saama äänet.



Liite 23. Regressioanalyysi kommenttien saamat äänet, residuals vs. fitted kuvaaja ja Q-Q kaavio.