



TAMPEREEN TEKNILLINEN YLIOPISTO

**Jarno Marttila**  
**Datalähtöinen sosiaalisten verkostojen analyysi:**  
**tapaus Suomen Lasten Parlamentti**

Diplomityö

Tarkastajat: Prof. Seppo Pohjolainen  
(TTY) ja tutkija Jukka Huhtamäki  
(TTY)

Tarkastajat ja aihe hyväksytty  
Tieto- ja sähkötekniikan tiedekuntaneu-  
voston kokouksessa 8.9.2010

# TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Tietotekniikan koulutusohjelma

**JARNO MARTTILA: Datalähtöinen sosiaalisten verkostojen analyysi:  
tapaus Suomen Lasten Parlamentti**

Diplomityö, 62 sivua, 6 liitesivua

Marraskuu 2010

Pääaine: Hypermedia

Tarkastajat: Prof. Seppo Pohjolainen (TTY) ja tutkija Jukka Huhtamäki (TTY)

Avainsanat: SNA, visualisointi, Wille, Suomen Lasten Parlamentti

Tieteellinen kiinnostus sosiaalisten verkostojen analysointiin on kehittynyt ja kasvanut tietokoneiden ja webin myötä. Verkkopalveluiden suuret käyttäjämäärät ja niiden sisällä tapahtuvat käyttäjien muodostamat sosiaaliset vuorovaikutussuhteet tekevät mielenkiintoiseksi, niin markkinataloudellisesti kuin tieteellisestikin, käyttäjien muodostamien sosiaalisten verkostojen analysoimisen.

Sosiaalisia verkostoja voidaan tulkita matemaattisin menetelmin ja visualisoida graafeilla. Sosiaalisten verkostojen matemaattiset analysointimenetelmät nojaavat graafiteoriaan ja matriisilaskentaan. Tietotekniikan avulla voidaan yhdistää ja automatisoida sosiaalisten verkostojen analyysin kannalta olennaiset vaiheet tiedon keräys, matemaattiset ja laskennalliset menetelmät sekä tulosten visuaalinen esittämisen graafeina. Graafien informatiivisuutta voidaan lisätä muokkaamalla visuaalisia elementtejä, kuten solmujen väriä, kokoa ja paikkaa, verkostoista laskettavien tunnuslukujen avulla.

Tässä diplomityössä esitetään Suomen Lasten Parlamentti tapauksessa web-pohjaisiin keskustelualueisiin sovellettuja sosiaalisten verkostojen laskennallisia menetelmiä sekä tulosten visualisointeja. Lisäksi esitetään uudenlainen työväline, joka yhdistää tiedonlouhinnan, matemaattisen analyysin ja visualisoinnin yhdeksi kontekstiherkäksi sovellukseksi.

Laskennallisilla menetelmillä on löydetty SLP-tapauksen keskustelualueista muodostetuista verkostoista erilaisia merkittävimpiä tekijöitä. Keskusteluiden mallintamistavalla tiedosta verkostoiksi huomattiin olevan merkittävä vaikutus laskennallisiin lopputuloksiin. Tutkimuksen tuloksena havaittiin, että keskustelualueista voidaan mallintaa verkostoja, joista voidaan löytää merkittäviä tekijöitä matemaattisilla menetelmillä. Näitä tekijöitä ja matemaattisia menetelmiä sovellettaessa joudutaan soveltamaan sisällöllistä analyysiä, jotta voidaan selvittää mitä laskennalliset menetelmät kertovat mallinnetusta verkostosta. Lisäksi havaittiin, että epäsovivalta mallintamisella voidaan päätyä tilanteeseen, jossa edistyneet laskennalliset menetelmät eivät tuota lisäinformaatiota verkostosta.

## ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Information Technology

**JARNO MARTTILA: Data Driven Social Network Analysis: Case Finnish Children's Parliament**

Master of Science Thesis, 62 pages, 6 Appendix pages

November 2010

Major: Hypermedia

Examiner: Prof. Seppo Pohjolainen (TUT) and researcher Jukka Huhtamäki (TUT)

Keywords: SNA, visualization, Wille, Finnish Children's Parliament

The scientific interest towards SNA (Social Network Analysis) has been developing and growing along the computers and the Web. The massive amount of users on Web services and the social interactions between the users on them make analysis of user generated social networks intriguing economically as well as scientifically.

Social networks can be modeled with mathematical methods and represented with graph visualizations. The mathematical methods of SNA focuses on graph theory and matrix algebra. By using computer science, the necessary phases of SNA data collection, data modeling, mathematical and computational methods, and visual representation of results with graphs can be combined and automated. Informativeness of graphs can be increased by modifying visual elements, such as size and position of nodes with respect to calculated metrics.

This Master of Science Thesis describes those SNA and visualization methods that are used on case Finnish Children's Parliament web based discussion forums. In addition, a new tool is represented which combines datamining, and mathematical analysis and visualization into a single context sensitive tool.

By using metrics, one can determine the most important actors in a modeled network. The method of modeling discussion forums into a network has been found to affect the results of computational metrics. Also, discussion forums have been modeled into networks, from which most important actors have been determined by using mathematical methods. When applying these metrics onto networks, qualitative research must be used to understand the context. Moreover, improper modeling of network may lead to situations where advanced metrics give no additional information about the network.

## ALKUSANAT

Tämä diplomityö on tehty Tampereen teknillisen yliopiston hypermedialaboratoriossa matematiikan laitoksella, sosiaalisten verkostojen tutkimusryhmässä. Diplomityön tutkimus tapaus liittyy Suomen Lasten Parlamentin kehittämiseen.

Kiitän hypermedialaboratoriota ja matematiikan laitosta kutkuttavan mielenkiintoisesta diplomityöaiheesta ja tunnen olevani etuoikeutettu päästessäni toteuttamaan itseäni ja luomaan tiedettä. Sosiaalisten verkostojen analyysi tuntui aiheena sopivan mukavasti joka suuntaan erikoistuneelle diplomityöntekijälle ja luovaa hulluutta sai käyttää vapaasti työn tekemisessä.

Haluan kiittää diplomityöni tarkastajia professori Seppo Pohjolaista ja tutkija Jukka Huhtämäkeä, jotka ovat ohjanneet tutkimustyötä ja ovat aktiivisesti olleet mukana molemminpuolisessa oppimisprosessissa sosiaalisten verkostojen analyysissä. Kiitokset Kirsi Siliukselle ja Anne-Maritta Tervakarille kvalitatiivisen tutkimukseen liittyvistä näkökulmista, Thumas Miilumäelle avusta matemaattisten menetelmien tulkitsemisessa sekä Jaakko Saloselle Wille-visualisointiympäristöön perehdyttämisestä.

Kiitos äidille ja isälle, jotka ovat aina kannustaneet minua opiskelemaan. Lopuksi haluan kiittää rakkaita ystäviäni, jotka ovat tukeneet ja motivoineet minua sekä viettäneet lukemattomia kahvihetkiä kanssani opintojen ohessa. Ilman teitä kaikkia en olisi tässä missä olen nyt.

Tampere, 24. marraskuuta 2010

Jarno Marttila  
Tekniikankatu 14 F 306  
33720 Tampere  
jarno.marttila@tut.fi

# SISÄLLYS

1. Johdanto . . . . .	1
2. Teoreettinen tausta . . . . .	3
2.1 Sosiaalisten verkostojen analysoinnista . . . . .	3
2.2 Sosiaalisten verkostojen visualisoinnista . . . . .	4
2.3 Sosiaalinen verkko . . . . .	4
2.3.1 Moodillisuus . . . . .	5
2.4 Graafiteoreettinen notaatio . . . . .	7
2.4.1 Sosiogrammi . . . . .	8
2.5 Sosiometrinen notaatio . . . . .	9
2.5.1 Sosiomatriisi . . . . .	10
2.5.2 Arvotettu sosiomatriisi . . . . .	10
2.6 Tunnusluvut . . . . .	13
2.6.1 Suuntaamattomien verkostojen tunnusluvut . . . . .	15
2.6.2 Suunnattujen verkostojen tunnusluvut . . . . .	17
3. Datalähtöinen analyysi . . . . .	26
3.1 Web 2.0 datalähtöisen analyysin kontekstina . . . . .	29
3.2 SNA-työvälineet . . . . .	30
4. Toteutus . . . . .	33
4.1 Visualisointityöväline ja visualisointiprosessi . . . . .	33
4.2 Toteutustekniikat . . . . .	37
5. TAPAUS: Lasten parlamentti . . . . .	42
5.1 Verkkoparlamentti . . . . .	42
5.2 Analysoitavan tiedon kerääminen . . . . .	44
5.3 Tunnusluvut kontekstissa . . . . .	44
5.4 Pohdintaa . . . . .	56
6. Yhteenveto ja johtopäätökset . . . . .	58
Lähteet . . . . .	60

## KUVAT

2.1	Kaksimoodinen yhteysverkosto . . . . .	6
2.2	Romahdutettu yhteysverkosto . . . . .	7
2.3	Sosiogrammi: ystävyysuhdeverkosto . . . . .	9
2.4	Sosiogrammi: tekstiviestiverkosto . . . . .	12
2.5	Vaikutusjoukko . . . . .	21
2.6	PageRank esimerkki . . . . .	24
3.1	Projektioiden ottaminen datajoukosta . . . . .	27
3.2	Informaation visualisointiprosessi . . . . .	28
3.3	Gephi . . . . .	31
3.4	NodeXL . . . . .	32
4.1	Työnkulun prosessikaavio . . . . .	34
4.2	Network dokumentti . . . . .	35
4.3	Pajek dokumentti . . . . .	35
4.4	Greasemonkey-paneeli . . . . .	36
4.5	Suomen Lasten Parlamentin yleistä keskustelua keskustelupalsta . . .	37
4.6	Greasemonkey-ikoni . . . . .	39
4.7	Greasemonkeyn skriptien asetusikkuna. . . . .	40
4.8	Yksittäisen keskustelun ATOM-syöte . . . . .	41
5.1	Suomen Lasten Parlamentin verkkoparlamentin etusivunäkymä . . . .	43
5.2	Suomen Lasten Parlamentin 2D-parlamenttiympäristö . . . . .	43
5.3	Yksimoodinen kouluruokakeskusteluverkosto . . . . .	45
5.4	Yksimoodisen verkoston toimijoiden astesummat. . . . .	46
5.5	Kaksimoodinen kouluruokakeskusteluverkosto . . . . .	47
5.6	Kaksimoodinen verkoston astesummat. . . . .	48
5.7	Yksimoodiseksi romahdutettu verkosto. . . . .	49
5.8	Yksimoodiseksi romahdutetun verkoston toimijoiden astesummat. . .	49
5.9	Yksimoodisen verkoston toimijoiden välillisyyksiä . . . . .	50
5.10	Yksimoodiseksi romahdutetun verkoston toimijoiden välillisyyksiä. .	51
5.11	Kaksimoodisen verkoston keskustelunaiheiden arvostusaste. . . . .	52
5.12	Yksimoodisen verkoston toimijoiden arvostusaste . . . . .	52
5.13	Kaksimoodisen verkoston keskustelunaiheiden arvostusläheisyys. . . .	53
5.14	Yksimoodisen verkoston toimijoiden arvostusläheisyys. . . . .	53
5.15	Kaksimoodisen verkoston toimijoiden PageRankit. . . . .	54
5.16	Yksimoodisen verkoston toimijoiden PageRankit. . . . .	54
5.17	Tunnuslukujen korrelaatio . . . . .	55

## TAULUKOT

2.1	Ystävyysverkoston listaesitys . . . . .	10
2.2	Ystävyysverkoston matriisiesitys . . . . .	11
2.3	Arvotetun ystävyysverkoston sosiomatriisi . . . . .	12
2.4	Arvotetun ystävyysverkoston listaesitys . . . . .	13
2.5	Ystävyysverkoston sosiogrammi . . . . .	15
2.6	Ystävyysverkoston toimijoiden välillisyydet . . . . .	16
2.7	Arvotetun ystävyysverkoston toimijoiden välillisyydet . . . . .	17
2.8	Ystävyysverkoston vienti- ja tuontiluvut . . . . .	18
2.9	Arvotetun ystävyysverkoston vienti- ja tuontiluvut . . . . .	18
2.10	Ystävyysverkoston toimijoiden arvostusasteet . . . . .	19
2.11	Arvotetun ystävyysverkoston toimijoiden arvostusasteet . . . . .	20
2.12	Ystävyysverkoston toimijoiden arvostusläheisyydet . . . . .	22
2.13	Ystävyysverkoston toimijoiden PageRankit . . . . .	25
2.14	Ystävyysverkoston toimijoiden PageRankit . . . . .	25

## KÄYTETYT LYHENTEET

BBS	Bulletin Board System
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
GPU	Graphics Processing Unit
SLP	Suomen Lasten Parlamentti
SNA	Social Network Analysis
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformations



## TERMIT JA SYMBOLIT

$C_B(n_i)$	Toimijan $n_i$ välillisuus
$C'_B(n_i)$	Toimijan $n_i$ normeerattu välillisuus
$d_O(n_i)$	Toimijan $n_i$ vientiluku
$d_I(n_i)$	Toimijan $n_i$ tuontiluku
$d(i, j)$	Toimijoiden $n_i$ ja $n_j$ välinen etäisyys
$d(n_i)$	Toimijan $n_i$ astesumma
$g$	tekijä/tapahtuma tai toimija
$I_i$	Vaikutusjoukon kokoa määrittävä luku
$l_i$	Kaari/yhteys $i$
$\mathcal{L}$	Solmuja yhdistävien yhteyksien/kaarien joukko
$m_i$	Tapahtumaa $i$ kuvaava solmu
$\mathcal{M}$	Tapahtumia kuvaavien solmujen joukko
$n_i$	Toimijaa $i$ kuvaava solmu
$\langle n_i, n_j \rangle$	Järjestetty pari
$n_i \rightarrow n_j$	Suunnattu yhteys toimijasta $n_i$ toimijaan $n_j$
$\mathcal{N}$	Toimijoita kuvaavien solmujen joukko
$P_D(n_i)$	Toimijan $n_i$ arvostusaste
$P'_D(n_i)$	Toimijan $n_i$ normeerattu arvostusaste
$\bar{P}_P$	Arvostusläheisyys
$PR(n_i)$	Toimijan $n_i$ PageRank
$P_R$	arvostusaste
$\sum_{j=1}^g x_{ij}$	Summa $x_{i1} + x_{i2} + \dots + x_{ig}$
$x_{ij}$	Sosiomatriisin alkio
$\mathbf{X}$	Sosiomatriisi

# 1. JOHDANTO

Sosiaaliset verkostot ja sosiaalisten verkostojen analysointimenetelmät ovat herättäneet kiinnostusta niin matemaatikoiden, sosiologien kuin kulttuuriantropologien keskuudessa. Erilaisia tutkijoita ovat viehättäneet sosiaalisten verkostojen analysoinnissa sosiaalisten entiteettien muodostamien yhteyksien analysointi sekä yhteyksien muodostamat hahmot ja niiden merkitykset. (Wasserman & Faust, 1994)

Sosiaalisen median palveluista on tullut pysyvä osa nykyaikaisia verkkopalveluita ja perinteisiin verkkopalveluihin pyritään integroimaan sosiaalisen median osia. Kalliala ja Toikkanen (2009) määrittävät sosiaalisen median seuraavasti: "Sosiaalinen media on prosessi, jossa yksilöt ja ryhmät rakentavat yhteisiä merkityksiä sisältöjen, yhteisöjen ja verkkoteknologioiden avulla".

Sosiaalisten verkkopalveluiden käyttäjät luovat, julkaisevat, muokkaavat ja jakavat sisältöä sekä vuorovaikuttavat toistensa kanssa. Yksittäiset käyttäjät muodostavat toimillaan yhteyksiä itsensä ja sisällön välille, sekä toisten käyttäjien välille. Näitä yhteyksiä voidaan analysoida, havainnollistaa ja tulkita sosiaalisten verkostojen analysointi- ja visualisointimenetelmillä.

Yksinomaan sosiaaliseen mediaan perustuvista verkkopalveluista on tullut erittäin suosittuja. Esimerkiksi erityisesti sosiaaliseen verkostoitumiseen keskittyvällä verkkopalvelulla Facebookilla<sup>1</sup> on jo yli 500 miljoonaa käyttäjää. Verkkopalveluiden suuret käyttäjämäärät ja niiden sisällä tapahtuvat käyttäjien muodostamat sosiaaliset vuorovaikutussuhteet tekevät mielenkiintoiseksi, niin markkinataloudellisesti kuin tieteellisestikin, käyttäjien muodostamien sosiaalisten verkostojen analysoimisen.

Tieteellinen kiinnostus sosiaalisten verkostojen analysointiin on kehittynyt ja kasvanut tietokoneiden ja Webin myötä. Erilaisten ihmisten muodostamien verkostojen määrä on räjähtänyt kasvuun Web 2.0 -teknologian mahdollistaman sosiaalisen median myötä. Tietokoneiden jatkuva laskentatehon kasvu on mahdollistanut kehittyneempiä ja automatisoidumpia menetelmiä sosiaalisen verkostodatan laskennalliselle käsittelylle.

Sosiaalisten verkostojen analysointi- ja visualisointimenetelmät tukeutuvat vahvasti kahteen matematiikan osa-alueeseen, matriisilaskentaan ja graafiteoriaan. Ihmisten välisistä vuorovaikutussuhteista muodostuvaa kaksikulotteista verkostodataa

---

<sup>1</sup>Facebook Tilastot. <http://www.facebook.com/apress/info.php?statistics>.

voidaan kuvata hyvin kompaktisti matriisien avulla. SNA-menetelmissä sosiaalisia verkostoja kuvaavia matriiseja nimitetäänkin sosiomatriiseiksi. (Wasserman & Faust, 1994)

Sosiaalisten verkostojen analyysissa graafiteoria ja matriisilaskenta nitoutuvat vahvasti yhteen. Kummallakin voidaan esittää sama asia, mutta tietokoneistetussa SNA-menetelmien käytössä matriisit ja matriisilaskenta vakiinnuttavat asemansa tiedon esitys ja analysointimuotoina, graafiteorian pysyessä tiedon – sosiomatriisien ja tunnuslukujen, visualisointimuotona. Freeman (2009) käsittelee erityisesti suuntaamattomien ja suunnattujen yksi- ja kaksimoodisten sosiaalisten verkostojen visualisointeja sekä niiden tuottamista erilaisista datajoukoista.

Tämän diplomityön tutkimuskysymyksiä olivat miten sosiaalisten verkostojen analysointi- ja visualisointimenetelmiä voidaan hyödyntää olemassa olevien sosiaalisten verkkopalveluiden visualisoimisessa ja mitä käytetyt tunnusluvut kertovat Suomen Lasten Parlamentin kontekstissa. Tutkimustyypiltään diplomityö on konstrukttiivinen tutkimus, jonka yhteydessä on luotu uudenlainen työväline keskustelun alueiden määrälliseen analysointiin ja visualisointiin.

## 2. TEOREETTINEN TAUSTA

### 2.1 Sosiaalisten verkostojen analysoinnista

Sosiaalisten verkostojen analyysi yhdistää monia eri tieteenaloja ja sen kehitykseen on osallistunut suuri joukko tutkijoita niin sosiologian, sosiaalipsykologian, sosiaali antropologian, matematiikan kuin organisaatiopsykologiankin alueilta (Moreno, 1934; Lewin, 1951).

Sosiaalisten verkostojen analyysi on suhteellisen tuore tieteenala ja se on saavuttanut nykyisen muotonsa tieteenä vasta 1900-luvun aikana. Sosiaalisia verkostoja kuvaavia visualisointeja on ollut olemassa jo ainakin tuhannen vuoden ajan. Eräs vanhimmista säilyneistä sosiaalisia verkostoja esittävistä kuvista on Isidore de Sévillen luoma sukupuu (Freeman, 2009).

Tässä diplomityössä esitettävät sosiaalisten verkostojen analyysimenetelmät perustuvat enimmäkseen Wasserman ja Faustin (1994) teokseen sekä Miilumäen (2010) suomennoksiin Wassermanin ja Faustin esittämistä termeistä ja menetelmistä. Wasserman ja Faust käsittelevät sosiaalisten verkostojen analyysiä selvästi erotettavissa olevana tieteenalanana, jonka tutkimusnäkökulmana ovat sosiaali- ja käyttäytymistieteet. Sosiaalisten verkostojen analyysi perustuu oletukseen vuorovaikuttavien yksilöiden tai yksiköiden muodostaman *suhteen* (relationship) vuorovaikutuksen tärkeydestä.

Suhteiden määrittäminen yksiköiden välisillä *yhteyksillä* (linkages) on perustavaa laatua oleva käsite verkostoteorioissa. Sosiaalisten verkostojen näkökulma sisältää teorioita, malleja ja sovelluksia jotka on ilmaistu suhteessa yhteyksien käsitteisiin ja prosesseihin. Wassermann ja Faust mainitsevat, että yhteyksien käsitteen lisäksi seuraavat käsitteet ovat tärkeitä:

- *Toimijoita* (actors) ja heidän *tekojansa* (actions) tulkitaan toisistaan riippuvina eikä toisistaan riippumattomina itsenäisinä yksikköinä (autonomous unit)
- Toimijoiden väliset *yhteydet* (relation) ovat virtaus- (flow) ja siirtokanavia aineellisille ja aineettomille resursseille
- Yksilöihin keskittyvät *verkostomallit* tarkastelevat miten verkoston rakenteellinen ympäristö tarjoaa tai rajoittaa yksilöiden toimintoja

- Verkostomallit käsitteellistävät erilaisia rakenteita; sosiaalisia, verkostollisia, poliittisia, pysyvinä toimijoiden välisien yhteyksien kuvioina

## 2.2 Sosiaalisten verkostojen visualisoinnista

Sosiaalisten verkostojen visualisointi on olennainen osa-alue sosiaalisten verkostojen analyysia. Wassermann ja Faust (1994) tarkastelevat sosiaalisten verkostojen analyysia hyvin matemaattis-teoreettisesta näkökulmasta, eikä heidän teoksessaan ole juuri jäänyt tilaa verkostojen visualisoinnille. Freeman (2009) puolestaan esittää yleistetyimmän mallin sosiaalisten verkostojen analyysille. Malli koostuu neljästä tekijästä jotka määrittävät modernin sosiaalisten verkkojen analysoinnin. Nämä tekijät ovat:

- SNA sisältää käsitteitä toimijoita sitovien sosiaalisten yhteyksien tärkeydestä
- SNA kerää sosiaalisiin yhteyksiin liittyvää dataa
- SNA:n kuuluu graafisten kuvien käyttö
- SNA käyttää matemaattisia ja laskennallisia malleja

Freemanin sekä Wassermanin ja Faustin määritelmässä on yhtymäkohtia, kuten toimijoiden välisien yhteyksien korostaminen ja matemaattisten sekä laskennallisten mallien käyttö, mutta Freeman painottaa määritelmässään graafisten kuvien käyttöä sosiaalisten verkostojen havainnollistamisen tukena.

## 2.3 Sosiaalinen verkko

Sosiaalinen verkko on *toimijoiden* (actors) muodostaman joukon rakenne missä osa toimijoista on yhdistynyt toisiinsa yhden tai useamman *yhteyden* (relation) kautta (Knocke & Yang, 2008). Nämä kaksi tekijää, toimijat ja yhteydet, esiintyvät yleisesti sosiaalisten verkostojen määritelmässä. Esimerkiksi sosiaalisen verkoston erilaisia rakenteita voidaan mallintaa ja esittää solmujen ja kaarien muodostamina verkkoina, joissa solmut kuvaavat sosiaalisen järjestelmän jäseniä ja kaaret näiden jäsenien välisiä suhteita (Wellman & Berkowitz, 1988).

Wassermann ja Faust (1994) esittelevät kolme erilaista matemaattista notatiotapaa sosiaalisten verkkojen kuvaamiseen: *graafiteoreettinen*, *sosiometrinen* ja *algebrallinen*. Jokaisella näistä kuvaustavoista voidaan esittää sama informaatio verkostoista. Tietyille datatyypeille ja verkostotyypeille jonkin tietyn kuvaustavan käyttö voi olla muita kuvaustapoja hyödyllisempää, selkeyden, tarkoituksenmukaisuuden tai laskennallisen tehokkuuden vuoksi.

Tässä diplomityössä käytetään erityisesti graafiteoreettista ja sosiometrista kuvaustapaa, joihin palataan myöhemmin uudestaan. Graafiteoreettisen kuvaustavan

avulla sosiaalisia verkostoja voidaan visualisoida ihmisille kognitiivisesti helpommassa muodossa. Sosiometrisella kuvaustavalla esitettyihin verkostoihin voidaan soveltaa erilaisia tunnuslukumenetelmiä. Algebrallista kuvaustapaa ei käsitellä tässä diplomityössä, koska sen käytölle sosiaalisten verkostojen laskennallisen analyysin tai visualisoinnin näkökulmasta ei ollut erityistä tarvetta saati hyötyä.

### 2.3.1 Moodillisuus

Sosiaalisia verkostoja luokitellaan toimijajoukkojen ja toimijoiden välisten vuorovaikutussuhteiden mukaan. Verkoston moodillisuus ilmaisee *entiteettijoukkojen* (set of entities) lukumäärän, jotka kuvaavat tarkasteltavan verkoston rakenteellisia ominaisuuksia. Entiteettijoukot voivat olla joko *tapahtuma-* (set of events) tai *toimijajoukkoja* (set of actors). Esimerkiksi yksimoodisissa verkostoissa tutkitaan vain yhtä tapahtuma- tai toimijajoukkoa.

Sosiaaliset verkostot ovat tyypillisesti yksi- tai kaksimoodisia. Yksimoodiset verkostot ovat tämän diplomityön laskennallisen analyysin kannalta vallitseva verkostotyyppi, mutta myös useampimoodisia kaksi-, kolme- tai nelimoodisia verkostoja voidaan tarkastella. Kaksimoodisia verkostoja suuremmat verkot ovat kuitenkin huomattavasti harvinaisempia, koska sosiaalisen verkoston analysointimenetelmiä ei ole suunniteltu käytettäväksi niin monimutkaisille tietorakenteille.

Tässä diplomityössä käsitellyt verkostot ovat pääsääntöisesti yksimoodisia verkostoja. Näitä verkostoja mallinnetaan sosiomatriiseina ja graafeina. Yksimoodisille verkostoille ovat voimassa kaikki sosiaalisten verkostojen laskennalliset menetelmät.

Kaksimoodista verkostoa voidaan mallintaa graafeilla ja sosiomatriiseilla kuten yksimoodisia verkostoja. Kaksimoodisissa verkostoissa tutkitaan tyypillisesti joko kahta *toimijajoukkoa* tai yhtä *toimijajoukkoa* ja yhtä *tapahtumajoukkoa*. Tämä jakaa kaksimoodiset verkostot kahteen eri tyyppiin eli toimijajoukkojen muodostamiin kaksimoodisiin verkostoihin ja toimija- sekä tapahtumajoukkojen muodostamiin verkostoihin eli *yhteysverkostoihin* (affiliation network).

Yhteysverkostot muodostuvat kun toimijajoukolla on yhteys tapahtumajoukkoon jonkin tietyn tapahtuman perusteella. Esimerkiksi opiskelijoilla on yhteys oppitunteihin osallistumisen perusteella. Opiskelijoiden muodostama toimijajoukko on siten yhteydessä oppituntien muodostamaan tapahtumajoukkoon. Toimijajoukon ja tapahtumajoukon välisen yhteyden ei tarvitse olla osallistumiseen perustuva yhteys, se voi olla muukin yhteys jonka perusteella joukot liittyvät toisiinsa.

### Verkostojen romahduttaminen

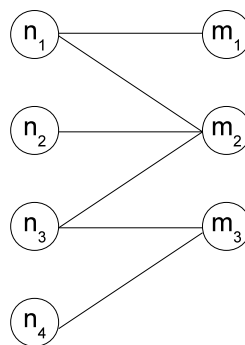
Kaksimoodiset verkostot voidaan muuntaa yksimoodisiksi verkostoiksi romahduttamalla ne toisen moodin suhteen esimerkiksi muodostamalla yksimoodinen verkosto

kaksimoodisesta verkostosta toimijoiden yhteenkuuluvuuksien suhteen. Romahduttaessa suurempimoodista verkostoa pienempimoodiseksi, tehdään aina tietynlainen tulkinta tai säännöt joiden perusteella uusi verkosto muodostetaan. Esimerkiksi kaksimoodinen verkosto jonka moodeina ovat elokuvat ja elokuvissa näytelleet henkilöt voidaan romahduttaa yksimoodiseksi näyttelijäverkostoksi tulkinnalla, jos näyttelijät ovat näytelleet samassa elokuvassa, niin heidän välilleen muodostuu yhteys (Breiger 1974, Watts 1999).

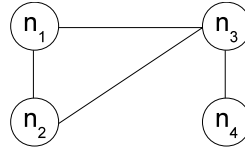
Koska kaksimoodisten verkostojen laskennalliset menetelmät ovat huomattavasti rajoitetumpia kuin yksimoodisten verkostojen, ne voidaan romahduttaa yksimoodiseksi verkostoiksi tietyn tulkinnan perusteella. Nämä tulkinnat ovat aina tapauskohtaisia. Erityisen tärkeä huomio verkostoa romahduttaessa on se, että korkeampimoodinen verkosto on aina tietomäärältään rikkaampi kuin pienempimoodinen verkosto. Romahduttaessa verkostoja informaatiota verkostosta yksinkertaistetaan – kadotetaan.

Kaksimoodisissa verkostoissa on kaksi eri solmujoukkoa. Toimijajoukkoja merkitään  $\mathcal{N} = \{n_1, n_2, \dots, n_g\}$  ja yhteysverkostoissa tapahtumajoukkoja merkitään  $\mathcal{M} = \{m_1, m_2, \dots, m_h\}$ . Kaksimoodisissa graafeissa joukkoja yhdistää kaarien joukko  $\mathcal{L} = \{l_1, l_2, \dots, l_L\}$

Kuvassa 2.1 on kaksimoodinen yhteysverkosto jossa toimijat  $n_1, n_2, n_3$  ja  $n_4$  ovat yhteydessä tapahtumiin  $m_1, m_2$  ja  $m_3$ . Kuvassa 2.2 samainen verkosto on romahduttu yhteisten tapahtumien perusteella yksimoodiseksi verkostoksi. Tällöin solmut jotka ovat olleet yhteydessä samaan tapahtumaan kaksimoodisessa verkostossa ovat nyt yhteydessä toisiinsa yksimoodisessa verkostossa. Olennainen huomio on, että romaautetussa verkostossa on kadotettu tieto siitä mihin tapahtumiin oltiin yhteydessä alunperin.



Kuva 2.1: Esimerkki kaksimoodisesta yhteysverkostosta. Verkoston toimijat  $n_i$  ovat yhteydessä verkoston tapahtumiin  $m_j$ .



Kuva 2.2: Kaksimoodinen yhteysverkosto romahdutettuna yksimoodiseksi verkostoksi toimijoille  $n_i$  yhteisten tapahtumien  $m_j$  suhteen.

## 2.4 Graafiteoreettinen notaatio

Verkostoja voidaan tarkastella useilla eri tavoilla. Yksi hyvin hyödyllinen näkymä on graafi. Graafi muodostuu solmuista, jotka on liitetty yhteen kaarilla. Kuvaamme seuraavaksi graafiteorian avulla toimijoiden (actor) eli solmujen ja suhteiden (relation) eli yhteyksien muodostaman verkostodatajoukon.

**Määritelmä 2.4.1.** *Olkoon joukko toimijoita (set of actors)  $\{n_1, n_2, \dots, n_g\}$ . Merkitsemme tätä joukkoa merkillä  $\mathcal{N}$ . Näin ollen joukko  $\mathcal{N}$  sisältää  $g$  toimijaa,*

$$\mathcal{N} = \{n_1, n_2, \dots, n_g\}.$$

*Esimerkiksi kuuden henkilön Allison, Drew, Eliot, Ross, Keith, Simon esittäminen graafiteorian notaatiolla  $\mathcal{N} = \{Allison, Drew, Eliot, Ross, Keith, Simon\}$ , jossa  $n_1 = Allison, n_2 = Drew, n_3 = Eliot, n_4 = Ross, n_5 = Keith, n_6 = Simon$ .*

Toimijoiden väliset yhteydet voivat olla suunnattuja tai suuntaamattomia. Nämä kaksi yhteystyyppiä jakavat graafit suunnattuihin ja suuntaamattomiin graafeihin. Suunnatuissa graafeissa yhteyden suunta ilmaistaan nuolella tai järjestetyllä parilla. Järjestettyä paria voidaan kuvitella toimijaparina  $\langle n_i, n_j \rangle$ , jossa ensimmäinen toimija  $n_i$  määrittää mistä toimijasta on olemassa yhteys ja toinen toimija  $n_j$  mihin toimijaan yhteys kohdistuu.

**Määritelmä 2.4.2.** *Olkoon yhteydet dikotomisissa ja suunnattuja ja olkoon yhteyksiä vain yksi. Näin tekijä  $n_i$  voi olla yhteydessä tekijään  $n_j$  tai sitten se ei ole yhteydessä siihen. Tämä esitysmuoto ei ota kantaa yhteyden voimakkuuteen tai siihen kuinka useasti toimija  $n_i$  on yhteydessä toimijan  $n_j$  kanssa.*

Koska yhteys on suunnattu niin toimijoiden  $n_i$  ja  $n_j$  muodostama yhteys on erilainen kuin toimijoiden  $n_j$  ja  $n_i$  muodostama yhteys. Esimerkiksi toimijoiden välinen suunnattu yhteys voi osoittaa toimijoiden välistä hierarkisuutta. Toimija  $n_i$  voi olla toimijan  $n_j$  esimies, mutta toimija  $n_j$  ei ole toimijan  $n_i$  esimies.



**Määritelmä 2.4.3.** Merkitsemme näitä kahden tekijän muodostamia yhteyksiä järjestetyllä parilla, jonka joukosta käytämme tunnusta  $\mathcal{L}$ . Jos jokin järjestetty pari kuuluu joukkoon  $\mathcal{L}$  niin kaksikon ensimmäinen tekijä on yhteydessä kaksikon toiseen tekijään. Joukossa  $\mathcal{L}$  voi olla enintään  $g(g - 1)$  järjestettyä paria eli yhteyttä ja vähintään 0.

**Määritelmä 2.4.4.** Olkoon järjestetty pari  $\langle n_i, n_j \rangle$  jolloin on olemassa yhteys  $n_i \rightarrow n_j$ . Merkitsemme joukkoon  $\mathcal{L}$  kuuluvia järjestettyjä pareja symbolilla  $l$ . Jos joukossa  $\mathcal{L}$  on  $M$  järjestettyä paria niin

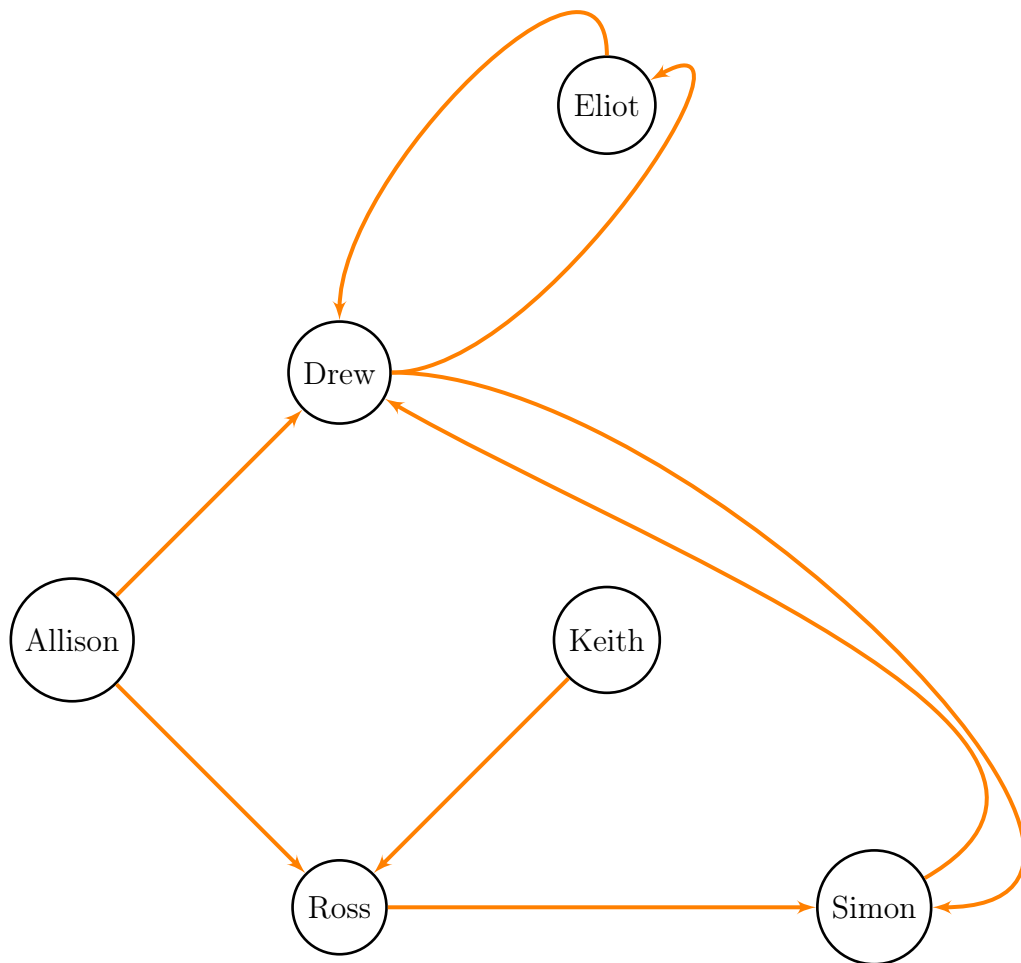
$$\mathcal{L} = \{l_1, l_2, \dots, l_M\}.$$

Tällaista graafia kutsutaan yleisesti suunnatuksi graafiksi, koska toimijoiden  $n_i$  ja  $n_j$  välillä on suunnattu yhteys. Graafi koostuu toimijoiden joukosta  $\mathcal{N}$  ja yhteyksien joukosta  $\mathcal{L}$ . Graafia voidaan matemaattisesti esittää kahdella joukolla  $(\mathcal{N}, \mathcal{L})$  jota merkitsemme symbolilla  $\mathcal{G}$ .

Tarkastellaan seuraavaksi kuuden henkilön Allison, Drew, Eliot, Keith, Ross ja Simon muodostamaa ystävyysverkostoa. Olkoon  $\mathcal{L} = 8$ , jonka parit ovat  $\langle Allison, Drew \rangle$ ,  $\langle Allison, Ross \rangle$ ,  $\langle Drew, Simon \rangle$ ,  $\langle Drew, Eliot \rangle$ ,  $\langle Eliot, Drew \rangle$ ,  $\langle Keith, Ross \rangle$ ,  $\langle Ross, Simon \rangle$  ja  $\langle Simon, Drew \rangle$ . Näin ollen  $\mathcal{L}$  tekijät ovat  $l_1 = \langle Allison, Drew \rangle$ ,  $l_2 = \langle Allison, Ross \rangle$ , ... ja  $l_8 = \langle Simon, Drew \rangle$ . Ystävyysuhteet voidaan tulkita seuraavalla tavalla. Allison on Drewin ystävä ja Allison on Rossin ystävä. Drew on Eliotin ystävä ja Eliot Drewin ystävä ja niin edelleen. Koska ystävyysverkosto on yksisuuntainen niin ystävyysuhteet eivät ole molemminpuolisia. Esimerkiksi Allison on Drewin ystävä mutta Drew ei ole Allisonin ystävä.

### 2.4.1 Sosiogrammi

Sosiogrammi on Morenon 1930-luvulla kehittämä tekniikka sosiaalisten verkoston hahmottamiseen. Sosiogrammit ovat sosiaalisten verkostojen yksinkertaisia visualisointeja. Sosiogrammi on kuva missä sosiaaliset yksiköt, kuten ihmiset on esitetty pisteinä kaksiulotteisessa avaruudessa. Yksikköparien väliset yhteydet kuvataan sosiogrammissa viivoilla, jotka yhdistävät molemmat yksiköt. Yhteydet voivat olla suuntaamattomia (undirected) tai suunnattuja (directed), jolloin viivan kärkeen lisätään nuolenpää osoittamaan yhteyden suuntaa.



Kuva 2.3: Sosiogrammi kuuden toimijan ystävyysverkostosta.

## 2.5 Sosiometrinen notaatio

Sosiometrisella notaatiolla on erityinen asema sosiaalisten verkostojen laskennallisessa analyysissä. Sosiometrisessä notaatiotavassa verkostoja kuvataan taulukoilla ja matriiseilla. Laskennallisesti tämä merkintätapa on luonnollisin kolmesta esiteltävästä merkintätävästä, koska sosiometrisellä notaatiolla kuvattuja verkostoja voidaan myös esittää matriiseilla joita kutsutaan *sosiomatriiseiksi*. Erilaisten matriisilaskentamenetelmien avulla sosiomatriisista voidaan selvittää verkostoon liittyviä tunnuslukuja. Wassermann ja Faust esittelevät laajan joukon matriisilaskentaa hyödyntäviä menetelmiä tunnuslukujen laskemiseen joista olennaisimpia tämän diplomityön kannalta on esitelty luvussa 2.6.

Sosiometriikaksi kutsutaan ihmisistä ja heidän välillä olevista mitattavista, yleensä dikotomisista *merkityksellisistä yhteyssuhteista* (affective relation) muodostuvaa sosiaalisen verkoston datajoukkoa. Sosiometrisia yhteyssuhteita kuvataan voimak-

kaasti. Tämänkaltaisia suhteita ovat muun muassa pitää/ei pidä ja ystävä/vihamies-suhteet.

Alkuperäinen määritelmä merkityksellisistä yhteyssuhteista on sosiaalitieteilijöiden näkemys yhteyksien muodostumisesta ihmisten välille, mutta yleisesti mikä tahansa määrällisesti mitattava ihmisiä yhdistävä tekijä voidaan lukea merkitykselliseksi yhteyssuhteeksi.

Yhteyksien muodostamaa dataa esitetään yleisesti matriiseilla, joita kutsutaan sosiomatriiseiksi. Sosiomatriisiin kaksi dimensiota, rivit ja sarakkeet, indeksoivat *lähtettäviä toimijoita* (sending actor) ja *vastaanottavia toimijoita* (receiving actor).

	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$
$n_1$	–	1	0	1	0	0
$n_2$	0	–	1	0	0	1
$n_3$	0	1	–	0	0	0
$n_4$	0	0	0	–	0	1
$n_5$	0	0	0	1	–	0
$n_6$	0	1	0	0	0	–

Taulukko 2.1: Kuuden toimijan  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  muodostaman ystävyssuhdeverkoston sosiogrammi listamuodossa. Listan rivit ilmaisevat lähteviä yhteyksiä ja sarakkeet saapuvia yhteyksiä. Esimerkiksi sarakkeelle  $n_1$  eli Allisonille ei tule yhtään yhteyttä mutta rivillä  $n_1$  Allison muodostaa yhteyden toimijoihin  $n_2$  ja  $n_4$ .

Taulukossa 2.1 ei ole sallittu että tekijä voi olla ystävä itsensä kanssa, jolloin toimijan yhteyden arvoa itsensä kansa esitetään – merkillä. Toinen esitystapa ystävyssuhdeverkoston sosiogrammille on mm. tunnuslukujen laskennallisessa käytetty sosiomatriisi.

### 2.5.1 Sosiomatriisi

Sosiomatriisi on tärkein sosiaalisten verkostojen analysoinnissa käytetty matriisi. Graafiteoreetikot kutsuvat tätä matriisia *vieruspistematriisiksi* (Ruohonen, 2006) tai *yhteysmatriisiksi* (adjacency matrix) koska matriisin arvot määrittävät ovatko solmut, eli tämän diplomityön kontekstissa toimijat, yhteydessä toisiinsa. Sosiaalisten verkostojen analyysissä tätä matriisia kutsutaan yleisesti sosiomatriisiksi.

Yksimoodillisille verkostoille, joissa on  $g$  tekijää on sosiomatriisin koko aina  $g \times g$  ( $g$  riviä ja  $g$  saraketta).

### 2.5.2 Arvotettu sosiomatriisi

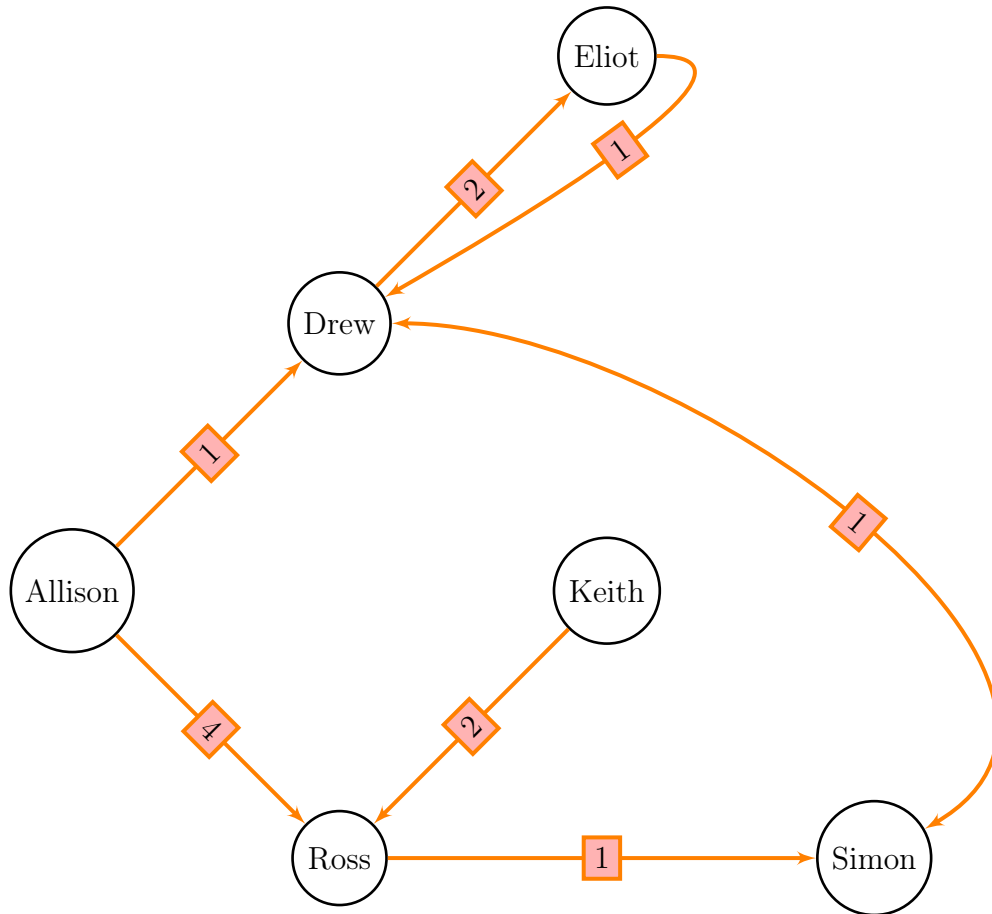
Sosiomatriisit ovat tyypillisesti dikotomisissa ts. sosiomatriisien alkioiden arvot ovat binäärisiä ja ne ilmaisevat yhteyden olemassaoloa. Arvotettu sosiomatriisi ilmaisee

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Taulukko 2.2: Ystävyysverkosto esitettynä siosimatriisina.

saman kuin tavallinen siosimatriisi eli yhteyden olemassaolon, mutta lisäksi se voi sisältää tulkinnan yhteyksien määrästä tai niiden voimakkuudesta. Arvotetussa siosimatriisissa nolla ilmaisee yhteydetttömyyttä ja sitä suuremmat luvut yhteyden olemassaoloa.

Tarkastellaan kappaleessa 2.4.1 esitettyä kuvan 2.3 ystävyysuhdeverkostoa. Muodostukoon nyt sosiaalisen verkoston yhteydet kuvan 2.4 mukaisesti lähettyjen tekstiviestien mukaan. Allisonin lähettäessä yhden viestin Drewille heidän välilleen muodostuu yhteys, jonka arvo on yksi. Drewin lähettäessä Elliotille kaksi viestiä, heidän välilleen muodostuu myös yhteys, jonka arvo on kaksi. Nyt suunnatussa verkostossa nuolen suunta osoittaa viestin päämäärän ja nuolen painoarvo viestien lukumäärän.



Kuva 2.4: Sosiogrammi kuuden toimijan lähettämistä tekstiviesteistä.

Tämän verkoston arvoitettu sosiomatriisi on esitetty kuvassa 2.5. Sosiomatriisin diagonaali on nolla, koska kukaan ei ole lähettänyt itselleen viestejä. Sosiomatriisin suurin arvo neljä, joka on myös suurin yksittäisen henkilön lähettämien viestien määrä toiselle henkilölle.

$$\mathbf{X} = \begin{bmatrix} 0 & 1 & 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Taulukko 2.3: Ystävien lähettämien tekstiviestin muodostaman arvoitetun verkoston sosiomatriisi.

	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$
$n_1$	–	1	0	4	0	0
$n_2$	0	–	2	0	0	1
$n_3$	0	1	–	0	0	0
$n_4$	0	0	0	–	0	1
$n_5$	0	0	0	2	–	0
$n_6$	0	1	0	0	0	–

Taulukko 2.4: Kuuden toimijan  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  toisilleen lähettämien tekstiviestin muodostama arvotettu ystävyys-suhdeverkosto listamuodossa. Listan rivit ilmaisevat lähettyjen viestien määrää ja sarakkeet saapuvien viestien määrää. Esimerkiksi sarakkeelle  $n_1$  eli Allisonille saa yhtään viestiä mutta rivillä  $n_1$  Allison lähettää  $n_2$  Drewille yhden ja  $n_4$  Rossille neljä viestiä.

Toinen yleinen esitystapa sosiomatriiseille on verkoston esitys listana. Lista poikkeaa sosiomatriisista esitysmuotona. Listassa esitetään sarakkeilla ja riveillä sekä tekijät että tekijöiden yhteyksien arvo. Jos sosiomatriisin tekijät eivät muodosta yhteyksiä itsensä kanssa, on niiden arvo tällöin nolla, listamuodossa nollan sijasta käytetään tyypillisesti viiva-merkkiä. Kuvan 2.4 sosiomatriisi on esitetty listamuodossa taulukossa 2.4.

Sosiomatriisissa matriisin rivit ja sarakkeet ilmaisevat lähetettyjen ja vastaanotettujen viestien lukumäärää. Laskemalla nämä rivien ja sarakkeiden arvot yhteen saadaan kokonaismäärä verkostossa olevan yksittäisen henkilön lähettämistä ja vastaanottamista viesteistä. Itseasiassa rivisummat ja sarakesummat ovat yksinkertaisimpia tunnuslukuja joita voidaan laskea sosiomatriiseista.

## 2.6 Tunnusluvut

Yksi olennaisimmista graafiteorian sovellutuksista sosiaalisten verkostojen analyysissä on verkoston tärkeimpien tekijöiden tunnistaminen tunnuslukujen avulla. Verkostojen tekijöitä voidaan etsiä laskennallisilla menetelmillä, eli tunnusluvuilla. Laskennallisia menetelmiä verkoston tärkeimpien tai huomattavimpien tekijöiden löytämiseen on lukuisia. Laskennalliset menetelmät eroavat toisistaan kompleksisuuden ja lopputuloksen mukaan.

Sosiaalisen verkoston rakenteella on merkitys tunnuslukujen käytössä. Sekä yksittäisiä kaksimoodisille verkostoille niin myös suunnatuille ja suuntaamattomille verkostoille on olemassa omanlaiset tunnusluvut.

Tärkeimpien tekijöiden löytäminen on aina kontekstista riippuvaista toimintaa. Tunnuslukuja laskettaessa on oleellista tiedostaa minkälainen verkosto on kyseessä, miten verkosto on muodostettu ja minkälaisia tietoja verkostosta tai verkoston

toimijoista halutaan selvittää, jotta voidaan valita sopivimmat menetelmät tunnuslukujen tuottamiseksi.

Käytännössä kaikkia näitä tietoja ei ole saatavilla tai vaihtoehtoisesti ei vain tiedä mitä verkostosta halutaan selvittää tai millä tavoin, jolloin tunnuslukuja joudutaan laskemaan niin sanotusti umpimähkäisesti. Tunnuslukujen liittäminen verkoston visualisointeihin edesauttaa tuloksien tulkitsemista. Tunnuslukujen merkitykseen SLP-tapauksen kontekstissa palataan luvussa 5.3.

## Geodeesi

Kahden solmun välistä lyhyintä etäisyyttä kutsutaan *geodeesiksi* (geodesic). Lyhin etäisyys on kaarien lukumäärä jotka yhdistävät kaksi solmua. Geodeesi ei ole varsinaisesti tunnusluku, mutta sitä käytetään eräiden tunnuslukujen laskennassa. Kahden solmun  $n_i$  ja  $n_j$  välistä geodeesista etäisyyttä tai lyhyemmin ilmaistuna etäisyyttä merkitään  $d(i, j)$ . Lyhin geodeesinen etäisyys on mikä tahansa lyhin etäisyys kahden solmun välillä. Jos kahden solmun välillä ei ole olemassa reittiä eli solmut eivät ole saavutettavissa merkitään niiden geodeesiä äärettömänä tai määrittelemättömänä.

Suuntaamattomille graafeille pätee  $d(i, j) = d(j, i)$ , mutta suunnatuille graafeille  $d(i, j)$  voi olla erisuuri kuin  $d(j, i)$ . Taulukossa 2.5 on esitetty luvun 2.4.1 ystävyysverkoston toimijoiden lyhimät etäisyydet.

	<i>etäisyys</i>
$d(n_1, n_2)$	1
$d(n_1, n_3)$	2
$d(n_1, n_4)$	1
$d(n_1, n_6)$	2
$d(n_2, n_3)$	1
$d(n_2, n_6)$	1
$d(n_3, n_2)$	1
$d(n_3, n_6)$	2
$d(n_4, n_2)$	2
$d(n_4, n_3)$	3
$d(n_4, n_6)$	1
$d(n_5, n_2)$	3
$d(n_5, n_3)$	4
$d(n_5, n_4)$	1
$d(n_5, n_6)$	2
$d(n_6, n_2)$	1
$d(n_6, n_3)$	2

Taulukko 2.5: Luvun 2.4.1 ystävyysverkoston toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  geodeesit. Verkoston lyhin etäisyys on yksi ja pisin neljä.

### 2.6.1 Suuntaamattomien verkostojen tunnusluvut

Tässä aliluvussa esitellyt tunnusluvut on tarkoitettu pääsääntöisesti suuntaamattomille verkostoille, mutta niitä voidaan soveltaa myös suunnatuille verkostoille.

#### Astesumma

*Astesumma* (nodal degree) ilmaisee solmun asteen  $d(n_i)$  eli kaarien lukumäärän jotka ovat yhteydessä kyseiseen solmuun sosiaalisessa verkostossa. Toisaalta astesumma ilmaisee solmun verkottumisen asteen eli solmujen lukumäärän jotka ovat yhteydessä kaarien avulla tarkasteltavaan solmuun. Astesumma on pienimmillään 0 ja suurimmillaan  $g - 1$ , jos solmun astesumma on nolla niin solmua kutsutaan isolaatiksi eli eristäytyneeksi toimijaksi.

Laskennallisesti astesumma on hyvin helppo toteuttaa ja suorittaa. Huolimatta astesumman laskennallisesta helppoudesta se on hyvin informatiivinen tunnusluku. Sosiaalisessa verkostossa solmu jolla on pieni asteluku ilmaisee että se ei ole liittynyt kovinkaan moneen solmuun verrattuna solmuun jolla on suhteellisen suuri asteluku,



jolloin se on liittynyt hyvin moneen solmuun.

**Määritelmä 2.6.1.** Suuntaamattoman dikotomisen graafin sosiomatriisin  $\mathbf{X}$  astesumma määritellään yhtälöllä

$$d(n_i) = \sum_{j=1}^g x_{ij} = \sum_{i=1}^g x_{ij} = x_{i+} = x_{+j}.$$

$i$ :n rivin tai sarakkeen arvojen summa muodostaa solmun  $n_i$  astesumman.

## Välillisuus

*Välillisuus* (betweenness centrality) ilmaisee kuinka keskeinen solmu on, jos kyseinen solmu sijaitsee muiden solmujen geodeeseissa eli lyhimpien etäisyyksien varrella. Yksittäisellä solmulla on suuri välillisuus jos kyseisen solmun lävitse kulkee monia muita solmuja yhdistävät lyhimmät etäisyydet.

**Määritelmä 2.6.2.** Verkoston toimijan  $n_i$  välillisuus  $C_B(n_i)$  määritellään yhtälöllä

$$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk},$$

missä  $g_{jk}$  on toimijoita  $n_j$  ja  $n_k$  yhdistävien geodeesien lukumäärä. Normeerattu välillisuus määritellään vastaavasti

$$C'_B(n_i) = \frac{C_B(n_i)}{\frac{(g-1)(g-2)}{2}},$$

ja se saa arvoja välillä  $[0, 1]$ .

Kuvassa 2.10 on esitetty ystävyysuhdeverkoston tekijöiden välillisuus. Kuvassa 2.11 on esitetty tekstiviestien lähettämisen painotetun ystävyysuhdeverkoston välillisuus.

	$C_B(n_i)$
$n_1$	0
$n_2$	5.5
$n_3$	0
$n_4$	3.5
$n_5$	0
$n_6$	4

Taulukko 2.6: Luvun 2.4.1 ystävyysverkoston toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  välillisyydet. Tuloksista voidaan huomata että välillisuus on nolla niillä toimijoilla, joiden kautta ei kulje yhtään yhteyttä. Verkostossa tällaisia toimijoita ovat Allison ja Keith. Välillisuus on lisäksi nolla niillä toimijoilla joiden kautta ei kulje yhtään lyhintä etäisyyttä kuten Elliotin tapauksessa.

	$C_B(n_i)$
$n_1$	0
$n_2$	5.2
$n_3$	0
$n_4$	3.8
$n_5$	0
$n_6$	4

Taulukko 2.7: Luvun 2.5.2 ystävyysverkoston toimijoiden välillisuus. Kuten yllä samat toimijat  $n_1 = Allison$ ,  $n_3 = Eliot$  ja  $n_5 = Keith$  saavat välillisyydeksi nollan. Tekstiviesteillä arvotettu verkosto ei vaikuta toimijoihin, jotka saavat arvoksi nolla, sillä välillisuus on tunnuslukuna riippuvainen verkoston muodosta eli toimijoiden sijoittumisesta verkostoon. Drewin ja Rossin välillisuus muuttuvat hieman verkoston toimijoiden välisten yhteyksien painotusten muuttuessa arvoiltaan ykköistä suuremmaksi.

## 2.6.2 Suunnattujen verkostojen tunnusluvut

Tässä aliluvussa esitellyt tunnusluvut on tarkoitettu vain suunnatuille verkostoille.

### Vienti- ja tuontiluvut

Suunnatuille graafeille on olemassa oma vastineensa astesummalle. Vienti- ja tuontiluvut ilmaisevat kuinka paljon tiettyyn solmuun tulee yhteyksiä ja kuinka paljon tietystä solmusta lähtee yhteyksiä. Suuntaamaton graafi on tavallaan erikoistapaus suunnatusta graafista, kun jokainen solmuun liittyvä yhteys on molempisuuntainen.

**Määritelmä 2.6.3.** Suunnatun verkoston toimijan  $n_i$  *vientiluku* (out degree) määritellään yhtälöllä

$$d_O(n_i) = \sum_{j=1}^g x_{ij} = x_{i+}.$$

Solmun vientiluku on solmusta lähtevien yhteyksien määrä.

**Määritelmä 2.6.4.** Suunnatun verkoston toimijan, solmun  $n_i$  *tuontiluku* (indegree) määritellään yhtälöllä

$$d_I(n_i) = \sum_{j=1}^g x_{ji} = x_{+i}.$$

Solmun tuontiluku on solmuun tulevien yhteyksien määrä.

	$d_I(n_i)$	$d_O(n_i)$
$n_1$	0	2
$n_2$	3	2
$n_3$	1	1
$n_4$	2	1
$n_5$	0	1
$n_6$	2	1

Taulukko 2.8: Luvun 2.4.1 ystävyysverkon toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  vienti- ja tuontiluvut. Sarakkeessa  $d_I(n_i)$  on toimijaan tulevien yhteyksien määrä ja sarakkeessa  $d_O(n_i)$  toimijasta lähtevien yhteyksien määrä.

	$d_I(n_i)$	$d_O(n_i)$
$n_1$	0	5
$n_2$	3	3
$n_3$	2	1
$n_4$	6	1
$n_5$	0	2
$n_6$	2	1

Taulukko 2.9: Luvun 2.5.2 ystävien lähettämien tekstiviestiverkon toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  vienti- ja tuontiluvut. Sarakkeessa  $d_I(n_i)$  on toimijan vastaanottamien viestien määrä ja sarakkeessa  $d_O(n_i)$  toimijan lähettämien viestien määrä.

Verkostot, joista taulukoiden 2.8 ja 2.9 vienti- ja tuontiluvut on laskettu, ovat hyvin samankaltaisia, sillä erotuksella että taulukon 2.9 tuloksiin pohjautuva verkosto sisältää tulkinnan siitä kuinka monta viestiä jokainen toimija on lähettänyt. Näin ollen tästä verkostosta laskettavat tunnusluvut eivät ilmaise kuinka monta sisään-tai ulosmenevää yhteyttä verkoston tekijöillä on. Sen sijaan tunnusluvut ilmaisevat verkoston toimijoiden lähettämien ja vastaanottamien viestien kokonaismäärät.

### Toimijan arvostusaste

*Toimijan arvostusaste* (actor degree prestige) on yksinkertaisin arvostuksen tunnusluku. Se määritellään toimijaan kohdistuvien yhteyksien summana.

**Määritelmä 2.6.5.** Suunnatun verkoston toimijan  $n_i$  arvostusaste  $P_D(n_i)$  määritellään yhtälöllä

$$P_D(n_i) = d_I(n_i) = \sum_{j=1}^g x_{ji} = x_{+i},$$

missä  $d_I(n_i)$  on toimijan  $n_i$  tuontiluku ja  $x_{+i}$  on graafin sosiomatriisin  $X$  sarekesumma sarakkeesta  $i$ .

Arvostusasteet voidaan normeerata, jolloin ne ovat keskenään vertailukelpoisia eri verkostojen välillä.

**Määritelmä 2.6.6.** Suunnatun verkoston toimijan  $n_i$  normeerattu arvostusaste  $P'_D(n_i)$  määritellään yhtälöllä

$$P'_D(n_i) = \frac{x_{+i}}{g-1},$$

missä  $g$  on verkoston tekijöiden lukumäärä ja  $x_{+i}$  on graafin sosiomatriisin  $X$  sarekesumma sarakkeesta  $i$ .

	$P'_D(n_i)$
$n_1$	0
$n_2$	0.6
$n_3$	0.2
$n_4$	0.4
$n_5$	0
$n_6$	0.4

Taulukko 2.10: Luvun 2.4.1 ystävyssyhdverkoston toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  arvostusasteet. Arvostetuimmat toimijat ovat  $n_3 = Eliot$  ja  $n_5 = Keith$ , jotka saavat suhteessa muihin toimijoihin eniten itseensä tulevia yhteyksiä.  $n_1 = Allison$  ja  $n_5 = Keith$  arvostusaste on nolla, koska heihin ei tule yhtään yhteyttä.

	$P'_D(n_i)$
$n_1$	0
$n_2$	0.6
$n_3$	0.4
$n_4$	1.2
$n_5$	0
$n_6$	0.4

Taulukko 2.11: Luvun 2.5.2 arvotetun ystävyssuhdeverkoston toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  arvostusasteet. Verkoston rakenteen pysyessä samana, mutta yhteyksien määrän muuttuessa  $n_3 = Eliot$  ja  $n_5 = Keith$  saavat yhä arvostusasteeksi nolla.  $n_4 = Ross$  muodostuu verkoston arvostetuin henkilö, sillä häneen tulevien yhteyksien määrä kasvaa suhteessa suuremmaksi kuin muiden toimijoiden.

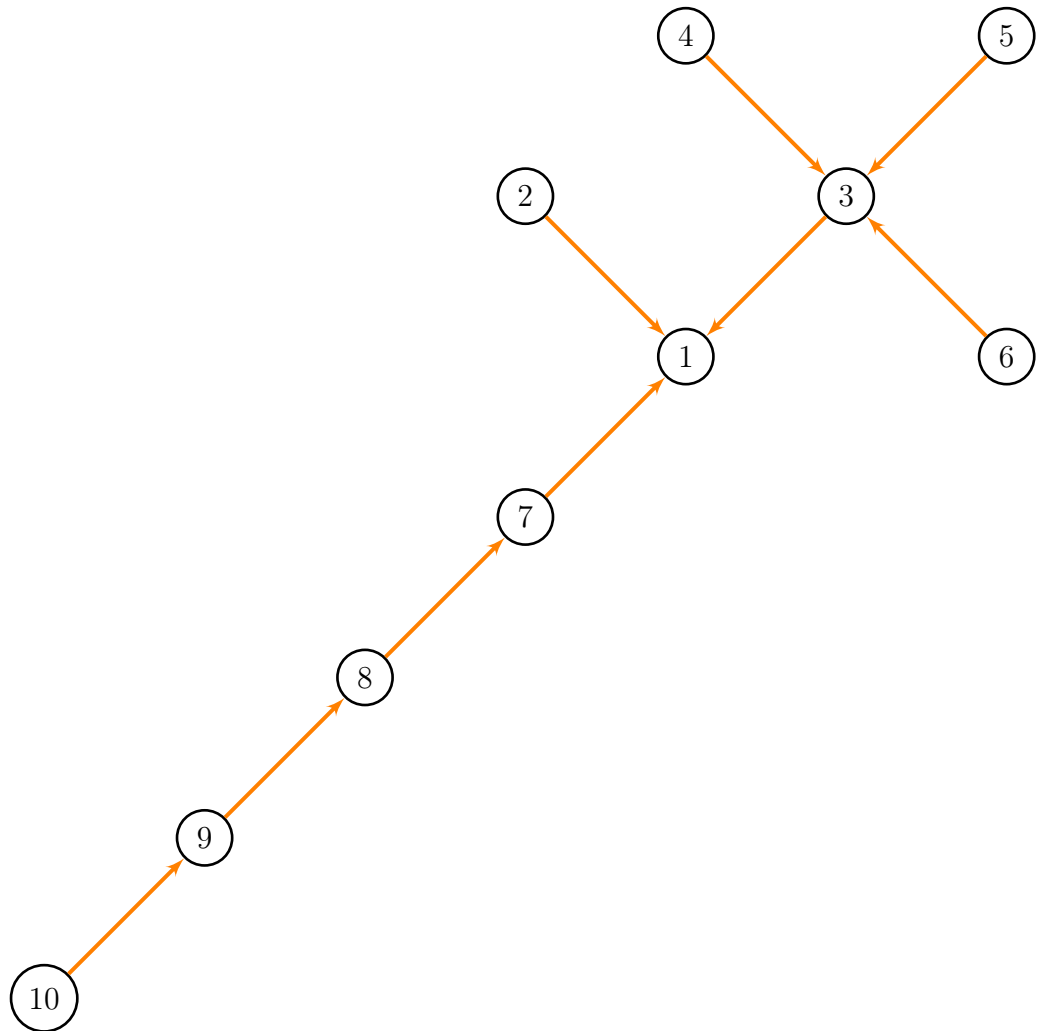
### Toimijan arvostusläheisyys

Arvostusaste  $P'_D(n_i)$  huomioi vain ne toimijat  $n_j$  jotka ovat suoraan yhteydessä toimijaan  $n_i$ . Määrittelemällä vaikutusjoukko (influence domain) voidaan laajentaa toimijan  $n_i$  arvostusaste toimijan *arvostusläheisyydeksi* (actor proximity prestige). Toimijan  $n_i$  vaikutusjoukko on niiden toimijoiden  $n_j$  joukko joista toimija  $n_i$  on suoraan tai epäsuorasti saavutettavissa. Toimijalle  $n_i$  voidaan määrittää luku  $I_i$  joka ilmoittaa toimijoiden  $n_j$  lukumäärän jotka voivat saavuttaa toimijan  $n_i$ . Tämän joukon mahtavuutta eli joukon toimijoiden lukumäärää nimitetään luvulla  $I_i$ . Luvun  $I_i$  avulla voidaan määrittää toimijan  $n_i$  vaikutusjoukon  $I_i$  toimijoiden  $n_j$  keskimääräinen etäisyys toimijaan  $n_i$ .

$$d(n_j, n_i)_{ave} = \frac{\sum_j d(n_j, n_i)}{I_i}$$

Kuvassa 2.5 on hahmoteltu Mrvarin (2010) esittämää esimerkkiä vaikutusjoukosta. Kuvassa kaikki toimijat ovat toimijan 1 vaikutusjoukossa. Toimijat 2, 3 ja 7 ovat tärkeämpiä toimijalle 1 kuin toimijat 4, 5, 6 ja 8. Toimijat 9 ja 10 vaikuttavat edellämainittuja toimijoita vähemmän toimijan yksi arvostusläheisyyteen. Toimijaan kohdistuvia välittömiä yhteyksiä voidaan painottaa suhteessa välillisiin yhteyksiin laskemalla keskimääräinen etäisyys toimijaan vaikutusjoukon sisällä.

Toimijan  $n_i$  arvostusläheisyys  $P_P(n_i)$  määritellään toimijan  $n_i$  saavuttavien toimijoiden  $n_j$  suhteena keskimääräiseen etäisyyteen toimijoista  $n_j$  toimijaan  $n_i$ .



Kuva 2.5: Kuvassa on kymmenen toimijan verkosto. Verkoston tärkeimmät toimijat vaikutuksen suhteen ovat toimijat yksi, kolme ja seitsemän. Toimijan yksi vaikutusjoukkoon kuuluvat kaikki verkoston toimijat. Toimijan kolme vaikutusjoukkoon kuuluvat toimijat neljä, viisi ja kuusi ja toimijan seitsemän vaikutusjoukkoon kuuluvat toimijat kahdeksan, yhdeksän ja kymmenen. Vaikutusjoukon hahmottamiseksi voidaan kuvitella käänteistä analogiaa hierarkiasta jossa hierarkia ketjun, eli toimijoiden suunnattuja yhteyksiä seuraamalla, päässä olevat toimijat ovat tärkeimpiä.

**Määritelmä 2.6.7.** Suunnatun verkoston toimijan  $n_i$  arvostusläheisyys määritellään yhtälöllä

$$P_P(n_i) = \frac{I_i/(g-1)}{\sum_j d(n_j, n_i)/I_i}.$$

Arvostusläheisyys saa arvoja väliltä  $[0, 1]$ . Jos kaikki toimijat  $n_j$  ovat suoraan yhteydessä toimijaan  $n_i$  saa toimija  $n_i$  arvostusläheisyydeksi  $P_P(n_i)$  arvon yksi. Jos toimija ei ole saavutettavissa, on vaikutusjoukon  $I_i$  koon ja arvostusläheisyyden  $P_P(n_i)$  arvo nolla.

Yksittäisten toimijoiden arvostusläheisyyksien vertailun helpottamiseksi voidaan määrittää verkoston keskiarvostusläheisyys. Keskiarvostusläheisyys kuvaa kuinka läheisesti toimijat arvostavat toinen toisiaan verkoston sisällä.

**Määritelmä 2.6.8.** Suunnatun verkoston keskiarvostusläheisyys määritellään yhtälöllä

$$\bar{P}_P = \sum_{i=1}^g \frac{P_P(n_i)}{g}$$

Taulukossa 2.12 on esitetty ystävyysuhdeverkoston toimijoiden arvostusläheisyydet.

	$P_P(n_i)$
$n_1$	0
$n_2$	0.625
$n_3$	0.417
$n_4$	0.4
$n_5$	0
$n_6$	0.625

Taulukko 2.12: Luvun 2.4.1 ystävyysverkoston toimijoiden  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  arvostusläheisyydet. Allisonin ja Keithin arvostusläheisyys on nolla koska ne sijaitsevat verkoston päässä. Kumpaankaan toimijaan ei kohdistu yhtään yhteyttä mutta molemmat ovat yhteydessä muihin toimijoihin. Arvostusläheisyys on suuri Drewilla ja Simonilla koska heidän kautta verkoston muut toimijat ovat yhteydessä muihin verkoston toimijoihin.

### Arvoasema

Edellä esitetyt keskeisyyden ja arvostuksen tunnusluvut huomioivat vain toimijoita ja toimijoihin kohdistuvien yhteyksien määrää sekä toimijoiden geodeesejä. Toimijan  $n_i$  arvoasema on riippuu häneen kohdistuvista toimijoista  $n_j$  ja toimijoiden  $n_j$  arvoasteista. Toimijoiden  $n_j$  arvoaste puolestaan on riippuvainen vastaavasti itseensä

kohdistuvista toimijoista  $n_k$  ja heidän arvoasteista ja niin edelleen. Tämä ongelma voidaan kuvata lineaarisella yhtälöryhmällä, jossa jokaisen toimijan  $n_i$  arvoaste on funktio kyseisen toimijaan kohdistuvien toimijoiden  $n_j$  arvoasteista.

Dikotominen sosiomatriisi  $\mathbf{X}$  esittää verkoston toimijoiden välisten yhteyksien olemassaoloa. Jos sosiomatriisin sarakkeen  $i$  ja rivin  $j$  arvo on yksi on toimijoiden  $i$  ja  $j$  välillä yhteys ja jos arvo on nolla yhteyttä ei ole olemassa. Kun verkostossa on  $g$  toimijaa voidaan toimijan  $n_i$  arvoasema esittää sosiomatriisin  $\mathbf{X}$  sarakkeen alkioden ja niitä vastaavien toimijoiden arvoasemien lineaarikombinaationa.

$$P_R(n_i) = x_{1i}P_R(n_1) + x_{2i}P_R(n_2) + \dots + x_{gi}P_R(n_g), i = 1, \dots, g$$

Toimijan  $n_i$  arvoaste on riippuvainen häneen kohdistuvien toimijoiden  $n_1, \dots, n_g$  toimijoista. Matemaattisesti ilmaistuna kyseessä on  $g$  lineaarisen yhtälön systeemi jossa on  $g$  tuntematonta muuttujaa. Ongelmaa voidaan esittää matriisiyhtälönä

$$\mathbf{p} = \mathbf{X}^T \mathbf{p}$$

missä  $\mathbf{p} = (P_R(n_1), P_R(n_2), \dots, (P_R(n_g)))'$  on pystyvektori ja  $\mathbf{X}$  on sosiomatriisi. Yhtälöä järjestelemällä se voidaan esittää muodossa

$$(\mathbf{I} - \mathbf{X}^T)\mathbf{p} = \mathbf{0}$$

missä  $\mathbf{I}$  on identiteettimatriisi ja  $\mathbf{0}$   $g$ -pituinen pystyvektori. Tämä yhtälö on matriisin ominaisarvoyhtälö ja sen laskemiseen voidaan käyttää PageRank-algoritmia.

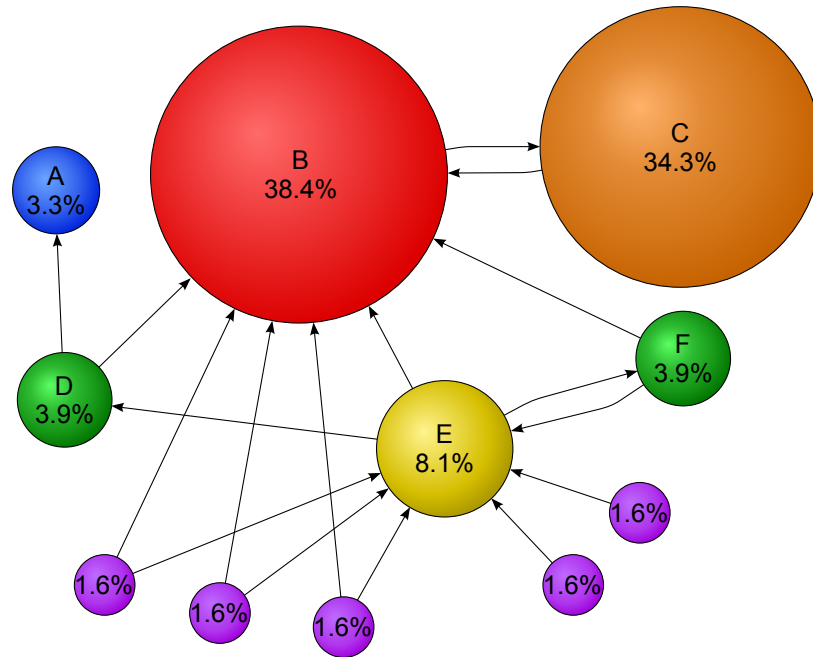
## PageRank

PageRank on Googlen perustajien Larry Pagen ja Sergey Brin (Page et al. 1999, Moler 2009) kehittämä algoritmi. PageRank-algoritmi on alunperin tarkoitettu web-sivustojen arvottamiseen niiden muodostamien linkkiverkostojen perusteella. Algoritmi noudattelee pääpiirteittäin arvoasema-tunnusluvun määritelmää.

PageRank on numeerinen arvo, joka esittää kuinka tärkeä toimija on. Pagerank algoritmia voidaan hahmottaa äänestys-analogialla (vrt. Google's PageRank Explained). Olkoon toimijaverkosto jossa toimijat äänestävät toisiaan muodostaen suunnattuja yhteyksiä. Toimijat voivat äänestää yhtä tai useampaa verkoston toimijaa. Näin ollen toimijan arvo verkostossa on riippuvainen hänen saamistaan äänistä sekä äänen laadusta. Äännet eivät ole tasavertaisia toimijoiden välillä. Äänestysmäärältään suosituimpien toimijoiden antamat äännet merkitsevät enemmän kuin pienempien toimijoiden äännet. PageRankia laskettaessa jokaisen äänen tärkeys otetaan huomioon. Kuvassa 2.6 on esimerkkikuva<sup>1</sup> äänestysverkoston toimijoiden lopputila.

<sup>1</sup>PageRank-Example. <http://en.wikipedia.org/wiki/File:PageRanks-Example.svg>





Kuva 2.6: Kuvitteellisen äänestystilanteen lopputila, jossa toimijoiden PageRank on ilmoitettu väliltä 0-100 %. Toimijoiden kokoa on paisutettu niiden suhteellisen merkittävyyden mukaisesti. Merkittävimmät toimijat ovat *B*, *C* ja *E*.

Alla esitetty yhtälö on alkuperäinen Googlen julkaisema PageRank algoritmi (Brin & Page, 1998). Laskettaessa toimijan PageRankia otetaan huomioon kaikki toimijaan kohdistuvat suunnatut yhteydet. Toimijan *A* PageRank  $PR(A)$  on summa toimijoiden  $t_1 \dots t_n$  PageRankeista, jotka muodostavat suunnatun yhteyden häneen.  $C$  on toimijasta lähtevien yhteyksien määrä ja  $d$  on vaimennuskerroin, jonka arvo on yleensä 0.85.

**Määritelmä 2.6.9.** Toimijan *A* PageRank määritellään yhtälöllä

$$PR(A) = (1 - d) + d * \left( \frac{PR(t_1)}{C(t_1)} + \dots + \frac{PR(t_n)}{C(t_n)} \right).$$

Kaava voidaan lukea muodossa: toimijan *A* PageRank on  $0.15 + 0.85$  kertaa osuus muiden toimijoiden PageRankeista, joista on suunnattu yhteys toimijaan *A*. Osuudella tarkoitetaan toimijan PageRankia jaettuna toimijasta lähtevien yhteyksien määrällä.

Toimijoiden äänillä on tietty painoarvo, joka on riippuvainen heidän PageRankeistaan. Toimijan äänestäessä hän välittää osan PageRankistaan eteenpäin, määritelmän 2.6.9 yhtälössä  $0.85$  kertaa toimijan oma PageRank. Tämä osuus PageRankeista jaetaan kaikkien toimijoiden kesken joita äänestetään, eli jaetaan toimijasta lähtevien yhteyksien määrällä. Koska toimijan välittämä PageRank on suhteellinen lähtevien yhteyksien määrään niin äänestettävälle toimijalle äänestävä toimija jonka  $PR = 4$  ja josta on 5 uloslähtevää yhteyttä on arvokkaampi kuin äänestävä toimija,

jonka PR = 10 ja josta 100 uloslähtevää yhteyttä. Toimijoiden PageRank ei vähene, vaikka ne välittivät äänestämällä omaa PageRankiaan eteenpäin muille toimijoille.

PageRankin tuottamiseen on erilaisia laskennallisia menetelmiä, joista ehkä suosituin on matriisilaskentaa käyttävä potenssiinkorotus-menetelmä (Power Method) (Langville & Meyer, 2004), jota ei tarkastella syvällisemmin tämän diplomityön puitteissa. Taulukoissa 2.13 ja 2.14 on esitetty ystävyyssuhdeverkostojen toimijoiden PageRankit. Laskentaan on käytetty Python-kielistä Igraph (The igraph project) kirjastoa.

	$P(n_i)$
$n_1$	0.025
$n_2$	0.429
$n_3$	0.208
$n_4$	0.057
$n_5$	0.025
$n_6$	0.256

Taulukko 2.13: Luvun 2.4.1 Kuuden toimijan  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  muodostaman ystävyyssuhdeverkoston PageRankit. Pienimmät PageRankit muodostuvat toimijoille  $n_1 = Allison$  ja  $n_5 = Keith$ , joihin ei tule ollenkaan yhteyksiä.  $n_2 = Drew$  on PageRankilla mitattuna verkoston merkittävin toimija ja seuraavaksi merkittävimmät ovat  $n_3 = Eliot$  ja  $n_6 = Simon$ .

	$P(n_i)$
$n_1$	0.025
$n_2$	0.423
$n_3$	0.265
$n_4$	0.063
$n_5$	0.025
$n_6$	0.199

Taulukko 2.14: Luvun 2.5.2 kuuden toimijan  $n_1 = Allison$ ,  $n_2 = Drew$ ,  $n_3 = Eliot$ ,  $n_4 = Ross$ ,  $n_5 = Keith$  ja  $n_6 = Simon$  lähetettävien tekstiviestien muodostaman ystävyyssuhdeverkoston PageRankit. Edelliseen taulukkoon 2.13 verrattuna PageRankit muuttuvat hyvin vähän. Verkoston rakenne ei muutu mutta toimijoiden välisten yhteyksien määrän muutoksella on vaikutus PageRankeihin. Verkoston reunalla sijaitsevat Allison ja Keith saavat edelleen pienimmät PageRankit. Eliotin ja Simonin PageRankit muuttuvat huomattavammin. Eliotista tulee Simonia merkittävämpi toimija, koska verkoston merkittävimmän toimijasta  $n_2 = Drew$  lähtee nyt kaksi yhteyttä tai viestiä Eliotille mutta vain yksi Simonille, tällöin määritelmän 2.6.9 mukaisesti Drew saa  $d * 2/3$  Drewin PageRankista ja Simon vain  $d * 1/3$  Drewin PageRankista.

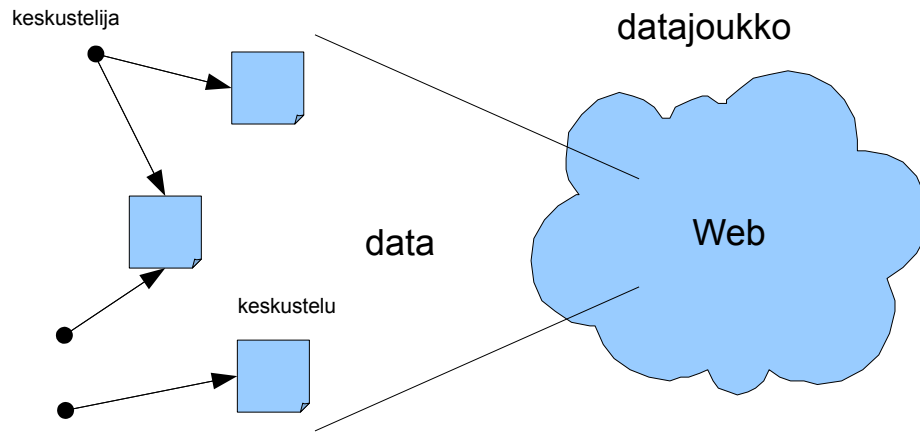
### 3. DATALÄHTÖINEN ANALYYSI

Kynä ja paperi ovat säilyttäneet asemansa sosiaalisten verkostojen tutkijoiden työvälineinä, mutta verkostojen koon kasvaessa muutamista solmuista satoihin ja tuhansiin solmuihin on tukeuduttava laskennallisiin menetelmiin. Tänä päivänä on olemassa useita erilaisia kehittyneitä ohjelmistoja tiedonlouhintaan ja sosiaalisten verkostojen analysointiin louhitusta tiedosta. Tämä kuitenkin tarkoittaa sitä, että hyvin harvat ohjelmistot osaavat tehdä sekä tiedonlouhintaa ja soveltaa SNA menetelmiä. Näin ollen tutkijat joutuvat käyttämään useita erilaisia ohjelmia, ensiksi kerätäkseen analyysin kannalta olennaisen tiedon ja toiseksi analysoidakseen kerätyt tiedot ja visualisoidakseen tulokset.

SNA-ohjelmistot voivat olla myös vaikeakäyttöisiä. Esimerkiksi Pajek on ehkä eräs monipuolisimmista sosiaalisten verkostojen analysointiohjelmistoista, mutta sen graafinen ulkoasu on hyvin pelkistetty ja ohjelmaan sisälle pääsemisessä voi olla korkea oppimiskynnys käyttäjälle.

Tässä diplomityössä informaation analyysi ja visualisointi suoritetaan datalähtöisesti (vrt. Nykänen et al. 2007). Informaation visualisointiprosessia mukaillen fyysinen ympäristö, Web 2.0 -kontekstissa oleva sosiaalisen median verkkopalvelu, josta data haetaan on dynaaminen alati muutoksessa oleva datan lähde. Työvälineet jotka diplomityön SLP-tapauksessa on kehitetty noudattelevat datalähtöistä analyysia ja visualisointia, kytkeytymällä datalähteeseen ja tuottamalla dynaamisesti visualisointeja. Työvälineen toteutuksessa on sovellettu kontekstiherkkää automatisoitua analyysia (vrt. Huhtamäki et al, 2010).

Kuvan 3.1 mukaisesti suuresta datajoukosta, Webistä, otetaan projektioita eli pienempiä datakokonaisuuksia, keskustelualueita, joista muodostetaan sosiaalisia verkostoja, joita analysoidaan ja visualisoidaan.



Kuva 3.1: Projektien ottaminen datajoukosta

### Datan kerääminen

Tämän diplomityön kontekstissa datan keräys tapahtuu Web 2.0 ympäristössä. Datan keräyksen kohteena ovat web-pohjaiset keskustelualueet. Näistä keskustelualueista on esiprosessoitu tietyn ennaltamäärätyn muodon omaavia ATOM-syötteitä, jotka kuvaavat yksittäisiä keskustelunaiheita. ATOM-syötteitä haetaan palvelimelta asynkronisesti ja käsitellään paikallisesti.

### Analysointimenetelmät

Verkkopalveluita voidaan arvostaa erilaisin analysointimenetelmin. Eräs menetelmä on käytönseuranta, jossa kerätään ja analysoidaan tietoa käyttäjien eri toimista verkkopalvelussa. Analysointimenetelmien avulla saadaan mm. tietoa navigointipoluista ja käyttäjien välisistä suhteista. Tässä diplomityössä analysointiin on käytetty sosiaalisten verkostojen analysointimenetelmiä.

### Visualisointimenetelmät

Datan visualisointimenetelmänä tai visualisointikeinona käytetään graafeja. Graafit ovat vain yksi tapa monien joukossa esittää sosiaalisia verkostoja mutta graafien etuina ovat niiden helppo ymmärrettävyys ja tulkittavuus (Ware, 2004).

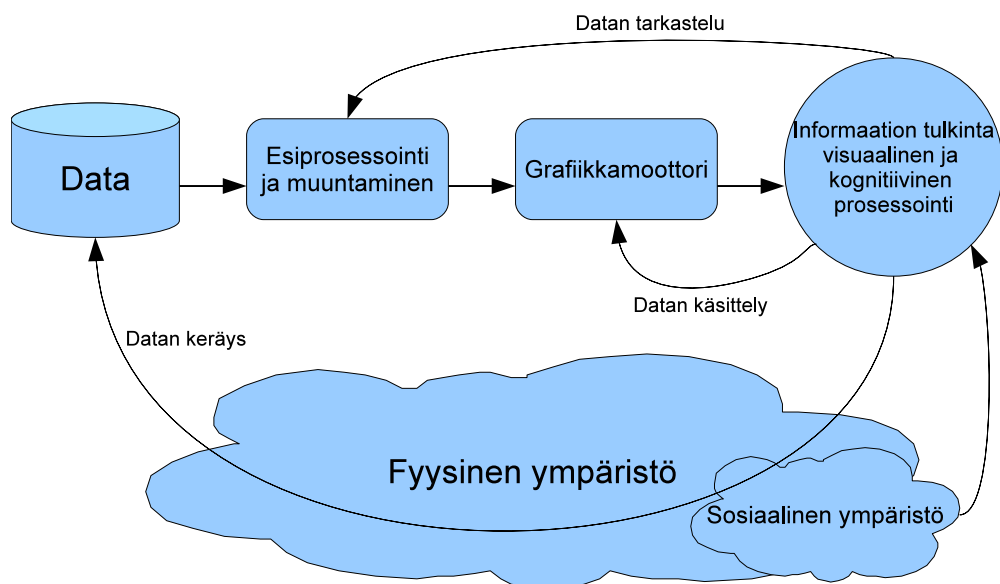
### Informaation visualisointi

Ihmisaivot ovat pitkälle erikoistuneet visuaalisen informaation vastaanottamiseen, käsittelyyn ja tulkitsemiseen. Aivot yrittävät löytää näköaistin välityksellä saadusta informaatiosta tuttuja piirteitä ja kuvioita. Käytössämme on siis erittäin monimutkainen ja tehokas hahmontunnistusjärjestelmä (Ware, 2004).

Informaation visualisoinnin avulla dataa voidaan esittää ihmisille kognitiivisesti helpommin tulkittavassa muodossa. Informaation visualisoinnilla on pieni mutta hyvin merkittävä osa meidän kognitiivisessa järjestelmässämme. Ihmiset keräävät näkönsä avulla enemmän informaatiota kuin kaikki muut aistit tuottavat yhteensä. Ware (2004) luettelee etuja joita voidaan saavuttaa informaatiota visualisoimalla: visualisointi tarjoaa keinon ymmärtää suuria määriä dataa, visualisointi paljastaa usein tietoa datan keräystavasta, visualisointi mahdollistaa yllättävien ja odottamattomien piirteiden hahmottamisen datajoukosta.

### Informaation visualisointiprosessi

Prosessia jolla dataa muokataan koneellisesti ihmisille tulkittavaksi informaatioksi – yleensä kuviksi, kutsutaan informaation visualisointiprosessiksi. Ware (2004) määrittelee informaation visualisointiprosessin koostuvan neljästä eri vaiheesta, jotka ovat datan keräys ja säilytys, datan esiprosessointi ja muuntaminen, tiedon esityslaitteisto ja graafiset algoritmit sekä ihmisen havainnointi ja kognitiivinen järjestelmä. Visualisointiprosessia on havainnollistettu kuvassa 3.2.



Kuva 3.2: Informaation visualisointiprosessi (Ware)

Kuvasta voidaan todeta, että informaation visualisointi on jatkuva prosessi. Dataa kerätään fyysisestä ja sosiaalisesta ympäristöstä, esiprosessoidaan ja muunnetaan helpommin tulkittavaksi muodoksi ja grafiikkamoottorin avulla esitetään käyttäjälle tulkittavaksi. Käyttäjä luo oman tulkinnan saamastaan informaatiosta visuaalisen ja kognitiivisen prosessin avulla, johon vaikuttaa myös vallitseva sosiaalinen ympäristö. Käyttäjän halutessa tarkastella dataa, esimerkiksi 3-ulotteista dataa jostain muusta näkökulmasta on se esiprosessoitava, muunnettava ja uudelleen muo-

dostettava grafiikkamoottorin avulla.

Seuraavissa aliluvuissa esitetään miten kolme ensimmäistä informaation visualisointiprosessin vaihetta, datan kerääminen, esiprosessointi ja muuntaminen sekä grafiikkamoottori ilmentyvät tässä diplomityössä. Neljäs ja viimeinen vaihe eli informaation tulkinta ja tulkinnan onnistuminen ovat riippuvaisia loppukäyttäjistä. Informaation visualisoinnin onnistumista voidaan mitata esimerkiksi empiirisesti käyttäjätutkimuksella, mutta tämän diplomityön puitteissa sitä ei ole suoritettu.

### 3.1 Web 2.0 datalähtöisen analyysin kontekstina

Termi Web 2.0 on vuonna 2004 Tim O'Reillyn ja Medialive International konferenssin yhteisen aivoriihen tuloksena luotu käsite World Wide Webin toiselle sukupolvelle (What is Web 2.0). Web 2.0 on vuorovaikutteisuuden ja käyttäjälähtöisyyden toteuttavien tietoteknisten ratkaisujen kokonaisuus jotka ovat mahdollistaneet sosiaalisen median kehityksen. Tekniset ratkaisut ovat perustuneet ajatukseen Internetin toiminnasta sisältöjen tallennuspaikkana ja sovelluksien alustana (Sanastokeskus, 2010).

Käytännössä termi Web 2.0 viittaa world wide webin perusrakenteiden muutokseen. Verrattaessa määritelmiä Web 1.0 ja Web 2.0 on niiden välillä tapahtunut siirtyminen staattisista HTML web-sivustoista dynaamisiin ja jäsenetympiin rakenteisiin web-sivustoihin sekä web-palveluiden tarjoamiseen käyttäjille (What is Web 2.0).

Web 2.0 palveluiden toteutustekniikat kuten XML-kieli ja sen johdannaiskielet ATOM ja RSS helpottavat tiedon jäsentämistä automatisoiduin menetelmin. XML-kielen (Extensible Markup Language) yksinkertainen määritelmä on joukko ennaltamäärättyjä sääntöjä joiden avulla dokumentteja voidaan muuntaa koneluettavaan muotoon. XML-kielisissä dokumenteissa kuvataan erikseen dokumentin rakenne ja sen sisältämä informaatio. Jakamalla dokumentin rakenne edellämainitulla tavalla mahdollistetaan mm. sisältöön liittyvät älykkäät kyselyt (Virk 2002).

Toisaalta Web 2.0 voidaan tarkastella konseptina jossa web-palveluiden kuluttajista tulee tuottajia sisällöntuotannon välityksellä. Web 2.0 mahdollistaa kuluttajien muuntumisen sisällöntuottajiksi antamalla heille keinot luoda, jakaa ja välittää tietoa. Esimerkiksi keskustelupalstat ja erilaiset wikit ovat Web 2.0:a. Käyttäjät voivat vapaasti luoda uutta tietoa ja kommentoida sekä toisinaan muokata muiden käyttäjien luomaa sisältöä.

#### Sosiaalinen media

Sosiaalinen media on Web 2.0 tavoin konsepti, jolle ei ole olemassa tarkkaa yksiselitteistä määritelmää. Sosiaalisen media käsitteenä on aiheuttanut huomattavasti

kritiikkiä mediatutkijoiden kesken, muun muassa koska se voi käsitteenä antaa mielikuvan että tavallinen media olisi jollain tavalla epäsosiaalista. Lietsala ja Sirkkunen (2008) ovat ehdottaneet sateenvarjotermiä sosiaaliselle medialle, joka kattaa verkossa tapahtuvat sisällöntuotantoon liittyvät toimintatavat, kuten blogaamisen, sosiaalisen verkottautumisen ja osallistumisen wikeihin, sekä sisällön tuotantoon liittyvät henkilöt.

Web 2.0 ei ole synonyymi sosiaaliselle medialle. Web 2.0 on löyhempi konsepti joka määrittelee palveluita ja teknologioita joilla ei välttämättä ole sosiaalisen median aspektia tai sosiaaliseen mediaan liittyviä aktiviteetteja.

Sosiaalisen median tyypillisiä piirteitä ovat käyttäjien siirtyminen kuluttajista (audience) tuottajiksi (producer), sisällön painottaminen, sosiaalisuus, profiilien tagien ja syötteiden sisällyttäminen ja median ketterä kuluttaminen sekä tuottaminen.

### **Keskustelualueet**

Keskustelualueet (Internet forum, message board) ovat webissä sijaitsevia keskustelusivustoja. Teknisesti ne ovat web-sovelluksia, jotka mahdollistavat käyttäjien luoman sisällön – viestien jakamisen ja hallitsemisen. Keskustelualueet ovat sähköinen versio perinteisestä ilmoitustaulusta. Ne ovat kehittyneet Internetia edeltäneistä BBS (Bulletin Board System) -tietokonejärjestelmistä, jotka mahdollistivat useiden käyttäjien samanaikaisen liittymisen puhelinlinjojen välityksellä samaan tietojärjestelmään.

## **3.2 SNA-työvälineet**

Sosiaalisen verkostojen analysointityövälineet ovat sosiaalisten verkostojen matemaattisiin analysointimenetelmiin, visualisointeihin sekä tiedonlouhintaan perustuvia ohjelmistoja. Nämä ohjelmistot voidaan jakaa karkeasti kahteen eri luokkaan, tiedonlouhinta- ja analysointityövälineisiin.

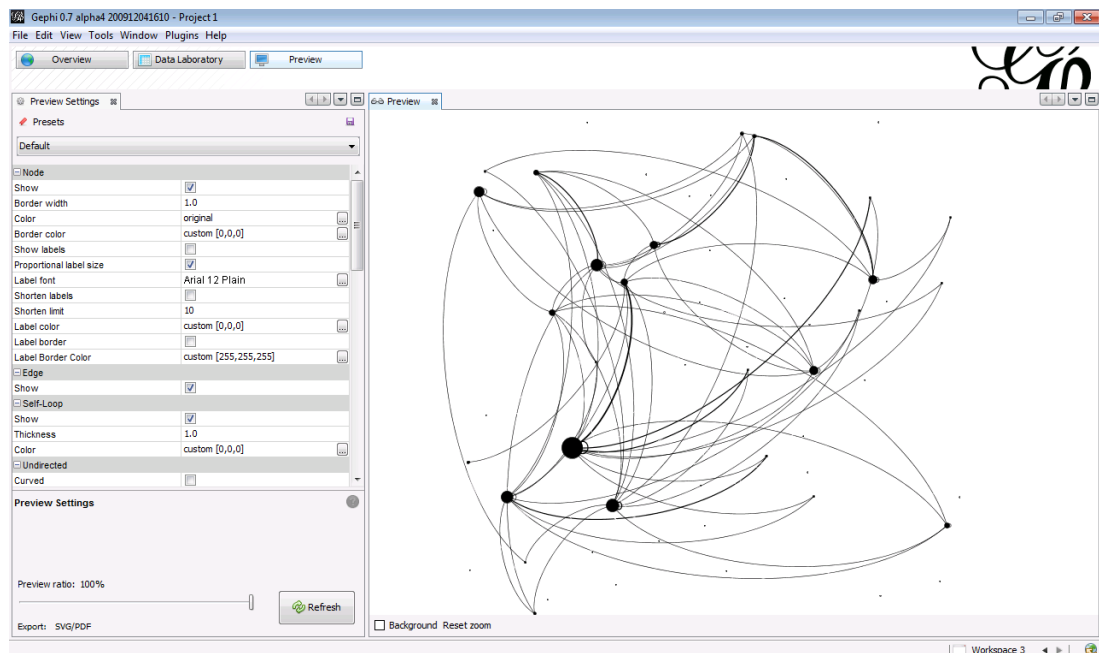
Tyypillisesti SNA-työvälineet ovat jakautuneet joko tiedonlouhintatyövälineisiin tai tiedon analysointi- ja visualisointityövälineisiin. Diplomityön SLP-tapauksessa kehitelty sosiaalisten verkostojen analysointi- ja visualisointityöväline pyrkii yhdistämään nämä molemmat työvälineet ja tarjoamaan rajapinnat ulkopuolisten työvälineiden käytölle.

Tiedonlouhintavälineet keskittyvät analysoitavan tiedon keräämiseen erityyppisistä lähteistä. Web 2.0 -kontekstissa tiedonlouhintaan käytetään erilaisia hakurobotteja (Web crawler). Hakurobotit ovat ohjelmistoja, jotka käyvät lävitse Webiä automatisoidusti tiettyjen ennalta määritettyjen sääntöjen mukaisesti. Web-pohjaiset hakupalvelut kuten Google käyttävät hakurobotteja ajantasaisen tiedon keräämiseen web-sivustoista.

Webin ominaiset piirteet, sen valtava koko, nopeasti muuttuva sisältö ja dynaaminen sivustojen luonti aiheuttavat haasteita hakuroboteille. Webin kokoa ei ole pystytty määrittämään eksaktisti, mutta varovaiset arviot sen koosta ovat noin 28-58 miljardin web-sivun välillä. Web on niin suuri että hakurobotit voivat ladata ja käsitellä vain osan siitä kerrallaan, näin ollen tiedonlouhinnan kannalta on tärkeää että SNA:n kannalta kiinnostava tieto, yleensä web-sivustot sekä niiden sisältö, on asetettu tutkittavan ongelman mukaisesti tärkeysjärjestykseen.

## Gephi

Gephi on avoimeen lähdekoodiin perustuva interaktiivinen graafien analysointi ja verkostojen visualisointiohjelma. Gephi on käyttäjien vapaasti laajennettavissa ja sen mukana tulevat moduulit tarjoavat erilaisten verkostojen tuonti ja vienti, manipulointi, suodatus ja tunnuslukujen laskentaomaisuudet. Monista muista visualisointiohjelmista poiketen Gephi tukee 3D-kiihdytystä. Gephi kykenee käyttämään tietokoneen grafiikkaprosessoria (GPU) sekä moniydinprosessori arkkitehtuuria verkostojen esittämisessä, tehostaen tuhansia solmuja sisältävien verkostojen käsittelyä (Bastain, Heymann & Jacomy, 2009).



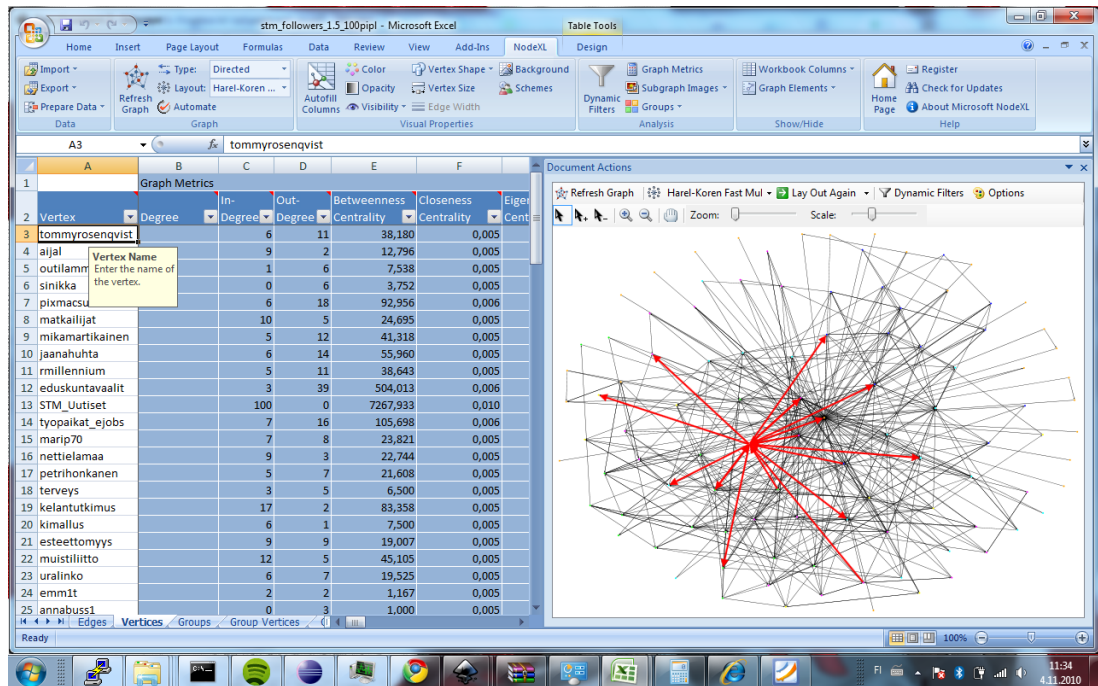
Kuva 3.3: Gephin näkemys yksimoodisesta verkostosta

## NodeXL

NodeXL on Microsoftin verkostoanalyysilaaajennus Excel 2007 ja 2010 ohjelmistoille. NodeXL on Excel sapluuna, johon voidaan tuoda eri tiedostformaateista tai louhia suoraan verkkopalveluista, kuten Twitteristä tai Flickristä verkstodataa. NodeXL



sisältää verkostojen visualisoinnin lisäksi suosituimmat laskentametriikat verkostojen analyyseille ja mahdollistaa muun muassa verkoston toimijoiden klusteroinnin (NodeXL).



Kuva 3.4: NodeXL metriikat ja verkostojen visualisointi.

## 4. TOTEUTUS

Tämän diplomityön SLP-tapauksessa on lähestytty Web 2.0 -kontekstin sisällä tapahtuvaa sosiaalisten verkostojen analysointia ja tiedonkeruuta tutkijoiden näkökulmasta. SLP-tapauksen tavoitteena on yhdistää ja automatisoida koko SNA:n kannalta tärkeät osuudet, tiedonkeruu, analyysi ja visualisointi, yhdeksi paketiksi.

Tämän paketin tarkoituksena on mahdollistaa ja tuottaa nopea vilkaisu (sneak peek) -tyyppinen ominaisuus verkkopalveluiden sosiaalisten verkostojen sisältöön. Tällainen nopea vilkaisu voi olla esimerkiksi jonkin SLP:n keskustelualueen keskustelijoiden muodostaman sosiaalisen verkoston visualisoiminen ja verkoston tärkeimpien henkilöiden toteaminen toteaminen visualisoinnista.

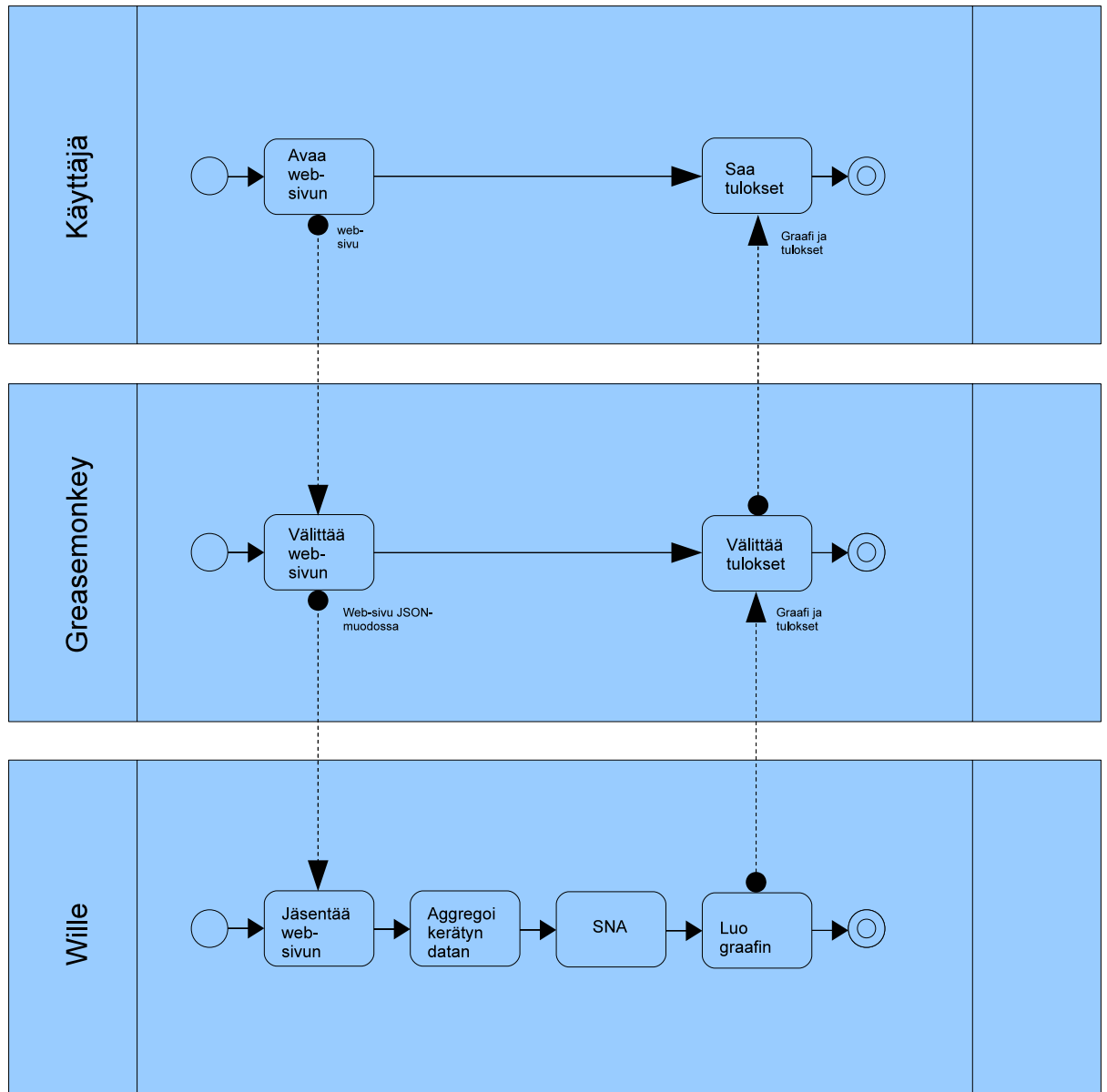
Tässä luvussa esitetään yleisesti diplomityön SLP-tapauksessa käytetyn visualisointityövälineen tekninen toteutus ja visualisointityöväline. Teknisessä toteutuksessa on painotettu valmiiden ohjelmistokirjastojen ja komponenttien käyttöä ohjelmistokehityksen nopeuttamiseksi. Taustalla on ollut tavoite 'pyörää ei kannata keksiä uudelleen', jota mukaillen toteutuksessa on käytetty hyödyksi sosiaalisten verkostojen analyysissä sovellettavia valmiita matemaattikka- ja tunnuslukulaskentakirjastoja.

### 4.1 Visualisointityöväline ja visualisointiprosessi

Diplomityön SLP-tapauksen käytetty visualisointiväline rakentuu kahdesta osasta, käyttäjän käyttöjärjestelmässä ajettavasta Wille-ohjelmistokehityksestä (Wille) sekä Firefox web-selaimessa toimivasta Greasemonkey-liitännäisestä (Greasemonkey). Wille on koko sosiaalisten verkostojen laskennallisen analyysin sydän, se jäsentää, analysoi ja luo tulokset Greasemonkeyn toimiessa rajapintana web-selaimen ja Wille välillä (vrt. Salonen & Huhtamäki, 2010), välittäen vuoroin dataa ja tuloksia. Kuvassa 4.1 on yleisesti hahmotettu työvälineen toiminta Suomen Lasten parlamentin kontekstissa. Työvälineen tiedonlouninta ja analysointiprosessi on kuvattu JHS 152 prosessien kuvaamismallin (JUHTA 2008) mukaisesti.

Kuvassa suorakulmioilla on rajattu toimijat ja toimijoihin suoraan kytköksissä olevat toiminnot. Nuoliviivat ilmaisevat prosessin suunnan ja pyöristetyt suorakulmiot kuvataan prosessin osatehtäviä. Katkoviivoitettu nuoli kuvaa toimijoiden välillä tapahtuvaa informaation välittymistä.

Kuvan 4.1 mukaisesti työvälineessä tapahtuva informaation kulku voidaan jakaa



Kuva 4.1: SLP-tapauksen sosiaalisten verkostojen analysoinnin työnkulun prosessikaavio

kolmen eri toimijan väliseen vuorovaikutukseen. Nämä toimijat ovat: Greasemonkey, käyttäjä sekä Wille-ohjelmistokehys. Työvälineen ollessa toiminnassa käyttäjälle prosessista näkyvät vain web-selaimessa oleva web-sivusto ja kuvan 4.4 mukainen Greasemonkey-paneeli. Wille-ohjelmointikehys toimii ns. taustalla – poissa käyttäjän näkyvistä.

Käyttäjän siirtyessä SLP verkkoparlamentin keskustelualueesta koostetulle ATOM-syötesivustolle Greasemonkey tunnistaa että kyseessä on web-osoite, jolle se on herkistetty. Greasemonkey saa herkistetyn web-osoitteen jota käyttäen se hakee web-selaimessa näkyvän web-sivuston eli rakenteisen ATOM-dokumentin ja välittää sen Willelle. Greasemonkey suorittaa web-sivuston haun HTTP-protokollaa käyttäen.

ATOM-dokumentista Wille jäsentää viestien kirjoittajien nimet, viestien sisäl-

löt sekä viestiketjun otsikon ja luo jokaisesta yksittäisestä Greasemonkeyn välittämästä ATOM-syötesivustosta (kts. kuva 4.8) XML-muotoisen network (kts. kuva 4.2) koostedokumentin johon edellämainitut tiedot kootaan. Network dokumentti on Willen sisällä tapahtuvassa tiedonsiirrossa käytetty tiedonvälitysmuoto. Yksittäisistä network-dokumenteista Wille luo sosiomatriisit sekä Pajek SNA-työvälineelle soveltuvan Pajek-muotoisen tiedoston (kts. kuva 4.3).

```
<?xml version="1.0" ?>
<network>
  <nodelist>
    <node id="Kouluruoka" />
    <node id="toimija1" />
  </nodelist>
  <edgelist>
    <edge posts="1" source="toimija1" target="Kouluruoka" />
  </edgelist>
</network>
```

Kuva 4.2: Esimerkkilistaus Willen sisäisessä informaationsiirrossa käytettävästä network muotoisesta rakenteisesta dokumentista. Dokumentin rakenteessa on ilmaistu toimijat node-tageilla ja toimijoiden väliset yhteydet edge-tageilla.

```
*Vertices 2
1 "Kouluruoka"
2 "toimija1"
*Arcs
2 1 1
```

Kuva 4.3: Esimerkki Willen tuottamasta Pajek-muotoisesta tiedostosta. Monet SNA-työvälineet käyttävät yhä Pajek-tiedostomuotoa tiedon tuontiformaattina. Pajek-formaatissa määritetään ensiksi verkoston toimijat juoksevalla numeroinnilla ja toimijoiden väliset yhteydet numeroinnin perusteella. Kuvassa toimijasta yksi on suunnattu voimakkuudeltaan yhden suuruinen yhteys toimijaan kouluruoka.

Verkostojen luominen soveltuvasta datasta on aina kontekstista riippuvaa ja tulokinnanvaraista toimintaa. Wille mallintaa SLP verkkoparlamentin keskusteludatas- ta jäsennettyjen tietojen avulla verkostoja yksi- ja kaksimoodisiksi. Yksimoodisessa verkostossa on tehty sosiaalisten verkostojen analyysin kannalta oletus että yksittäisen viestiketjun ensimmäisen viestin kirjoittaja on henkilö, joka ns. omistaa keskustelun ja johon muut keskustelijat liittyvät. Näin ollen muut keskustelijat muodostavat yksisuuntaisen yhteyden ensimmäisen viestikeskustelun viestin kirjoittajaan.

Tämä yhteys on arvoltaan yhteys, jolloin yhteyden arvo on verrannollinen yksittäisen kirjoittajan lähettämien viestien määrään.

Kaksimoodisessa verkostossa on tehty oletus että keskustelun aiheet muodostavat yhden moodin ja keskustelijat toisen moodin. Keskustelijat liittyvät täten kommentoihinsa keskustelunaiheisiin yksisuuntaisen suunnattujen yhteyksien avulla. Yhteyksien arvot muodostuvat keskustelijoiden keskustelun aiheisiin lähettämien viestien määrästä.

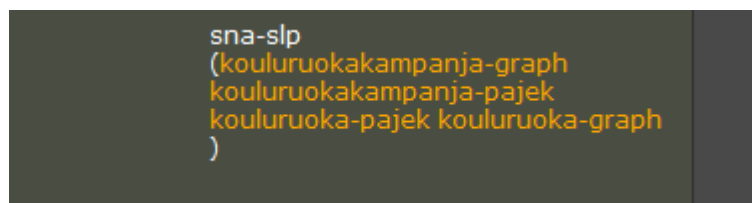
Wille vastaanottaa ATOM-dokumentteja yksitellen ja se koostaa vastaanottamista ja mallinnetuista verkostoista aggregoituja verkostoja. Yksittäiset dokumentit kuvastavat SLP-tapauksessa yhtä keskustelua kts. kuva 4.2.

Tunnuslukujen laskennan kannalta yksittäisistä keskusteluista muodostetut verkostot eivät ole laskennallisesti niin mielenkiintoisia kuin 'täysi' verkosto jossa on mallinnettu jokainen Willen vastaanottama keskustelu.

Edellämainitun kaltainen 'täysi' verkosto voidaan mallintaa aggregoimalla eli yhdistämällä vastaanotetut yksittäiset keskustelut. Yksittäisistä keskusteluista mallinnetuista yksi- ja kaksimoodisista verkostoista muodostetaan iteratiivisesti aggregoituja yksi- ja kaksimoodisia verkostoja. Aggregoidut verkostot mallintavat yksittäisistä verkostoista yhdistämällä toteutettua verkostojen kokonaisuutta.

Aggregoidusta kaksimoodisesta verkostosta Wille luo romahdetun yksimoodisen verkoston. Kaksimoodinen verkosto romahdutetaan yksimoodiseksi koska kaksimoodisten verkostojen tunnuslukujen laskentamenetelmät ovat huomattavasti rajoitettuja kuin yksimoodisten. Romahdutetussa verkostossa keskustelijoiden välille on muodostettu suuntaamaton dikotominen yhteys jos keskustelijat ovat osallistuneet kirjoittamalla samaan keskustelun aiheeseen.

Analysoinnin tulosten valmistuttua kuvan 4.4 mukainen paneeli ilmestyy web-selaimen ylänurkkaan. Greasemonkey-paneelissa näkyvät linkit analysoinnin tulokseen. Greasemonkey toimii käyttäjän ja Willen välillä eräänlaisena rajapintana. Greasemonkey 'haistelee' millä web-sivustoilla käyttäjä liikkuu ja käyttäjän mennessä Greasemonkeylle herkistetylle sivustolle Greasemonkey välittää sivuston edelleen Willelle.



Kuva 4.4: Web-selaimessa näkyvä Greasemonkey-paneeli

The screenshot shows the website of the Suomen Lasten Parlamentti (Finnish Children's Parliament). The main section is 'Yleinen keskustelu' (General discussion). It includes a sidebar with navigation links and a main content area with a list of discussion topics.

Vilmeisin muutos T	Aihe	Tekijä	Vs.	Vilmeisin vastaus
18.06.2010, 08:29	poikien liikunnan suosiminen	Helmi Tulonen	0	Ei vielä vastauksia.
18.06.2010, 08:18	kesäloman huomattava lyhennys	Helmi Tulonen	4	Vs: kesäloman huomattava lyhennys (helmi)
17.06.2010, 17:44	LEMMIKKIPÄIVÄT!	Tilda Rauhanen	16	Vs: LEMMIKKIPÄIVÄTI (meerira)
17.06.2010, 17:33	numerotodistukset	Helmi Tulonen	8	Vs: numerotodistukset (minttu)
17.06.2010, 17:26	Haave ammattisi?	Kamilla Willis	37	Vs: Haave ammattisi? (minttu)
16.06.2010, 16:21	Missä on fyysinen istunto?	Jenny Impiö	5	Vs: Missä on fyysinen istunto? (jenkkü)

Kuva 4.5: Suomen Lasten Parlamentin yleistä keskustelua keskustelupalsta

## 4.2 Toteutustekniikat

### Python ohjelmointikieli

Python (versio 2.6) on tämän diplomityön toteutuksessa käytetty ensisijainen toteutuskiehi. Diplomityön toteutuksessa on käytetty Pythonin lisäksi Java-ohjelmointikieltä sekä XML- ja XSLT-merkkäuskieliä. Rakenteeltaan Python on monikäyttöinen korkean tason tulkattava ohjelmointikieli. Tulkattavuudella tarkoitetaan, että Python kielellä kirjoitettuja ohjelmia ei käännetä kuten esimerkiksi C- tai C++ -kielellä toteutettuja ohjelmia vaan Python-kielellä toteutut ohjelmat toimivat eräänlaisen virtuaaliympäristön päällä joka tulkitsee Python-kielellä kirjoitettuja käskyjä ja muuntaa ne konekielisiksi. Yleinen tulkattavien ohjelmointikielien ominaisuus on niillä tehtyjen ohjelmien (väitetty) suhteellinen hitaus verrattuna käännettävillä ohjelmointikielillä luotuihin ohjelmistoihin. Tulkattavuus ei ole pelkästään huono asia, sillä sen ansiosta Python on hyvin ilmaisuvoimainen kieli, ohjelmia voidaan kirjoittaa tehokkaammin, lisäksi se on suunniteltu erityisesti ajatellen koodin luettavuutta.

Syntaksiltaan Python muistuttaa Matlab-kieltä. Monikäyttöisyyttä edesauttavat Pythonin mukana tulevat ja erikseen ladattavat kirjastot, jotka laajentavat Python kieltä. Esimerkiksi NumPy-kirjasto (Numerical Python) tarjoaa tieteellisen lasken-

nan työvälineet Pythonille.

Python on valittu SLP-tapauksessa käytetyn työvälineen toteutuskieleksi koska se tarjoaa hyvät laajennusmahdollisuudet ja se on ilmaisuvoimainen sekä ilmainen. Lisäksi Wille-visualisointi ohjelmistokehys on luotu Python-kielillä. Python laajennettuna NumPy kirjaston avulla luovat yhdessä varteenotettavan tieteellistä laskentaa tukevan ohjelmointiympäristön – vaihtoehdon MATLABille. Pythonin etuna MATLABiin nähden ovat avoin lähdekoodi ja sen luonne modernina, laajennettavana ohjelmointikielenä. Willen kautta tulee myös mahdollisuus sovellusohjelmointiin Python-kielillä.

### **NumPy**

NumPy on kirjastolaajennus Python-kielille. NumPy lisää tuen muun muassa laajoille moniulotteisille taulukoille ja matriiseille, joita usein tarvitaan tieteellisessä laskennassa. NumPy tarjoaa lisäksi suuren joukon numeerisia algoritmeja, funktioita ja operaattoreita taulukoiden ja matriisien käsittelyyn. NumPy lähentää Python ja MATLAB-kieliä lisäämällä Pythoniin MATLAB-kielen syntaksin kaltaisia ominaisuuksia ja funktioita, esimerkiksi matriisien indeksoinnin ja transponoinnin, näin helpottaen MATLABilla kirjoitettujen ohjelmien muuntamista Python-kielille (NumPy).

### **Igraph**

Igraph usealle eri ohjelmointikielille mm. Ruby, Python ja R, saatavissa oleva suunnattujen ja suuntaamattomien graafien tuottamiseen sekä käsittelyyn erikoistunut kirjastolaajennus. Igraph lisää funktioita ja algoritmeja graafien tunnuslukujen laskemiseksi sekä graafien muodostamiseksi. Igraphin avulla verkostoja voidaan muodostaa eri keinoin kuten sosiomatriiseista tai erilaisten työvälineiden esim. Pajek-tiedostomuodosta, lisäksi verkostoja voidaan muuntaa ja tallentaa vastaaviin tiedostomuotoihin. Igraphilla verkostoja voidaan lataa ja visualisoida erilaisin algoritmein esimerkiksi voimaohjatulla Kamada-Kawai algoritmilla (The igraph project).

### **Wille visualisointiympäristö**

Wille-visualisointiympäristö (Wille Visualisation System) on Tampereen teknillisen yliopiston Hypermedialaboratorion kehittämä kevyt ohjelmistokehys, jonka tarkoituksena on edistää informaation visualisointiohjelmistojen tuottamista. Wille on ohjelmistokehys joka on suunniteltu erityisesti ajatellen informaation prosessointia putkilinjastomaisesti. Informaatioputkilinjastoissa dataa jalostetaan informaatioksi ja sitä voidaan prosessoida linjaston eri vaiheissa erilaisten kyselyiden ja muunnoksien avulla lisäksi linjastosta voidaan tuottaa itsenäisiä visualisointeja ohjelmiston

käyttäjille. Vaihtoehtoisesti Willeä voidaan käyttää informaation jalostukseen ulkoisille visualisointityövälineille esimerkiksi Pajekille (Wille).

### Greasemonkey

Greasemonkey (Greasemonkey) on Aaron Boodmanin suunnittelema liitännäinen (add-on) Firefox web-selaimelle, jonka avulla käyttäjät voivat luoda skripteja (scripts) jotka muuntavat web-sivustojen näkymää. Greasemonkeyn avulla web-sivuista voidaan tehdä luettavampia ja käytettävämpiä tai niille voidaan luoda kokonaan uusia käyttömahdollisuuksia. Web-sivustojen näkymä voidaan muuntaa puheeksi tai esimerkiksi Braaille kirjoitusta tukevalle laitteelle yhteensopivaksi informaatioksi.

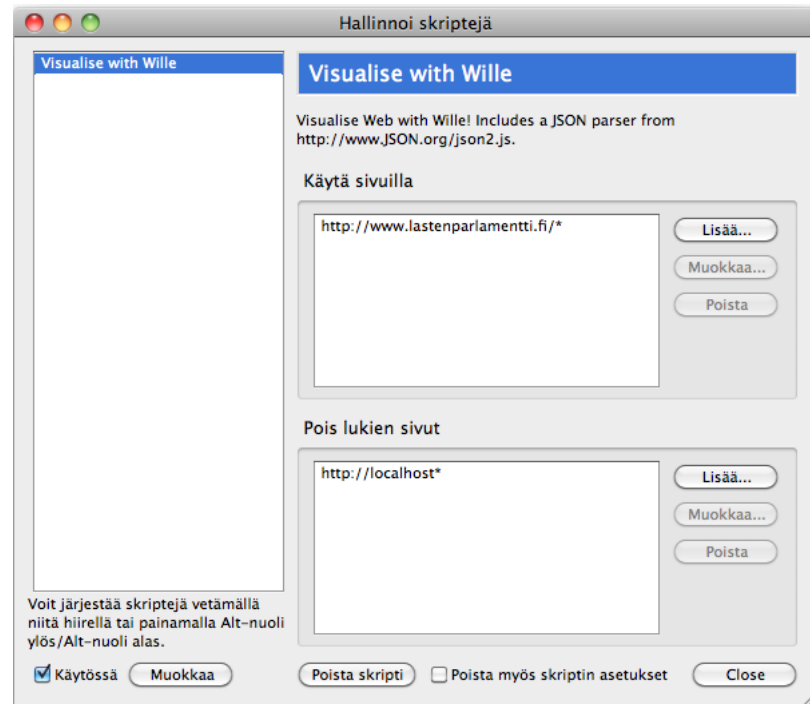
Käyttäjien luomat skriptit ovat Greasemonkeyn sydän. Niitä ilman se ei tee mitään. Skriptit ovat Java-koodia, jotka kertovat Greasemonkeylle missä ja miten sen pitää toimia. Skripteillä voidaan tehdä mitä tahansa, mitä Java-kielellä voidaan luoda. Jokainen skripti on mahdollista kohdistaa yksittäiselle sivustolle www-osoitteen mukaan tai suuremmalle joukolle sivustoja.

Kuvissa 4.6 ja 4.7 on esitetty Greasemonkeyn skriptien hallinnointi-ikkuna Firefox selaimessa Mac OS X käyttöjärjestelmässä. Wille-visualisointiympäristöskripti on asetettu käyttöön Suomen Lasten Parlamentin sivustolla.



Kuva 4.6: Firefox web-selaimen alapalkissa näkyvä Greasemonkey-ikoni osoittaa, että Greasemonkey on käytössä.





Kuva 4.7: Greasemonkeyn skriptien asetuskikkuna.

## ATOM-syöte

ATOM-syöte (feed, RFC 4287) on IETF:n (Internet Engineering Task Force) vuonna 2005 esittelemä määrittely XML-muotoisesta dokumenttityypistä, joka määrittää yhteenkuuluvat informaatiot syötteinä. Syötteet koostuvat metadatan sisältävistä määrällisistä osioista, joita kutsutaan elementeiksi. ATOM-syötteen pääkäyttökohteita ovat web-sisällön, kuten blogien ja uutissivuston otsikoiden välittäminen käyttäjille ja web-sivustoille. Kuvan 4.8 esimerkkilistauksessa on havainnollistettu ATOM-syöte.

```

<?xml version="1.0" encoding="utf-8" ?>
<feed xmlns="http://www.w3.org/2005/Atom">
  <title>Lasten Parlamentti</title>
  <subtitle>Kouluruoka</subtitle>
  <link rel="self" type="application/atom+xml"
        href="http://www.lastenparlamentti.fi/intra/Kirjasto-
        ja-kokouhuone/hallitus/keskustelu/kouluruoka/
        thread-atom.xml"/>
  <updated>2008-04-18T12:47:37+03:00</updated>
  <id>tag:www.lastenparlamentti.fi,2008:shuone/hallitus/
    keskustelu/kouluruoka#atom</id>
  <generator>PrimaControl</generator>

  <entry>
    <title>Kouluruoka</title>
    <link href="http://www.lastenparlamentti.fi/intra/
      Kirjasto-ja-kokouhuone/hallitus/keskustelu/kouluruoka
      "/>
    <id>tag:www.lastenparlamentti.fi,2008:shuone/hallitus/
      keskustelu/kouluruoka</id>
    <updated>2008-02-20T15:23:20+02:00</updated>
    <content type="text">Toimijan1 kirjoittama viesti</
      content>
    <author>
      <name>toimija1</name>
      <uri>http://www.lastenparlamentti.fi/intra/Members/
        toimija1</uri>
    </author>
  </entry>
</feed>

```

Kuva 4.8: Esimerkki Suomen Lasten Parlamentin keskustelupalstan ATOM-syötteestä

## 5. TAPAUK: LASTEN PARLAMENTTI

Suomen Lasten Parlamentti on 9-13 vuotiaiden lasten oma eduskunta, jonka tarkoituksena on antaa lapsille vaikuttamiskanava sekä tarjota viranomaisille ja päättäjäille mahdollisuus kysyä ja selvittää erilaisia asioita suoraan lapsilta. SLP perustettiin avajaisistunnossa Tampereella 9.11.2007 ja vuonna 2010 alkoi järjestyksessä toinen lasten parlamentin kaksivuotiskausi. Edustuskaudelle 2010-2011 valittiin noin 360 edustajaa noin 200 kunnasta (Suomen Lasten parlamentti, 2010).

Suomen Lasten Parlamentin perustamisen taustalla on vuonna 2006 hyväksytty nuorisolaki. Nuorisolain kahdeksas pykälä määrää että nuorille tulee järjestää mahdollisuus osallistua paikallista ja alueellista nuorisotyötä ja -politiikkaa koskevien asioiden käsittelyyn. Lisäksi nuoria on kuultava heitä koskevissa asioissa.

Kunnat ovat vahvasti mukana lasten parlamentin toiminnassa. Jokaisella kunnalla on oikeus lähettää yksi virallinen edustaja ja yksi varaedustaja lasten parlamentin toimintaan. Molemmilla edustajilla on yhtäläiset oikeudet osallistua verkkoparlamentin toimintaan.

Lasten parlamentin toiminta voidaan jakaa kahteen osaan, fyysiseen parlamenttitoimintaa ja verkkoparlamenttitoimintaa. Fyysinen parlamenttitoiminta toteutuu kahdesti vuodessa järjestettävissä fyysisessä täysistunnoissa. Verkkoparlamenttitoiminta on ympärivuotista, lasten parlamentin verkkopalvelussa tapahtuvaa vaikuttamistoimintaa.

### 5.1 Verkkoparlamentti

Verkkoparlamentti, [www.lastenparlamentti.fi](http://www.lastenparlamentti.fi) on jatke ja kokoontumisalusta Suomen Lasten Parlamentille ja sen jäsenille. Verkkoparlamentti mahdollistaa lasten virtuaalisen yhteenkokoontumisen maantieteellisesti hyvin laajalta alueelta. Torniossa ja Helsingissä asuvat jäsenet voivat tavata virtuaalisesti verkossa ilman että heidän matkustaa ja tavata fyysisesti. Verkkoparlamentti koostuu erilaisista vuorovaikuttamisen työvälineistä, keskustelupalstoista, gallupeista, 2D-verkkoparlamentista sekä peleistä.

Verkkoparlamentin erilaiset vuorovaikuttamisen toiminnot voidaan jakaa keskustelupalstoihin, kyselyihin ja chatteihin. Keskustelupalstoilla lapset voivat aloittaa uusia keskusteluita sekä osallistua olemassaoleviin keskusteluihin. Keskustelupalstat jakautuvat aihealueittain ja kohderyhmittäin verkkoparlamentin sisällä, jotkin



Kuva 5.1: Suomen Lasten Parlamentin verkkoparlamentin etusivunäkymä

keskustelupalstat ovat yleisiä kaikille Lasten Parlamentin jäsenille ja jotkin tietyille ryhmälle Lasten Parlamentin sisällä kuten puheenjohtajille tai valiokunnille.



Kuva 5.2: Suomen Lasten Parlamentin 2D-parlamenttiympäristö

## 5.2 Analysoitavan tiedon kerääminen

Diplomityön puitteissa analysoitavan tiedon määrää on mielekästä rajata joillain keinoilla, siksi sosiaalisten verkostojen analyysia varten on valittu erityisesti verkkoparlamentin keskustelualueiden analysointi. Keskustelualueista tutkittavaksi valittiin lasten parlamentin kevään 2010 verkkotäysistuntoon liittyvät kouluruokakeskustelut. Kouluruokakeskustelut olivat määrällisesti mitattuna yksi rikkaimmista keskustelualueista. Vaihtoehtoisia analysointikohteita olisivat olleet lastenparlamentin chat-keskustelut ja kyselyt, tai jokin näiden yhdistelmästä.

Kouluruokakampanja ja siihen liittyvät kouluruokakeskustelut olivat yksi kevään 2010 verkkotäysistunnon teemoista. Kaikkiaan kahdeksasta teemasta kouluruokakeskustelut olivat selvästi suosituin. Kouluruokakeskustelut keräsivät yhdeksälle eri keskustelualueelle 117 viestiä ja 58 keskustelijaa. Sisällöllisesti keskustelunaiheet vaihtelivat kouluruoan nykytilanteesta kouluruokakulttuuriin.

Keskusteluiden jäsentämisen helpottamiseksi Suomen Lasten Parlamentin web-sivustoille on toteutettu ominaisuus ATOM-syötteiden luomiseen, joka koostaa ja jäsentää yksittäiset keskustelunaiheet ennalta määritetyn muotoiseksi syötteiksi. Kun viestikeskusteluista jäsennetään automaattisesti ATOM-syötteitä palvelinpäässä niin asiakaspään ei tarvitse erikseen ryömiä hakurobottien avulla keskustelualueita. Näin ollen asiakaspäähän riittää yksinkertaisesti WWW-osoite, mistä jäsenneetyt dokumentit löytyvät.

## 5.3 Tunnusluvut kontekstissa

Tunnuslukujen tarkoituksena on löytää verkostoista laskennallisesti niiden merkittävimmät tekijät. Yksittäistä verkostoa tarkasteltaessa erityyppiset tunnusluvut tuottavat samalle verkostolle erilaisia lopputuloksia, nostaan esiin verkostosta eri tekijöitä. Tunnuslukujen avulla verkoston toimijoita voidaan arvostaa ja asettaa merkittävyysjärjestykseen.

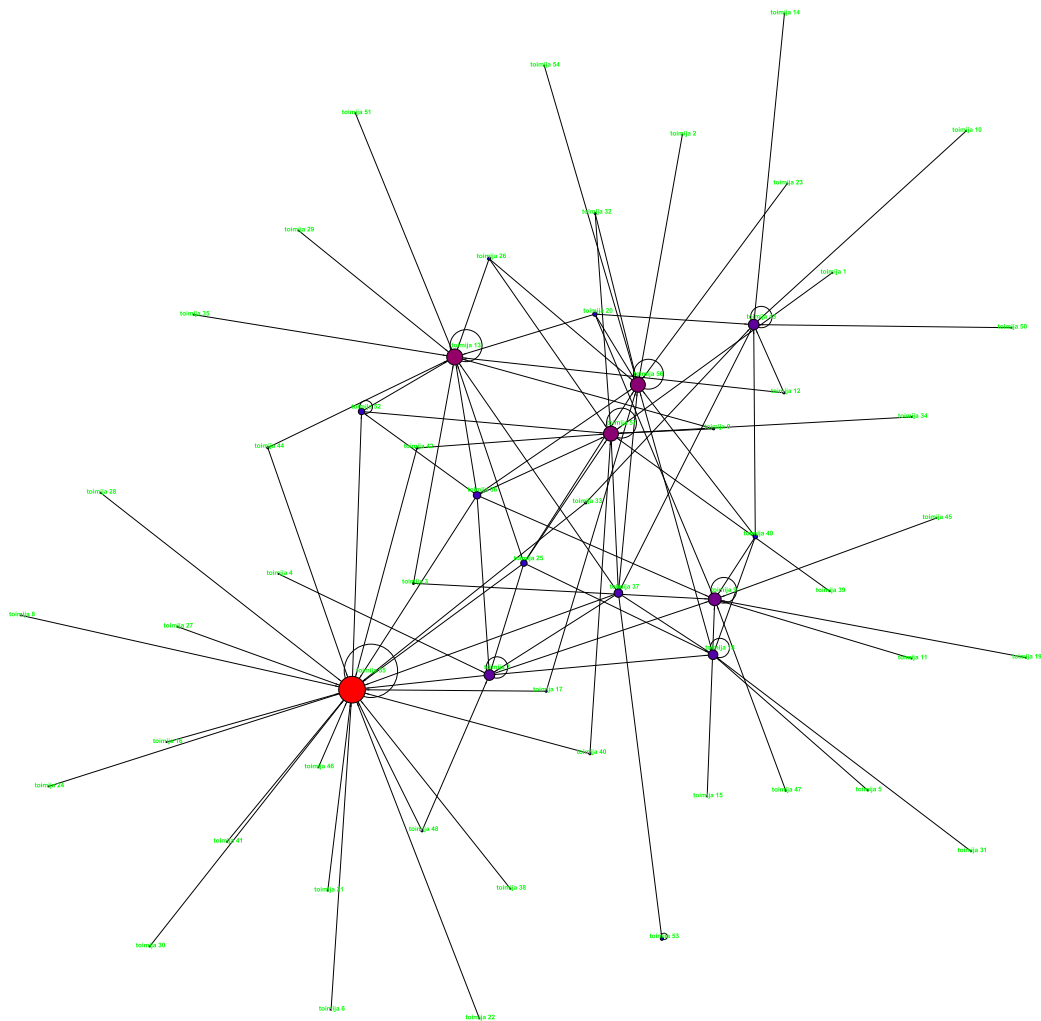
Verkostonmuodostustapa sisältää jonkinlaisen subjektiivisen tulkinnan siitä, miten verkoston tekijät liittyvät yhteen. Sosiaalisen median palvelun keskusteluryhmittä muodostettua sosiaalista verkostoa arvostettaessa erilaisin tunnusluvuin nousee tutkimuskysymykseksi se, mikä tunnusluvun merkitys on kyseisessä kontekstissa ja miten tunnuslukujen tuottamia tuloksia tulisi tulkita. Kontekstin merkitys tunnuslukuja laskettaessa on merkittävä koska datasta, diplomityön SLP-tapauksessa keskusteludatasta, voidaan muodostaa yksi- ja kaksimoodisia verkostoja usealla eri tavalla.

Tarkasteltavasta keskusteludatasta muodostettiin yksi- ja kaksimoodisia verkostoja. Molemmista verkostoista mallinnettiin SLP:n kevään verkkotäysistunnon kouluruokakeskusteluja. Verkostojen muodostamisessa käytettiin kahta erilaista tulkin-

taa.

Yksimoodisessa verkostossa yksittäisen keskustelun aloittajan ja kyseiseen keskusteluun liittyneiden henkilöiden välille muodostettiin suunnattu yhteys siten, että jälkimmäiset henkilöt muodostavat suunnatun yhteyden keskustelun aloittajaan. Kaksimoodisessa verkostossa kaikki yksittäiseen keskustelun aloittajaan liittyvät henkilöt muodostavat suunnatun yhteyden keskustelun aloittajaan, muodostaen näin verkoston kaksi moodia keskustelijat ja keskustelunaiheet.

Seuraavissa aliluvuissa käsitellään teoriaosuuden mukaisessa järjestyksessä tunnuslukujen merkitystä SLP-keskusteludatasta muodostetuissa verkostoissa.



Kuva 5.3: Kouluruokakeskustelun aloittaneita ja keskusteluihin vastanneita toimijoita kuvaava yksimoodinen verkosto. Verkostossa muita solmuja suurempina solmuina ovat suosituimpien keskustelunaiheiden aloittaneet toimijat.

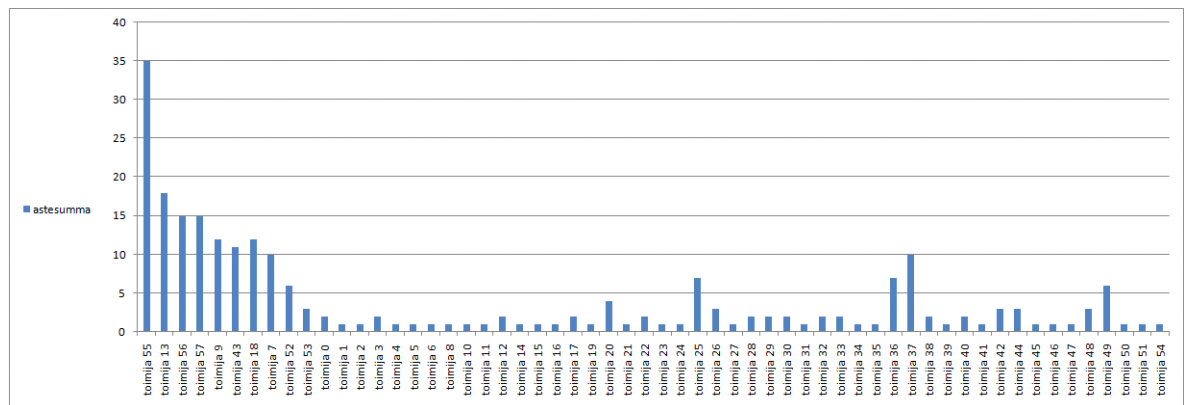
## Astesumma

Astesumma ilmaisee toimijaan yhteydessä olevien tekijöiden eli toimijan yhteyksien määrän. Astesumma on sama suunnatuille ja suuntaamattomille verkostoille. SLP-keskusteluista muodostetuissa yksi- ja kaksimoodisissa verkostoissa astesumma korostaa määrällisesti eniten viestejä lähettäneitä ja vastaanottaneita toimijoita.

Kuvissa 5.3, 5.5 ja 5.7 on esitetty yksi- ja kaksimoodiset verkostot sekä kaksimoodisesta verkostosta yksimoodiseksi verkostoksi romahdutettu verkosto. Kuvien solmujen koko ja väri ovat verrannollisia graafin solmujen eli toimijoiden astesummaan. Mitä suurempi solmu sitä isompi astesumma kyseisellä toimijalla on. Solmujen värit on skaalattu sinisestä punaiseen vastaamaan graafeista laskettujen astesummia. Sininen väri merkitsee pientä astesummaa ja punainen väri suurta astesummaa.

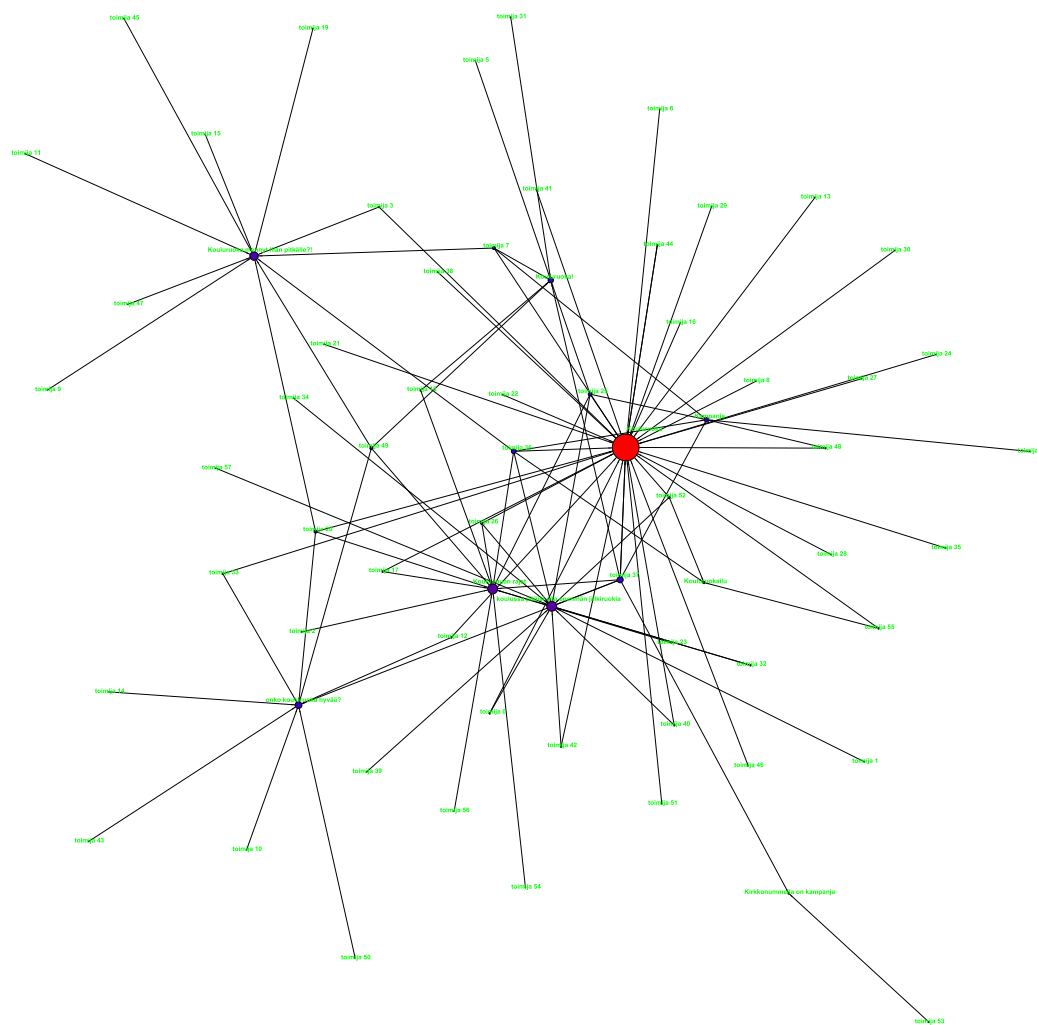
Solmujen koko ja väri ovat suhteellisia, mutta verrannollisia solmujen astesummiin. Värien käytön ja solmujen koon muutoksilla pyritään havainnollistamaan ja nostamaan esiin yhteysmäärällisesti merkittävimpiä toimijoita verkostosta.

Kuvan 5.3 yksimoodisessa sosiograafissa esiin nousevat ensinnäkin ne toimijat, jotka kirjoittivat yksittäisiin keskusteluihin ensimmäisen viestin ja toiseksi ne toimijat, jotka eivät aloittaneet viestikeskustelua mutta kommentoivat keskustelunaiheita. Kuvassa on havaittavissa yksi muista poikkeava suuri punainen solmu ja muutamia keskisuuria lilan värisiä solmuja ja joitakin hyvin pieniä sinisiä solmuja. Kuvassa 5.4 on esitetty toimijoiden astesummat.



Kuva 5.4: Yksimoodisen verkoston toimijoiden astesummat.

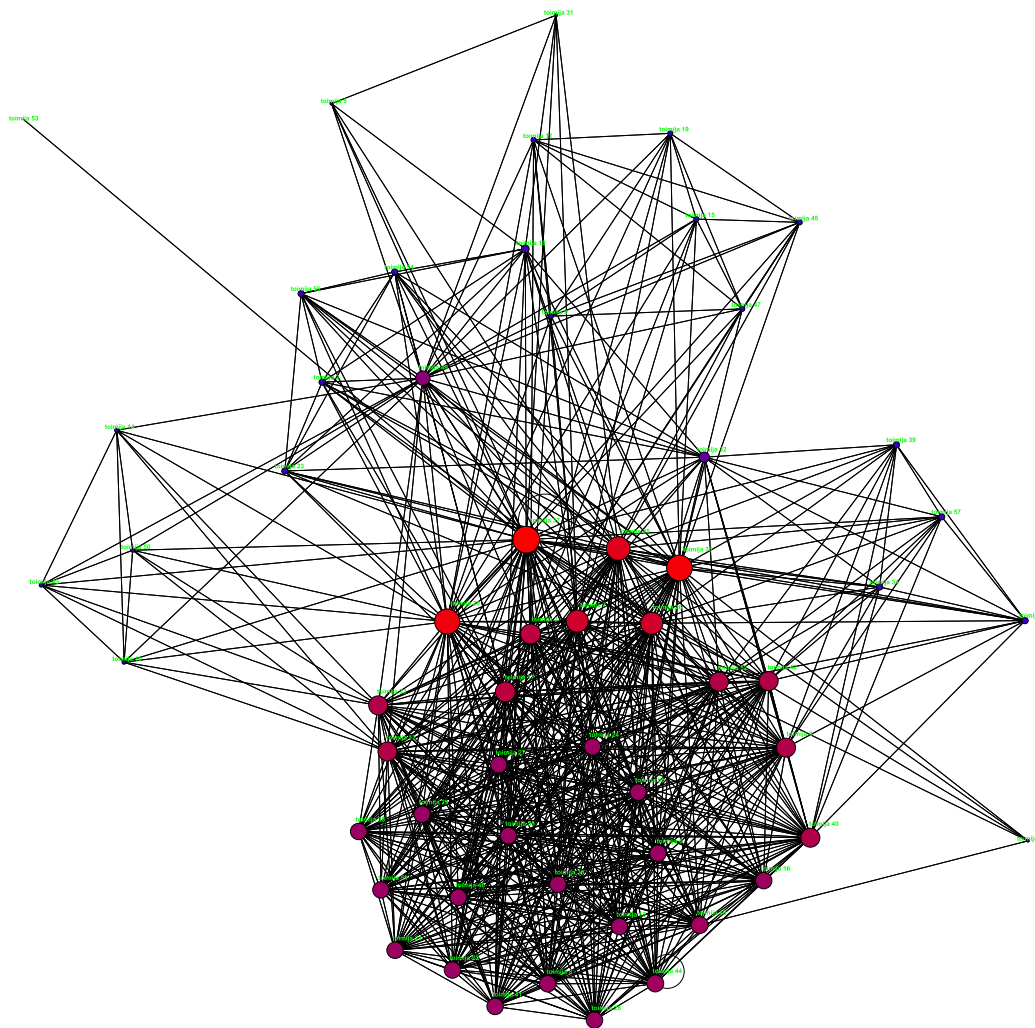
Kuvan 5.5 kaksimoodisessa sosiograafissa verkoston toinen moodi eli keskustelunaiheet korostuvat. Määrällisesti mitattuna suosituimmat keskustelunaiheet ovat nähtävissä graafissa muita solmuja suurempina. Yksi keskustelunaihe, suuri punainen solmu sosiograafissa, kouluruokakeskustelu oli muihin keskustelunaiheisiin verrattuna osallistujamäärällisesti mitattuna selvästi suurin. Kuvassa 5.6 esitetty toimijoiden astesummat.



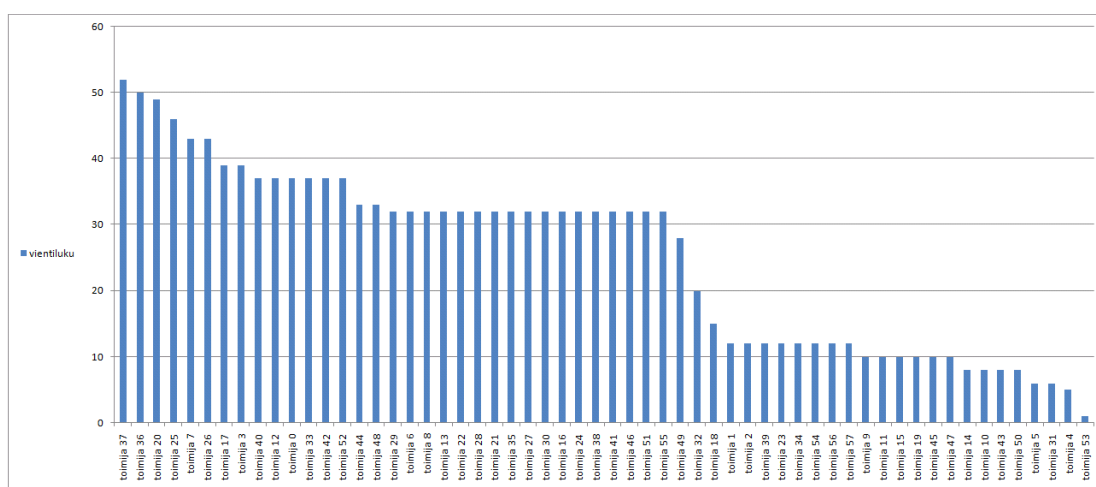
Kuva 5.5: Kouluruokakeskusteluiden kaksimoodinen verkosto, jonka moodeina ovat keskustelunaiheet ja keskusteluun osallistuneet toimijat. Kuvassa toimijoita yhdistävinä solmuina ovat keskustelunaiheet. Määrällisesti suosituimmat keskustelunaiheet ilmenevät suurempina solmuina ja ovat väriltään punertavampia.







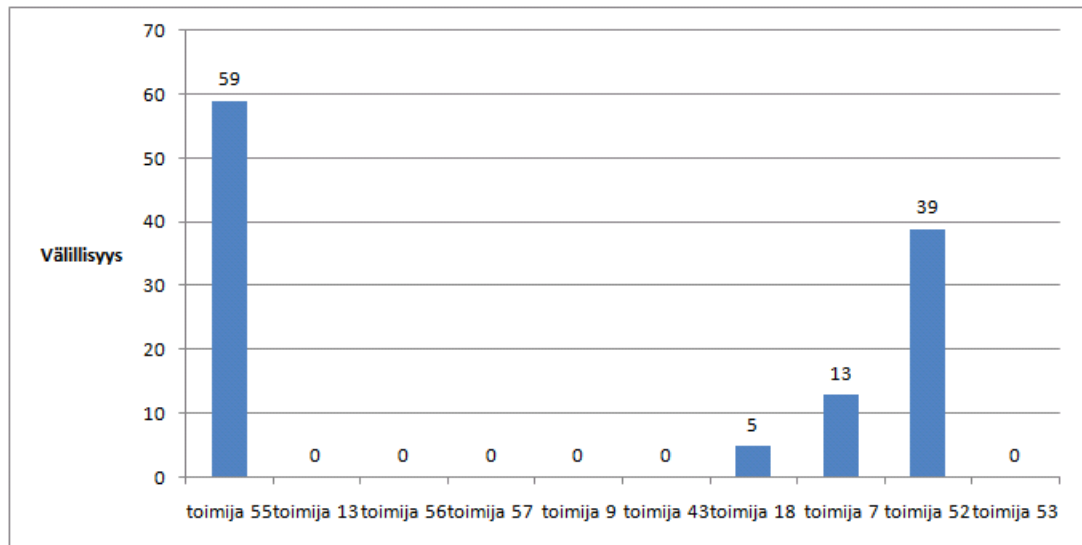
Kuva 5.7: Yksimoodiseksi romahdutettu verkosto.



Kuva 5.8: Yksimoodiseksi romahdutetun verkoston toimijoiden astesummat.

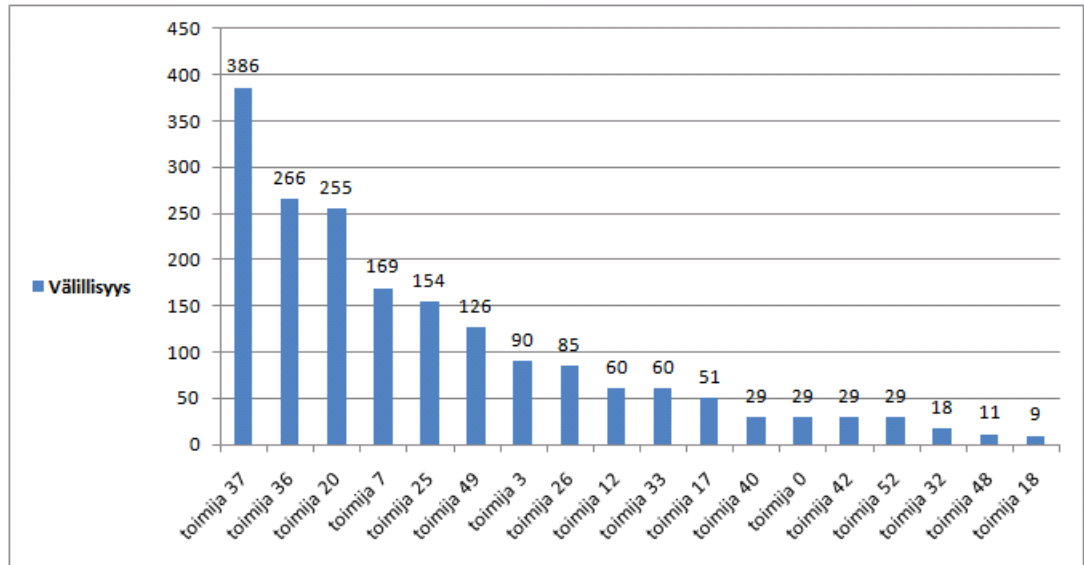
## Välillisuus

Välillisuus ilmaisee toimijan keskeisyyttä verkostossa. Toimijan välillisuus on suuri jos monen toimijan täytyy kulkea sen kautta saavuttaakseen verkoston muut toimijat. Kuvassa 5.9 on esitetty kouluruokakeskusteluiden aloitusviestien kirjoittajien välillisyydet. Välillisuus on nolla niillä toimijoilla, jotka ovat kirjoittaneet viestejä vain omaan keskusteluunsa. Nollasta poikkeavan välillisyyden ovat saaneet ne toimijat, jotka ovat kommentoineet muitakin keskusteluita kuin aloittamaansa keskustelunaihetta.



Kuva 5.9: Yksimoodisen verkoston toimijoiden välillisyyksiä. Kuvan toimijat ovat astesummien perusteella merkittävimmät, mutta vain toimijoiden 55, 18, 7 ja 52 välillisyydet ovat nollasta poikkeavia. Vain muita keskustelunaloittaneita kommentoineet keskustelunaloittaneet toimijat saivat nollasta eroavan välillisyyden.

Kuvassa 5.10 on kaksimoodisesta, keskustelunaihe keskustelija, yksimoodiseksi romahdutetun verkoston keskustelijoiden välillisyydet. Kuvassa on ne toimijat joiden välillisuus on nollasta poikkeava. SLP-kontekstissa tämän romahdutetun verkoston välillisyyttä voidaan ajatella näkyvyyden mittarina. Astesummaan verrattuna ne toimijat jotka ovat kommentoineet mahdollisimman monta viestien määrän mukaan suosittua eri keskustelun aiheita nousevat merkittäviksi.

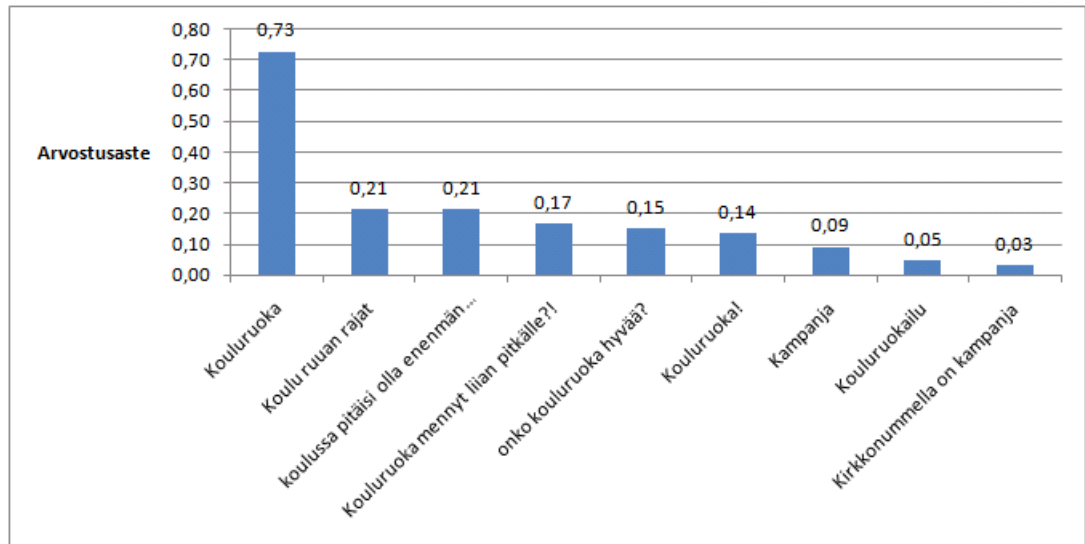


Kuva 5.10: Yksimoodiseksi romahduttetun verkoston toimijoiden välillisyyksiä.

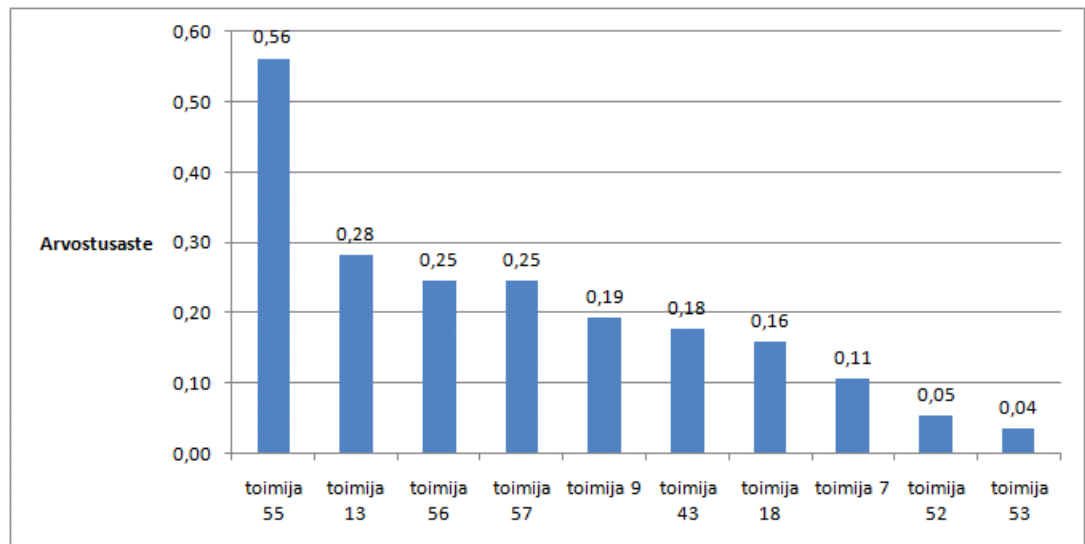
### Arvostusaste

Arvostusaste ilmaisee toimijaan kohdistuvien eli graafin solmuun tulevien yhteyksien määrää. SLP-keskusteluista muodostetussa kaksimoodisissa verkostossa astesumma korostaa verkoston toista moodia eli keskustelun aiheita ja erityisesti määrällisesti mitattuna suosituimpia keskustelunaiheita.

Yksimoodisessa verkostossa arvostusaste korostaa verkoston niitä toimijoita, jotka ovat verkoston mallinnuksen yhteydessä tehdyn tulkinnan mukaisesti kirjoittaneet keskusteluaiheen ensimmäisen viestin eli aloittaneet keskustelun. Verkoston luomissa käytetty tulkinta ja verkostosta saatavat tunnusluvut selittävät miksi kuvien 5.3 ja 5.5 yksi- ja kaksimoodiset verkostot ja niistä saatavat kuvien 5.11 ja 5.12 tunnusluvut muistuttavat toisiaan niin merkittävästi.



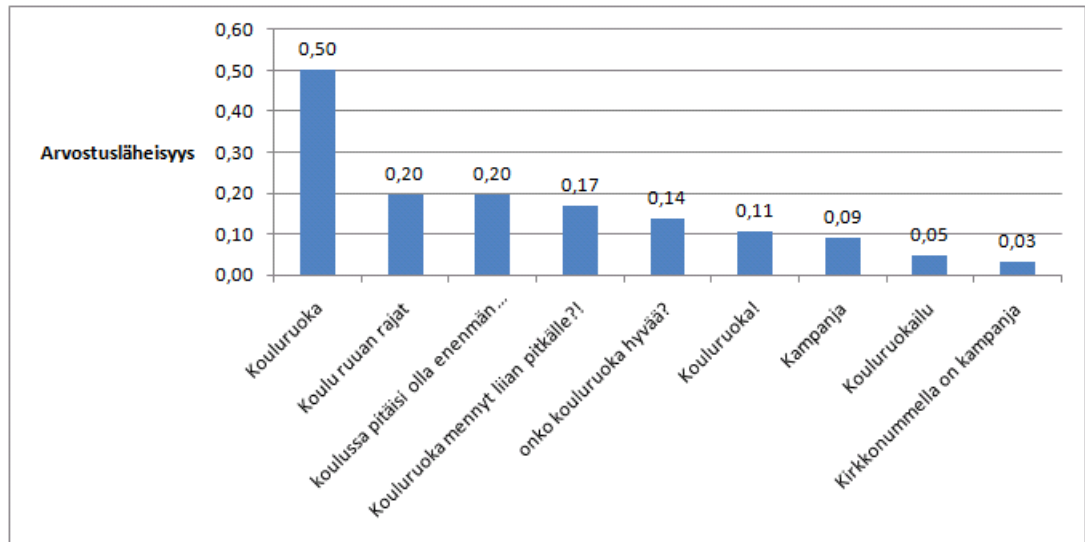
Kuva 5.11: Kaksimoodisen verkoston keskustelunaiheiden arvostusaste.



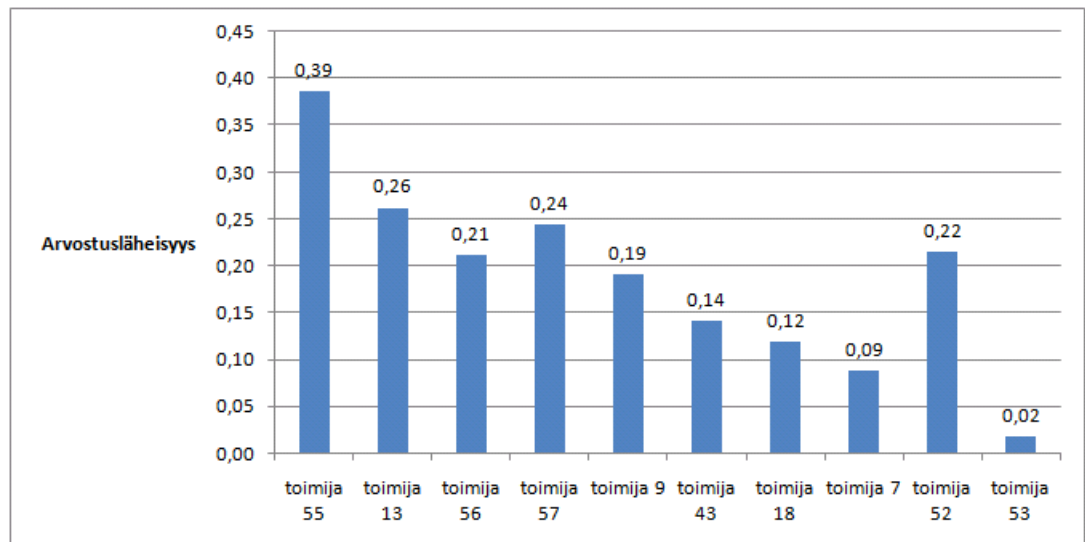
Kuva 5.12: Yksimoodisen verkoston toimijoiden arvostusaste

### Arvostusläheisyys

Arvostusläheisyyttä voidaan kuvata edistyneempänä versiona arvostusasteesta, missä toimijaan kohdistuvien yhteyksien lisäksi verkoston rakenteella on vaikutus arvostukseen. Arvostusläheisyydessä toimijaan tulevien yhteyksien määrä nostaa arvostusta verkostossa, mutta tämän lisäksi arvostusläheisyyteen vaikuttavat verkoston toimijoiden muodostama vaikutusjoukko, joista on yhteys yhden tai useamman solmun välityksellä arvotettavaan toimijaan.



Kuva 5.13: Kaksimoodisen verkoston keskustelunaiheiden arvostusläheisyys.



Kuva 5.14: Yksimoodisen verkoston toimijoiden arvostusläheisyys.

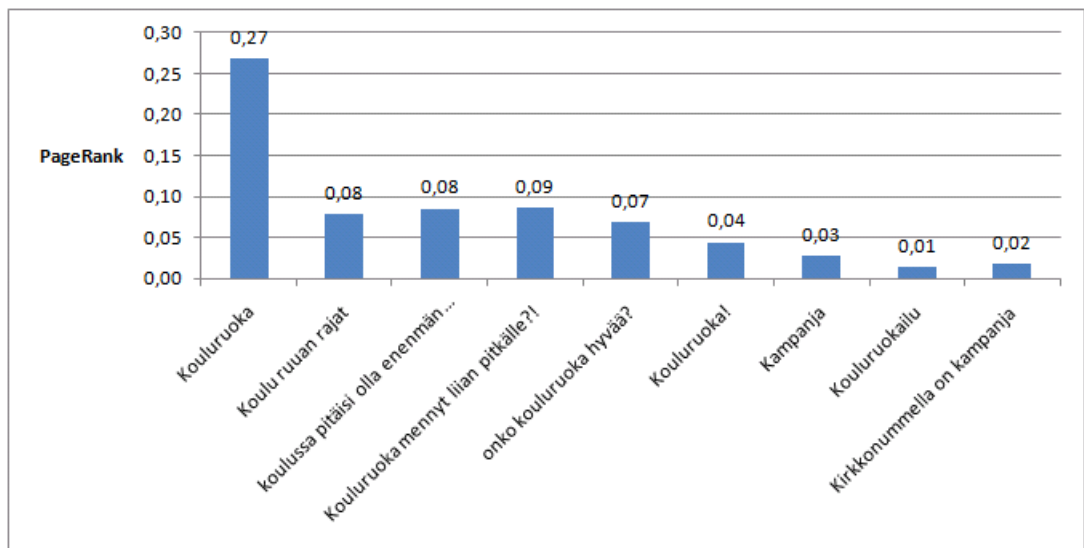
## PageRank

PageRank ilmaisee toimijan arvostusastetta. PageRank ja arvostusläheisyys mittaavat toimijan arvostusta, mutta ne painottavat eri tekijöitä. Arvostusläheisyys painotetaan kuinka monta muuta toimijaa tunnetaan kun PageRankissa tärkeämpää on mikä toimija tunnetaan. PageRankia voidaankin ajatella laadullisena ja arvostusläheisyyttä määrällisenä toimijan arvostuksen mittarina.

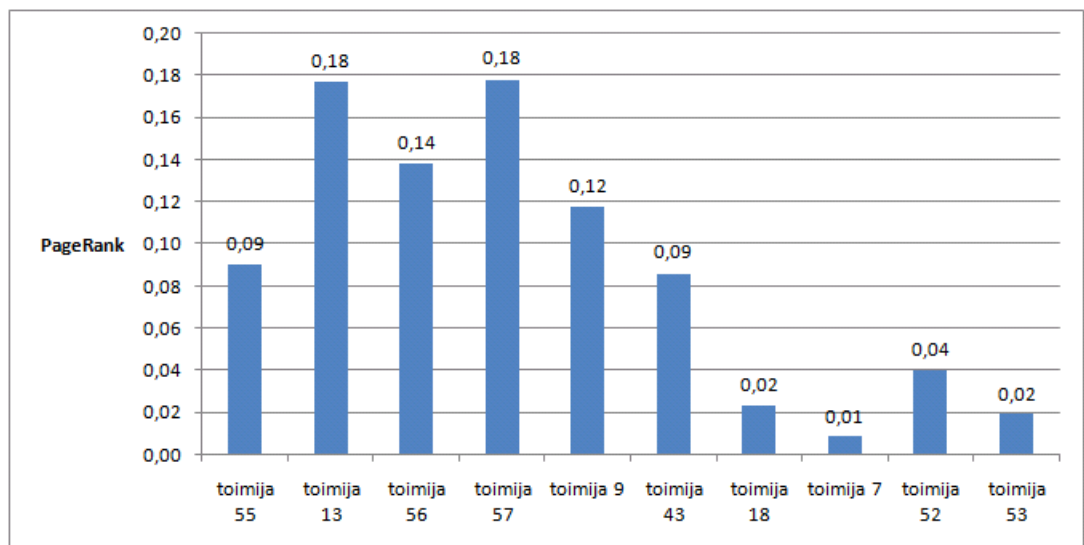
Verrattaessa kuvissa 5.14 ja 5.16 yksimoodisista verkostoista laskettuja arvostusläheisyyttä ja PageRankia voidaan havaita ero arvostetuimpien toimijoiden välillä. Kuvassa 5.14 ylivoimaisesti arvostetuin toimija on toimija 55 mutta PageRankin

mukaan kyseinen toimija on vain melko arvostettu muihin toimijoihin verrattaessa. Tämä tunnuslukujen antama erotus voidaan selittää tarkastelemalla kuvaa 5.3. Kuvassa toimija 55 on suuri punainen solmu, mihin on yhdistynyt useita muita pienempiä toimijoita jotka eivät ole yhteydessä kehenkään muuhun.

Arvostusläheisyys nostaa tämän tyyppiset solmut hyvin merkittäviksi koska niihin on monta tulevaa yhteyttä ja lyhyt etäisyys solmusta. PageRank arvostaa tämäntyyppisiä solmuja arvokkaimmiksi solmuiksi ne joihin edellämämainitun kaltaiset solmut muodostavat suunnatun yhteyden. Toimija 13 on tämänkaltaisen solmu, johon toimija 55 muodostaa yhteyden. Kuvan 5.16 toimija 13 on hyvin merkittävä PageRankilla mitattuna ja vain melko merkitsevä kuvan 5.14 arvostusläheisyyden mukaan.



Kuva 5.15: Kaksimoodisen verkoston toimijoiden PageRankit.



Kuva 5.16: Yksimoodisen verkoston toimijoiden PageRankit.

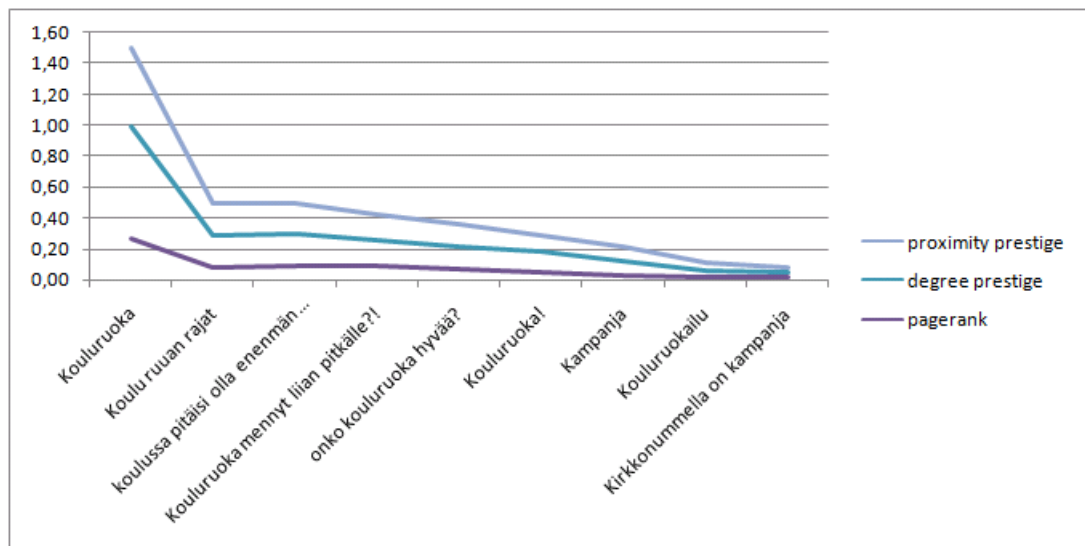
## Verkoston rakenne ja tunnusluvut

Verkostojen laskennallisessa analyysissä keskityttiin toimijoista laskettavien tunnuslukujen vienti- ja tuontilukujen, PageRankin, arvostusasteen, arvostusläheisyyden ja välillisyyden tuottamiseen mallinnetuista verkostoista. Tunnuslukujen merkitystä ja käytettävyyttä on arvioitu SLP-tapauksen verkostojen kontekstissa.

Kuvassa 5.17 on esitetty kaksimoodisesta verkostosta laskettujen PageRank, arvostusläheisyys sekä arvostusaste tunnuslukujen suhteellinen käyttäytyminen eri toimijoiden välillä. Kuvan x-akselilla on verkoston toinen moodi eli keskustelun aiheet ja y-akselilla lasketut tunnusluvut.

Tunnuslukujen välisen korrelaation todentamiseksi on laskettu tunnuslukujen korrelaatio Pearsonin otoskorrelaatiota käyttäen. Pearsonin otoskorrelaatio, tai lyhyemmin korrelaatio, ilmaisee jos kaksi joukkoa korreloivat toisiaan. Näiden suhdetta kuvataan otoskorrelaatiokerroimella  $r$ . Korrelaatiokerroin  $r$  saa arvoja väliltä  $[-1, 1]$ . Plus yhden arvoinen korrelaatio tarkoittaa täydellistä lineaarista riippuvuutta kahden populaation välillä, nolla arvoinen korrelaatio täydellistä riippumattomuutta populaatioiden välillä ja miinus yhden arvoinen korrelaatio täydellistä negatiivista riippuvuutta. (Ruohonen, 2008)

PageRankin  $P_{PageRank}$ , arvostusläheisyyden  $P_P$  sekä arvostusasteen  $P_D$  korrelaatiot on laskettu liitteen A arvoista. Korrelaatioksi saatiin  $r_{P_{PageRank}, P_P} = 0,995$ ,  $r_{P_{PageRank}, P_D} = 0,989$  ja  $r_{P_D, P_P} = 0,990$ . Tunnuslukujen korrelaatiot ovat hyvin lähellä yhtä ja ne siis korreloivat hyvin voimakkaasti keskenään.



Kuva 5.17: Kuvassa on kolmella eri tunnuslukumenetelmällä keskustelunaiheille lasketut tunnusluvut. Käyrät muistuttavat merkittävästi muodoltaan toisiaan ja ne myös korreloivat voimakkaasti keskenään.

Tunnuslukujen vahvaa korrelaatiota selittää verkoston malli, josta ne on laskettu. Kaksimoodiset verkostot (kuva 5.5) joissa toimijat ja tapahtumat liittyvät suoraan



toisiinsa eivätkä muodosta yhteyksiä keskenään muodostavat verkstorakenteen, josta lasketut tunnusluvut korreloivat voimakkaasti keskenään. Tunnukslukujen vahva korrelaatio voidaan myös ymmärtää niin että erilaiset tunnusluvut eivät lisää verkostosta laskennallisista menetelmin saatavan informaation määrää kun eri tunnusluvut korostavat samoja toimijoita samassa suhteessa.

Mallinnetun verkoston rakenteella voidaan lisäksi selittää miksi kaksimoodisesta verkostosta laskettu välillisuus on jokaiselle verkoston toimijalle nolla. Välillisuus ilmaisee kuinka keskeinen solmu on kun se sijaitsee muiden solmujen lyhimpien etäisyyksien varrella. Verkostossa ei kuitenkaan ole solmuja jotka sijaitisivat kahden solmun välissä, jokainen verkoston solmu on kytkeytynyt suoraan yhdellä suunnatulla yhteydellä toiseen solmuun. Näin ollen verkoston kaikkien toimijoiden välillisuus on nolla.

Juuri tämän tyyppinen kaksimoodinen verkosto soveltuukin tunnuslukulaskentaa paremmin verkoston yleisen rakenteen hahmottamiseen visualisointien avustuksella. Tunnukslukujen tuottamista varten verkostosta voidaan ottaa projektioita romahduttamalla se tiettyjen sääntöjen mukaisesti yksimoodiseksi verkostoksi.

### **Verkostojen visualisointi ja tunnusluvut**

Tunnukslukuja voidaan käyttää sosiaalisten verkostojen visualisoinnin tukena. Määrittämällä verkostosta piirrettävän sosiogrammin visuaalisia piirteitä tunnukslukujen avulla (kts. kuvat 5.3, 5.5, 5.7), voidaan lisätä sosiaalista verkostoa esittävän kuvan informatiivisuutta. Tapoja joilla graafin visuaalista ulkoasua voidaan muokata on monia. Esimerkiksi solmujen väriä ja kokoa voidaan muokata, kaarien paksuutta muuttaa ja solmujen asemaa graafissa voidaan vaihtaa.

## **5.4 Pohdintaa**

Diplomityössä web-pohjaisista keskustelualueista mallinnetut ja analysoidut sosiaaliset verkostot ovat pelkistetty näkemys ihmisten muodostamista sosiaalisista verkostoista. Analyysissä ei ole esimerkiksi mallinnettu erikseen tapauksia, joissa käyttäjät ovat lainanneet toistensa kirjoittamia viestejä keskustelualueilla. SLP-tapauksen analyysissä ei ole pyritty mallintamaan käyttäjien välistä verkostoa täydellisesti, vaan ensisijaisesti on pyritty testaamaan toteutusta verkostanalyysin eri osien tiedonlouhinnan, matemaattisen analyysin ja visualisoinnin automatisoinnista.

Alunperin pääasiassa sosiaalitieteilijöiden kehittämiä sekä Wassermanin & Faustinin esittämiä tunnukslukujen laskennallisia menetelmiä ei ole välttämättä toteutusteknisesti tehokasta toteuttaa. Aikana jolloin nämä tunnusluvut on pääasiassa kehitelty ei nykyaikaisia tietojärjestelmiä eikä tietokoneita ole ollut olemassa. Tämän vuoksi diplomityössä on käsitelty myös nykyaikaisempia tunnukslukumenetelmiä ku-

ten PageRankia, jonka toimintaperiaate on loogisella tasolla hyvin samankaltainen Wassemanin & Faustin teoksessa kuvatun Katzin arvoaseman (rank prestige) kanssa.

Toteutuskielenä käytettyä Pythonia on kritisoitu sen hitaudesta, mutta yleisesti ohjelmistojen hitaus erityisesti näin vaatimattomissa datamäärissä on riippuvainen enemmän ohjelmoijan osaamisesta ja algoritmeista kuin ohjelmistokielen hitaudesta. Toisaalta käsiteltäessä hyvin suuria, kymmeniä tuhansia alkioita sisältäviä matriiseja, tulevat vastaan jo tiedon väliaikaisen säilyttämisen ongelmat, jolloin on harkittava miten ja minkälaisessa ympäristössä tunnuslukuja halutaan tuottaa.

Valmiita ohjelmistokomponentteja käytettäessä ei voida aina olla varmoja, dokumentoinnin puutteen tms. seurauksena, miten esimerkiksi tunnuslukujen matemaattiset algoritmit on toteutettu. Verkostosta laskettaviin tunnuslukuihin onkin syytä suhtautua varauksellisesti ja avomielisesti ja niitä käytettäessä on hyvä tiedostaa että ne ovat verkostosta rakenteesta ja laskennallisista menetelmistä riippuvaisia. Verkoston visualisoinnit tunnuslukujen avulla eivät ole absoluuttisia vaan subjektiivisia, verkon rakenteesta ja tulkitsijasta riippuvaisia.

SNA-työvälineitä ja niiden yleistämistä ajatellen datalähtöisessä analyysissä, kuten tämän diplomityön SLP-tapauksessa ei ole olemassa hopealuoti-ratkaisua. Kaikkea ei voida automatisoida, verkostoja joudutaan muodostamaan sekä mallintamaan aina jonkin tulkinnan perusteella ja nämä tulkinnat vaikuttavat lopputulokseen. Kuitenkin soveltamalla menetelmiä harkinnan mukaan ja tiedostamalla mitä eri laskennalliset menetelmät tarkoittavat eri konteksteissa voidaan SLP-tapauksessa sovellettua työvälinettä ja menetelmiä yleistää ja hyödyntää muissa vastaanvanlaisissa konteksteissa.

## 6. YHTEENVETO JA JOHTOPÄÄTÖKSET

Tässä diplomityössä on esitetty sosiaalisten verkostojen matemaattisia analysointimenetelmiä, sovellettu niitä web-pohjaisiin keskustelualueisiin ja visualisoitu tuloksia automatisoidusti. Verkostoanalyysimenetelmät ovat sovellettavissa minkä tahansa tekijöiden muodostamiin verkostoihin. Yksi sovellusala on ihmisten välisten vuorovaikutteisten tapahtumien mallintaminen verkostoiksi ja niiden analysoiminen eli sosiaalisten verkostojen analysointi.

Sosiaalisten verkostojen matemaattiset ja laskennalliset menetelmät pohjautuvat pääosin graafiteoriaan ja matriisilaskentaan. Laskennallisen toteutuksen kannalta pääasiassa sosiaalitieteilijöiden ja matemaatikkojen kehittämiä verkostosta laskettavia tunnuslukumenetelmiä ei ole välttämättä aina tehokasta toteuttaa. Laskennallisten menetelmien lisäksi on sovellettava sisällöllistä analyysia tulosten tulkitsemiseksi. Sisällöllisen analyysin ja laskennallisen analyysin tukena on käytetty sosiaalisten verkostojen visualisoimista graafeilla.

Erilaisia laskennallisia menetelmiä on sovellettu diplomityön Suomen Lasten Parlamentti -tapauksessa. SLP-tapauksessa on analysoitu web-pohjaisille keskustelualueille kirjoittaneiden lasten muodostamia sosiaalisia verkostoja. Analyysissa on sovellettu diplomityössä esiteltyjä verkostoanalyysin matemaattisia menetelmiä. Osana tapausta on luotu prototyyppi visualisointi-työvälineestä, jonka avulla verkostoanalyysin eri osat, tiedonlouhinta, matemaattinen analyysi ja visualisointi, on voitu paketoita yhdeksi kontekstiherkäksi kokonaisuudeksi.

SLP-tapauksen keskustelunalueista mallinnettiin kolme verkostoa keskustelijoiden välinen yksimoodinen verkosto (kuva 5.3), keskustelijoiden ja keskustelunaiheiden välinen kaksimoodinen verkosto (kuva 5.5) sekä edellämäinitusta verkostosta keskustelunalueiden suhteen romahdutettu verkosto (kuva 5.7). Verkostojen mallintaminen edellä mainituilla tavoilla oli tietoinen päätös, mutta jälkeenpäin tarkasteltuna ei välttämättä paras vaihtoehto, koska keskusteluiden mallinnuksessa ei otettu huomioon muun muassa viestien lainausmahdollisuutta ja kaksimoodisen verkoston tapauksessa havaittiin, että tunnusluvut korreloivat hyvin voimakkaasti keskenään.

Tunnusluvuilla löydettiin ja visuaalisesti korostettiin keskusteluista mallinnetuista verkostoista erilaisia toimijoita (luku 5.3). Astesumman avulla löydettiin viestien määrällä mitattuna suosituimmat keskustelunaiheet kaksimoodisesta keskustelija keskustelunaihe verkostosta, yksimoodisesta verkostosta suosituimpien keskus-

telunaiheiden aloittaneet toimijat sekä romahdetusta verkostosta näkyvimvät toimijat.

Välillisyyden avulla yksimoodisesta verkostosta sekä romahdetusta verkostosta löydettiin toimijat, jotka ovat kommentoineet muiden toimijoiden viestejä. Arvos-  
tusasteen avulla verkostosta löydettiin toimijat, jotka ovat kommentointien kohteena. PageRankin avulla löydettiin ne toimijat, jotka ovat määrällisesti mitattuna merkittävimpien kommentoijien kohteina.

Tulevaisuudessa haasteena on mallintaa verkostoja tarkemmin esimerkiksi ottamalla verkostomallissa huomioon keskusteluiden lainaamisen sekä niihin viittaamisen ja soveltaa tässä työssä käytettyjä laskennallisia menetelmiä tulosten analysoimiseksi ja vertaamiseksi. Toinen mielenkiintoinen haaste on tutkia erityisesti viestien välittymistä verkoston sisällä toimijalta toiselle. Esimerkiksi Zhou (Zhou et al, 2010) on käsitellyt informaation välittymistä Twitter-palvelun verkostoissa. Seuraavaksi voitaisiin analysoida informaation leviämistä SLP-tapauksen verkostoissa ja kehittää malli, jolla ennustetaan viestien määrällistä leviävyyttä verkostossa toimijoiden välityksellä.

## LÄHTEET

- Bastian M. & Heymann S. & Jacomy M. 2009. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. Association for the Advancement of Artificial Intelligence, 2009. Saatavilla  
<http://gephi.org/publications/gephi-bastian-feb09.pdf>
- Breiger, R. 1974. *The duality of persons and groups*. Social Forces. 53, 181-190. University of North Carolina Press.
- Brin, S. & Page, L. 1998 *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. Computer Science Department, Stanford University.
- Craven, P. *Google's PageRank Explained and how to make most of it*.  
 Lainattu 3.11.2010. Saatavilla  
<http://www.webworkshop.net/pagerank.html>
- Freeman, L. C. 2009. *Methods of Social Network Visualization*. Teoksessa R. A. Meyers (toim.) *Encyclopedia of Complexity and Systems Science*. Berlin: Springer.
- Greasemonkey*. Lainattu 6.9.2010. Saatavilla  
<http://www.greasespot.net/>
- Huhtamäki, J. & Salonen, J. & Marttila, J. & Nykänen, O. To appear 2010. *Towards Context-Sensitive, Visual Social Network Analysis: Case Wiki Co-Creation*. Tampere University of Technology.
- The igraph project. *The igraph library for complex network research*. Lainattu 15.9.2010. Saatavilla  
<http://igraph.sourceforge.net/>
- Kalliala, E. & Toikkanen T. 2009. *Sosiaalinen media opetuksessa*. Helsinki: Finn Lectura.
- Knoke, D. & Yang, S. 2008. *Social Network Analysis*. Second edition. Los Angeles: Sage Publications.
- Langville, A. N. & Meyer ,C. D. 2004. *The Use of Linear Algebra by Web Search Engines*. North Carolina State University. Raleigh: North Carolina State University. Saatavilla  
[http://meyer.math.ncsu.edu/Meyer/PS\\_FILES/IMAGE.pdf](http://meyer.math.ncsu.edu/Meyer/PS_FILES/IMAGE.pdf)
- Johanson, J-E. , Uusikylä P. & Mattila, M. 1998. *Johdatus verkostoanalyysiin*. Helsinki: Kuluttajatutkimuskeskus.

Lainattu 12.10.2010. Saatavilla

<http://www.valt.helsinki.fi/vol/kirja/>

JUHTA - Julkisen hallinnon tietohallinnon neuvottelukunta. 2008. *JHS 152 Prosessien kuvaaminen*. Saatavilla

<http://www.jhs-suositukset.fi/suomi/jhs152>

Lietsala, K., & Sirkkunen, E. 2008. *Social Media Introduction to the tools and processes of participatory economy*. Tampere: Tampere University Press.

Miilumäki, T. 2010. *Web-pohjaisten sosiaalisten verkostojen analyysimenetelmät*. Tampereen teknillinen yliopisto. Juvenes Print TTY.

Moler, C. 2009. *Google PageRank*. Saatavilla

<http://www.mathworks.com/moler/exm/chapters/PageRank.pdf>

Mrvar, A. *Network Analysis using Pajek*. Saatavilla

<http://mrvar.fdv.uni-lj.si/sola/info4/>

NodeXL. *NodeXL: Network Overview, Discovery and Exploration for Excel*. Saatavilla

<http://nodexl.codeplex.com>

Numpy. *Scientific Computing Tools For Python*. Saatavilla

<http://numpy.scipy.org/>

Nykänen, O. & Mannio, M. & Huhtamäki J. & Salonen J. 2007. *A Socio-technical framework for visualising an open knowledge space*. Tampere University of Technology

Page, L. & Brin, S. & Motwani, R & Winograd, T. 1999. *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford InfoLab.

Salonen, J. & Huhtamäki, J. 2010. *Launching Context-Aware Visualisations*. 3rd International OPAALS Conference on Digital Ecosystems. Aracaju, Brazil, 22.-23.3.2010.

Sanastokeskus TSK. 2010. *Sosiaalisen median sanasto*. Helsinki: Sanastokeskus TSK ry. Saatavilla

[http://www.tsk.fi/tiedostot/pdf/Sosiaalisen\\_median\\_sanasto](http://www.tsk.fi/tiedostot/pdf/Sosiaalisen_median_sanasto)

*Suomen Lasten parlamentti*. 2010. Saatavilla

[www.lastenparlamentti.fi](http://www.lastenparlamentti.fi)

- O'Reilly, T. *What is Web 2.0*. Saatavilla  
<http://oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1>
- Ruohonen, K. 2008. *Tilastomatematiikka*. Saatavilla  
<http://math.tut.fi/ruohonen/TM.pdf>
- Ruohonen, K. 2006. *Graafiteoria*. Saatavilla  
<http://math.tut.fi/ruohonen/GT.pdf>
- Virk, R. 2002. *Why Convert Documents into XML?*. Saatavilla  
<http://www.cambridgedocs.com/resources/whitepapers/id35.htm>
- Wasserman, S. & Faust, K. 1994. *Social Network Analysis Methods and Applications*. New York: Cambridge University Press.
- Watts, D. 1999. *Small Worlds*. Princeton, NJ: Princeton University Press.
- Wellman, B. & Berkowitz, S.D. 1988. *Structural analysis: From method and metaphor to theory and substance*. Wellman, B. & Berkowitz, S.D (toim.) *Social Structures: A Network Approach 19-61*. Cambridge, England: Cambridge University Press.
- Wille. Tampereen teknillinen yliopisto. Saatavilla  
<http://wiki.tut.fi/Wille/WebHome>
- Zhou, Z. & Bandari, R. & Kong, J. & Qian, H. & Roychowdhury, V. 2010. *Information Resonance on Twitter: Watching Iran*. 1st Workshop on Social Media Analytics (SOMA '10). Washington, DC, USA.

## A. KAKSIMOODISEN KOULURUOKAKESKUSTELUVERKOSTON TUNNUSLUVUT

toimija	$d_{igraph}$	$d_I$	$d_O$	$P_{PageRank}$	$P_D$	$C_B$	$P_P$
Kouluruoka	35	48	0	0,26768	0,72727	0	0,5
Koulu ruuan rajat	13	14	0	0,078239	0,21212	0	0,19697
koulussa pitäisi olla enen- män jälkiruokia	13	14	0	0,084929	0,21212	0	0,19697
Kouluruoka mennyt liian pitkälle?!	11	11	0	0,085926	0,16667	0	0,16667
onko kouluruoka hyvää?	9	10	0	0,068008	0,15152	0	0,13636
Kouluruoka!	7	9	0	0,043873	0,13636	0	0,10606
Kampanja	6	6	0	0,027083	0,090909	0	0,090909
Kouluruokailu	3	3	0	0,014307	0,045455	0	0,045455
Kirkkonummella on kam- panja	2	2	0	0,017507	0,030303	0	0,030303
toimija 0	2	0	2	0,005387	0	0	0
toimija 1	1	0	1	0,005387	0	0	0
toimija 2	1	0	1	0,005387	0	0	0
toimija 3	2	0	2	0,005387	0	0	0
toimija 4	1	0	1	0,005387	0	0	0
toimija 6	1	0	1	0,005387	0	0	0
toimija 7	4	0	4	0,005387	0	0	0
toimija 8	1	0	1	0,005387	0	0	0
toimija 9	1	0	1	0,005387	0	0	0
toimija 10	1	0	1	0,005387	0	0	0
toimija 11	1	0	1	0,005387	0	0	0
toimija 12	2	0	2	0,005387	0	0	0
toimija 13	1	0	2	0,005387	0	0	0
toimija 14	1	0	1	0,005387	0	0	0
toimija 15	1	0	1	0,005387	0	0	0
toimija 16	1	0	1	0,005387	0	0	0
toimija 17	2	0	2	0,005387	0	0	0
toimija 18	2	0	3	0,005387	0	0	0
toimija 19	1	0	1	0,005387	0	0	0
toimija 35	1	0	1	0,005387	0	0	0
toimija 20	4	0	4	0,005387	0	0	0



toimija 21	1	0	1	0,005387	0	0	0
toimija 22	1	0	2	0,005387	0	0	0
toimija 23	1	0	1	0,005387	0	0	0
toimija 24	1	0	1	0,005387	0	0	0
toimija 25	5	0	7	0,005387	0	0	0
toimija 26	3	0	3	0,005387	0	0	0
toimija 27	1	0	1	0,005387	0	0	0
toimija 28	1	0	2	0,005387	0	0	0
toimija 29	1	0	2	0,005387	0	0	0
toimija 30	1	0	2	0,005387	0	0	0
toimija 31	1	0	1	0,005387	0	0	0
toimija 32	2	0	2	0,005387	0	0	0
toimija 33	2	0	2	0,005387	0	0	0
toimija 34	1	0	1	0,005387	0	0	0
toimija 36	6	0	7	0,005387	0	0	0
toimija 37	8	0	10	0,005387	0	0	0
toimija 38	1	0	2	0,005387	0	0	0
toimija 39	1	0	1	0,005387	0	0	0
toimija 40	2	0	2	0,005387	0	0	0
toimija 46	1	0	1	0,005387	0	0	0
toimija 41	1	0	1	0,005387	0	0	0
toimija 42	2	0	3	0,005387	0	0	0
toimija 43	1	0	1	0,005387	0	0	0
toimija 44	2	0	3	0,005387	0	0	0
toimija 45	1	0	1	0,005387	0	0	0
toimija 47	1	0	1	0,005387	0	0	0
toimija 48	2	0	3	0,005387	0	0	0
toimija 5	1	0	1	0,005387	0	0	0
toimija 49	4	0	6	0,005387	0	0	0
toimija 50	1	0	1	0,005387	0	0	0
toimija 51	1	0	1	0,005387	0	0	0
toimija 52	3	0	3	0,005387	0	0	0
toimija 53	1	0	1	0,005387	0	0	0
toimija 54	1	0	1	0,005387	0	0	0
toimija 55	2	0	3	0,005387	0	0	0
toimija 56	1	0	1	0,005387	0	0	0
toimija 57	1	0	1	0,005387	0	0	0

## B. YKSIMOODISEN KOULURUOKAKESKUSTELUVERKOSTON TUNNUSLUVUT

toimija	$d_{astesumma}$	$d_O$	$d_I$	$P_{PageRank}$	$P_D$	$C_B$	$P_P$
toimija 55	35	3	32	0,089843	0,5614	59	0,3867
toimija 13	18	2	16	0,17701	0,2807	0	0,26172
toimija 56	15	1	14	0,13805	0,24561	0	0,21126
toimija 57	15	1	14	0,17823	0,24561	0	0,24434
toimija 9	12	1	11	0,1175	0,19298	0	0,19103
toimija 43	11	1	10	0,085876	0,17544	0	0,14035
toimija 18	12	3	9	0,023001	0,15789	5	0,11842
toimija 7	10	4	6	0,008083	0,10526	13	0,087719
toimija 52	6	3	3	0,039566	0,052632	39	0,21501
toimija 53	3	1	2	0,018707	0,035088	0	0,017544
toimija 0	2	2	0	0,002586	0	0	0
toimija 1	1	1	0	0,002586	0	0	0
toimija 2	1	1	0	0,002586	0	0	0
toimija 3	2	2	0	0,002586	0	0	0
toimija 4	1	1	0	0,002586	0	0	0
toimija 5	1	1	0	0,002586	0	0	0
toimija 6	1	1	0	0,002586	0	0	0
toimija 8	1	1	0	0,002586	0	0	0
toimija 10	1	1	0	0,002586	0	0	0
toimija 11	1	1	0	0,002586	0	0	0
toimija 12	2	2	0	0,002586	0	0	0
toimija 14	1	1	0	0,002586	0	0	0
toimija 15	1	1	0	0,002586	0	0	0
toimija 16	1	1	0	0,002586	0	0	0
toimija 17	2	2	0	0,002586	0	0	0
toimija 19	1	1	0	0,002586	0	0	0
toimija 20	4	4	0	0,002586	0	0	0
toimija 21	1	1	0	0,002586	0	0	0
toimija 22	2	2	0	0,002586	0	0	0
toimija 23	1	1	0	0,002586	0	0	0
toimija 24	1	1	0	0,002586	0	0	0
toimija 25	7	7	0	0,002586	0	0	0
toimija 26	3	3	0	0,002586	0	0	0

toimija 27	1	1	0	0,002586	0	0	0
toimija 28	2	2	0	0,002586	0	0	0
toimija 29	2	2	0	0,002586	0	0	0
toimija 30	2	2	0	0,002586	0	0	0
toimija 31	1	1	0	0,002586	0	0	0
toimija 32	2	2	0	0,002586	0	0	0
toimija 33	2	2	0	0,002586	0	0	0
toimija 34	1	1	0	0,002586	0	0	0
toimija 35	1	1	0	0,002586	0	0	0
toimija 36	7	7	0	0,002586	0	0	0
toimija 37	10	10	0	0,002586	0	0	0
toimija 38	2	2	0	0,002586	0	0	0
toimija 39	1	1	0	0,002586	0	0	0
toimija 40	2	2	0	0,002586	0	0	0
toimija 41	1	1	0	0,002586	0	0	0
toimija 42	3	3	0	0,002586	0	0	0
toimija 44	3	3	0	0,002586	0	0	0
toimija 45	1	1	0	0,002586	0	0	0
toimija 46	1	1	0	0,002586	0	0	0
toimija 47	1	1	0	0,002586	0	0	0
toimija 48	3	3	0	0,002586	0	0	0
toimija 49	6	6	0	0,002586	0	0	0
toimija 50	1	1	0	0,002586	0	0	0
toimija 51	1	1	0	0,002586	0	0	0
toimija 54	1	1	0	0,002586	0	0	0

## C. YKSIMOODISEKSI ROMAHDUTETUNKOU- LURUOKAKESKUSTELUVERKOSTON TUNNUSLUVUT

toimija	$d_O$	$d_I$	$d_I scaled$	$P_{pagernak}$	$P_D$	$C_B$	$P_P$
toimija 37	52	52	1	0,036975	0,91228	385,93	0,90476
toimija 36	50	50	0,961538462	0,033255	0,87719	265,93	0,89063
toimija 20	49	49	0,942307692	0,033045	0,85965	254,95	0,87692
toimija 25	46	46	0,884615385	0,030241	0,80702	154,33	0,83824
toimija 7	43	43	0,826923077	0,028867	0,75439	168,91	0,80282
toimija 26	43	43	0,826923077	0,027446	0,75439	84,929	0,80282
toimija 17	39	39	0,75	0,024705	0,68421	50,551	0,76
toimija 3	39	39	0,75	0,025357	0,68421	90,327	0,76
toimija 40	37	37	0,711538462	0,023166	0,64912	29,29	0,74026
toimija 12	37	37	0,711538462	0,023913	0,64912	60,127	0,74026
toimija 0	37	37	0,711538462	0,023166	0,64912	29,29	0,74026
toimija 33	37	37	0,711538462	0,023913	0,64912	60,127	0,74026
toimija 42	37	37	0,711538462	0,023166	0,64912	29,29	0,74026
toimija 52	37	37	0,711538462	0,023166	0,64912	29,29	0,74026
toimija 44	33	33	0,634615385	0,02031	0,57895	0	0,69512
toimija 48	33	33	0,634615385	0,020698	0,57895	11,2	0,7037
toimija 29	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 6	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 8	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 13	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 22	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 28	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 21	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 35	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 27	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 30	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 16	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 24	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 38	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 41	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 46	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 51	32	32	0,615384615	0,019784	0,5614	0	0,69512
toimija 55	32	32	0,615384615	0,019784	0,5614	0	0,69512

toimija 49	28	28	0,538461538	0,0227	0,49123	126,49	0,66279
toimija 32	20	20	0,384615385	0,015133	0,35088	17,676	0,60638
toimija 18	15	15	0,288461538	0,012647	0,26316	9,3667	0,57576
toimija 1	12	12	0,230769231	0,009688	0,21053	0	0,55882
toimija 2	12	12	0,230769231	0,010181	0,21053	0	0,55882
toimija 39	12	12	0,230769231	0,009688	0,21053	0	0,55882
toimija 23	12	12	0,230769231	0,010181	0,21053	0	0,55882
toimija 34	12	12	0,230769231	0,009688	0,21053	0	0,55882
toimija 54	12	12	0,230769231	0,010181	0,21053	0	0,55882
toimija 56	12	12	0,230769231	0,010181	0,21053	0	0,55882
toimija 57	12	12	0,230769231	0,009688	0,21053	0	0,55882
toimija 9	10	10	0,192307692	0,00963	0,17544	0	0,54286
toimija 11	10	10	0,192307692	0,00963	0,17544	0	0,54286
toimija 15	10	10	0,192307692	0,00963	0,17544	0	0,54286
toimija 19	10	10	0,192307692	0,00963	0,17544	0	0,54286
toimija 45	10	10	0,192307692	0,00963	0,17544	0	0,54286
toimija 47	10	10	0,192307692	0,00963	0,17544	0	0,54286
toimija 14	8	8	0,153846154	0,008149	0,14035	0	0,53774
toimija 10	8	8	0,153846154	0,008149	0,14035	0	0,53774
toimija 43	8	8	0,153846154	0,008149	0,14035	0	0,53774
toimija 50	8	8	0,153846154	0,008149	0,14035	0	0,53774
toimija 5	6	6	0,115384615	0,006671	0,10526	0	0,52778
toimija 31	6	6	0,115384615	0,006671	0,10526	0	0,52778
toimija 4	5	5	0,096153846	0,005419	0,087719	0	0,52294
toimija 53	1	1	0,019230769	0,003191	0,017544	0	0,47899