Kompleksisten verkostojen fysiikkaa

Jari Saramäki[1], Jukka-Pekka Onnela[1,2], Kimmo Kaski[1]
[1] Laskennallisen tekniikan laboratorio, Teknillinen korkeakoulu
[2] Department of Physics, Clarendon Laboratory, University of Oxford

Johdanto

Luonnontieteellisessä tutkimuksessa on viime vuosina ollut käynnissä paradigman murros. Tieteenhaarojen pari vuosisataa jatkunut eriytyminen on vaihtumassa poikki-tieteisyydeksi. Fyysikoiden osa tässä murroksessa on merkittävä johtuen erityisesti fysiikan teoreettisten ja tilastollisten menetelmien yleispätevyydestä. Poikkitieteisyys ulottuu myös kovien luonnontieteiden ulkopuolelle, eikä esimerkiksi tilastollisen fysiikan soveltaminen sosiaalisten systeemien tutkimukseen ole tänä päivänä tavatonta. Tavallaan kyse on ympyrän sulkeutumisesta – esimerkiksi **Maxwell** näki yhtäläisyyksiä molekyylien ja ihmisyksilöiden muodostamien systeemien tilastollisissa lainalaisuuksissa [1]. Yksi tärkeä poikkitieteellisen tutkimuksen ja fysiikan yhtymäkohta on kompleksisten verkostojen tutkimus [2].

Sanaa ”kompleksinen” käytetään usein esimerkiksi biologisten ja sosiaalisten järjestelmien yhteydessä. Käsite ei ole tarkasti määritelty – ei ole olemassa yksikäsitteistä mittaria, jonka avulla voitaisiin sanoa jonkin systeemin olevan toista kompleksisempi. Yleensä kompleksisella systeemillä tarkoitetaan suurta joukkoa (ei-identtisiä) elementtejä, joiden väliin vuorovaikutuksiin voi liittyä satunnaisuutta ja monimutkaista rakennetta. Tällaisen systeemin kollektiivinen käytös ei ole nähtävissä yksittäisten elementtien ominaisuuksista. Tyypillisiä esimerkkejä ovat talous, liikennevirrat, ihmisyhteisöt sekä lähes kaikki biologiset systeemit, kuten ekosysteemit, solut ja geenien toiminta.

Verkostotarkastelussa kompleksisista systeemeistä pyritään redusoimaan esiin olennainen. Yksinkertaisimmassa kuvauksessa elementtejä kuvataan solmuilla ja niiden välisiä vuorovaikutuksia kaarilla. Vuorovaikutusten verkostorakenteella on suuri vaikutus systeemin dynamiikkaan. Systeemien käyttäytymistä voidaankin ymmärtää analysoimalla niiden rakennetta: esimerkiksi metaboliaverkostojen tiheästi kytketyt klusterit voidaan liittää solubiologisiin toimintoihin. Myös mallinnuksessa ja simuloinnissa verkostorakenteen tunteminen on olennaista.

Verkostojen lyhyt historia

Vaikka matemaattista graafiteoriaa onkin tutkittu jo **Eulerin** ajoista lähtien, on kompleksisten verkostojen tutkimus tieteenalana nuori. Ensimmäiset askeleet kompleksisten verkostojen tutkimuksessa ottivat 1950-luvulla unkarilaiset matemaatikot **Pál Erdõs** ja **Alfréd Rényi** yksinkertaisella satunnaisverkon mallillaan [3]. Erdõs-Rényi-satunnaisverkko sisältää N solmua siten, että jokaista solmuparia yhdistää kaari todennäköisyydellä p. Tällaisten satunnaisverkkojen muodostama *ensemble* on varsin rikas ja useat sen ominaisuuksista ovat eksaktisti ratkaistavissa ”termodynaamisella” rajalla $N \to \infty$ s.e. pN = vakio. Esimerkiksi pienillä p:n arvoilla verkosto koostuu häviävän pienistä epäyhtenäisistä komponenteista, ja p:n arvon kasvattaminen johtaa perkolaatiotransition kautta tilaan, jossa äärellinen osuus solmuista kuuluu yhtenäiseen nk. jättiläiskomponenttiin. Kytkös tilastolliseen fysiikkaan on ilmeinen.

Seuraava suurempi edistysaskel oli **Wattsin** ja **Strogatzin** ”pieni maailma”-tutkimus vuonna 1998 [4]. Aineistona olivat mm. Internet Movie Database -tietokannasta johdettu näyttelijöiden yhteistyöverkosto ja *C. Elegans* -madon hermoverkko. ”Pieni maailma” viittaa siihen, että solmut ovat suuressakin verkossa keskimäärin vain pienen määrän kaaria päässä toisistaan. Havainto ei sinänsä ollut yllättävä, ja kirjailija **Frigyes Karinthy** esittikin tämänsuuntaisia ajatuksia ihmisyhteisöön liittyen jo vuonna 1929. Sosiologi **Stanley Milgram** totesikin 1960-luvulla kuuluisan ketjukirjekokeensa perusteella, että kaksi satunnaisesti valittua henkilöä ovat keskimäärin kuuden tuttavuuden päässä toisistaan. Myös Erdõs-Rényi-verkkojen lyhimmät solmuja yhdistävät polut ovat lyhyitä, ja niiden pituudet skaalautuvat logaritmisesti koon funktiona: $l \propto \ln(N)$. Lyhyiden polkujen voidaan siis ajatella olevan satunnaisten verkostojen ominaisuus. Watts ja Strogatz havaitsivat kuitenkin aineistossaan merkittävän poikkeaman tästä ominaisuudesta, koska empiirisissä verkostoissa kaaret ovat klusteroituneita: yksittäisen solmun naapuritkin ovat todennäköisesti kytkettyjä kaarella. Satunnaisissa verkoissa tämä todennäköisyys on pieni.

Klusteroituminen ilmentää säännöllisyyttä, jolloin empiiristen verkostojen voidaan ajatella sijoittuvan satunnaisen ja säännöllisen verkkojen välimaastoon. Watts ja Strogatz esittivätkin mallin, jossa säännöllisen, klusteroituneen verkoston kaaria uudelleenkytketään satunnaisiin solmuihin todennäköisyydellä p. Jo pienillä p:n arvoilla polun pituus romahtaa klusteroitumisen pysyessä korkeana. Tällä "pieni maailma"-alueella verkostossa tapahtuva dynamiikka muuttuu: esimerkiksi oskillaattoreiden synkronisoituminen nopeutuu. Verkostojen rakenteella on siis keskeinen vaikutus niissä esiintyviin ilmiöihin. Tämä on merkittävä havainto, sillä useat luonnossa esiintyvät ilmiöt tapahtuvat verkostoissa.

Merkittävä oli myös **Albertin** ja **Barabásin** havainto empiiristen verkostojen astejakauman muodosta [5]. Astejakauma P(k) on todennäköisyys(tiheys)jakauma sille, että satunnaisesti valittuun solmuun liittyy k kaarta. Esimerkiksi Erdõs-Rényi-verkkojen asteet ovat Poisson-jakautuneita, mutta Albert ja Barabási havaitsivat useiden luonnollisten verkostojen astejakauman olevan leveähäntäinen ja approksimoitavissa potenssilailla $P(k) \propto k^{-\gamma}$. Tällaisia verkostoja kutsutaan mittakaavattomiksi (engl. *scale-free*). Mittakaavattomat verkot ovat niissä esiintyvien erittäin suuren asteen solmujen ansiosta vikasietoisia: satunnaisten solmujen poisto ei johda verkon sirpaloitumiseen [6]. Lisäksi suuren asteen solmut vaikuttavat merkittävästi verkostoissa esiintyviin ilmiöihin, esimerkiksi stokastiseen infektioiden leviämiseen. Jo häviävän pieni tartuntatodennäköisyys johtaa verkon laajuiseen epidemiaan äärettömissä verkoissa, joille $P(k) \propto k^{-\gamma}$ ja $\gamma \in [2,3]$. Äärellisissäkin verkoissa kynnystodennäköisyys on hyvin pieni. Tällä keskeisellä tuloksella [7] on perusteltu tietokonevirusten laajaa levinneisyyttä. Barabási ja Albert pystyivät myös selittämään mittakaavattomien jakaumien synnyn yksinkertaisella mallilla [5]. Olennaista on, että luonnolliset verkostot eivät ole tasapainotilassa, vaan kasvavat aikaskaaloilla jotka vaihtelevat evoluutiosta kotisivujen päivitystahtiin. Astejakauma noudattaa asymptoottisesti potenssilakia, kun verkkoon liittyvät uudet solmut kytkeytyvät mieluummin sellaisiin solmuihin, joilla on jo ennestään paljon kaaria – esimerkiksi www-sivut linkitetään jo ennestään suosittuihin sivustoihin.

Todellisilla verkostoilla on siis yleisiä ominaispiirteitä: lyhyet polunpituudet, korkea klusteroitumisaste ja leveät astejakaumat. Näiden alle mahtuu toki paljon vaihtelua – esimerkiksi sosiaalisissa verkostoissa naapurisolmujen asteet korreloivat positiivisesti (suositut ihmiset ovat tekemisissä toistensa kanssa), kun taas biologisissa verkoissa korrelaatio on negatiivinen. Lisäksi kaarten kuvaamien vuorovaikutusten vahvuudet vaihtelevat – tämä voidaan verkostokuvauksessa huomioida kaarten painoina. Tällä hetkellä tutkimuksen suunta on siirtymässä painotettujen verkostojen lisäksi kohti "mesoskooppista" tarkastelua, jossa pyritään ymmärtämään esimerkiksi vahvasti klusteroituneiden solmujoukkojen vaikutusta verkostojen toimintaan. Tällaisista joukoista käytetään nimitystä *yhteisö* – nimitys juontaakin juurensa verkostososiologiasta.

Sosiaaliset verkostot

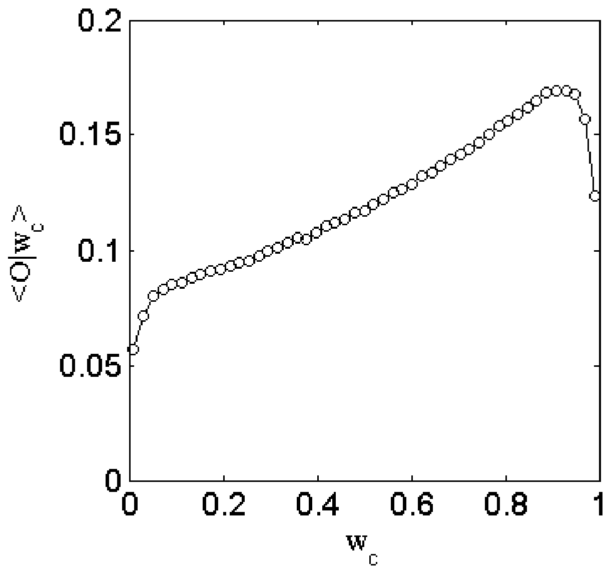
Sosiaalisia verkostoja on yhteiskuntatieteissä tutkittu jo 1930-luvulta lähtien [8]. Yhteiskuntatieteiden näkökulmasta yksilön sosiaalinen elämä koostuu vuorovaikutusten ja ihmissuhteiden verkostosta, joka kanavoi informaation kulkua [9]. Informaatio voidaan tässä yhteydessä ymmärtää laajasti normien, arvojen ja erilaisten sosiaalisten ja kulttuurillisten resurssien virtana, jolloin yksilön asema verkostossa on suoraan sidoksissa hänen sosiaaliseen toimintaansa [10].

Myös fyysikoille sosiaaliseen kanssakäymiseen liittyvät verkostot ovat olleet tärkeitä. Yksi syy tähän on elektronisen datan monipuolinen saatavuus. Esimerkkejä ovat yhteisjulkaisujen perusteella muodostetut tieteelliset yhteistyöverkostot sekä sähköpostiliikenteen ja Internet-palveluiden käyttötietojen pohjalta rakennetut verkostot. Mutta mihin fyysikoita tarvitaan sosiaalisten systeemien tutkimuksessa? Yhteiskuntatieteellisen verkostotutkimuksen yksi avainkysymys on selvittää, kuinka yksittäisten mikroskooppisten vuorovaikutusten tuloksena syntyy makroskooppinen sosiaalinen järjestelmä. Tämän valossa fyysikoiden kiinnostus näyttääkin perustellulta, sillä kyseessä on statistisesta mekaniikasta tuttu kysymyksenasettelu. Fyysikoiden lähestymistapa sosiaalisiin systeemeihin noudattaakin luonnontieteellisen tutkimuksen perusparadigmaa, missä lähdetään liikkeelle kokeellisesta tutkimuksesta, analysoidaan saadut tulokset ja muodostetaan teoria tai malli selittämään havaintoja.

Yhteiskuntatieteilijöiden kiinnostus on usein rajoittunut pieniin sosiaalisiin verkostoihin (solmujen lukumäärä N~100), jolloin aineisto voidaan kerätä esimerkiksi haastattelututkimuksilla. Näin voidaan tavoittaa laaja kirjo erilaisia sosiaalisen vuorovaikutuksen muotoja. Fyysikot ovat puolestaan kiinnostuneita suuremmista systeemeistä (N~10^3-10^6), joista on valmiiksi saatavilla sähköisiä aineistoja. Tässä lähestymistavassa tutkittava sosiaalisen vuorovaikutuksen muoto on yleensä rajoitettu, esimerkiksi em. tieteellinen yhteistyö. Toisaalta datan perusteella voidaan joissakin tapauksissa kvantifioida sosiaalisten vuorovaikutusten voimakkuutta. Molemmilla lähestymistavoilla on siis puolensa, eivätkä ne ole toisensa poissulkevia vaan täydentävät toisiaan.

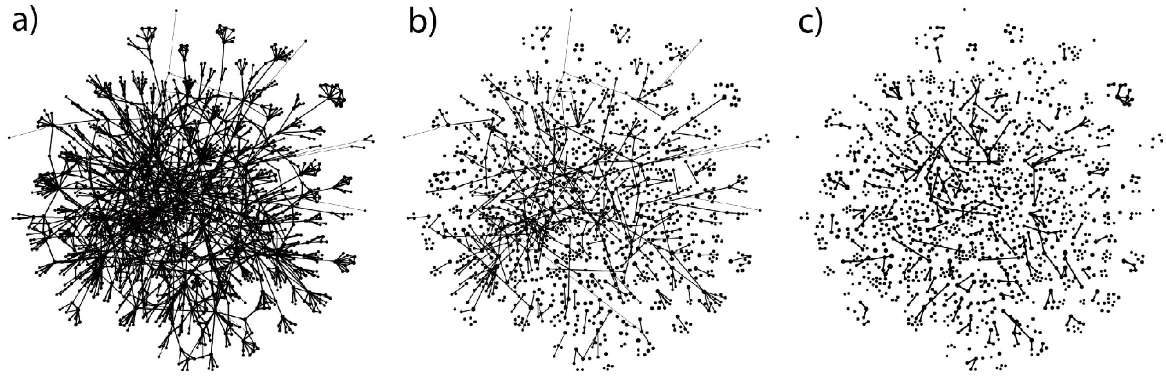
Kirjoittajat osallistuivat tutkimukseen, jossa rekonstruoitiin suuri sosiaalisen kanssakäymisen verkosto [11,12]. Aineistona käytettiin tietokantaa seitsemän miljoonan henkilökohtaisen matkapuhelinliittymän soitoista 18 viikon aikana. Puhelu- ja soittajatiedot oli luonnollisesti anonymisoitu. Verkostoa muodostettaessa kaksi henkilöä yhdistettiin kaarella mikäli kumpikin oli soittanut toiselle seurantajakson aikana. Näin saatu verkosto koostui $N = 4.6\times10^6$ solmusta ja $L = 7.0\times10^6$ kaaresta. Vuorovaikutuksille määriteltiin voimakkuus siten, että henkilöiden *i* ja *j* välisen kaaren paino $w_{ij} = w_{ji}$ on henkilöiden seurantajakson aikana keskenään puhelimessa viettämä kokonaisaika sekunteina. Paino heijastaa siis sosiaalisen siteen vahvuutta. Keskitymme jatkossa puheluverkoston suurimman yhtenäisen komponentin ominaisuuksiin – tämän jokaisesta 3.9 miljoonasta henkilöstä on olemassa reitti kehen tahansa toiseen henkilöön komponentin sisällä.

Sosiologi **Mark Granovetter** kiinnostui 1970-luvulla sosiaalisten siteiden voimakkuuden vaikutuksesta näitä välittömästi ympäröivän sosiaalisen verkoston rakenteeseen [13]. Granovetterin ns. ”heikkojen linkkien” hypoteesin mukaan kahden yksilön tuttavapiirin suhteellinen päällekkäisyys on verrannollinen heitä yhdistävän vuorovaikutuksen voimakkuuteen. Vahvasta vuorovaikutuksesta seuraa siis, että tuttavapiirit (eli verkostossa lähinaapurustot) ovat paljolti päällekkäisiä. Tällöin yhteisöjen sisäiset linkit ovat luonteeltaan vahvoja, kun taas niiden väliset linkit ovat heikkoja. Vaikka hypoteesi on verkostososiologiassa keskeinen, ei sitä ole aiemmin tutkittu muutamaa sataa yksilöä suuremmilla aineistoilla. Hypoteesin testaamiseksi määrittelimme puheluverkon kaarille päällekkäisyyssuureen $O_{ij} = n_{ij}/[(k_i-1)+(k_j-1)-n_{ij}]$, missä n_{ij} on solmujen *i* ja *j* yhteisten naapurien lukumäärä ja k_i (k_j) on solmun *i* (*j*) aste. Suure on normalisoitu välille $[0,1]$: $O_{ij}=0$ jos solmuilla ei ole yhteisiä naapureita, kun taas $O_{ij}=1$, jos kaikki solmujen naapurit ovat yhteisiä. Hypoteesin mukaan kaarten *keskimääräinen* päällekkäisyys $\langle O_{ij}|w\rangle$ kasvaa monotonisesti painon funktiona. Empiirinen tulos on esitetty kuvassa 1: $\langle O_{ij}|w\rangle$ kasvaa painon w funktiona noin 95%:lle kaarista. Poikkeuksen muodostavat vahvimmat 5% kaarista, joille $\langle O_{ij}|w\rangle$ *laskee* painon kasvaessa, toisin sanoen mitä enemmän nämä raskaan sarjan puhujat viettävät puhelimessa aikaa, sitä varmemmin he puhuvat vain yhden henkilön kanssa. Kuitenkin kokonaisuutena tulos tukee vahvasti heikkojen linkkien hypoteesia.



Kuva 1. Kaarien naapuruston keskimääräinen suhteellinen päällekkäisyys $<O|w_c>$ kumulatiivisen painon w_c funktiona.

Kaarten vahvuus on siis selvässä yhteydessä niitä paikallisesti ympäröivän verkon rakenteeseen, mutta millainen merkitys sillä on verkon globaalin rakenteen kannalta? Tätä voidaan tutkia poistamalla verkon kaaria systemaattisesti niiden painon perusteella. Tässä perkolaatioteoriaan perustuvassa tarkastelussa kontrolliparametri $f \in [0,1]$ on poistettujen kaarten osuus kaikista kaarista ja järjestysparametri $R_{GC}(f) \in [0,1]$ määritellään siten, että kun kaarista on poistettu osuus *f*, on $R_{GC}(f)$ suurimpaan jäljellä olevaan yhtenäiseen komponenttiin (sen hetkiseen ”jättiläiskomponenttiin”) kuuluvien solmujen suhteellinen osuus verkoston kaikista solmuista. Yleisesti ottaen verkko osoittautuu varsin sitkeäksi. Kun kaaria poistetaan heikoimmasta vahvimpaan, verkko sirpaloituu täysin ($R_{GC}(f) \approx 0$) vasta, kun $f \approx 0.8$. Päinvastaisessa järjestyksessä kaaria poistettaessa $R_{GC}(f) \approx 0$ vasta, kun lähes kaikki kaaret on poistettu. Tarkempi perkolaatioanalyysi, jossa analysoidaan transitiopisteessä divergoituvia suureita sekä erikokoisia näytteitä verkostosta viittaa siihen, että aloittamalla kaarien poisto vahvoista kaarista perkolaatiotransitio tapahtuu vasta kun $f \approx 1$. Näin ollen heikoilla ja vahvoilla kaarilla on erilainen globaali rooli verkossa. Vahvojen kaarien poistaminen rikkoo verkoston paikalliset yhteisöt vahingoittamatta globaalia kytkeytyneisyyttä. Heikkojen kaarien poisto sen sijaan katkoo yhteisöjen välisiä linkkejä. Kun riittävä määrä tällaisia ”siltoja” poistetaan, jäljellä on yhtenäisen verkoston sijaan vain joukko toisistaan eristettyjä saarekkeita.



Kuva 2. a) Vajaa tuhannen solmun näyte verkostosta visualisoituna. Näytteen reunoilla on havaittavissa ”yhteisöjä”, joiden sisällä on tiheälti kaaria. b) Sama näyte, kun 80% kaarista on poistettu järjestyksessä vahvimmasta heikompaan. Verkostossa on edelleen suuri yhtenäinen komponentti. c) Näyte, kun 80% kaarista on poistettu järjestyksessä heikoimmasta vahvimpaan. Verkosto on selkeästi fragmentoitunut.

Yllä kuvatulla rakenteella on vaikutuksensa informaation diffuusioon ja leviämiseen verkostossa. Tätä voidaan tarkastella yksinkertaisella simulaatiolla, jossa satunnainen solmu i ”tartutetaan” informaatiolla hetkellä $t = 0$, joka leviää sen kuhunkin naapuriin j todennäköisyydellä p_{ij}. Tämän todennäköisyyden voidaan olettaa riippuvan vuorovaikutuksen voimakkuudesta, jolloin $p_{ij}^M = w_{ij}$. Vertailukohtana olkoon $p_{ij}^R = c$, jolloin todennäköisyys kaikille kaarille sama vakio c. Osoittautuu, että informaatio leviää vuorovaikutusten voimakkuudet huomioivassa mallissa ($p_{ij} = p_{ij}^M = w_{ij}$) huomattavasti hitaammin kuin referenssissä ($p_{ij} = p_{ij}^R = c$). Syynä tähän on, että kyseisessä mallissa informaatio ”jää loukkuun” tiiviisiin sosiaalisiin yhteisöihin, koska sen leviäminen heikkojen linkkien yli on vaikeaa. Tällä lienee vastineensa myös todellisuudessa; sosiaalinen verkostomme näyttäisikin tämän perusteella soveltuvan paremmin paikalliseen tiedon prosessointiin kuin mahdollisimman tehokkaaseen pitkän kantaman tiedonsiirtoon.

Entä millaiset kaaret ovat tässä yksinkertaistetussa mallissa tärkeimpiä informaation kulun kannalta? Jos $p = p_{ij}^R = c$, solmut saavat informaation ensimmäistä kertaa pääasiassa heikkojen kaarien kautta, mutta kun $p = p_{ij}^M = w_{ij}$, ovat keskivahvat kaaret tärkeimmässä asemassa. Vahvat kaaret siirtävät informaatiota tehokkaasti, mutta koska ne sijaitsevat tiiviiden yhteisöjen sisällä, niitä pitkin kulkee harvoin *uutta* informaatiota. Toisaalta heikot linkit yhteisöjen välillä ovat hyvin potentiaalisia uuden informaation välityskanavia, mutta tämän kääntöpuolena on matala informaation siirtokyky. Keskivahvat linkit osuvat välimaastoon: niillä on sopiva sijainti ja ne myös siirtävät tietoa riittävän usein.

Edellä esitetty on vain pintaraapaisu sosiaalisten verkostojen erittäin rikkaaseen rakenteeseen ja dynamiikkaan. Käyttämässämme ”ensimmäisessä approksimaatiossa” verkostoa on käsitelty staattisena, kun todellisuudessa se kuitenkin elää ja muuttuu jatkuvasti. Tämän aikakehityksen tutkiminen on varsin mielenkiintoinen ongelma. Lisäksi empiirisen tutkimuksen tulosten pohjalta kehitetään synteettisten sosiaalisten verkostojen malleja [14]. Nämä mahdollistavat erilaisten sosiodynaamisten ilmiöiden simuloimisen; esimerkiksi ryhmäpaineen vaikutusta mielipiteisiin voidaan lähestyä verkottuneen Ising-mallin johdannaisten kautta. Fyysikoille riittääkin tekemistä tämäntapaisten ongelmien parissa vielä pitkälle tulevaisuuteen.

Viitteet

1. Philip Ball: *Critical mass – how one thing leads to another*. Heinemann, UK, 2004.
2. R. Albert ja A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002); S.N. Dorogovtsev ja J.F.F. Mendes, *Adv. Phys.* **51**, 1079 (2002); M.E.J. Newman, *SIAM Review* **45**,167 (2003); S. Boccaletti *et al.*, *Physics Reports* **424**, 175 (2006).
3. P. Erdõs ja A. Rényi, *Publ. Math. Debrecen* **6**, 290 (1959).
4. D.J. Watts ja S. Strogatz, *Nature* **393**, 440 (1998).
5. A.-L. Barabási ja R. Albert, *Science* **286**, 509 (1999).
6. R. Albert, H. Jeong ja A.-L. Barabási, *Nature* **406**, 378 (2000.)
7. R. Pastor-Satorras ja A. Vespignani, *Phys. Rev. Lett.* **86**, 3200, (2001).
8. S. Wasserman ja K. Faust, Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences), Cambridge University Press, 1994.
9. H. C. White, S. A. Boorman, R. R. Breiger, *American Journal of Sociology* **81**, 730 (1976).
10. M. Granovetter in: Decision Making: Alternatives to Rational Choice Models (ed. M. Zey), 304–333. SAGE Publications, 1992.
11. Preprint: http://arxiv.org/abs/physics/0610104 (2006)
12. Kts. esim *Science* News Focus 10.11. 2006, Social science: Tracking People’s Electronic Footprints ja Physics World Newsflash: http://physicsweb.org/articles/news/10/10/18
13. M. Granovetter, The Strength of Weak Ties, *The American Journal of Sociology* **78**, 1360–1380 (1973).
14. R. Toivonen *et al.*, Physica A **71**, 851 (2006).