

Wattsin ja Strogatzin satunnaisverkkomallin
klusteroituneisuus

Henrik Meyer

Matematiikan pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
Kevät 2015

Tiivistelmä: Henrik Meyer, *Wattsin ja Strogatzin satunnaisverkkomallin klusterointineisuus*, (engl. *Clustering of the Watts-Strogatz random graph model*), matematiikan pro gradu -tutkielma, 82 s., Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, kevät 2015.

Duncan J. Watts ja Steven H. Strogatz [10] määrittivät vuonna 1998 verkon *klusterointikertoimen*, joka on yleisesti käytetty tilastollinen mitta verkon klusteroituneisuuden analysointiin. Esimerkiksi sosiaalisen verkoston henkilöiden tuttavapiireillä on taipumus klusteroitua siten, että monet henkilön tuttavista ovat myös keskenään tuttavina. Verkon klusterointikerroin laskee keskiarvoisen murtoluvun henkilön tuttavapareista, jotka ovat myös keskenään tuttavina. Alain Barrat ja Martin Weigt [11] sekä Mark E. J. Newman, Duncan J. Watts ja Steven H. Strogatz [13] määrittävät verkon *transitiivisuuskertoimen* vaihtoehtoisena klusterointikertoimena. Sosiaalisessa verkostossa se laskee millä todennäköisyydellä kaksi henkilöä, jotka ovat saman henkilön tuttavina ovat myös keskenään tuttavina.

Satunnaisverkkomalleja on pitkään käytetty isojen monimutkaisten verkostojen analysointiin. Klusterointikertoimen käyttöönotosta alkaen on ollut mahdollista osoittaa tilastollisesti, että monet sosiaaliset verkostot sekä ihmisten tekemät verkostot ovat *pieni maailma -verkkoja*. Esimerkiksi Facebook ja Internet ovat pieni maailma -verkkoja, eli harvoja ja hyvin klusteroituneita verkkoja, joissa verkon solmuparien keskimääräinen etäisyys on lyhyt. On osoittautunut, että pieni maailma -verkko-rakennetta ei ole mahdollista mallintaa perinteisillä satunnaisverkoilla tai säännöllisillä hiloilla. D. J. Watts ja S. H. Strogatz [10] esittivät vuonna 1998 uuden satunnaisverkkomallin, jolla on mahdollista generoida pieni maailma -verkkoja. Mallillaan he tutkivat muun muassa epidemian leviämistä sekä signaalin etenemisnopeutta pieni maailma -verkoissa.

Tämän tutkielman tarkoitus on laskea matemaattiset lausekkeet Wattsin ja Strogatzin verkkomallin satunnaisverkon klusterointikertoimelle ja transitiivisuuskertoimelle. Tutkielmassa määritellään matemaattisesti verkon tilastolliset klusterointi- ja transitiivisuuskertoimien eri muodot sekä esitetään lauseita niiden yhteyksistä ja eroista. Lisäksi määritellään *satunnaisverkon* klusterointi- ja transitiivisuuserroin matemaattisesti tarkkaan sekä esitetään kaavoja näiden laskemiseen. Wattsin ja Strogatzin verkkomallin satunnaisverkolle lasketaan solmuparien kytkentätodennäköisyydet, joiden avulla lasketaan matemaattiset lausekkeet verkkomallin satunnaisverkon klusterointi- ja transitiivisuuskertoimelle.

Sisältö

Johdanto	1
Luku 1. Satunnaisverkkomalli	3
1.1. Yleinen verkko ja satunnaisverkko	3
1.2. Harva satunnaisverkkomalli	6
1.3. Verkon solmujen asteet	7
1.4. Verkon karakteristisen polun pituus	9
Luku 2. Yleisen verkon klusteroituneisuus	11
2.1. Kytetty kolmikko ja kolmio	11
2.2. Verkon klusterointikerroin	14
2.3. Verkon transitiivisuuserroin	15
2.4. Klusterointi- ja transitiivisuuskertoimen yhteys	16
Luku 3. Satunnaisverkon klusteroituneisuus	18
3.1. Odotusarvo, varianssi ja ehdollinen odotusarvo	19
3.2. Kytetty kolmikko ja kolmio satunnaisverkossa	24
3.3. Satunnaisverkon klusterointikerroin	28
3.4. Satunnaisverkon transitiivisuuserroin	31
3.5. Klusterointi- ja transitiivisuuskertoimen yhteys satunnaisverkossa	32
Luku 4. Empiirisiä tutkimuksia sosiaalisista verkostoista	35
4.1. Poolin ja Kochenin tutkimus ja verkkomalli	35
4.2. Milgramin empiirinen koe	37
4.3. Wattsin ja Strogatzin tilastollinen tutkimus	38
4.4. Facebookin sosiaalisen verkoston tutkimus	39
Luku 5. Wattsin ja Strogatzin pieni maailma -verkkomalli	40
5.1. Pieni maailma -verkkomallin kytkentämenetelmä	40
5.2. Pieni maailma -verkkomallin satunnaisverkko	42
5.3. Pieni maailma -verkkomallin kytkentätodennäköisyydet	44
5.3.1. Kaukaisten solmujen kytkentätodennäköisyys	44
5.3.2. Läheisten solmujen kytkentätodennäköisyys.	46
5.4. Solmujen asteiden jakauma pieni maailma -verkkomallissa	49
5.5. Karakteristisen polun pituus pieni maailma -verkkomallissa	51
Luku 6. Pieni maailma -verkkomallin klusteroituneisuus	52
6.1. Pieni maailma -verkkomallin satunnaisverkon transitiivisuuserroin	52
6.1.1. Solmun kytkettyjen kolmikoiden odotusarvo	52
6.1.2. Solmun kolmioiden odotusarvo	59
6.1.3. Satunnaisverkon transitiivisuuserroin	64

6.2. Pieni maailma -verkkomallin satunnaisverkon klusterointikerroin	65
Luku 7. Johtopäätös	69
Liite A. Pieni maailma -verkkomallille tehty approksimaatio	70
Liite B. Linkkien ehdolliset todennäköisyydet pieni maailma -verkkomallissa	72
Liite C. Summia	77
Liite D. Käytetty koodi	79
Liite E. Merkintöjä	80
Kirjallisuutta	82

Johdanto

Isojen monimutkaisten verkostojen tutkimukset ja niiden mallintaminen satunnaisverkkomalleilla voidaan karkeasti jakaa seuraaviin ryhmiin [12]:

- Empiiriset mittaukset verkostojen rakenteellisista ominaisuuksista, kuten solmujen asteet, karakteristisen polun pituus, klusteroituneisuus jne.
- Ehdotukset uusista satunnaisverkkomalleista, jotka pyrkivät kuvaamaan mitatut ominaisuudet.
- Uusien satunnaisverkkomallien ominaisuuksien mittaaminen tietokonesimuloinnin avulla.
- Uusien satunnaisverkkomallien heuristiset analyysit, jotka ennustavat niiden ominaisuudet.
- Uusien satunnaisverkkomallien tarkka matemaattinen tutkimus, jolla todistetaan lauseita mallin satunnaisverkon ominaisuuksista.

Tässä tutkielmassa esitetään lyhyesti neljän ensimmäisen kohdan tärkeimmät tulokset Wattsin ja Strogatzin satunnaisverkkomalliin liittyen ja keskitytään pääasiallisesti laskemaan mallin satunnaisverkolle matemaattisia tuloksia.

Valtiotieteilijä Ithiel de Sola Pool ja matemaatikko Manfred Kochen [8] yrittivät 1950-luvulla empiirisesti estimoida Yhdysvaltain väestön henkilöiden tuttavien lukumäärää ja henkilöiden tuttavapiirin klusteroituneisuutta kyselylomakkeiden avulla. 1960-luvulla yhteiskuntapsykologi Stanley Milgram [9] teki empiirisen kokeen kirjeiden avulla, joilla hän yritti estimoida lyhimmän tuttavaketjun pituutta, jolla keskimäärin voidaan yhdistää kaksi satunnaisesti valittua henkilöä Yhdysvaltain väestöstä. Tutkimukset olivat työläitä ja otokset hyvin pieniä (alle 50), mutta tulokset viittasivat silti siihen, että sosiaalista verkostoa karakterisoi kolme ominaisuutta:

1. Verkosto on harva, eli henkilön tuttavien keskimääräinen lukumäärä on pieni verrattuna verkoston kokoon.
2. Verkoston kaksi henkilöä voidaan keskimäärin yhdistää hyvin lyhyellä tuttavaketjulla (Milgram arvioi, että tuttavaketjussa oli keskimäärin 5 välikättä).
3. Verkosto on vahvasti klusteroitunut siten, että henkilön tuttavista monet ovat myös keskenään tuttavina.

Milgramin tulos tuli monille yllätyksenä, mutta Pool ja Kochen [8] olivat jo osoittaneet heuristisin perusteluin perinteisellä satunnaisverkkomallilla, että ominaisuuksien 1-2 olemassaolo samanaikaisesti ei ole yllättävää matemaattisesta näkökulmasta. Pool ja Kochen yrittivät seuraavan kahdenkymmenen vuoden ajan keksiä satunnaisverkkomallia, jolla voitaisiin mallintaa (tai laskea) kaikki kolme ominaisuutta samanaikaisesti, mutta joutuivat lopulta luovuttamaan. On osoittautunut, että mikään perinteisistä satunnaisverkkomalleista ei pysty mallintamaan ominaisuuksia 1-3 samanaikaisesti.

Fyysikko Duncan J. Watts ja matemaatikko Steven H. Strogatz [10] esittivät vuonna 1998 satunnaisverkkomallin, jolla on mahdollista generoida *pieni maailma -verkkoja*, eli verkkoja, joilla on samat rakenteelliset ominaisuudet kuin kohdissa 1-3. He määrittivät samalla *verkon klusterointikertoimen* ja *verkon karakteristisen polun pituuden*, joilla voidaan tilastollisesti laskea verkon klusteroituneisuus ja verkon solmuparien keskimääräinen etäisyys. Watts ja Strogatz laskivat kyseiset arvot kolmelle verkostolle, joihin heillä oli tiedossa täydelliset kytkentäkaaviot: Internet Movie Database:n elokuvanäyttelijöiden yhteistyöverkosto, läntisen Yhdysvaltain sähköverkosto ja Caenorhabditis elegans -madon hermoverkosto. Ne osoittautuivat kaikki pieni maailma -verkoiksi.

Yleiselle verkolle on esitetty kaksi tunnettua tilastollista klusterointikerrointa. Ensimmäisen määrittämiseen esitetään ensin verkon *solmun i klusterointikerroin* [10]

$$C_i = \frac{\text{linkkien lukumäärä solmun } i \text{ naapurien välillä}}{\binom{k_i}{2}},$$

missä $\binom{k_i}{2}$ on linkkien lukumäärä, joka voi korkeintaan olla solmun i naapurien välillä, kun solmulla i on k_i naapuria. D. J. Watts ja S. H. Strogatz [10] määrittävät *verkon klusterointikertoimen*

$$C = \frac{1}{n} \sum_{i=1}^n C_i,$$

joka on keskiarvo verkon solmujen klusterointikertoimista. On esitetty eri määritelmiä klusterointikertoimelle C , kun verkossa on solmuja, joilla on vähemmän kuin kaksi naapuria. Fyysikko Mark E. J. Newman [15] määrittää solmun klusterointikertoimen $C_i = 0$, kun solmun naapurien lukumäärä $k_i < 2$. Tietojenkäsittelytieteilijät Thomas Schank ja Dorothea Wagner [16] laskevat solmujen klusterointikertoimien keskiarvon ainoastaan sellaisille solmuille, joilla on vähintään kaksi naapuria. Fyysikot Alain Barrat ja Martin Weigt [11] määrittävät verkolle toisen klusterointikertoimen, josta käytetään nimitystä *verkon transitiivisuuskeroin*

$$C' = \frac{\frac{1}{n} \sum_{i=1}^n \text{linkkien lukumäärä solmun } i \text{ naapurien välillä}}{\frac{1}{n} \sum_{i=1}^n \binom{k_i}{2}}.$$

Tässä tutkielmassa määritellään *yleisen verkon* klusterointi- ja transitiivisuuskertoimien eri muodot matemaattisesti, jolloin saadaan todistettua lauseita niiden yhteyksistä ja eroista. Tutkielmassa määritellään myös *satunnaisverkon* klusterointi- ja transitiivisuuskeroin matemaattisesti tarkasti sekä esitellään kaavoja niiden laskemiseen. Niiden avulla lasketaan matemaattiset lausekkeet Wattsin ja Strogatzin verkkomallin satunnaisverkon klusterointi- ja transitiivisuuskertoimelle.

Satunnaisverkkomalli

1.1. Yleinen verkko ja satunnaisverkko

Tutkittaessa isoa ja monimutkaista verkostoa, kuten ihmisten sosiaalista verkostoa, on usein tarpeellista tehdä *satunnaisverkkomalli*, jolla voidaan analysoida sitä helpommin. Satunnaisverkkomallilla voidaan satunnaisesti generoida *verkkoja*, jotka muodostuvat *solmuista* ja niitä yhdistävistä *linkeistä*. Sosiaalista verkostoa mallinnattaessa solmu edustaa ihmistä ja linkki kahta henkilöä yhdistävää sosiaalista kontaktia. Mallista riippuen sosiaalinen kontakti voi esimerkiksi olla ystävyys-, tuttavuus- tai yhteistyösuhde. Määritellään seuraavaksi *suuntaamaton verkko*.

MÄÄRITELMÄ 1.1. Suuntaamaton verkko [4] G on järjestetty pari $G = (V, E)$, missä E on joukon V järjestämättömien parien $V^{(2)}$ osajoukko. Joukon $V = V(G)$ alkioita kutsutaan *solmuiksi* ja joukon $E = E(G)$ alkioita *linkeiksi*. Verkon *koko* on sen solmujen lukumäärä $|V(G)|$. Merkitään verkon solmuparia lyhyesti $ij = \{i, j\}$, jolloin ij ja ji tarkoittavat samaa solmuparia. Kun $ij \in E(G)$ sanotaan, että linkki ij kytkee solmut i ja j .

Numeroilla merkittynä n -kokoisen verkon solmujen joukko on $V = V_n = \{1, 2, \dots, n\}$ (ks. kuva 1.1). Merkitään solmujoukolla V_n määriteltyjen verkkojen kokoelmaa

$$(1.1) \quad \mathcal{G}^n = \{(V, E) : V = V_n = \{1, 2, \dots, n\}, E = E_n \subset V_n^{(2)}\},$$

missä $V_n^{(2)}$ on solmujoukon V_n järjestämättömien parien joukko. Verkkojoukko \mathcal{G}^n on äärellinen, sillä jokaisessa äärellisessä solmujoukossa V_n on $|V_n^{(2)}| = \binom{n}{2} = n(n-1)/2$ järjestämättömää solmuparia, joten $|\mathcal{G}^n| = 2^{n(n-1)/2}$ [4]. Tarkastellaan jatkossa ainoastaan solmujoukolla V_n määriteltyjä verkkomalleja.

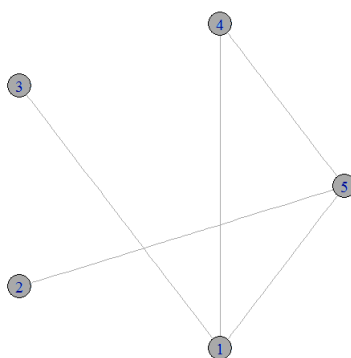
Sosiaalisen verkoston kontaktien muodostumisessa on jonkin verran satunnaisuutta. Mallinnetaan tällaista verkostoa solmujoukolla V_n määritellyllä satunnaisverkkomallilla, missä linkkien kytkentämenetelmässä esiintyy satunnaisuutta. Satunnaisverkkomallin satunnaisesti generoitu verkko on äärellisen verkkojoukon \mathcal{G}^n alkio, joten kyseisen satunnaiskokeen *todennäköisyysavaruus* on *diskreetti*.

MÄÄRITELMÄ 1.2. Joukko Ω on *numeroituva* [4], jos se on äärellinen tai on olemassa bijektiokuvaus $f : \mathbb{N} \rightarrow \Omega$, jolloin se on *numeroituvasti ääretön*.

MÄÄRITELMÄ 1.3. Numeroituvan joukon Ω *todennäköisyysfunktio* tai *diskreetti todennäköisyysjakauma* [4] on kuvaus $P : \Omega \rightarrow \mathbb{R}$, jolle

- (i) $P(\omega) \geq 0$ kaikilla $\omega \in \Omega$,
- (ii) $\sum_{\omega \in \Omega} P(\omega) = 1$.

MÄÄRITELMÄ 1.4. Kokoelma \mathcal{F} perusjoukon Ω osajoukkoja on σ -*algebra* [3], jos



KUVA 1.1. Numeroilla merkitty verkko $G = (V, E) \in \mathcal{G}^5$, jonka solmujen joukko on $V(G) = V_5 = \{1, 2, 3, 4, 5\}$ ja linkkien joukko $E(G) = \{\{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 5\}, \{4, 5\}\}$.

- (i) $\Omega \in \mathcal{F}$,
- (ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$,
- (iii) $A_n \in \mathcal{F}, n \geq 1 \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

MÄÄRITELMÄ 1.5. Parille (Ω, \mathcal{F}) , jossa Ω on joukko ja kokoelma sen osajoukkoja \mathcal{F} on σ -algebra käytetään nimitystä *mitallinen avaruus* [3].

MÄÄRITELMÄ 1.6. Kuvaus $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ on *todennäköisyyssmitta* [3], jos

- (i) $\mathbb{P}(A) \geq 0$, kaikilla $A \in \mathcal{F}$,
- (ii) $\mathbb{P}(\Omega) = 1$,
- (iii) Olkoot joukot $\{A_n, n \geq 1\}$ erillisiä joukkoja $A_n \in \mathcal{F}$. Tällöin

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

MÄÄRITELMÄ 1.7. Kolmikko $(\Omega, \mathcal{F}, \mathbb{P})$ on *diskreetti todennäköisyysavaruus* [3] [4], jos

- (i) Ω on alkeistapausten muodostama numeroituva perusjoukko (otosavaruus),
- (ii) Kokoelma \mathcal{F} perusjoukon Ω osajoukkoja (tapahtumia) on σ -algebra,
- (iii) $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ on todennäköisyyssmitta, joka on määritelty kaavalla

$$\mathbb{P}(A) = \sum_{\omega \in A} P(\omega), \text{ kaikilla } A \subset \mathcal{F},$$

missä funktio $P : \Omega \rightarrow \mathbb{R}$ on joukon Ω todennäköisyysfunktio.

MÄÄRITELMÄ 1.8. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus. Kokoelma joukkoja $\{H_i : 1 \leq i \leq n\}$ on perusjoukon Ω *ositus* [3], jos

$$\Omega = \bigcup_{i=1}^n H_i, \text{ missä } H_i \cap H_j = \emptyset, \text{ kun } 1 \leq i, j \leq n, i \neq j.$$

MÄÄRITELMÄ 1.9. Olkoot (Ω, \mathcal{F}) ja (S, \mathcal{S}) mitallisia avaruuksia. S -arvoinen *sattunnaismuuttuja* [3] on kuvaus

$$X : \Omega \rightarrow S,$$

jolle pätee

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}, \text{ kaikilla } A \in \mathcal{S}.$$

MÄÄRITELMÄ 1.10. Diskreetillä todennäköisyysavaruudella $(\Omega, \mathcal{F}, \mathbb{P})$ määritellyn satunnaismuuttujan $X : \Omega \rightarrow S$ *jakauma* [4] on kuvaus $P_X : S \rightarrow \mathbb{R}$, joka määritellään kaavalla

$$P_X(s) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = s\}) = \sum_{\omega: X(\omega)=s} P(\omega), \text{ kaikilla } s \in S,$$

missä \mathbb{P} on todennäköisyysfunktion P määräämä todennäköisyysmitta.

Solmujoukolla V_n määritelty *satunnaisverkkomalli* saadaan nyt määriteltyä diskreettisenä todennäköisyysavaruutena.

MÄÄRITELMÄ 1.11. *Satunnaisverkkomalli* [6] on diskreetti todennäköisyysavaruus $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, missä

- (i) $\Omega_n \subseteq \mathcal{G}^n$ on jokin solmujoukolla V_n määriteltyjen verkkojen perusjoukko,
- (ii) kokoelma \mathcal{F}_n perusjoukon Ω_n osajoukkoja on σ -algebra,
- (iii) $\mathbb{P}_n : \mathcal{F}_n \rightarrow [0, 1]$ on todennäköisyysmitta, joka on määritelty kaavalla

$$\mathbb{P}_n(A) = \sum_{\omega \in A} P_n(\omega), \text{ kaikilla } A \subset \mathcal{F}_n,$$

missä funktio $P_n : \Omega_n \rightarrow \mathbb{R}$ on joukon Ω_n diskreetti todennäköisyysjakauma.

Merkitään satunnaisverkkomallia lyhyesti $\Omega_n = (\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)$, missä jatkossa valitaan todennäköisyysavaruuden $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ σ -algebraksi *potenssijoukko* [2] $\mathcal{F}_n = \mathcal{P}(\Omega_n)$. Määritellään seuraavaksi satunnaisverkkomallin Ω_n *satunnaisverkko* G_n ja sen jakauma.

MÄÄRITELMÄ 1.12. Satunnaisverkkomallilla $(\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)$ määritelty *satunnaisverkko* [4] on \mathcal{G}^n -arvoinen satunnaismuuttuja

$$G_n : \Omega_n \rightarrow \mathcal{G}^n.$$

Satunnaisverkon G_n *jakauma* [4] on verkkojoukon \mathcal{G}^n todennäköisyysfunktio

$$P_{G_n}(g) = \mathbb{P}(G_n = g), \text{ kaikilla } g \in \mathcal{G}^n.$$

Käytetään jatkossa satunnaisverkkomallin Ω_n satunnaisverkkona identtistä kuvausta $G_n : \Omega_n \rightarrow \Omega_n$,

$$G_n(\omega) = \omega \text{ kaikilla } \omega \in \Omega_n.$$

Tämä määrittää satunnaisverkkomallia Ω_n yksikäsitteisesti, sillä sen jakauma P_{G_n} on sama kuin satunnaisverkkomallin Ω_n diskreetti todennäköisyysjakauma P_n ,

$$P_{G_n}(\omega) = \mathbb{P}(G_n = \omega) = \mathbb{P}_n(\{\omega\}) = P_n(\omega) \text{ kaikilla } \omega \in \Omega_n,$$

missä \mathbb{P}_n on todennäköisyysfunktion P_n määräämä todennäköisyysmitta.

ESIMERKKI 1.13. Satunnaisverkkomallin $\mathcal{G}_{n, M(n)} = (\mathcal{G}_{n, M(n)}, \mathcal{P}(\mathcal{G}_{n, M(n)}), \mathbb{P}_{n, M(n)})$ [6] perusjoukko $\mathcal{G}_{n, M(n)} = \{(V_n, E_n) : V_n = \{1, 2, \dots, n\}, |E_n| = M(n)\} \subset \mathcal{G}^n$ koostuu kaikista solmujoukolla V_n määritetyistä verkoista, joiden linkkien lukumäärä parametrin n funktiona on $M(n)$. Satunnaisverkkomallissa $\mathcal{G}_{n, M(n)}$ jokaisen perusjoukon

verkolla on sama todennäköisyys, jolloin sanotaan, että satunnaisverkkomallin satunnaisverkko $G_{n,M(n)}$ on *tasajakautunut*,

$$P_{G_{n,M(n)}}(g) = \mathbb{P}(G_{n,M(n)} = g) = \frac{1}{|\mathcal{G}_{n,M(n)}|} = \binom{\binom{n}{2}}{M(n)}^{-1} \text{ kaikilla } g \in \mathcal{G}_{n,M(n)},$$

missä $\binom{n}{2} = n(n-1)/2$ on mahdollisten linkkien lukumäärä n -kokoisessa verkossa.

1.2. Harva satunnaisverkkomalli

Isojen sosiaalisten verkostojen rakenteita karakterisoi, että kontakteja on suhteellisen vähän. Verkon linkkien suhteellista lukumäärää kutsutaan verkon *tiheydeksi*.

MÄÄRITELMÄ 1.14. Yleisen verkon $G = (V, E)$ *tiheys* [20] on

$$D(G) = \frac{|E|}{\binom{n}{2}} = \frac{2|E|}{n(n-1)},$$

missä $\binom{n}{2} = n(n-1)/2$ verkon mahdollisten linkkien maksimimäärä, kun verkon koko on $|V| = n$. Satunnaisverkon G_n tiheys määritellään samalla tavalla, missä linkkien lukumäärä $|E(G_n)|$ on satunnaismuuttuja.

Erityisesti sosiaalisten verkostojen kasvaessa kontaktien lukumäärä kasvaa vain hitaasti. Sanotaan, että satunnaisverkkomalli Ω_n on *harva*, jos parametrin n kasvaessa satunnaisverkon G_n linkkien lukumäärä $|E(G_n)|$ on *alempaa kertaluokkaa* kuin n^2 . Ensin tulee määritellä *pieni o-estimaatti*.

MÄÄRITELMÄ 1.15. Olkoot $f(n)$ ja $g(n)$ muuttujan $n \in \mathbb{N}$ funktioita. Funktio $f(n)$ on *alempaa kertaluokkaa* kuin funktio $g(n)$ [1], merkitään $f(n) = o(g(n))$, jos funktio $g(n) \neq 0$ isoilla n ja

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0.$$

LAUSE 1.16. *Olkoot $f(n)$ ja $g(n)$ muuttujan $n \in \mathbb{N}$ funktioita, joille pätee $f(n) = o(g(n))$. Olkoon lisäksi $b > 0$ positiivinen vakio. Tällöin $f(n) = o(bg(n))$.*

TODISTUS. Olkoon $b > 0$ vakio ja $f(n) = o(g(n))$, jolloin $f(n)/g(n) \rightarrow 0$, kun $n \rightarrow \infty$. Väitteenä on, että myös $f(n)/(bg(n)) \rightarrow 0$, kun $n \rightarrow \infty$. Tehdään antiteesi. Oletetaan, että $f(n)/(bg(n)) \rightarrow a \neq 0$, kun $n \rightarrow \infty$ tai $f(n)/(bg(n))$ ei suppene. Jos $f(n)/(bg(n)) = 1/b \cdot f(n)/g(n) \rightarrow a \neq 0$, kun $n \rightarrow \infty$, niin $f(n)/g(n) \rightarrow b \cdot a \neq 0$, kun $n \rightarrow \infty$. Tämä on ristiriidassa oletuksen $f(n) = o(g(n))$ kanssa. Jos $f(n)/(bg(n))$ hajaantuu, niin $f(n)/g(n)$ selvästi myös hajaantuu, kun $b > 0$ on vakio. Tämä on myös ristiriidassa oletuksen $f(n) = o(g(n))$, joten väite $f(n) = o(bg(n))$ on tosi. \square

MÄÄRITELMÄ 1.17. Olkoon (Ω_n) jono satunnaisverkkomalleja ja (G_n) jono sitä vastaavia satunnaisverkkoja. Sanotaan, että satunnaisverkkomalli Ω_n on *harva* [10], jos $|E(G_n)| = o(n^2)$, eli jos

$$(1.2) \quad \frac{|E(G_n)|}{n^2} \rightarrow 0, \text{ kun } n \rightarrow \infty.$$

LAUSE 1.18. *Olkoon Ω_n harva satunnaisverkkomalli. Tällöin satunnaisverkon G_n tiheydelle pätee*

$$D(G_n) \rightarrow 0, \text{ kun } n \rightarrow \infty.$$

TODISTUS. Määritelmän 1.14 mukaan satunnaisverkon G_n tiheys on $D(G_n) = |E(G_n)|/\binom{n}{2}$, missä linkkien lukumäärä $|E(G_n)|$ on satunnaismuuttuja. Nimittäjälle $\binom{n}{2}$ pätee $\binom{n}{2} = \frac{1}{2}n(n-1) > \frac{1}{4}n^2 \Leftrightarrow n-1 > n/2$, kun $n > 2$. Oletuksen mukaan harvalle satunnaisverkkomallille Ω_n pätee $|E(G_n)| = o(n^2)$. Lauseesta 1.16 seuraa, että $|E(G_n)| = o(bn^2)$ kaikilla vakiolla $b > 0$, jolloin satunnaisverkon G_n tiheydelle pätee $0 \leq D(G_n) = |E(G_n)|/\binom{n}{2} < |E(G_n)|/(\frac{1}{4}n^2) \rightarrow 0$, kun $n \rightarrow \infty$. \square

Joissain satunnaisverkkomalleissa kaavan (1.2) varma suppeneminen voi olla liian vahva käsite. Määritellään seuraavaksi heikompia suppenemiskäsitteitä.

MÄÄRITELMÄ 1.19. Satunnaismuuttujajono (X_n) *suppenee melkein varmasti* (m.v) [3] kohti satunnaismuuttujaa X , jos

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \rightarrow X(\omega) \text{ kun } n \rightarrow \infty\}) = 1,$$

jolloin merkitään $X_n \xrightarrow{m.v.} X$.

MÄÄRITELMÄ 1.20. Satunnaismuuttujajono (X_n) *suppenee stokastisesti* [3] kohti satunnaismuuttujaa X , jos kaikilla $\epsilon > 0$ pätee

$$(1.3) \quad \mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0, \text{ kun } n \rightarrow \infty,$$

jolloin merkitään $X_n \xrightarrow{P} X$.

Sanotaan, että satunnaisverkkomalli Ω_n on (stokastisesti) harva, jos satunnaisverkkomallijonon (Ω_n) vastaavalle satunnaisverkkojonolle (G_n) pätee

$$(1.4) \quad \frac{|E(G_n)|}{n^2} \xrightarrow{P} 0, \text{ kun } n \rightarrow \infty.$$

Merkitään tällöin matemaatikkojen Béla Bollobás ja Oliver Riordan [7] tapaan $|E(G_n)| = o_P(n^2)$, missä *pieni o_P -estimaatti* tarkoittaa stokastista suppenemistä.

ESIMERKKI 1.21. Olkoon $\mathcal{G}_{n,p(n)}$ satunnaisverkkomalli [6], jonka perusjoukko koostuu kaikista solmujoukolla V_n määritetyistä verkoista, joissa solmuparit on riippumattomasti kytketty kytkentätodennäköisyydellä $0 < p(n) < 1$. Asetetaan parametri $p(n) = b/n$, vakiolla $b > 0$. Satunnaisverkon $G_{n,p(n)}$ linkkien lukumäärä on tällöin satunnaismuuttuja $|E(G_{n,p(n)})| \sim \text{Bin}(n(n-1)/2, b/n)$, jolloin satunnaismuuttujajono $|E(G_{n,p(n)})|/n^2$ ei suppene kohti nollaa varmasti. *Markovin epäyhtälön* [2] mukaan kaikilla $\epsilon > 0$ pätee

$$\mathbb{P}(|E(G_{n,p(n)})|/n^2 \geq \epsilon) \leq \frac{\mathbb{E}[|E(G_{n,p(n)})|/n^2]}{\epsilon} = \frac{b(n-1)}{2\epsilon n^2} \rightarrow 0, \text{ kun } n \rightarrow \infty,$$

joten määritelmästä 1.20 nähdään, että $|E(G_{n,p(n)})|/n^2 \xrightarrow{P} 0$, kun $n \rightarrow \infty$.

1.3. Verkon solmujen asteet

Sosiaalisen verkoston tiheys voidaan arvioida empiirisesti tutkimalla kuinka monta sosiaalista kontaktia verkoston henkilöillä on. Jos jokaisen henkilön kontaktien lukumäärä on tiedossa, voidaan tilastollisesti laskea verkoston tiheys henkilöiden kontaktien keskimääräisen lukumäärän avulla.

MÄÄRITELMÄ 1.22. Yleisen verkon $G = (V, E)$ solmun $i \in V$ *naapurusto* [4] on joukko

$$N_i = \{j \in V : ij \in E\},$$

jonka alkioita kutsutaan solmun i *naapureiksi*.

Solmun i naapuruston alkioden lukumäärää kutsutaan solmun *asteeksi*.

MÄÄRITELMÄ 1.23. Yleisen verkon G solmun i *aste* [5] on

$$(1.5) \quad k_i = |N_i|,$$

missä $|N_i|$ on solmun i naapuruston alkioden lukumäärä. Verkon solmujen *asteiden keskiarvo* on n -kokoisessa verkossa

$$(1.6) \quad \bar{k} = \bar{k}(G) = \frac{1}{n} \sum_{i=1}^n k_i.$$

MÄÄRITELMÄ 1.24. Solmu i on *eristetty*, jos $k_i = 0$, ja solmu i on *loppusolmu*, jos $k_i = 1$.

Matemaatikko Leonhard Euler esitti lauseen, joka usein helpottaa yleisten verkkojen analysointia.

LAUSE 1.25 ("Handshaking theorem" [5]). *Yleiselle verkolle* $G = (V, E)$ *pätee*

$$(1.7) \quad \sum_{i=1}^n k_i = 2|E|,$$

jossa n on verkon koko ja $|E|$ linkkien lukumäärä verkossa.

TODISTUS. Yleisessä verkossa solmun aste k_i kertoo kuinka monen linkin päätesolmu solmu i on. Yleisessä verkossa jokaisella linkillä on kaksi päätesolmua, joten summaamalla verkon solmujen asteet tulee jokainen linkki laskettua kahdesti. Verkossa G on $|E|$ linkkiä, jolloin solmujen asteiden summa on $\sum_{i=1}^n k_i = 2|E|$. \square

Kun verkon solmujen joukon V ja linkkien joukon E alkioden lukumäärät ovat tiedossa, saadaan laskettua verkon solmujen asteiden keskiarvo \bar{k} ilman, että tiedetään verkon yksittäisten solmujen asteista mitään.

SEURAUS 1.26. *Yleisen verkon* $G = (V, E)$ *solmujen asteiden keskiarvo* \bar{k} *on*

$$\bar{k} = \frac{2|E|}{|V|}.$$

TODISTUS. Verkolle, jonka koko on n pätee $|V| = n$, jolloin kaavoista (1.6) ja (1.7) saadaan laskettua $\bar{k} = \frac{1}{n} \sum_{i=1}^n k_i = 2|E|/|V|$. \square

Vastaavasti, jos verkon koko n ja solmujen asteiden keskiarvo \bar{k} on tiedossa, saadaan helposti laskettua linkkien lukumäärä verkossa.

SEURAUS 1.27. *Olkoon* $G = (V, E)$ *yleinen verkko, jonka koko on n ja solmujen asteiden keskiarvo on \bar{k} . Tällöin verkon linkkien lukumäärä on*

$$|E| = \frac{\bar{k}n}{2}.$$

TODISTUS. Verkon G solmujen joukon V alkioden lukumäärä on $|V| = n$, jolloin tulos saadaan suoraan seurauksesta 1.26. \square

Kun n -kokoisen verkon solmujen asteiden keskiarvo \bar{k} on tiedossa, voidaan ilmaista verkon tiheys myös verkon solmujen asteiden keskiarvon \bar{k} avulla.

SEURAUUS 1.28. Yleisen verkon $G = (V, E)$ tiheys on

$$D = \frac{\bar{k}}{n-1},$$

missä \bar{k} on verkon solmujen asteiden keskiarvo ja n verkon koko.

TODISTUS. Määritelmän (1.14) mukaan $D = 2|E|/(n(n-1))$. Seurauksesta 1.27 saadaan $2|E| = \bar{k}n$, joten $D = \bar{k}/(n-1)$. \square

1.4. Verkon karakteristisen polun pituus

Sosiaalisen verkoston rakenteellinen ominaisuus, joka monesti on vaikeampi mitata empiirisesti, on lyhimmän kontaktiketjun pituus, jota keskimäärin tarvitaan yhdistämään verkon kahta henkilöä. Kahta henkilöä yhdistävää kontaktiketjua kutsutaan verkkoteoriassa kahta solmua yhdistäväksi *poluksi*.

MÄÄRITELMÄ 1.29. Yleisen verkon $G = (V, E)$ polku [5] solmusta u solmuun v on jono

$$u = v_0, e_1, v_1, e_2, v_2, \dots, e_m, v_m = v,$$

jossa on vuorotellen verkon solmuja $v_i \in V$ ja verkon linkkejä $e_i = \{v_{i-1}, v_i\} \in E$. Verkon polku on *suora*, jos $v_i \neq v_j$, kun $i \neq j$, ja polun pituus on jonon linkkien lukumäärä m .

Kahta henkilöä yhdistäviä kontaktiketjuja voi olla useita, mutta sosiaalisten verkostojen analysoinnissa kiinnostaa erityisesti lyhin näistä. Verkkomallissa tämä vastaa kahden tietyn solmun lyhintä yhdistävää polkua, jonka pituutta kutsutaan solmujen *etäisyydeksi*.

MÄÄRITELMÄ 1.30. Yleisen verkon $G = (V, E)$ solmujen $i, j \in V$ välinen *etäisyys* [15], d_{ij} , on lyhin polun pituus, joka yhdistää solmuja i ja j . Jos solmuja i ja j ei yhdistä mikään polku, määritellään etäisyys äärettömäksi, $d_{ij} = \infty$.

Eräs globaali mitta, joka karakterisoi sosiaalisen verkoston rakennetta on lyhimmän kontaktiketjun pituus, jota keskimäärin tarvitaan yhdistämään verkoston kahta henkilöä. Tämä kertoo verkon solmuparien keskimääräisen etäisyyden, jota D. J. Watts ja S. H. Strogatz [10] kutsuvat verkon *karakteristisen polun pituudeksi*.

MÄÄRITELMÄ 1.31. Yleisen verkon $G = (V, E)$ karakteristisen polun pituus [10] on

$$(1.8) \quad L = L(G) = \frac{1}{\frac{1}{2}n(n-1)} \sum_{j>i} d_{ij},$$

missä d_{ij} on solmuparin $\{i, j\} \in V_n^{(2)}$ välinen etäisyys ja $\binom{n}{2} = n(n-1)/2$ on eri solmuparien lukumäärä n -kokoisessa verkossa.

Toinen globaali mitta on verkon *halkaisija*, joka ilmaisee verkon solmuparien isoimman etäisyyden.

MÄÄRITELMÄ 1.32. Yleisen verkon $G = (V, E)$ *halkaisija* [15] on

$$(1.9) \quad \text{diam}(G) = \max\{d_{ij} : \{i, j\} \in V_n^{(2)}\}.$$

LAUSE 1.33. Yleiselle verkolle G pätee

$$(1.10) \quad L(G) \leq \text{diam}(G).$$

TODISTUS. Olkoon $d_{max} = \max\{d_{ij} : \{i, j\} \in V_n^{(2)}\}$, jolloin $d_{ij} \leq d_{max}$ kaikilla $\{i, j\} \in V_n^{(2)}$. Kaavasta (1.8) saadaan

$$L(G) = \frac{1}{\frac{1}{2}n(n-1)} \sum_{j>i} d_{ij} \leq \frac{1}{\frac{1}{2}n(n-1)} \sum_{j>i} d_{max} = \frac{\frac{1}{2}n(n-1)d_{max}}{\frac{1}{2}n(n-1)} = d_{max} = \text{diam}(G).$$

□

Verkon karakteristisen polun pituus $L(G)$ ja halkaisija $\text{diam}(G)$ ovat äärettömiä, jos verkossa on solmupari, jota ei yhdistä mikään polku. Seuraavaksi tarkastellaan minkä tyyppisissä verkkorakenteissa näin käy, ja kuinka se voidaan ottaa huomioon verkon karakteristisen polun pituuden ja halkaisijan laskemisessa.

MÄÄRITELMÄ 1.34. Yleisen verkon solmupari on *yhdistetty* [5], jos niiden välillä on polku.

MÄÄRITELMÄ 1.35. Yleinen verkko G on *yhtenäinen* [5], jos verkon jokainen solmupari on yhdistetty.

Monesti verkko ei ole yhtenäinen, jolloin se on *epäyhtenäinen*, mutta sillä on olemassa *aliverkkoja*, jotka ovat yhtenäisiä.

MÄÄRITELMÄ 1.36. Olkoot $G = (V, E)$ ja $H = (W, F)$ yleisiä verkkoja, joille $W \subseteq V$ ja $F \subseteq E$. Silloin verkko H on verkon G *aliverkko* [5]. Jos lisäksi $W \subset V$ tai $F \subset E$, niin verkko H on verkon G *aito aliverkko*.

MÄÄRITELMÄ 1.37. Verkon G aliverkkoa H sanotaan verkon G *maksimaaliseksi aliverkoksi* [5] jonkin ominaisuuden B suhteen, jos

- (i) verkolla H on ominaisuus B ja
- (ii) aina, kun H on verkon G aliverkon F aito aliverkko, $H \subset F \subseteq G$, verkolla F ei ole ominaisuutta B .

MÄÄRITELMÄ 1.38. Verkon $G = (V, E)$ *komponentti* [5] on verkon maksimaalinen yhtenäinen aliverkko $H = (W, F)$.

HUOMAUTUS 1.39. (i) Epäyhtenäisen verkon G karakteristisen polun pituus $L(G)$ ja halkaisija $\text{diam}(G)$ ovat äärettömiä. Tästä huolimatta voidaan kuvailla verkon rakennetta kyseisillä mitoilla laskemalla ainoastaan verkon yhdistettyjen solmuparien etäisyydet, eli samassa komponentissa sijaitsevien solmuparien etäisyydet [15]. (ii) Toinen tapa välttää äärettömät etäisyydet epäyhtenäisen verkon G karakteristisen polun pituuden $L(G)$ ja halkaisijan $\text{diam}(G)$ laskuissa on laskea etäisyys ainoastaan solmupareille, jotka sijaitsevat verkon isoimmassa yhtenäisessä komponentissa [10].

LUKU 2

Yleisen verkon klusteroituneisuus

Verkon koon n ja solmujen asteiden keskiarvon \bar{k} avulla saadaan laskettua millä todennäköisyydellä verkon kaksi satunnaisesti valittua solmua on kytketty linkillä. Harvassa verkkomallissa kyseinen todennäköisyys on hyvin pieni, jos verkko on iso.

LAUSE 2.1. *Olkoon $G = (V, E)$ yleinen verkko, jonka koko on n . Tällöin todennäköisyys, että verkon satunnaisesti valittu solmupari $\{i, j\}$ on kytketty linkillä, on*

$$(2.1) \quad \mathbb{P}(ij \in E) = \frac{\bar{k}}{n-1},$$

missä \bar{k} on verkon solmujen asteiden keskiarvo.

TODISTUS. Seurauksen 1.27 mukaan n -kokoisen verkon linkkien lukumäärä on $|E| = \bar{k}n/2$, jossa \bar{k} on solmujen asteiden keskiarvo. Verkossa on yhteensä $\binom{n}{2} = n(n-1)/2$ eri solmuparia, joten todennäköisyys, että satunnaisesti valittu solmupari $\{i, j\}$ on kytketty linkillä, on $\mathbb{P}(ij \in E) = |E|/\binom{n}{2} = (\bar{k}n/2)/(n(n-1)/2) = \bar{k}/(n-1)$. \square

Esimerkiksi sosiaaliset verkostot ovat hyvin klusteroituneita siten, että kaavan 2.1 todennäköisyys kasvaa huomattavasti tiedolla, että valituilla henkilöillä on yhteinen tuttava. Tämä viittaa siihen, että he todennäköisesti liikkuvat samoissa piireissä, jolloin on todennäköistä, että he ovat myös keskenään tuttavina.

2.1. Kytketty kolmikko ja kolmio

Verkossa kaksi solmua ja niiden yhteinen naapuri muodostavat *kytketyn kolmikon*. Kyseiset solmut muodostavat lisäksi *kolmion*, jos ne kaikki ovat keskenään kytkettyjä linkillä (ks. kuva 2.1).

MÄÄRITELMÄ 2.2. *Olkoon $G = (V, E)$ yleinen verkko. Kytketty kolmikko [16] on verkon kolmen eri solmun $i, j, k \in V$ muodostama yhtenäinen aliverkko*

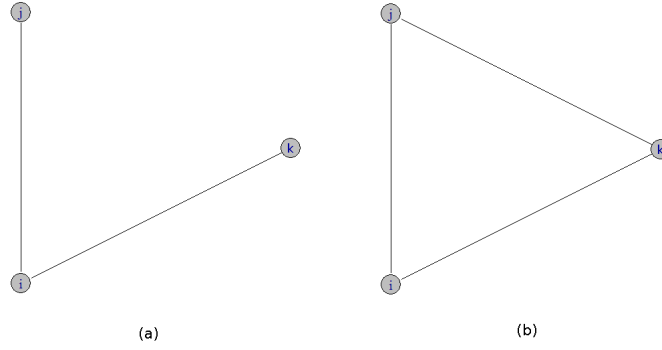
$$\Upsilon = (W_\Upsilon, F_\Upsilon),$$

missä $W_\Upsilon = \{i, j, k\} \subset V$ ja $F_\Upsilon = \{\{i, j\}, \{i, k\}\} \subset E$. Solmua i , joka on joukon F_Υ alkioiden yhteinen päätesolmu $\{i, j\} \cap \{i, k\} = \{i\}$, kutsutaan kytketyn kolmikon keskiisolmuksi. Merkitään verkon G kytkettyjen kolmikkojen joukkoa

$$\Upsilon(G) = \{\Upsilon : \Upsilon \subset G\}$$

ja keskiisolmun $i \in V$ kytkettyjen kolmikkojen joukkoa

$$\Upsilon_i = \{\Upsilon \in \Upsilon(G) : i \in W_\Upsilon \text{ on kytketyn kolmikon } \Upsilon \text{ keskiisolmu}\}.$$



KUVA 2.1. (a): Keskisolmun i kytketty kolmikko $\Upsilon = (W_\Upsilon, F_\Upsilon) \in \Upsilon_i$, missä $W_\Upsilon = \{i, j, k\}$ ja $F_\Upsilon = \{\{i, j\}, \{i, k\}\}$. (b): Keskisolmun i kytketyn kolmikon Υ solmut muodostavat myös kolmion $\nabla = (W_\nabla, F_\nabla)$, missä $W_\nabla = \{i, j, k\}$ ja $F_\nabla = \{\{i, j\}, \{i, k\}, \{j, k\}\}$, jolloin kolmio ∇ kuuluu joukkoihin Δ_i, Δ_j ja Δ_k .

LAUSE 2.3. Olkoon $G = (V, E)$ yleinen verkko, jonka koko on n . Solmun $i \in V$ kytkettyjen kolmikoiden joukon lukumäärä on

$$(2.2) \quad |\Upsilon_i| = \binom{k_i}{2} = \frac{k_i(k_i - 1)}{2},$$

missä k_i on solmun i aste. Kytkettyjen kolmikoiden lukumäärä verkossa G on

$$(2.3) \quad |\Upsilon(G)| = \sum_{i=1}^n |\Upsilon_i|.$$

TODISTUS. Määritelmän 2.2 mukaan keskisolmun i muodostama kytketty kolmikko Υ , jossa $W_\Upsilon = \{i, j, k\}$ ja $F_\Upsilon = \{\{i, j\}, \{i, k\}\}$, muodostuu solmusta i ja sen naapuriparista $\{j, k\} \subset N_i$. Joukon Υ_i eri kytketyt kolmikot muodostuvat solmusta i ja sen eri naapuripareista, joten joukon Υ_i alkuiden lukumäärä on $|\Upsilon_i| = \binom{|N_i|}{2} = \binom{k_i}{2} = k_i(k_i - 1)/2$. Määritelmän 2.2 mukaan verkon kytketyllä kolmikolla Υ , jolle $W_\Upsilon = \{i, j, k\}$ ja $F_\Upsilon = \{\{i, j\}, \{i, k\}\}$, on yksi keskisolmu, sillä $\{i, j\} \cap \{i, k\} = \{i\}$. Joukot $(\Upsilon_i : i \in V)$ ovat siis erillisiä joukkoja, joten kytkettyjen kolmikoiden kokonaismäärä n -kokoisessa verkossa saadaan summasta $|\Upsilon| = \sum_{i=1}^n |\Upsilon_i|$. \square

Määritellään seuraavaksi kolmio, joko on verkon kolmen solmun muodostama *täydellinen aliverkko*, eli *klikki*.

MÄÄRITELMÄ 2.4. Yleinen verkko $G = (V, E)$ on *täydellinen* [5], jos verkon jokainen solmupari on kytketty linkillä.

MÄÄRITELMÄ 2.5. Olkoon G verkko ja H sen aliverkko. Jos H on täydellinen, niin H on verkon G *klikki* [5].

MÄÄRITELMÄ 2.6. Olkoon $G = (V, E)$ yleinen verkko. Kolmio [16] on verkon kolmen solmun $i, j, k \in V$ muodostama klikki

$$\nabla = (W_\nabla, F_\nabla),$$

missä $W_\nabla = \{i, j, k\} \subset V$ ja $F_\nabla = \{\{i, j\}, \{i, k\}, \{j, k\}\} \subset E$. Merkitään verkon G kolmioiden joukkoa

$$\Delta(G) = \{\nabla : \nabla \subset G\}$$

ja solmun i sisältävien kolmioiden joukkoa

$$\Delta_i = \{\nabla \in \Delta(G) : i \in V_\nabla\}.$$

LEMMA 2.7. *Olkoon $G = (V, E)$ yleinen verkko ja $\nabla \in \Delta(G)$ sen kolmio. Tällöin kolmio ∇ kuuluu kolmen eri solmun kolmioiden joukkoon.*

TODISTUS. Olkoon $\nabla = (W_\nabla, F_\nabla) \in \Delta(G)$ kolmio, missä $W_\nabla = \{i, j, k\} \subset V(G)$ ja $F_\nabla = \{\{i, j\}, \{i, k\}, \{j, k\}\} \subset E(G)$. Tällöin määritelmän 2.6 mukaan $\nabla \in \Delta_i$, $\nabla \in \Delta_j$ ja $\nabla \in \Delta_k$, sillä $i, j, k \in V_\nabla$ (ks. kuva 2.1). \square

LAUSE 2.8. *Olkoon $G = (V, E)$ yleinen verkko, jonka koko on n . Tällöin*

$$(2.4) \quad \sum_{i=1}^n |\Delta_i| = 3|\Delta(G)|,$$

missä $|\Delta(G)|$ on kolmioiden lukumäärä verkossa ja $|\Delta_i|$ solmun i sisältävien kolmioiden lukumäärä kaikilla $i \in V$.

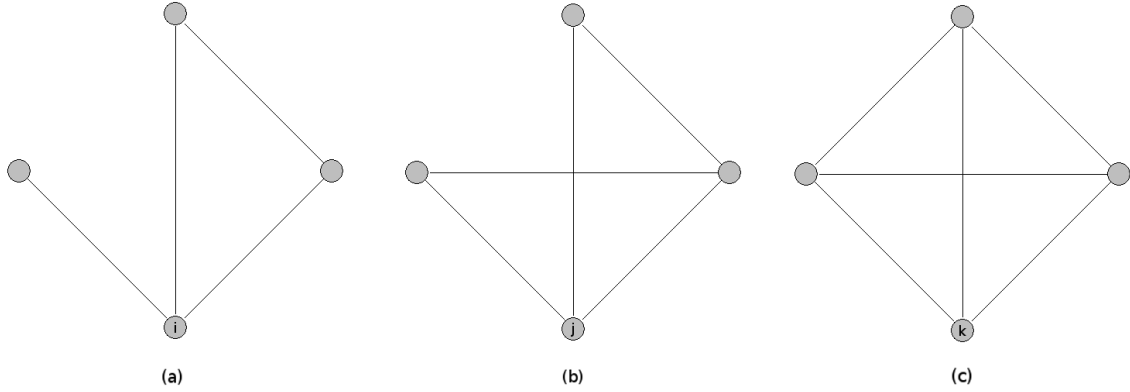
TODISTUS. Verkon G kolmio $\nabla \in \Delta(G)$ sisältyy lemmän 2.7 mukaan kolmen eri solmun kolmioiden joukkoon. Summaamalla jokaisen solmun $i \in V$ sisältävien kolmioiden lukumäärä $|\Delta_i|$ tulee laskettua jokainen verkon kolmio kolme kertaa. \square

MÄÄRITELMÄ 2.9. Olkoon $G = (V, E)$ yleinen verkko ja $F \subseteq E$ sen epätyhjä linkkijoukko. Tällöin $H = (W, F)$, missä $W \subseteq V$ on joukon F linkkien päätesolmujen joukko, on joukon F *linkki-indusoima* [5] verkon G aliverkko.

MÄÄRITELMÄ 2.10. Olkoon $G = (V, E)$ yleinen verkko ja $W \subseteq V$ sen epätyhjä solmujoukko. Tällöin $H = (W, F)$ on joukon W *solmuindusoima* [5] verkon G aliverkko, jos $F \subseteq E$ muodostuu niistä joukon E linkeistä, joiden päätesolmut kuuluvat joukkoon W .

LAUSE 2.11. *Olkoon $G = (V, E)$ yleinen verkko ja $\Delta(G)$ verkon G kolmioiden joukko. Tällöin jokaisella kolmiolla $\nabla \in \Delta(G)$ on olemassa kolme eri linkki-indusoitua kytkettyä kolmikkoa.*

TODISTUS. Olkoon $\nabla \in \Delta(G)$ verkon kolmio. Määritelmän 2.6 mukaan kolmio $\nabla = (V_\nabla, E_\nabla)$ muodostuu verkon kolmesta solmusta $V_\nabla = \{i, j, k\} \subset V(G)$ ja linkkijoukosta $E_\nabla = \{\{i, j\}, \{i, k\}, \{j, k\}\}$. Olkoon $F_i = \{\{i, j\}, \{i, k\}\} \subset E_\nabla$ linkkijoukon E_∇ epätyhjä osajoukko. Joukon F_i linkkien päätesolmujen joukko on V_∇ , joten $H_i = (V_\nabla, F_i)$ on joukon F_i linkki-indusoima kolmion ∇ aliverkko. Määritelmän 2.2 mukaan verkko H_i on kytketty kolmikko, jonka keskisolmu on $\{i\} = \{i, j\} \cap \{i, k\}$, joten $H_i \in \Upsilon_i$. Samoin $H_j = (V_\nabla, F_j)$ ja $H_k = (V_\nabla, F_k)$ muodostavat joukkojen $F_j = \{\{i, j\}, \{j, k\}\} \subset E_\nabla$ ja $F_k = \{\{i, k\}, \{j, k\}\} \subset E_\nabla$ linkki-indusoimat kolmion ∇ aliverkot, jotka määritelmän 2.2 mukaan muodostavat keskisolmujen j ja k kytketyt kolmikot $H_j \in \Upsilon_j$ ja $H_k \in \Upsilon_k$. \square



KUVA 2.2. Kuvissa nähdään solmun i , j ja k naapurustot, joissa kaikissa on kolme naapurisolmua, joten $|\Upsilon_i| = |\Upsilon_j| = |\Upsilon_k| = \binom{3}{2} = 3$. (a): Yksi solmun i kytketyistä kolmikoista muodostaa myös kolmion ∇ , joten solmun i klusterointikerroin on $C_i = |\Delta_i|/|\Upsilon_i| = 1/3$. (b): $C_j = 2/3$ ja (c): $C_k = 3/3 = 1$.

2.2. Verkon klusterointikerroin

Fyysikko Duncan J. Watts ja matemaatikko Steven H. Strogatz [10] määrittävät vuoden 1998 julkaisussaan *solmun klusterointikertoimen*, joka laskee osuuden keski-solmun i kytketyistä kolmikoista, jotka muodostavat myös kolmion (ks. kuva 2.2).

MÄÄRITELMÄ 2.12. Yleisessä verkossa $G = (V, E)$ solmun $i \in V$ klusterointikerroin [10] on

$$(2.5) \quad C_i = \frac{|\Delta_i|}{|\Upsilon_i|} \in [0, 1],$$

kun $|\Upsilon_i| > 0$, missä $|\Delta_i|$ on solmun i sisältävien kolmioiden lukumäärä ja $|\Upsilon_i|$ on keski-solmun i kytkettyjen kolmikoiden joukon lukumäärä. Solmulle i , jonka aste on $k_i \in \{0, 1\}$, määritellään solmun klusterointikerroin $C_i = 0$ [15].

D. J. Watts ja S. H. Strogatz [10] määrittävät myös *verkon klusterointikertoimen*, joka laskee verkon solmujen klusterointikertoimien keskiarvon.

MÄÄRITELMÄ 2.13. Olkoon $G = (V, E)$ yleinen verkko, jonka koko on n . Verkon klusterointikerroin on

$$(2.6) \quad C(G) = \frac{1}{n} \sum_{i=1}^n C_i \in [0, 1],$$

missä C_i on solmun $i \in V$ klusterointikerroin.

Tietojenkäsittelytieteilijät Thomas Schank ja Dorothea Wagner [16] määrittävät vaihtoehtoisen muodon verkon klusterointikertoimelle, joka laskee keskiarvon solmujen klusterointikertoimista ainoastaan niille solmuille, joilla on vähintään kaksi naapuria.

MÄÄRITELMÄ 2.14. Olkoon $G = (V, E)$ yleinen verkko. Määritellään joukko

$$(2.7) \quad V' = \{i \in V : k_i \geq 2\} \subseteq V,$$

joka sisältää verkon G solmut, joilla on vähintään kaksi naapuria. Verkon klusterointikerroin laskettuna joukon V' solmuille on [11]

$$(2.8) \quad C_{V'}(G) = \frac{1}{|V'|} \sum_{i \in V'} C_i \in [0, 1],$$

kun $|V'| > 0$, missä C_i on solmun $i \in V'$ klusterointikerroin.

Seuraavaksi esitetään lause, millä tavoin klusterointikertoimet C ja $C_{V'}$ eroavat toisistaan.

LAUSE 2.15. *Olkoon $G = (V, E)$ yleinen verkko ja $V' = \{i \in V : k_i \geq 2\}$ solmujoukon V epätyhjä osajoukko. Kaavojen (2.6) ja (2.8) klusterointikertoimille pätee*

$$C(G) = \frac{|V'|}{|V|} C_{V'}(G).$$

TODISTUS. Olkoon $G = (V, E)$ yleinen verkko ja $V' = \{i \in V : k_i \geq 2\}$ solmujoukon epätyhjä osajoukko. Määritelmän 2.12 mukaan solmun i klusterointikerroin $C_i = 0$, kun $k_i \in \{0, 1\}$, jolloin $\sum_{i \in V} C_i = \sum_{\{i \in V : k_i \geq 2\}} C_i + \sum_{\{i \in V : k_i < 2\}} C_i = \sum_{i \in V'} C_i$. Määritelmistä 2.13 ja 2.14 saadaan nyt $C(G) = \frac{1}{n} \sum_{i \in V} C_i = \frac{1}{|V|} \sum_{i \in V'} C_i = \frac{|V'|}{|V|} \frac{1}{|V'|} \sum_{i \in V'} C_i = \frac{|V'|}{|V|} C_{V'}(G)$. \square

2.3. Verkon transitiivisuuserroin

Fyysikot Alain Barrat ja Martin Weigt [11] esittävät vuoden 2000 julkaisussaan vaihtoehtoisen verkon klusterointikertoimen, joka laskee suhdelukujen $(|\Delta_i|/|\Upsilon_i| : 1 \leq i \leq n)$ keskiarvon sijaan suhdeluvun keskiarvoista $\frac{1}{n} \sum_{i=1}^n |\Delta_i|$ ja $\frac{1}{n} \sum_{i=1}^n |\Upsilon_i|$.

MÄÄRITELMÄ 2.16. A. Barratin ja M. Weigtin klusterointikerroin [11] yleiselle verkolle G on

$$(2.9) \quad \tilde{C}(G) = \frac{\frac{1}{n} \sum_{i=1}^n |\Delta_i|}{\frac{1}{n} \sum_{i=1}^n |\Upsilon_i|},$$

kun $|\Upsilon_i| > 0$ jollakin $i \in V$, missä $|\Upsilon_i|$ on solmun i kytkettyjen kolmikoiden lukumäärä ja $|\Delta_i|$ solmun i sisältävien kolmikoiden lukumäärä.

Fyysikot Mark E. J. Newman ja Duncan J. Watts ja matemaatikko Steven H. Strogatz [13] esittävät vuoden 2002 julkaisussaan toisen klusterointikertoimen verkolle, joka on yleisemmin käytetty. Tämä laskee suhdeluvun verkon kytketyistä kolmioista, jotka myös muodostavat kolmion. Käytetään sille nimitystä *transitiivisuuserroin*.

MÄÄRITELMÄ 2.17. Yleisen verkon G *transitiivisuuserroin* [13] on

$$(2.10) \quad C'(G) = \frac{3|\Delta(G)|}{|\Upsilon(G)|} \in [0, 1],$$

kun $|\Upsilon(G)| > 0$, missä $|\Delta(G)|$ on kolmioiden lukumäärä verkossa ja $|\Upsilon(G)|$ kytkettyjen kolmikoiden lukumäärä verkossa. Lauseen 2.11 mukaan kolmio $\nabla \in \Delta(G)$ sisältää kolmen eri keskisolmun linkki-indusoidun kytketyn kolmikon, joten verkon kytkettyjen kolmikoiden lukumäärä, joiden solmut myös muodostavat kolmion on $3|\Delta(G)|$.

Osoitetaan, että määritelmän 2.17 transitiivisuuserroin on sama kuin A. Barratin ja M. Weigtin esittämä klusterointikerroin määritelmästä 2.16.

LAUSE 2.18. *Olkoon $G = (V, E)$ yleisen verkon, jossa $|\Upsilon_i| > 0$ jollakin $i \in V$. Tällöin $\tilde{C}(G) = C'(G)$.*

TODISTUS. Sievennetään kaavan (2.9) klusterointikerrointa muotoon $\tilde{C}(G) = \sum_{i=1}^n |\Delta_i| / (\sum_{i=1}^n |\Upsilon_i|)$. Verkon kytkettyjen kolmikoiden lukumäärä on kaavan (2.3) mukaan $|\Upsilon(G)| = \sum_{i=1}^n |\Upsilon_i|$ ja verkon kolmioiden lukumäärä on kaavan (2.4) mukaan $|\Delta(G)| = \frac{1}{3} \sum_{i=1}^n |\Delta_i|$. Verkon transitiivisuuserroin kaavasta (2.10) saadaan muotoon

$$(2.11) \quad C'(G) = \frac{3|\Delta(G)|}{|\Upsilon(G)|} = \frac{\sum_{i=1}^n |\Delta_i|}{\sum_{i=1}^n |\Upsilon_i|}.$$

□

2.4. Klusterointi- ja transitiivisuuskertoimen yhteys

Tietojenkäsittelytieteilijät Thomas Schank ja Dorothea Wagner [16] määrittävät verkon *painotetun klusterointikertoimen* C_w painofunktiolla $w : V' \rightarrow \mathbb{R}_+$, jonka avulla he osoittavat klusterointikertoimen $C_{V'}$ ja transitiivisuuskertoimen C' yhteyden. Samat tulokset todistetaan yleisemmin solmujoukon V painofunktiolla $w : V \rightarrow \mathbb{R}_+$.

MÄÄRITELMÄ 2.19. *Olkoon $G = (V, E)$ yleinen verkko ja $w : V \rightarrow \mathbb{R}_+$ verkon solmujen painokerroinfunktio. Verkon *painotettu klusterointikerroin* [16] on*

$$(2.12) \quad C_w(G) = \frac{1}{\sum_{i \in V} w(i)} \sum_{i \in V} w(i) C_i,$$

kun $w(i) > 0$ jollakin $i \in V$, missä C_i on solmun $i \in V$ klusterointikerroin.

LEMMA 2.20. *Olkoon $G = (V, E)$ yleinen verkko ja solmujen painokerroinfunktio $w : V \rightarrow a > 0$ vakiokuvaus. Tällöin $C_w(G) = C(G)$.*

TODISTUS. *Olkoon painokerroinfunktio $w : V \rightarrow a > 0$ vakio, $w(i) = a$ kaikilla $i \in V = \{1, 2, \dots, n\}$. Kaavoista (2.6) ja (2.12) saadaan laskettua $C_w(G) = \sum_{i \in V} w(i) C_i / \sum_{i \in V} w(i) = a \sum_{i=1}^n C_i / \sum_{i=1}^n a = \frac{1}{n} \sum_{i=1}^n C_i = C(G)$.* □

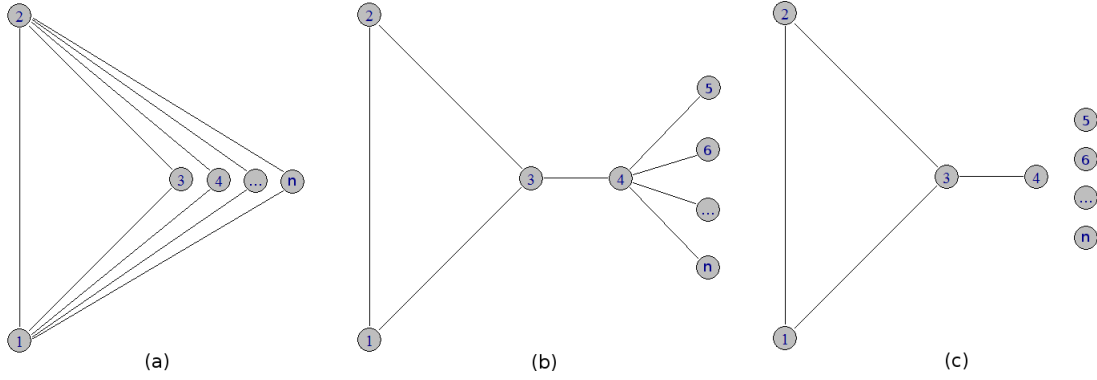
Schank ja Wagner [16] osoittavat, että verkon painotettu klusterointikerroin C_w on sama kuin verkon transitiivisuuserroin C' , kun painokerroinfunktioksi $w : V' \rightarrow \mathbb{R}_+$ valitaan $w(i) = |\Upsilon_i|$. Todistetaan tämä väite solmujoukolle V .

LAUSE 2.21. *Olkoon $G = (V, E)$ yleinen verkko, jossa $|\Upsilon_i| > 0$ jollakin $i \in V$. Olkoon $w : V \rightarrow \mathbb{R}_+$ verkon solmujen painokerroinfunktio, jolle $w(i) = |\Upsilon_i|$ kaikilla $i \in V$. Tällöin $C_w(G) = C'(G)$.*

TODISTUS. Solmun i klusterointikerroin on määritelmän 2.12 mukaan $C_i = |\Delta_i| / |\Upsilon_i|$. Olkoon painofunktio $w(i) : V \rightarrow \mathbb{R}_+$, $w(i) = |\Upsilon_i|$ kaikilla $i \in V$. Kun $|\Upsilon_i| > 0$ jollakin $i \in V$ saadaan kaavoista (2.12) ja (2.11) laskettua $C_w(G) = \sum_{i \in V} w(i) C_i / \sum_{i \in V} w(i) = \sum_{i=1}^n |\Upsilon_i| C_i / \sum_{i=1}^n |\Upsilon_i| = \sum_{i=1}^n |\Delta_i| / \sum_{i=1}^n |\Upsilon_i| = C'(G)$. □

Schank ja Wagner [16] esittävät kaksi tilannetta, jossa verkon klusterointikerroin $C_{V'}$ ja transitiivikerroin C' ovat samoja. Todistetaan, että ne ovat kyseisissä tilanteissa myös sama kuin verkon klusterointikerroin C .

LAUSE 2.22. *Olkoon $G = (V, E)$ yleisen verkon, jossa $k_i \geq 2$ jollakin $i \in V$. Tällöin $C(G) = C_{V'}(G) = C'(G)$, jos*



KUVA 2.3. Esimerkkejä klusterointikertoimien $C = \frac{1}{n} \sum_{i=1}^n |\Delta_i|/|\Upsilon_i|$, $C_{V'} = \frac{1}{|V'|} \sum_{i \in V'} |\Delta_i|/|\Upsilon_i|$ ja $C' = 3|\Delta|/|\Upsilon|$ eroista, kun $n \rightarrow \infty$, missä $V' = \{i \in V : k_i \geq 2\}$. (a): $C = C_{V'} = \frac{1}{n} \left(2 \frac{n-2}{(n-1)(n-2)/2} + (n-2) \right) \rightarrow 1$ ja $C' = 3(n-2)/(2 \binom{n-1}{2} + n-2) = 3/n \rightarrow 0$. (b): $C = \frac{1}{n} \cdot \frac{7}{3} \rightarrow 0$, $C_{V'} = \frac{1}{4} \cdot \frac{7}{3} = \frac{7}{12}$ ja $C' = 3/(5 + \binom{n-3}{2}) \rightarrow 0$. (c): $C = \frac{1}{n} \cdot \frac{7}{3} \rightarrow 0$, $C_{V'} = \frac{1}{3} \cdot \frac{7}{3} = \frac{7}{9}$ ja $C' = \frac{3}{5}$.

- (1) *verkon solmujen asteet ovat samoja tai*
- (2) *verkon solmujen klusterointikertoimet ovat samoja.*

TODISTUS. (1): Olkoon $k_i = a \geq 2$ kaikilla $i \in V$. Kaavan (2.2) mukaan $|\Upsilon_i| = \binom{k_i}{2} = \binom{a}{2} = a' \geq 1$ kaikilla $i \in V$. Määritelmien 2.12 ja 2.13 ja kaavan (2.11) mukaan $C(G) = \frac{1}{n} \sum_{i=1}^n |\Delta_i|/|\Upsilon_i| = \sum_{i=1}^n |\Delta_i|/(na') = \sum_{i=1}^n |\Delta_i|/\sum_{i=1}^n |\Upsilon_i| = C'(G)$. Joukko $V' = \{i \in V : k_i \geq 2\} = V$, joten lauseen 2.15 mukaan $C(G) = C_{V'}|V'|/|V| = C_{V'}$. (2): Olkoon $C_i = b \in [0, 1]$, kaikilla $i \in V$ ja $k_i \geq 2$ jollakin $i \in V$. Tällöin määritelmän 2.13 mukaan $C(G) = \frac{1}{n} \sum_{i=1}^n C_i = \frac{1}{n} \sum_{i=1}^n b = b$. Olkoon $w(i) = |\Upsilon_i|$ kaikilla $i \in V$, jolloin lauseen 2.21 mukaan $C_w(G) = C'(G)$. Toisaalta saadaan kaavasta (2.12) laskettua $C_w(G) = \sum_{i \in V} w(i)C_i / \sum_{i \in V} w(i) = b \sum_{i \in V} |\Upsilon_i| / \sum_{i \in V} |\Upsilon_i| = b = C(G)$. Määritelmän 2.14 mukaan $C_{V'}(G) = \frac{1}{|V'|} \sum_{i \in V'} C_i = b|V'|/|V'| = b = C(G)$. \square

Matemaatikot Béla Bollobás ja Oliver M. Riordan [12] huomauttavat, että yleisessä verkossa C (tai $C_{V'}$) ja C' ovat eri klusterointikertoimia, joka nähdään äärimmillään kuvassa 2.3(a). Lemmasta 2.20 ja lauseesta 2.21 nähdään, että C (tai $C_{V'}$) painottaa solmujen klusterointikertoimia C_i yhtä paljon, kun taas C' painottaa enemmän sellaisia solmujen klusterointikertoimia C_i , joiden aste k_i on iso. Lauseesta 2.15 nähdään, että klusterointikertoimet C ja $C_{V'}$ antavat eri arvot verkossa, jossa $|V'| = |\{i \in V : k_i \geq 2\}| < |V|$ (ks. kuva 2.3(b),(c)). Lauseesta 2.22 seuraa, että verkossa, jossa solmujen asteet eroavat paljon toisistaan, $C_{V'}$ ja C' antavat samankaltaisia arvoja, jos solmujen klusterointikertoimet C_i ovat lähes riippumattomia kyseisten solmujen asteista, muulloin niiden arvot saattavat erota.

Satunnaisverkon klusteroituneisuus

Määritellään seuraavaksi satunnaisverkon klusterointi- ja transitiivisuuserroin sekä kaavoja niiden laskemiseen. Tätä varten tarvitaan todennäköisyysteorian käsitteitä, kuten *riippumattomuus*, *odotusarvo*, *ehdollinen odotusarvo*, *varianssi* ja *Kolmogorovin vahva suurten lukujen laki*.

MÄÄRITELMÄ 3.1. Reaaliarvoiset satunnaismuuttujat X_1, X_2, \dots, X_m ovat *riippumattomia* [3], jos mielivaltaisille *Borel-mitallisille* joukoille A_1, A_2, \dots, A_m pätee

$$\mathbb{P} \left(\bigcap_{t=1}^m \{X_t \in A_t\} \right) = \prod_{t=1}^m \mathbb{P}(X_t \in A_t).$$

Hieman heikompi riippumattomuuskäsite on *pareittainen riippumattomuus*.

MÄÄRITELMÄ 3.2. Satunnaismuuttujat X_1, X_2, \dots, X_m ovat *pareittain riippumattomia* [3], jos kaikki parit ovat riippumattomia.

Satunnaisverkon analysointia usein helpottaa tieto, että satunnaisverkon solmujen asteet ovat riippumattomia. Määritellään seuraavaksi satunnaisverkon G_n solmun $i \in V_n$ aste $k_i(G_n)$, joka on satunnaismuuttuja.

MÄÄRITELMÄ 3.3. Satunnaisverkon G_n solmun $i \in V_n$ aste on satunnaismuuttuja

$$(3.1) \quad k_i(G_n) = \sum_{\{i,j\} \in V_i^{(2)}} X_{ij},$$

missä

$$(3.2) \quad V_i^{(2)} = \{\{i', j'\} \in V_n^{(2)} : i \in \{i', j'\}\} \subset V_n^{(2)}$$

on solmujoukon V_n järjestämättömien parien $V_n^{(2)}$ osajoukko, jonka kaikki alkiot sisältävät solmun i . X_{ij} on indikaattorisatunnaismuuttuja

$$(3.3) \quad X_{ij} = \begin{cases} 1, & \text{jos solmupari } \{i, j\} \in V_n^{(2)} \text{ on kytketty linkillä,} \\ 0, & \text{muulloin.} \end{cases}$$

Solmun $i \in V_n$ aste $k_i(G_n)$ riippuu siitä, kuinka monta solmuparia on kytketty linkillä joukossa $V_i^{(2)}$, jonka koko on $|V_i^{(2)}| = n - 1$. Béla Bollobás [6] huomauttaa, että joukot $(V_i^{(2)} : i \in V_n)$ ovat melkein erillisiä osajoukkoja, sillä leikkaus $V_i^{(2)} \cap V_j^{(2)} = \{i, j\}$, $i, j \in V_n, i \neq j$ koostuu ainoastaan yhdestä alkioista.

3.1. Odotusarvo, varianssi ja ehdollinen odotusarvo

Seuraavaksi määritellään satunnaismuuttujan odotusarvo ja ehdollinen odotusarvo sekä niiden laskusääntöjä.

MÄÄRITELMÄ 3.4. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus. Positiivisen reaaliarvoisen satunnaismuuttujan $X : \Omega \rightarrow S \subset \mathbb{R}_+$ odotusarvo [4] määritellään kaavalla

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) P(\omega),$$

joka voi olla äärellinen tai ääretön.

LAUSE 3.5. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus ja $X : \Omega \rightarrow S \subset \mathbb{R}_+$ sen positiivinen reaaliarvoinen satunnaismuuttuja, jonka jakauma on P_X . Olkoon lisäksi $f : S \rightarrow \mathbb{R}_+$ positiivinen kuvaus. Tällöin [4]

(i) satunnaismuuttujan X odotusarvo saadaan kaavasta

$$\mathbb{E}[X] = \sum_{s \in S} s P_X(s),$$

(ii) satunnaismuuttujan $f(X)$ odotusarvo saadaan kaavasta

$$\mathbb{E}[f(X)] = \sum_{s \in S} f(s) P_X(s).$$

TODISTUS. (ii) [4] Oletetaan, että $f : S \rightarrow \mathbb{R}_+$. Olkoon $X^{-1}\{s\} = \{\omega \in \Omega : X(\omega) = s\}$ tapahtuma, että X saa arvon $s \in S$. Tällöin $(X^{-1}\{s\})_{s \in S}$ on numeroituva kokoelma erillisiä tapahtumia, jotka määritelmän 1.8 mukaan on avaruuden Ω ositus $\bigcup_{s \in S} X^{-1}\{s\} = \Omega$. Koska positiivisten termien summa ei riipu summausjärjestyksestä, havaitaan, että yhdistetyn kuvauksen $h : \Omega \rightarrow \mathbb{R}_+$, $h = f(X)$ odotusarvolle pätee

$$\mathbb{E}[h] = \sum_{\omega \in \Omega} h(\omega) P(\omega) = \sum_{\omega \in \Omega} f(X(\omega)) P(\omega) = \sum_{s \in S} \sum_{\omega \in X^{-1}\{s\}} f(X(\omega)) P(\omega).$$

Koska $X(\omega) = s$ kaikilla $\omega \in X^{-1}\{s\}$ saadaan

$$\sum_{s \in S} \sum_{\omega \in X^{-1}\{s\}} f(X(\omega)) P(\omega) = \sum_{s \in S} f(s) \sum_{\omega \in X^{-1}\{s\}} P(\omega) = \sum_{s \in S} f(s) \mathbb{P}(X^{-1}\{s\}),$$

missä määritelmän 1.7 mukaan $\mathbb{P}(A) = \sum_{\omega \in A} P(\omega)$ kaikilla $A \subset \mathcal{F}$. Määritelmän 1.10 mukaan $P_X(s) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = s\}) = \mathbb{P}(X^{-1}\{s\})$, joten

$$\mathbb{E}[f(X)] = \sum_{s \in S} f(s) \mathbb{P}(X^{-1}\{s\}) = \sum_{s \in S} f(s) P_X(s).$$

(i) Väite seuraa suoraan kohdasta (ii) asettamalla $f(s) = s$ kaikilla $s \in S$. \square

SEURAUS 3.6. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja $\mathbb{1}_A$ indikaattori tapahtumalle $A \in \mathcal{F}$. Tällöin indikaattorin $\mathbb{1}_A$ odotusarvo on

$$\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(\mathbb{1}_A = 1),$$

jossa $\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A)$.

TODISTUS. Indikaattori $\mathbb{1}$ on $\{0, 1\}$ -arvoinen satunnaismuuttuja, joten lauseen 3.5 mukaan $\mathbb{E}[\mathbb{1}_A] = 0 \cdot \mathbb{P}(\mathbb{1}_A = 0) + 1 \cdot \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(\mathbb{1}_A = 1)$. \square

LAUSE 3.7. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus ja $X, Y : \Omega \rightarrow \mathbb{R}_+$ sen positiivisia reaaliarvoisia satunnaismuuttujia. Tällöin [4]*

$$(3.4) \quad \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] \quad \text{kaikilla } a, b \geq 0.$$

TODISTUS. Olkoon $a, b \geq 0$ vakioita. Määritelmästä (3.4) ja positiivitermisen summan summausjärjestystä vaihtamalla saadaan [4]

$$\begin{aligned} \mathbb{E}[aX + bY] &= \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega)) P(\omega) = \sum_{\omega \in \Omega} aX(\omega) P(\omega) + \sum_{\omega \in \Omega} bY(\omega) P(\omega) \\ &= a \sum_{\omega \in \Omega} X(\omega) P(\omega) + b \sum_{\omega \in \Omega} Y(\omega) P(\omega) = a\mathbb{E}[X] + b\mathbb{E}[Y] \end{aligned}$$

\square

LAUSE 3.8. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus ja $X, Y : \Omega \rightarrow \mathbb{R}_+$ sen positiivisia reaaliarvoisia satunnaismuuttujia. Jos X ja Y ovat riippumattomia satunnaismuuttujia, niin [4]*

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y].$$

TODISTUS. Sivuutetaan (ks. [4]). \square

Lasketaan satunnaisverkon G_n solmun asteen odotusarvo käyttämällä positiivisen satunnaismuuttujan odotusarvon lineaarisuutta lauseesta 3.7.

LAUSE 3.9. *Satunnaisverkon G_n solmun $i \in V_n$ asteen $k_i(G_n)$ odotusarvo on*

$$(3.5) \quad \mathbb{E}[k_i(G_n)] = \sum_{\{i,j\} \in V_i^{(2)}} \mathbb{P}(X_{ij} = 1),$$

missä X_{ij} on indikaattori tapahtumalle, että solmupari $\{i, j\}$ on kytketty linkillä.

TODISTUS. Positiivisille reaaliarvoisille satunnaismuuttujille ($X_{ij} : \{i, j\} \in V_i^{(2)}$) saadaan lauseen 3.7 ja seurauksen 3.6 mukaan

$$\mathbb{E}[k_i(G_n)] = \mathbb{E} \left[\sum_{\{i,j\} \in V_i^{(2)}} X_{ij} \right] = \sum_{\{i,j\} \in V_i^{(2)}} \mathbb{E}[X_{ij}] = \sum_{\{i,j\} \in V_i^{(2)}} \mathbb{P}(X_{ij} = 1).$$

\square

MÄÄRITELMÄ 3.10. Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ todennäköisyysavaruus ja $X : \Omega \rightarrow \mathbb{R}$ reaaliarvoinen satunnaismuuttuja, jolle $\mathbb{E}[|X^2|] < \infty$. Satunnaismuuttujan X *varianssi* [2] määritellään kaavalla

$$\text{Var}(X) = \mathbb{E}[X - \mathbb{E}[X]]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

LAUSE 3.11. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus ja $X, Y : \Omega \rightarrow \mathbb{R}_+$ sen positiivisia reaaliarvoisia satunnaismuuttujia, joiden varianssit ovat äärellisiä. Jos X ja Y ovat riippumattomia satunnaismuuttujia, niin [3]*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

TODISTUS. Määritelmän 3.10 mukaan $\text{Var}(X) = \mathbb{E}[(X)^2] - \mathbb{E}[X]^2$, joten lauseista 3.7 ja 3.8 saadaan laskettua

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= \mathbb{E}[X^2 + 2XY + Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - (\mathbb{E}[X]^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\ &= \text{Var}(X) + \text{Var}(Y). \end{aligned}$$

□

Ehdollisen odotusarvon määrittämiseen tarvitaan *ehdollinen todennäköisyys* ja sen laskusääntöjä.

MÄÄRITELMÄ 3.12. *Olkoot $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus ja $A, B \in \mathcal{F}$ tapahtumia, jossa $\mathbb{P}(B) > 0$. Tapahtuman A ehdollinen todennäköisyys [4] ehdolla B on*

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

MÄÄRITELMÄ 3.13. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus. Tapahtumat $A, B \in \mathcal{F}$ ovat riippumattomia [2], jos*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

MÄÄRITELMÄ 3.14. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus. Tapahtumat $A, B \in \mathcal{F}$ ovat ehdollisesti riippumattomia ehdolla $C \in \mathcal{F}$, $\mathbb{P}(C) > 0$ jos*

$$\mathbb{P}(A \cap B | C) = \mathbb{P}(A | C)\mathbb{P}(B | C).$$

LAUSE 3.15. *Olkoot $A, B \in \mathcal{F}$ diskreetin todennäköisyysavaruuden $(\Omega, \mathcal{F}, \mathbb{P})$ riippumattomia tapahtumia, jossa $\mathbb{P}(B) > 0$. Tällöin*

$$\mathbb{P}(A | B) = \mathbb{P}(A).$$

TODISTUS. *Olkoon $\mathbb{P}(B) > 0$. Määritelmän 3.12 mukaan $\mathbb{P}(A | B) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$. Koska tapahtumat A ja B ovat riippumattomia, saadaan määritelmästä 3.13, että $\mathbb{P}(A | B) = \mathbb{P}(A \cap B) / \mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B) / \mathbb{P}(B) = \mathbb{P}(A)$. □*

LAUSE 3.16. *Olkoot $A_1, \dots, A_n \in \mathcal{F}$ tapahtumia ja $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$. Tällöin*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}).$$

TODISTUS. [2] *Olkoon $n = 2$. Väite seuraa tällöin suoraan määritelmästä 3.12. Tehdään induktio-oletus, että väite pätee $n - 1$ tapahtumille $A_1 \cap \dots \cap A_{n-1}$, jolloin*

$$\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_{n-1} | A_1 \cap \dots \cap A_{n-2}).$$

Seuraavaksi todistetaan, että väite pätee myös n tapahtumille. Merkitän tapahtuma $B = A_1 \cap \dots \cap A_{n-1}$, jolloin määritelmästä 3.12 saadaan $\mathbb{P}(B \cap A_n) = \mathbb{P}(A_n | B) \mathbb{P}(B)$. Induktio-oletuksesta seuraa

$$\begin{aligned} \mathbb{P}(A_1 \cap \dots \cap A_n) &= \mathbb{P}(B \cap A_n) = \mathbb{P}(A_n | B) \mathbb{P}(B) \\ &= \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_{n-1} | A_1 \cap \dots \cap A_{n-2}) \\ &\quad \cdot \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}). \end{aligned}$$

□

LAUSE 3.17. *Olkoon $(\Omega, \mathcal{F}, \mathbb{P})$ diskreetti todennäköisyysavaruus ja $\{H_i : 1 \leq i \leq n\}$ perusjoukon Ω ositus. Tällöin kaikille $A \in \mathcal{F}$ pätee*

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A | H_i) \mathbb{P}(H_i).$$

TODISTUS. [2] Olkoon $\{H_i : 1 \leq i \leq n\}$ perusjoukon Ω ositus. Tapahtumalle $A \in \mathcal{F}$ pätee tällöin $A = A \cap \Omega = A \cap (\cup_{i=1}^n H_i) = \cup_{i=1}^n (A \cap H_i)$. Joukot $\{H_i : 1 \leq i \leq n\}$ ovat erillisiä joukkoja, joten sama pätee joukoille $\{(A \cap H_i) : 1 \leq i \leq n\}$. Tällöin todennäköisyysmitan määritelmän 1.6 mukaan $\mathbb{P}(A) = \mathbb{P}(\cup_{i=1}^n (A \cap H_i)) = \sum_{i=1}^n \mathbb{P}(A \cap H_i) = \sum_{i=1}^n \mathbb{P}(A | H_i) \mathbb{P}(H_i)$. □

Nyt voidaan määritellä diskreetin satunnaismuuttujan *ehdollinen odotusarvo* ja sen laskusääntöjä.

MÄÄRITELMÄ 3.18. Olkoot $X : \Omega \rightarrow \{x_1, x_2, \dots\}$ ja $Y : \Omega \rightarrow \{y_1, y_2, \dots\}$, $Y \geq 0$, diskreettejä satunnaismuuttujia. Satunnaismuuttujan Y *ehdollinen odotusarvo* [2] ehdolla $\{X = x_j\}$, kun $\mathbb{P}(X = x_j) > 0$ on

$$\mathbb{E}[Y | X = x_j] = \sum_{k=1}^{\infty} y_k \mathbb{P}(Y = y_k | X = x_j).$$

Määritellään seuraavaksi diskreetin satunnaismuuttujan Y ehdollinen odotusarvo, kun ehtona on *satunnaismuuttuja* X tapahtuman $\{X = x_j\}$ sijaan. Tällöin ehdollinen odotusarvo $\mathbb{E}[Y | X]$ on satunnaismuuttuja.

MÄÄRITELMÄ 3.19. Olkoot $X : \Omega \rightarrow \{x_1, x_2, \dots\}$ ja $Y : \Omega \rightarrow \{y_1, y_2, \dots\}$, $Y \geq 0$, diskreettejä satunnaismuuttujia. Satunnaismuuttujan Y *ehdollinen odotusarvo* [2] ehdolla X on

$$\mathbb{E}[Y | X] = f(X),$$

missä funktio $f : \{x_1, x_2, \dots\} \rightarrow \mathbb{R}$ on määritelty

$$(3.6) \quad f(x) = \begin{cases} \mathbb{E}[Y | X = x], & \text{kun } \mathbb{P}(X = x) > 0, \\ \text{mielivaltainen arvo,} & \text{kun } \mathbb{P}(X = x) = 0 \end{cases}$$

kun satunnaismuuttuja Y on integroitava todennäköisyysmitan $\mathbb{P}(\cdot | X = x_j)$ suhteen kaikille j , joille $\mathbb{P}(X = x_j) > 0$.

LAUSE 3.20. *Olkoot $X : \Omega \rightarrow \{x_1, x_2, \dots\}$ ja $Y : \Omega \rightarrow \{y_1, y_2, \dots\}$, $Y \geq 0$, diskreettejä satunnaismuuttujia. Tällöin*

$$\mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}[Y].$$

TODISTUS. Määritelmän 3.19 mukaan satunnaismuuttujan X muunnos $f(X) = \mathbb{E}[Y | X]$ on satunnaismuuttuja, jonka odotusarvo saadaan laskettua satunnaismuuttujan X jakauman avulla lauseesta 3.5, $\mathbb{E}[f(X)] = \sum_{j=1}^{\infty} f(x_j) \mathbb{P}(X = x_j)$. Kaavan (3.6) mukaan $f(x) = \mathbb{E}[Y | X = x]$, kun $\mathbb{P}(X = x) > 0$, joten saadaan

$$\mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}[f(X)] = \sum_{j=1}^{\infty} f(x_j) \mathbb{P}(X = x_j) = \sum_{j:\mathbb{P}(X=x_j)>0} \mathbb{E}[Y | X = x_j] \mathbb{P}(X = x_j).$$

Määritelmästä 3.18 saadaan nyt laskettua

$$\sum_{j:\mathbb{P}(X=x_j)>0} \mathbb{E}[Y | X = x_j] \mathbb{P}(X = x_j) = \sum_{j:\mathbb{P}(X=x_j)>0} \left(\sum_{k=1}^{\infty} y_k \mathbb{P}(Y = y_k | X = x_j) \right) \mathbb{P}(X = x_j).$$

Määritelmän 3.12 mukaan $\mathbb{P}(Y = y_k | X = x_j) = \mathbb{P}(Y = y_k, X = x_j) / \mathbb{P}(X = x_j)$, joten saadaan

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y | X]] &= \sum_{j:\mathbb{P}(X=x_j)>0} \left(\sum_{k=1}^{\infty} y_k \mathbb{P}(Y = y_k | X = x_j) \right) \mathbb{P}(X = x_j) \\ &= \sum_{j:\mathbb{P}(X=x_j)>0} \left(\sum_{k=1}^{\infty} y_k \frac{\mathbb{P}(Y = y_k, X = x_j)}{\mathbb{P}(X = x_j)} \right) \mathbb{P}(X = x_j) \\ &= \sum_{j:\mathbb{P}(X=x_j)>0} \sum_{k=1}^{\infty} y_k \mathbb{P}(Y = y_k, X = x_j) \\ &= \sum_{k=1}^{\infty} y_k \left(\sum_{j:\mathbb{P}(X=x_j)>0} \mathbb{P}(Y = y_k, X = x_j) \right) \\ &= \sum_{k=1}^{\infty} y_k \left(\sum_{j=1}^{\infty} \mathbb{P}(Y = y_k, X = x_j) \right). \end{aligned}$$

Olkoon $X^{-1}\{x_j\} = \{\omega \in \Omega : X(\omega) = x_j\} = \{X = x_j\}$ tapahtuma. Tällöin tapahtumat $(X^{-1}\{x_j\})_{j=1}^{\infty}$ ovat numeroituva kokoelma erillisiä tapahtumia, jotka määritelmän 1.8 mukaan muodostavat avaruuden Ω osituksen, $\bigcup_{j=1}^{\infty} X^{-1}\{x_j\} = \Omega$. Tällöin tapahtumat $(\{Y = y_k \cap X = x_j\})_{j=1}^{\infty}$ ovat myös erillisiä tapahtumia, joten todennäköisyysmitan määritelmän 1.6 mukaan

$$\begin{aligned} &\sum_{j=1}^{\infty} \mathbb{P}(Y = y_k, X = x_j) \\ &= \mathbb{P} \left(\bigcup_{j=1}^{\infty} (Y = y_k \cap X = x_j) \right) = \mathbb{P} \left((Y = y_k) \cap \left(\bigcup_{j=1}^{\infty} X = x_j \right) \right) \\ &= \mathbb{P}((Y = y_k) \cap \Omega) = \mathbb{P}(Y = y_k). \end{aligned}$$

Lauseesta 3.5 saadaan nyt laskettua $\mathbb{E}[\mathbb{E}[Y | X]] = \sum_{k=1}^{\infty} y_k \mathbb{P}(Y = y_k) = \mathbb{E}[Y]$. \square

LAUSE 3.21. *Olkoot X, Y ja Z diskreettejä satunnaismuuttujia, jossa $X, Y \geq 0$. Tällöin*

$$\mathbb{E}[aX + bY \mid Z] = a\mathbb{E}[X \mid Z] + b\mathbb{E}[Y \mid Z], \text{ kaikilla } a, b \geq 0.$$

TODISTUS. [3] Olkoot $a, b \geq 0$ vakioita. Lauseista 3.20 ja 3.7 saadaan

$$\begin{aligned} \mathbb{E}[\mathbb{E}[aX + bY \mid Z]] &= \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y] = a\mathbb{E}[\mathbb{E}[X \mid Z]] + b\mathbb{E}[\mathbb{E}[Y \mid Z]] \\ &= \mathbb{E}[a\mathbb{E}[X \mid Z] + b\mathbb{E}[Y \mid Z]]. \end{aligned}$$

□

3.2. Kytketty kolmikko ja kolmio satunnaisverkossa

Seuraavaksi tarkastellaan satunnaisverkon G_n solmun $i \in V_n$ kytkettyjen kolmikoiden lukumäärä $|\Upsilon_i|$, joka on satunnaismuuttuja. Merkitään symbolilla $V_n^{(3)}$ solmujoukon V_n järjestämättömien kolmikoiden joukkoa.

MÄÄRITELMÄ 3.22. Satunnaisverkon G_n solmun i kytkettyjen kolmikoiden lukumäärä on satunnaismuuttuja

$$(3.7) \quad |\Upsilon_i(G_n)| = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{1}_{\{ij,ik\}},$$

missä

$$(3.8) \quad V_i^{(3)} = \{\{i',j',k'\} \in V_n^{(3)} : i \in \{i',j',k'\}\} \subset V_n^{(3)},$$

on solmujoukon V_n järjestämättömien kolmikoiden $V_n^{(3)}$ osajoukko, jonka kaikki alkiot sisältävät solmun $i \in V_n$. $\mathbb{1}_{\{ij,ik\}}$ on indikaattorisatunnaismuuttuja

$$(3.9) \quad \mathbb{1}_{\{ij,ik\}} = \begin{cases} 1, & \text{jos solmuparit } \{i,j\}, \{i,k\} \in V_n^{(2)} \text{ on kytketty linkillä,} \\ 0, & \text{muulloin.} \end{cases}$$

Kaavassa 3.7 indikaattorit ($\mathbb{1}_{\{ij,ik\}} : \{i,j,k\} \in V_i^{(3)}$) eivät ole riippumattomia (ks. kuva 3.1), mutta siitä huolimatta saadaan helposti laskettua solmun kytkettyjen kolmikoiden lukumäärän odotusarvo.

LAUSE 3.23. *Satunnaisverkon G_n solmun $i \in V_n$ kytkettyjen kolmikoiden lukumäärän odotusarvo on*

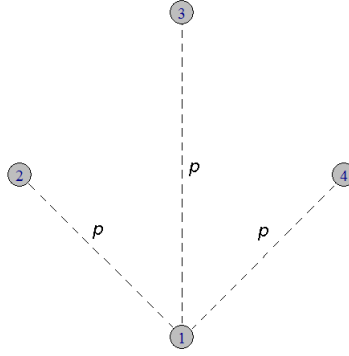
$$(3.10) \quad \mathbb{E}[|\Upsilon_i(G_n)|] = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik\}} = 1),$$

missä $\mathbb{1}_{\{ij,ik\}}$ on indikaattori tapahtumalle, että solmuparit $\{i,j\}, \{i,k\} \in V_n^{(2)}$ on kytketty linkillä.

TODISTUS. Kaavoista (3.4), (3.7) ja seurauksesta 3.6 saadaan laskettua

$$\mathbb{E}[|\Upsilon_i(G_n)|] = \mathbb{E} \left[\sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{1}_{\{ij,ik\}} \right] = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{E}[\mathbb{1}_{\{ij,ik\}}] = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik\}} = 1).$$

□



KUVA 3.1. Satunnaisverkossa $G_{4,p}$ solmun 1 kytkettyjen kolmikoiden lukumäärä on satunnaismuuttuja $|\Upsilon_1(G_n)| = \sum_{\{i,j,k\} \in V_1^{(3)}} \mathbb{1}_{\{ij,ik\}}$. Indikaattorit $(\mathbb{1}_{\{ij,ik\}} : \{i,j,k\} \in V_1^{(3)})$ eivät ole riippumattomia, sillä $\mathbb{P}(\mathbb{1}_{\{\{1,2\},\{1,3\}\}} = 1, \mathbb{1}_{\{\{1,3\},\{1,4\}\}} = 1) = p^3 \neq p^4 = \mathbb{P}(\mathbb{1}_{\{\{1,2\},\{1,3\}\}} = 1) \mathbb{P}(\mathbb{1}_{\{\{1,3\},\{1,4\}\}} = 1)$. Solmun 1 kytkettyjen kolmikoiden lukumäärän odotusarvo saadaan laskettua kaavasta (3.10), $\mathbb{E}[|\Upsilon_1(G_{n,p})|] = \sum_{\{i,j,k\} \in V_1^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik\}} = 1) = \binom{3}{2} p^2 = 3p^2$.

SEURAUUS 3.24. *Satunnaisverkon G_n kytkettyjen kolmikoiden kokonaismäärän odotusarvo on*

$$\mathbb{E}[|\Upsilon(G_n)|] = \sum_{i=1}^n \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik\}} = 1),$$

missä $\mathbb{1}_{\{ij,ik\}}$ on indikaattori tapahtumalle, että solmuparit $\{i,j\}, \{i,k\} \in V_n^{(2)}$ on kytketty linkillä.

TODISTUS. Lauseen 2.3 mukaan kytkettyjen kolmikoiden kokonaismäärä on $|\Upsilon| = \sum_{i=1}^n |\Upsilon_i|$, joten satunnaisverkon G_n kytkettyjen kolmikoiden kokonaismäärän odotusarvo saadaan laskettua lauseesta 3.7, $\mathbb{E}[|\Upsilon(G_n)|] = \mathbb{E}[\sum_{i=1}^n |\Upsilon_i|] = \sum_{i=1}^n \mathbb{E}[|\Upsilon_i|]$. Väite seuraa lauseesta 3.23. \square

Kaavasta (2.2) huomataan, että solmun i kytkettyjen kolmikoiden lukumäärä riippuu ainoastaan solmun i asteesta $k_i(G_n)$, sillä $|\Upsilon_i(G_n)| = \binom{k_i(G_n)}{2} = k_i(G_n)(k_i(G_n) - 1)/2$. Seuraavaksi esitetään lause, jonka avulla voidaan laskea solmun i kytkettyjen kolmikoiden lukumäärän odotusarvo solmun i asteen $k_i(G_n)$ jakauman tunnuslukujen avulla.

LAUSE 3.25. *Satunnaisverkon G_n solmun $i \in V_n$ kytkettyjen kolmikoiden lukumäärän odotusarvo on*

$$\mathbb{E}[|\Upsilon_i(G_n)|] = \frac{1}{2} (\mathbb{E}[k_i(G_n)]^2 - \mathbb{E}[k_i(G_n)] + \text{Var}(k_i(G_n))),$$

missä $k_i(G_n)$ on solmun i aste ja $\text{Var}(k_i(G_n))$ asteen $k_i(G_n)$ varianssi.

TODISTUS. Kaavan (2.2) mukaan satunnaisverkon G_n solmun $i \in V_n$ kytkettyjen kolmikoiden lukumäärä on $|\Upsilon_i(G_n)| = \binom{k_i(G_n)}{2} = k_i(G_n)(k_i(G_n) - 1)/2$, missä $k_i(G_n)$

on solmun i aste. Lauseesta 3.7 saadaan nyt laskettua

$$\mathbb{E}[|\Upsilon_i(G_n)|] = \frac{1}{2}\mathbb{E}[k_i(G_n)(k_i(G_n) - 1)] = \frac{1}{2}(\mathbb{E}[k_i(G_n)^2] - \mathbb{E}[k_i(G_n)]).$$

Määritelmän 3.10 mukaan asteen $k_i(G_n)$ varianssi on $\text{Var}(k_i(G_n)) = \mathbb{E}[k_i(G_n)^2] - \mathbb{E}[k_i(G_n)]^2$, joten solmun i kytkettyjen kolmikoiden lukumäärän odotusarvo saadaan muotoon

$$\begin{aligned}\mathbb{E}[|\Upsilon_i(G_n)|] &= \frac{1}{2}(\mathbb{E}[k_i(G_n)^2] - \mathbb{E}[k_i(G_n)]) \\ &= \frac{1}{2}(\text{Var}(k_i(G_n)) + \mathbb{E}[k_i(G_n)]^2 - \mathbb{E}[k_i(G_n)]).\end{aligned}$$

□

Seuraavaksi tarkastellaan satunnaisverkon G_n solmun $i \in V_n$ kolmioiden lukumäärää $|\Delta_i|$, joka on satunnaismuuttuja.

MÄÄRITELMÄ 3.26. Satunnaisverkon G_n solmun $i \in V_n$ kolmioiden lukumäärä on satunnaismuuttuja

$$(3.11) \quad |\Delta_i(G_n)| = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{1}_{\{ij,ik,jk\}},$$

missä $\mathbb{1}_{\{ij,ik,jk\}}$ on indikaattorisatunnaismuuttuja

$$(3.12) \quad \mathbb{1}_{\{ij,ik,jk\}} = \begin{cases} 1, & \text{jos solmuparit } \{i,j\}, \{i,k\}, \{j,k\} \in V_n^{(2)} \text{ on kytketty linkillä,} \\ 0, & \text{muulloin.} \end{cases}$$

LEMMA 3.27. *Satunnaisverkon G_n solmun i kolmioiden lukumäärälle $|\Delta_i(G_n)|$ pätee*

$$|\Delta_i(G_n)| \leq |\Upsilon_i(G_n)|,$$

missä $|\Upsilon_i(G_n)|$ on solmun i kytkettyjen kolmikoiden lukumäärä.

TODISTUS. Indikaattorin $\mathbb{1}_{\{ij,ik,jk\}}$ määritelmästä nähdään, että $\mathbb{1}_{\{ij,ik,jk\}} = \mathbb{1}_{\{ij,ik\}} X_{jk}$, missä $X_{j,k}$ on $\{0,1\}$ -arvoinen satunnaismuuttuja, joten

$$|\Delta_i(G_n)| = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{1}_{\{ij,ik\}} X_{jk} \leq \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{1}_{\{ij,ik\}} \cdot 1 = |\Upsilon_i(G_n)|.$$

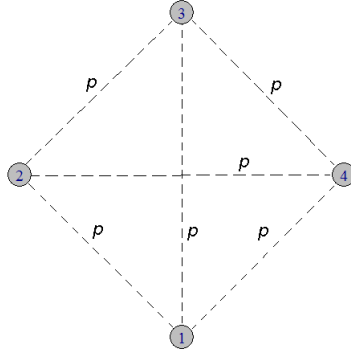
□

Kaavan (3.11) indikaattorit ($\mathbb{1}_{\{ij,ik,jk\}} : \{i,j,k\} \in V_i^{(3)}$) eivät ole riippumattomia (ks. kuva 3.2), mutta siitä huolimatta saadaan helposti laskettua solmun kolmioiden lukumäärän odotusarvo.

LAUSE 3.28. *Satunnaisverkon G_n solmun i kolmioiden lukumäärän odotusarvo on*

$$(3.13) \quad \mathbb{E}[|\Delta_i(G_n)|] = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1),$$

missä $\mathbb{1}_{\{ij,ik,jk\}}$ on indikaattori tapahtumalle, että solmuparit $\{i,j\}, \{i,k\}, \{j,k\} \in V_n^{(2)}$ on kytketty linkeillä.



KUVA 3.2. Satunnaisverkossa $G_{4,p}$ solmun 1 kolmioiden lukumäärä on satunnaismuuttuja $|\Delta_1(G_n)| = \sum_{\{i,j,k\} \in V_1^{(3)}} \mathbb{1}_{\{ij,ik,jk\}}$. Indikaattorit $(\mathbb{1}_{\{ij,ik,jk\}} : \{i,j,k\} \in V_1^{(3)})$ eivät ole riippumattomia, sillä $\mathbb{P}(\mathbb{1}_{\{\{1,2\},\{1,3\},\{2,3\}\}} = 1, \mathbb{1}_{\{\{1,3\},\{1,4\},\{3,4\}\}} = 1) = p^5 \neq p^6 = \mathbb{P}(\mathbb{1}_{\{\{1,2\},\{1,3\},\{2,3\}\}} = 1) \mathbb{P}(\mathbb{1}_{\{\{1,3\},\{1,4\},\{3,4\}\}} = 1)$. Solmun 1 kolmioiden lukumäärän odotusarvo saadaan laskettua kaavasta (3.13), $\mathbb{E}[|\Delta_1(G_{n,p})|] = \sum_{\{i,j,k\} \in V_1^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1) = \binom{3}{2} p^3 = 3p^3$.

TODISTUS. Kaavoista (3.4), (3.11) ja seurauksesta 3.6 saadaan

$$\mathbb{E}[|\Delta_i|] = \mathbb{E} \left[\sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{1}_{\{ij,ik,jk\}} \right] = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{E}[\mathbb{1}_{\{ij,ik,jk\}}] = \sum_{\{i,j,h\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1).$$

□

SEURAUS 3.29. *Satunnaisverkon G_n kolmioiden kokonaismäärän odotusarvo on*

$$\mathbb{E}[|\Delta(G_n)|] = \sum_{\{i,j,k\} \in V_n^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1),$$

missä $\mathbb{1}_{\{ij,ik,jk\}}$ on indikaattori tapahtumalle, että solmuparit $\{i,j\}, \{i,k\}, \{j,k\} \in V_n^{(2)}$ on kytketty linkillä.

TODISTUS. Kaavasta (2.4) saadaan $|\Delta(G_n)| = \frac{1}{3} \sum_{i=1}^n |\Delta_i(G_n)|$, joten satunnaisverkon G_n kolmioiden kokonaismäärän odotusarvo on lauseen 3.7 mukaan $\mathbb{E}[|\Delta(G_n)|] = \mathbb{E}[\frac{1}{3} \sum_{i=1}^n |\Delta_i|] = \frac{1}{3} \sum_{i=1}^n \mathbb{E}[|\Delta_i(G_n)|]$. Lauseesta 3.28 saadaan tämä muotoon

$$\frac{1}{3} \sum_{i=1}^n \mathbb{E}[|\Delta_i(G_n)|] = \frac{1}{3} \sum_{i=1}^n \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1).$$

Jokainen järjestämätön kolmen solmun joukko $\{i,j,k\} \in V_n^{(3)}$ kuuluu kolmen eri solmun $i,j,k \in V_n$ joukkoihin $V_i^{(3)}, V_j^{(3)}$ ja $V_k^{(3)}$. Käymällä jokaisen solmun $i \in V_n$ joukot $\{i,j,k\} \in V_i^{(3)}$ läpi tulee todennäköisyys $\mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1)$ laskettua kolme

kertaa, josta seuraa, että

$$(3.14) \quad \mathbb{E}[|\Delta(G_n)|] = \frac{1}{3} \sum_{i=1}^n \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1) = \sum_{\{i,j,k\} \in V_n^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1).$$

□

3.3. Satunnaisverkon klusterointikerroin

Solmun $i \in V$ klusterointikerroin yleisessä verkossa G on määritelmän 2.12 mukaan $C_i = |\Delta_i|/|\Upsilon_i|$, kun $|\Upsilon_i| \geq 1$ ja $C_i = 0$ muulloin. Koko verkon G klusterointikerroin on määritelmän 2.13 mukaan $C(G) = \frac{1}{n} \sum_{i=1}^n C_i$. Satunnaisverkossa G_n solmun $i \in V_n$ kytkettyjen kolmikoiden ja kolmioiden lukumäärät $|\Upsilon_i(G_n)|$ ja $|\Delta_i(G_n)|$ ovat satunnaismuuttujia, joten satunnaisverkon G_n solmun i klusterointikerroin on myös satunnaismuuttuja,

$$(3.15) \quad C_i(G_n) = \begin{cases} |\Delta_i(G_n)|/|\Upsilon_i(G_n)|, & \text{kun } |\Upsilon_i(G_n)| > 0, \\ 0, & \text{muulloin.} \end{cases}$$

Samoin $\frac{1}{n} \sum_{i=1}^n C_i(G_n)$ on satunnaismuuttujien summana myös satunnaismuuttuja.

Symmetrisesti jakautuneessa satunnaisverkossa, kuten $G_{n,p}$, solmujen klusterointikertoimet ovat samoin jakautuneita satunnaismuuttujia. Tällöin voidaan esittää seuraava määritelmä satunnaisverkon G_n klusterointikertoimelle.

MÄÄRITELMÄ 3.30. Olkoon G_n symmetrisesti jakautunut satunnaisverkko, jossa solmujen klusterointikertoimet $(C_i(G_n))_{i \in V_n}$ ovat samoin jakautuneita satunnaismuuttujia. Määritellään satunnaisverkon G_n klusterointikerroin

$$(3.16) \quad c(G_n) = \mathbb{E}[C_i(G_n)],$$

joka ei riipu solmun $i \in V_n$ valinnasta.

LAUSE 3.31. *Satunnaisverkon G_n solmun i klusterointikertoimen odotusarvo on*

$$\mathbb{E}[C_i(G_n)] = \sum_{k': \mathbb{P}(k_i(G_n)=k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k'),$$

missä $k_i(G_n) \in \{0, 1, \dots, n-1\}$ on solmun i aste.

TODISTUS. Satunnaismuuttujat $C_i(G_n)$ ja $k_i(G_n)$ ovat diskreettejä, joten lauseesta 3.20 saadaan solmun i klusterointikertoimen odotusarvo muotoon

$$\mathbb{E}[C_i(G_n)] = \mathbb{E}[\mathbb{E}[C_i(G_n) \mid k_i(G_n)]].$$

Määritelmän 3.19 mukaan $\mathbb{E}[C_i(G_n) \mid k_i(G_n)] = f(k_i(G_n))$ on satunnaismuuttuja, jolle $f(k') = \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k']$, kun $\mathbb{P}(k_i(G_n) = k') > 0$. Lauseesta 3.5 saadaan nyt laskettua satunnaismuuttujan $k_i(G_n)$ muunnoksen $f(k_i(G_n))$ odotusarvo

$$\mathbb{E}[C_i(G_n)] = \mathbb{E}[f(k_i(G_n))] = \sum_{k': \mathbb{P}(k_i(G_n)=k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k').$$

□

LEMMA 3.32. *Satunnaisverkon G_n solmun i klusterointikertoimen ehdollinen odotusarvo ehdolla $k_i(G_n) = k' \in \{0, 1\}$, kun $\mathbb{P}(k_i(G_n) = k') > 0$ on*

$$\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = 0.$$

TODISTUS. Olkoon $k_i(G_n) = k' \in \{0, 1\}$ ja $\mathbb{P}(k_i(G_n) = k') > 0$. Tällöin kaavan (2.2) mukaan $|\Upsilon_i(G_n)| = 0$. Kaavasta (3.15) nähdään, että $C_i(G_n) = 0$, kun $|\Upsilon_i(G_n)| < 1$, joten $\mathbb{E}[C_i(G_n, p) \mid k_i(G_n) = k'] = \mathbb{E}[0] = 0$, kun $k' \in \{0, 1\}$. \square

SEURAUUS 3.33. *Satunnaisverkon solmun i klusterointikertoimen odotusarvo on*

$$\mathbb{E}[C_i(G_n)] = \sum_{k' \geq 2: \mathbb{P}(k_i(G_n) = k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k'),$$

missä $k_i(G_n) \in \{0, 1, \dots, n-1\}$ on solmun i aste.

TODISTUS. Määritelmän 3.3 mukaan solmun i asteelle pätee $k_i(G_n) \in \{0, 1, \dots, n-1\}$, joka on jakautunut sen jakauman $P_{k_i}(k') = \mathbb{P}(k_i(G_n) = k')$, $k' \in \{0, 1, \dots, n-1\}$ mukaan. Lauseesta 3.31 ja lemmasta 3.32 saadaan

$$\begin{aligned} \mathbb{E}[C_i(G_n)] &= \sum_{k' < 2: \mathbb{P}(k_i(G_n) = k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k') \\ &+ \sum_{k' \geq 2: \mathbb{P}(k_i(G_n) = k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k') \\ &= 0 + \sum_{k' \geq 2: \mathbb{P}(k_i(G_n) = k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k'). \end{aligned}$$

\square

LEMMA 3.34. *Satunnaisverkon solmun i klusterointikertoimen ehdollinen odotusarvo ehdolla $k_i(G_n) = k' \geq 2$, kun $\mathbb{P}(k_i(G_n) = k') > 0$ on*

$$\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = \frac{1}{\binom{k'}{2}} \sum_{\{i, j, k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij, ik, jk\}} = 1 \mid k_i(G_n) = k').$$

TODISTUS. Olkoon $k' \geq 2$ ja $\mathbb{P}(k_i(G_n) = k') > 0$. Kaavan (2.2) mukaan ehdosta $k_i(G_n) = k'$ seuraa $|\Upsilon_i(G_n)| = \binom{k_i(G_n)}{2} = \binom{k'}{2}$, joten kaavasta (3.15) ja lauseesta 3.21 saadaan

(3.17)

$$\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = \mathbb{E}\left[\frac{|\Delta_i(G_n)|}{|\Upsilon_i(G_n)|} \mid k_i(G_n) = k'\right] = \frac{1}{\binom{k'}{2}} \mathbb{E}[|\Delta_i(G_n)| \mid k_i(G_n) = k'].$$

Määritelmän 3.26 ja lauseen 3.21 ehdollisen odotusarvon lineaarisuuden mukaan

$$\mathbb{E}[|\Delta_i(G_n)| \mid k_i = k'] = \mathbb{E}\left[\sum_{\{i, j, k\} \in V_i^{(3)}} \mathbb{1}_{\{ij, ik, jk\}} \mid k_i = k'\right] = \sum_{\{i, j, k\} \in V_i^{(3)}} \mathbb{E}[\mathbb{1}_{\{ij, ik, jk\}} \mid k_i = k'].$$

Määritelmästä 3.18 ja seurauksesta 3.6 saadaan laskettua

$$\mathbb{E}[\mathbb{1}_{\{ij, ik, jk\}} \mid k_i(G_n) = k'] = 0 + 1 \cdot \mathbb{P}(\mathbb{1}_{\{ij, ik, jk\}} = 1 \mid k_i(G_n) = k')$$

kaikilla $\{i, j, k\} \in V_i^{(3)}$, joten

$$\mathbb{E} [|\Delta_i(G_n)| \mid k_i(G_n) = k'] = \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1 \mid k_i(G_n) = k').$$

□

LAUSE 3.35. *Satunnaisverkon G_n solmun i klusterointikertoimen odotusarvo on*

$$\mathbb{E} [C_i(G_n)] = \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \frac{1}{\binom{k'}{2}} \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1, k_i(G_n) = k').$$

TODISTUS. Seurauksen 3.33 mukaan satunnaisverkon G_n solmun i klusterointikertoimen odotusarvo on

$$\mathbb{E}[C_i(G_n)] = \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k').$$

Lemmasta 3.34 saadaan laskettua satunnaisverkon solmun i klusterointikertoimen ehdollinen odotusarvo ehdolla $k_i(G_n) = k' \geq 2$, kun $\mathbb{P}(k_i(G_n) = k') > 0$

$$\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = \frac{1}{\binom{k'}{2}} \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1 \mid k_i(G_n) = k').$$

Määritelmän 3.12 mukaan $\mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1 \mid k_i(G_n) = k') = \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1, k_i(G_n) = k') / \mathbb{P}(k_i(G_n) = k')$ kaikilla $\{i, j, k\} \in V_i^{(3)}$, kun $\mathbb{P}(k_i(G_n) = k') > 0$, joten satunnaisverkon solmun i klusterointikertoimen odotusarvo saadaan nyt muotoon

$$\mathbb{E}[C_i(G_n)] = \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \frac{1}{\binom{k'}{2}} \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1, k_i(G_n) = k').$$

□

Lopuksi määritellään symmetrisesti jakautuneelle satunnaisverkolle G_n klusterointikerroin $c_{V'}(G_n)$, joka vastaa yleisen verkon $G = (V, E)$ klusterointikerrointa $C_{V'}(G)$. Yleisessä verkossa $C_{V'}(G)$ laskee joukon $V' = \{i \in V : k_i \geq 2\}$ solmujen klusterointikertoimien C_i keskiarvon. Satunnaisverkossa klusterointikerroin $c_{V'}(G_n)$ laskee solmun i klusterointikertoimen $C_i(G_n)$ odotusarvon ehdolla, että solmun i aste on $k_i \geq 2$.

MÄÄRITELMÄ 3.36. Olkoon G_n symmetrisesti jakautunut satunnaisverkko, jossa solmujen klusterointikertoimet $(C_i(G_n))_{i \in V_n}$ ovat samoin jakautuneita satunnaismuuttujia. Määritellään satunnaisverkon G_n klusterointikerroin laskettuna joukon $V' = \{i \in V_n : k_i(G_n) \geq 2\}$ solmulle $i \in V'$,

$$(3.18) \quad c_{V'}(G_n) = \mathbb{E}[C_i(G_n) \mid k_i(G_n) \geq 2],$$

joka ei riipu solmun $i \in V'$ valinnasta.

Seuraavaksi esitellään satunnaisverkon G_n klusterointikertoimien $c(G_n)$ ja $c_{V'}(G_n)$ yhteys. Lauseen 2.15 mukaan yleiselle verkolle $G = (V, E)$ pätee $C(G) = C_{V'}(G) |V'| / |V|$. Satunnaisverkossa G_n solmujoukon $V'(G_n) = \{i \in V_n : k_i(G_n) \geq 2\}$ koko $|V'(G_n)|$ on satunnaismuuttuja, jolloin $|V'(G_n)| / |V_n| = \mathbb{P}(k_i(G_n) \geq 2)$.

LAUSE 3.37. *Olkoon G_n satunnaisverkko, jossa $\mathbb{P}(k_i(G_n) \geq 2) > 0$. Tällöin*

$$c(G_n) = \mathbb{P}(k_i(G_n) \geq 2)c_{V'}(G_n).$$

TODISTUS. Olkoon G_n satunnaisverkko, jossa $\mathbb{P}(k_i(G_n) \geq 2) > 0$. Määritelmien 3.30 ja 3.36 mukaan $c(G_n) = \mathbb{E}[C_i(G_n)]$ ja $c_{V'}(G_n) = \mathbb{E}[C_i(G_n) \mid k_i(G_n) \geq 2]$. Määritelmistä 1.6 ja 3.18 sekä seurauksesta 3.33 saadaan laskettua

$$\begin{aligned} \mathbb{E}[C_i(G_n)] &= \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k') \\ &= \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \sum_{a=0}^{\infty} a \mathbb{P}(C_i(G_n) = a \mid k_i(G_n) = k') \mathbb{P}(k_i(G_n) = k') \\ &= \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \sum_{a=0}^{\infty} a \mathbb{P}(C_i(G_n) = a, k_i(G_n) = k') \\ &= \sum_{a=0}^{\infty} a \mathbb{P}\left(\bigcup_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} C_i(G_n) = a, k_i(G_n) = k'\right) \\ &= \sum_{a=0}^{\infty} a \mathbb{P}(C_i(G_n) = a, k_i(G_n) \geq 2) \\ &= \sum_{a=0}^{\infty} a \mathbb{P}(C_i(G_n) = a \mid k_i(G_n) \geq 2) \mathbb{P}(k_i(G_n) \geq 2) \\ &= \mathbb{E}[C_i(G_n) \mid k_i(G_n) \geq 2] \mathbb{P}(k_i(G_n) \geq 2) \\ &= c_{V'}(G_n) \mathbb{P}(k_i(G_n) \geq 2). \end{aligned}$$

□

3.4. Satunnaisverkon transitiivisuuserroin

Yleisen verkon $G = (V, E)$ transitiivisuuserroin on määritelmän 2.16 mukaan

$$C'(G) = \frac{\frac{1}{n} \sum_{i=1}^n |\Delta_i|}{\frac{1}{n} \sum_{i=1}^n |\Upsilon_i|}, \text{ kun } |\Upsilon_i| > 0 \text{ jollakin } i \in V.$$

Satunnaisverkossa G_n solmun i kolmioiden ja kytkettyjen kolmikoiden lukumäärät $|\Delta_i(G_n)|$ ja $|\Upsilon_i(G_n)|$ ovat satunnaismuuttujia. Samoin summat $\frac{1}{n} \sum_{i=1}^n |\Delta_i(G_n)|$ ja $\frac{1}{n} \sum_{i=1}^n |\Upsilon_i(G_n)|$ ovat myös satunnaismuuttujia.

Seuraavaksi määritellään transitiivisuuserroin symmetrisesti jakautuneelle satunnaisverkolle G_n , jossa kokoelmat $(|\Upsilon_i(G_n)|)_{i \in V_n}$ ja $(|\Delta_i(G_n)|)_{i \in V_n}$ koostuvat samoin jakautuneista satunnaismuuttujista.

MÄÄRITELMÄ 3.38. Olkoon G_n symmetrisesti jakautunut satunnaisverkko, jossa kokoelmat $(|\Upsilon_i(G_n)|)_{i \in V_n}$ ja $(|\Delta_i(G_n)|)_{i \in V_n}$ koostuvat samoin jakautuneista satunnaismuuttujista. Määritellään satunnaisverkon G_n transitiivisuuserroin, kun $\mathbb{E}[|\Upsilon_i(G_n)|] > 0$,

$$(3.19) \quad c'(G_n) = \frac{\mathbb{E}[|\Delta_i(G_n)|]}{\mathbb{E}[|\Upsilon_i(G_n)|]},$$

joka ei riipu solmun $i \in V_n$ valinnasta.

Remco van der Hofstad [18] määrittelee satunnaisverkon G_n transitiivisuuskertoimen suhdelukuna $3\mathbb{E} [|\Delta(G_n)|] / \mathbb{E} [|\Upsilon(G_n)|]$. Todistetaan, että tämä on sama kuin määritelmän 3.38 transitiivisuuskerroin.

LAUSE 3.39. *Olkoon G_n symmetrisesti jakautunut satunnaisverkko, jossa kokoelmat $(|\Upsilon_i(G_n)|)_{i \in V_n}$ ja $(|\Delta_i(G_n)|)_{i \in V_n}$ koostuvat samoin jakautuneista satunnaismuuttujista. Satunnaisverkon G_n transitiivisuuskerroin on tällöin*

$$(3.20) \quad c'(G_n) = \frac{3\mathbb{E} [|\Delta(G_n)|]}{\mathbb{E} [|\Upsilon(G_n)|]},$$

kun $\mathbb{E} [|\Upsilon(G_n)|] > 0$.

TODISTUS. Lauseen 2.3 mukaan $|\Upsilon(G_n)| = \sum_{i=1}^n |\Upsilon_i(G_n)|$. Samoin jakautuneille satunnaismuuttujille $(|\Upsilon_i(G_n)|)_{i \in V_n}$ saadaan lauseesta 3.7 laskettua

$$n\mathbb{E} [|\Upsilon_i(G_n)|] = \sum_{i=1}^n \mathbb{E} [|\Upsilon_i(G_n)|] = \mathbb{E} \left[\sum_{i=1}^n |\Upsilon_i(G_n)| \right] = \mathbb{E} [|\Upsilon(G_n)|].$$

Lauseen 2.8 mukaan $\sum_{i=1}^n |\Delta_i(G_n)| = 3|\Delta(G_n)|$. Samoin jakautuneille satunnaismuuttujille $(|\Delta_i(G_n)|)_{i \in V_n}$ saadaan lauseesta 3.7 laskettua

$$n\mathbb{E} [|\Delta_i(G_n)|] = \sum_{i=1}^n \mathbb{E} [|\Delta_i(G_n)|] = \mathbb{E} \left[\sum_{i=1}^n |\Delta_i(G_n)| \right] = 3\mathbb{E} [|\Delta(G_n)|],$$

joten tapauksessa $\mathbb{E} [|\Upsilon_i(G_n)|] > 0$ saadaan määritelmästä 3.38 laskettua

$$c'(G_n) = \frac{\mathbb{E} [|\Delta_i(G_n)|]}{\mathbb{E} [|\Upsilon_i(G_n)|]} = \frac{n\mathbb{E} [|\Delta_i(G_n)|]}{n\mathbb{E} [|\Upsilon_i(G_n)|]} = \frac{3\mathbb{E} [|\Delta(G_n)|]}{\mathbb{E} [|\Upsilon(G_n)|]}.$$

□

3.5. Klusterointi- ja transitiivisuuskertoimen yhteys satunnaisverkossa

Seuraavaksi esitetään tilanne, jossa satunnaisverkon G_n klusterointikertoimet $c(G_n)$ ja $c_{V'}(G_n)$ antavat saman arvon kuin transitiivisuuskerroin $c'(G_n)$.

LAUSE 3.40. *Olkoon G_n symmetrisesti jakautunut satunnaisverkko, jossa solmujen klusterointikertoimet $(C_i(G_n))_{i \in V_n}$ ovat samoin jakautuneita satunnaismuuttujia ja $\mathbb{P}(k_i(G_n) \geq 2) > 0$.*

- (1) *Jos satunnaisverkon solmujen asteet ovat samoja, niin $c(G_n) = c_{V'}(G_n) = c'(G_n)$.*
- (2) *Jos satunnaisverkon solmun i klusterointikertoimen ehdolliselle odotusarvolle pätee $\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = a \in [0, 1]$ kaikilla $k' \geq 2$, jolle $\mathbb{P}(k_i(G_n) = k') > 0$, niin $c_{V'}(G_n) = c'(G_n)$ riippumatta solmun $i \in V_n$ valinnasta.*

TODISTUS. (1): Olkoon satunnaisverkon G_n solmujen asteet $k_i(G_n) = a \geq 2$ kaikilla $i \in V_n$. Tällöin $\mathbb{P}(k_i(G_n) = a) = 1$ ja $\mathbb{P}(k_i(G_n) \neq a) = 0$, jolloin $\mathbb{P}(k_i(G_n) \geq 2) = 1$. Lauseesta 3.37 saadaan nyt laskettua $c(G_n) = \mathbb{P}(k_i(G_n) \geq 2)c_{V'}(G_n) = c_{V'}(G_n)$. Oletuksesta seuraa lauseen 2.3 mukaan $|\Upsilon_i(G_n)| = \binom{k_i(G_n)}{2} = \binom{a}{2} \geq 1$ kaikilla $i \in V_n$,

jolloin $\mathbb{E}[|\Upsilon_i(G_n)|] = \mathbb{E}\left[\binom{a}{2}\right] = \binom{a}{2}$. Kaavasta (3.15) ja odotusarvon lineaarisuudesta lauseesta 3.7 saadaan nyt laskettua

$$c(G_n) = \mathbb{E}[C_i(G_n)] = \mathbb{E}\left[\frac{|\Delta_i(G_n)|}{|\Upsilon_i(G_n)|}\right] = \frac{\mathbb{E}[|\Delta_i(G_n)|]}{\binom{a}{2}} = \frac{\mathbb{E}[|\Delta_i(G_n)|]}{\mathbb{E}[|\Upsilon_i(G_n)|]} = c'(G_n).$$

(2): Olkoon $\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = b \in [0, 1]$ kaikilla $k' \geq 2$, jolle $\mathbb{P}(k_i(G_n) = k') > 0$. Määritelmästä 3.18 saadaan laskettua $c_{V'}(G_n) = \mathbb{E}[C_i(G_n) \mid k_i(G_n) \geq 2] = b$. Kaavan (3.15) mukaan $|\Delta_i(G_n)| = |\Upsilon_i(G_n)|C_i(G_n)$, kun $k_i(G_n) \geq 2$ ja nolla muulloin. Lauseesta 3.21 saadaan laskettua kaikille $k' \geq 2$, jolle $\mathbb{P}(k_i(G_n) = k') > 0$, $\mathbb{E}[|\Delta_i(G_n)| \mid k_i(G_n) = k'] = \binom{k'}{2}\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = b\binom{k'}{2}$, jolloin seurauksen 3.33 ja lauseen 3.5 mukaan

$$\begin{aligned} \mathbb{E}[|\Delta_i(G_n)|] &= \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \mathbb{E}[|\Delta_i(G_n)| \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k') \\ &= b \sum_{k' \geq 2: \mathbb{P}(k_i(G_n)=k') > 0} \binom{k'}{2} \mathbb{P}(k_i(G_n) = k') = b\mathbb{E}[|\Upsilon_i(G_n)|]. \end{aligned}$$

Määritelmästä 3.38 saadaan nyt laskettua $c'(G_n) = \mathbb{E}[|\Delta_i(G_n)|] / \mathbb{E}[|\Upsilon_i(G_n)|] = b$. \square

ESIMERKKI 3.41. Olkoon $G_{n,p}$ satunnaisverkko, missä $\mathbb{P}(X_{ij} = 1) = p > 0$ kaikilla $\{i, j\} \in V_n^{(2)}$ riippumattomasti toisistaan. Seurauksista 3.24 ja 3.29 saadaan laskettua $\mathbb{E}[|\Upsilon(G_{n,p})|] = n(n-1)(n-2)/2 \cdot p^2$ ja $\mathbb{E}[|\Delta(G_{n,p})|] = n(n-1)(n-2)/6 \cdot p^3$, jolloin satunnaisverkon $G_{n,p}$ transitiivisuuskertoimen on lauseen 3.39 mukaan

$$c'(G_{n,p}) = 3\mathbb{E}[|\Delta(G_{n,p})|] / \mathbb{E}[|\Upsilon(G_{n,p})|] = p.$$

Satunnaisverkon $G_{n,p}$ klusterointikertoimen $c(G_{n,p})$ laskemista varten lasketaan ensin $\mathbb{E}[C_i(G_{n,p}) \mid k_i(G_n) = k']$. Lemman 3.32 mukaan $\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] = 0$, kun $k' \in \{0, 1\}$. Kaavasta (3.17) saadaan laskettua $\mathbb{E}[C_i(G_{n,p}) \mid k_i(G_{n,p}) = k'] = \mathbb{E}[|\Delta_i(G_{n,p})| \mid k_i(G_{n,p}) = k'] / \binom{k'}{2}$ kaikilla $k' \in \{2, 3, \dots, n-1\}$. Satunnaismuuttuja $\mathbb{E}[|\Delta_i(G_{n,p})| \mid k_i(G_{n,p})]$ noudattaa binomijakaumaa $\text{Bin}\left(\binom{k_i}{2}, p\right)$, joten saadaan laskettua $\mathbb{E}[C_i(G_{n,p}) \mid k_i(G_{n,p}) = k'] = \binom{k'}{2}p / \binom{k'}{2} = p$ kaikilla $k' \in \{2, 3, \dots, n-1\}$. Seurauksesta 3.33 saadaan laskettua satunnaisverkon $G_{n,p}$ klusterointikerroin

$$c(G_{n,p}) = \mathbb{E}[C_i(G_{n,p})] = p \sum_{k'=2}^{n-1} \mathbb{P}(k_i(G_{n,p}) = k') = p\mathbb{P}(k_i(G_{n,p}) \geq 2).$$

Lauseesta 3.37 saadaan laskettua satunnaisverkon $G_{n,p}$ klusterointikerroin

$$c_{V'}(G_{n,p}) = \mathbb{E}[C_i(G_{n,p}) \mid k_i(G_{n,p}) \geq 2] = c(G_{n,p}) / \mathbb{P}(k_i(G_{n,p}) \geq 2) = p.$$

Satunnaisverkon klusterointikertoimen laskuissa esiintyy virheitä tieteellisissä julkaisuissa. Tietojenkäsittelytieteilijät Matthias Grossglauser ja Patrick Thiran [17] laskevat satunnaisverkon $G_{n,p}$ solmun i klusterointikertoimen odotusarvoksi $\mathbb{E}[C_i(G_{n,p})] = p$, vaikka tämän oikea arvo on $p \cdot \mathbb{P}(k_i(G_{n,p}) \geq 2)$. Fyysikko Mark E. J. Newman [14] väittää, että satunnaisverkon $G_{n,M}$ klusterointikerroin on $c(G_{n,M}) = \bar{k}/n$, missä $\bar{k} = 2M/n$ on solmujen asteiden keskiarvo. Satunnaisverkot $G_{n,M}$ ja $G_{n,p}$ ovat kuitenkin keskenään lähes vaihtokelpoiset, kun $M \sim p\binom{n}{2}$, eli $\bar{k}/(n-1) \sim p$ [12]. Newman

laskee parametreilla $n = 4941$ ja $\bar{k} = 2M/n = 2,67$ klusterointikertoimen arvoksi $c(G_{n,M}) = \bar{k}/n$, vaikka $c(G_{n,M}) \approx c(G_{n,p}) = \mathbb{P}(k_i \geq 2) \bar{k}/(n-1) \approx 0,746 \cdot \bar{k}/(n-1)$.

Transitiivisuuskerroin on yleisemmin käytetty satunnaisverkon analyysissä. Määritelmän 2.17 transitiivisuuskerroin yleiselle verkolle $C'(G) = 3|\Delta(G)|/|\Upsilon(G)|$ ei sovi satunnaisverkon analyysiin, mutta kolmioiden ja kytkettyjen kolmikoiden odotusarvot $\mathbb{E}[|\Delta(G_n)|]$ ja $\mathbb{E}[|\Upsilon(G_n)|]$ on analysoitu aikaisemminkin [6]. Tieteellisissä julkaisuissa on kuitenkin epäselvää, miten transitiivisuuskerroin määritellään satunnaisverkolle. Matemaatikot Béla Bollobás ja Oliver M. Riordan [12] määrittävät satunnaisverkolle G_n *transitiivisuuskerroimen odotusarvon* $\mathbb{E}[C'(G_n)] = \mathbb{E}[3|\Delta(G_n)|/|\Upsilon(G_n)|]$, mutta laskevat $\mathbb{E}[C'(G_n)] \sim 3\mathbb{E}[|\Delta(G_n)|]/\mathbb{E}[|\Upsilon(G_n)|]$. Lauseessa 3.39 on todistettu transitiivisuuskerroimen olevan $c'(G_n) = 3\mathbb{E}[|\Delta(G_n)|]/\mathbb{E}[|\Upsilon(G_n)|]$.

Empiirisiä tutkimuksia sosiaalisista verkostoista

4.1. Poolin ja Kochenin tutkimus ja verkkomalli

Valtiotieteilijä Ithiel de Sola Pool ja matemaatikko Manfred Kochen [8] alkoivat 1950-luvulla tutkia Yhdysvaltain väestön sosiaalisen verkoston rakennetta. He pyysivät 27 koehenkilöä pitämään listaa heidän sosiaalisesta kontaktiverkostosta 100 päivän ajan. Koehenkilöiden kontaktien lukumäärä oli $k_{\text{koehenkilö}} \in [72, 1043]$. Sen jälkeen he kysyivät koehenkilön kontaktilistassa olevilta henkilöiltä tunsivatko he muut henkilöt kontaktilistalta. Viidelle koehenkilölle he saivat laskettua klusterointikertoimet, joissa $C_{\text{koehenkilö}} = |\Delta_{\text{koehenkilö}}|/|\Upsilon_{\text{koehenkilö}}| \in [0.08, 0.36]$.

Pool ja Kochen pyrkivät seuraavaksi satunnaisverkkomallin avulla laskemaan Yhdysvaltain väestön sosiaalisen verkoston karakteristisen polun pituuden. He päättivät yksinkertaistaa satunnaisverkkomallia asettamalla verkon kaikille solmuille saman asteen r , josta seurasi, että se on *satunnainen r -säännöllinen verkkomalli*.

MÄÄRITELMÄ 4.1. Yleinen verkko $G = (V, E)$ on *r -säännöllinen* [5], jos verkon jokaisen solmun aste on r .

MÄÄRITELMÄ 4.2. Olkoon $3 \leq r < n$, jossa rn on parillinen kokonaisluku. Merkitään

$$\mathcal{G}_{n,r\text{-reg}} = \{g \in \mathcal{G}^n : k_i(g) = r \text{ kaikilla } i \in V_n\}$$

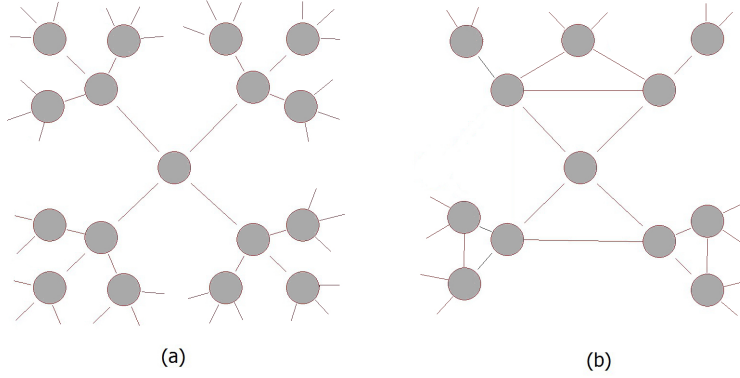
kaikkien n -kokoisten r -säännöllisten verkkojen joukkoa. Asettamalla joukon $\mathcal{G}_{n,r\text{-reg}}$ jokaiselle alkionle sama todennäköisyys, $P_{n,r\text{-reg}}(g) = 1/|\mathcal{G}_{n,r\text{-reg}}|$ kaikilla $g \in \mathcal{G}_{n,r\text{-reg}}$ saadaan todennäköisyysavaruus $(\mathcal{G}_{n,r\text{-reg}}, \mathcal{P}(\mathcal{G}_{n,r\text{-reg}}), \mathbb{P}_{n,r\text{-reg}})$, josta käytetään nimitystä *satunnainen r -säännöllinen verkkomalli* [6]. Satunnaisverkkomallin $\mathcal{G}_{n,r\text{-reg}}$ satunnaisverkko on tasaisesti jakautunut *satunnainen r -säännöllinen verkko*, $G_{n,r\text{-reg}}$.

Lasketaan satunnaisverkon $G_{n,r\text{-reg}}$ transitiivisuuserroin. Lauseesta 2.3 saadaan laskettua satunnaisverkon $G_{n,r\text{-reg}}$ kytkettyjen kolmikoiden kokonaismäärä $|\Upsilon(G_{n,r\text{-reg}})| = \sum_{i=1}^n |\Upsilon_i(G_{n,r\text{-reg}})| = \sum_{i=1}^n \binom{k_i(G_{n,r\text{-reg}})}{2} = n \binom{r}{2} = nr(r-1)/2$, jolloin $\mathbb{E}[|\Upsilon(G_n)|] = nr(r-1)/2$. Béla Bollobás [6] osoittaa kirjassaan, että satunnaisen r -säännöllisen verkon kolmioiden lukumäärä $|\Delta(G_{n,r\text{-reg}})|$ on *Poisson-jakautunut* [2] parametrilla $\lambda = (r-1)^3/6$, kun $n \rightarrow \infty$, jolloin $\mathbb{E}[|\Delta(G_{n,r\text{-reg}})|] = (r-1)^3/6$. Satunnaisverkon $G_{n,r\text{-reg}}$ klusterointi- ja transitiivisuuserroin on tällöin lauseiden 3.39 ja 3.40 mukaan

$$c(G_{n,r\text{-reg}}) = c'(G_{n,r\text{-reg}}) = \frac{3\mathbb{E}[|\Delta(G_{n,r\text{-reg}})|]}{\mathbb{E}[|\Upsilon(G_{n,r\text{-reg}})|]} = \frac{\frac{3}{6}(r-1)^3}{\frac{1}{2}nr(r-1)} = \frac{(r-1)^2}{nr} \sim \frac{r-1}{n},$$

kun satunnaisverkon koko n on tarpeeksi iso.

Tutkitaan seuraavaksi satunnaisverkon $G_{n,r\text{-reg}}$ karakteristisen polun pituutta, jolle lauseen 1.33 mukaan pätee $L(G_{n,r\text{-reg}}) \leq \text{diam}(G_{n,r\text{-reg}})$. Béla Bollobás [6] määrittää satunnaisverkon $G_{n,r\text{-reg}}$ halkaisijalle ylärajan d siten, että $\mathbb{P}(\text{diam}(G_{n,r\text{-reg}}) \leq d) \rightarrow 1$,



KUVA 4.1. (a): Satunnaisverkon $G_{n,r\text{-reg}}$ tyypillisen solmun naapurusto, kun $r = 4$ ja $n = 10^8$, jolloin verkon klusterointikerroin on $c(G_{n,r\text{-reg}}) = \mathbb{E}[C_i(G_{n,r\text{-reg}})] \sim (r-1)/n \approx 0$. Verkon tyypilliselle solmulle i pätee $\mathbb{E}[|\{j \in V_n : d_{ij} = 2\}|] \approx 12 = r(r-1)$. (b): Hypoteettinen 4-säännöllinen verkko $g \in \mathcal{G}_{n,r\text{-reg}}$, $n = 10^8$, jonka klusterointikerroin on $c(g) = \mathbb{E}[C_i(g)] = 1/3$. Verkon tyypilliselle solmulle i pätee $\mathbb{E}[|\{j \in V_n : d_{ij} = 2\}|] \leq 8 < r(r-1)$.

kun $n \rightarrow \infty$. Sanotaan tällöin, että *melkein kaikilla* satunnaisverkkomallin $\mathcal{G}_{n,r\text{-reg}}$ verkoilla on halkaisija, joka on korkeintaan d .

LAUSE 4.3. *Olkoot $r \geq 3$ ja $\epsilon > 0$ vakioita ja $d = d(n)$ pienin kokonaisluku, joka toteuttaa*

$$(r-1)^{d-1} \geq (2+\epsilon)rn \log n,$$

Tällöin melkein kaikilla satunnaisen r -säännöllinen verkkomallin verkoilla on halkaisija, joka on korkeintaan d .

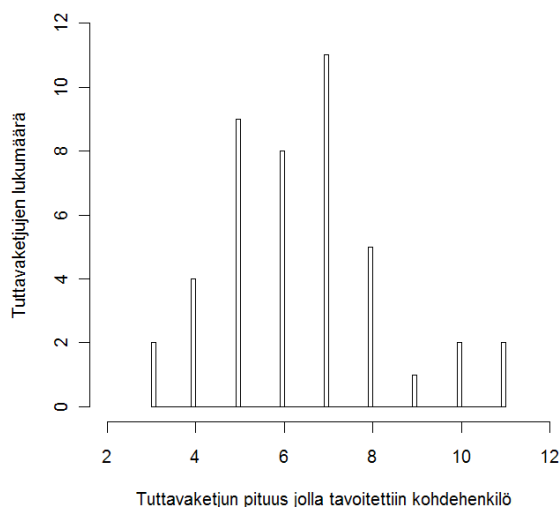
TODISTUS. Katso [6]. □

ESIMERKKI 4.4. Olkoon $\mathcal{G}_{n,r\text{-reg}}$ satunnaisverkkomalli, jossa $n = 2 \cdot 10^8$ on samaa suuruusluokkaa, kuin Yhdysvaltain väestö 1950- ja 1960-luvulla [22]. Olkoon väestön henkilöllä keskimäärin $r = 1000$ kontaktia. Ratkaistaan pienin kokonaisluku d , joka toteuttaa lauseen 4.3 kaavan $(r-1)^{d-1} \geq (2+\epsilon)rn \log n$, missä r ja $\epsilon > 0$ pysyvät vakioina. Saadaan

$$\begin{aligned} 999^{d-1} &\geq (2+\epsilon)10^3(2 \cdot 10^8) \log(2 \cdot 10^8) \\ \Leftrightarrow (d-1) \log 999 &\geq \log(2(2+\epsilon)) + \log 10^{11} + \log(\log(2 \cdot 10^8)) \\ \Leftrightarrow (d-1) &\geq \frac{\log(4+2\epsilon) + 11 + \log(\log 2 + 8)}{\log 999} \approx \frac{\log(4+2\epsilon)}{\log 999} + 3,97. \end{aligned}$$

Lauseen 4.3 mukaan melkein kaikilla satunnaisverkkomallin $\mathcal{G}_{n,r\text{-reg}}$ verkoilla on halkaisija, joka on korkeintaan $d = 6$, kun $\epsilon \leq 497,5$, jolloin sen melkein kaikkien verkkojen karakteristisen polun pituus on lauseen 1.33 mukaan korkeintaan $L \leq 6$.

Satunnaisverkkomalli $\mathcal{G}_{n,r\text{-reg}}$ osoittaa, että isolla harvalla verkostolla voi teoriassa ole hyvin pieni karakteristisen polun pituus, jos kontaktit ovat täysin satunnaisia. Se ei kuitenkaan pysty mallintamaan ison sosiaalisen verkoston klusteroitunutta rakennetta, jossa kontaktit eivät ole täysin satunnaisia.



KUVA 4.2. Stanley Milgramin [9] empiirisen kokeen tuloksessa nähdään tuttavaketjujen pituudet, jotka tarvittiin yhdistämään 44 satunnaisesti valittua koehenkilöä Nebraskasta tiettyyn valittuun koehenkilöön Bostonissa. Tuttavaketjujen pituuksien keskiarvo oli 6,43.

Pool ja Kochen tekevät heuristisen arvioinnin, miten isompi klusterointikerroin vaikuttaisi satunnaisverkkomallilla $\mathcal{G}_{n,r\text{-reg}}$ laskettuun karakteristisen polun pituuteen. Merkitään verkon solmujen joukkoa, joiden etäisyys solmuun i on m , $\Gamma_i(m) = \{j \in V_n : d_{ij} = m\}$. Ison harvan satunnaisverkon $G_{n,r\text{-reg}}$ klusterointikerroin $c(G_{n,r\text{-reg}}) \sim (r-1)/n \approx 0$, joten Pool ja Kochen päättelivät heuristisin perusteluin, että tyypillisellä solmulla i on noin $|\Gamma_i(m)| \approx r(r-1)^{m-1}$ solmua, joiden etäisyys solmuun i on m (ks. kuva 4.1(a)). Kasvattaessa hypoteettisen verkon $g \in \mathcal{G}_{n,r\text{-reg}}$ klusterointikerrointa, verkon g tyypilliselle solmulle i pätee $|\Gamma_i(m)| < r(r-1)^{m-1}$ kaikilla $m \geq 2$ (ks. kuva 4.1(b)). Tämän seurauksena Pool ja Kochen olettavat, että verkon karakteristisen polun pituus kasvaa verkon klusterointikerroimen kasvaessa, eli C kasvaa $\Rightarrow L$ kasvaa.

4.2. Milgramin empiirinen koe

Yhteiskuntapsykologi Stanley Milgram [9] julkaisi vuonna 1967 tutkimuksen Yhdysvaltain väestön sosiaalisen verkoston karakteristisen polun pituudesta, jonka lähtökohtana oli Poolin ja Kochenin [8] tutkimus. Milgram päätti matemaattisen mallintamisen sijaan tehdä empiirisen kokeen, jolla hän pyrki määrittämään kahden satunnaisesti valitun henkilön lyhimmän tuttavaketjun pituuden Yhdysvaltojen väestössä. Kokeen suorittamiseen hänen ei tarvinnut tietää mitään henkilöiden kontaktien lukumäärästä tai verkon klusteroituneisuudesta. Milgram jakoi 160 kirjettä satunnaisesti valituille henkilöille Nebraskasta ja käski heidän toimittaa kirjeet valitulle koehenkilölle Bostonissa sillä säännöllä, että kirjeen saisi ainoastaan antaa eteenpäin henkilölle, jonka lähettäjä tunsi etunimellä. Tarkoituksena oli, että kirjeet päätyisivät jossain vaiheessa valitulle koehenkilölle Bostonissa kuljettuaan lähettäjän ja koehenkilön välistä tuttavaketjua pitkin. Näistä 160 kirjeestä 44 löysi perille. Milgram laski, että kirjeet olivat kulkeneet keskimäärin ainoastaan noin viiden välikäden kautta ennen kuin ne saapuivat perille. Tuttavaketjujen pituuksien keskiarvo oli 6,43 (ks.

kuva 4.2), joka tuli yllätyksenä monille tutkijoille. Hän oletti, että sama tulos pätee, jos valittaisiin kaksi satunnaista henkilöä Yhdysvaltojen 200 miljoonan kokoisesta väestöstä, jolloin Yhdysvaltojen väestön sosiaalisen verkoston karakteristisen polun pituus olisi ainoastaan $L_{\text{Yhdysvaltain väestö}} \approx 6,43$.

4.3. Wattsin ja Strogatzin tilastollinen tutkimus

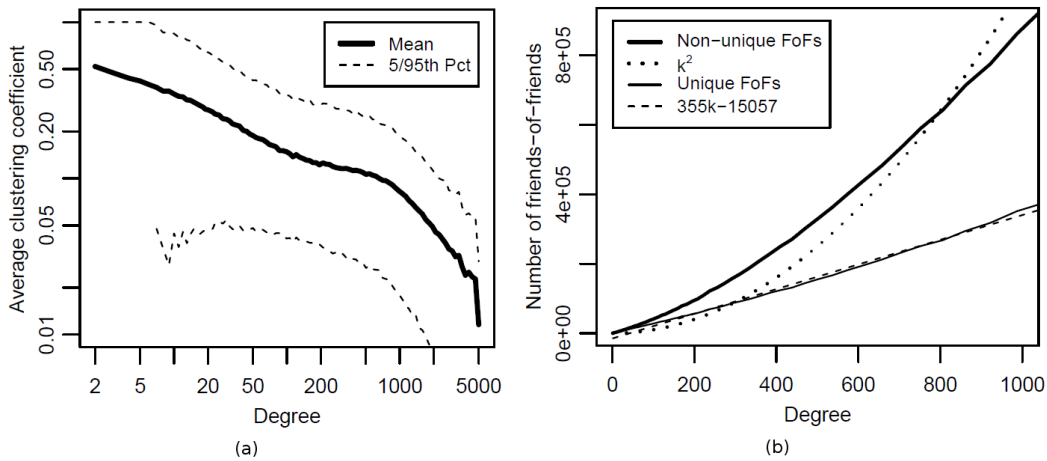
Vuonna 1998 Duncan J. Watts ja Steven H. Strogatz julkaisivat tutkimuksen [10] verkostoista, joiden täydelliset kytkentäkaaviot oli tallennettu tietokantoihin. Tietokantojen datasta he saivat tilastollisesti laskettua verkoston klusterointikerroimen C ja karakteristisen polun pituuden L .

Eräs heidän tutkittavista verkostoistaan, jonka koko kytkentärakenne oli täysin tiedossa, oli Internet Movie Database:n (sivustolta <http://us.imdb.com>) elokuvanäyttelijöiden yhteistyöverkosto, joka edustaa tietynlaista sosiaalista verkostoa. Watts ja Strogatz rajoittivat verkoston tutkimisen huomautuksen 1.39(ii) tapaan ainoastaan verkoston isoimpaan yhtenäiseen komponenttiin, joka muodosti huhtikuussa 1997 noin 90% yhteistyöverkoston näyttelijöistä. He laskivat, että elokuvanäyttelijöiden yhteistyöverkosto oli hyvin klusteroitunut samalla, kun verkoston karakteristisen polun pituus oli vain vähän pitempi kuin saman kokoisessa satunnaisverkossa (ks. taulukko 1). Watts ja Strogatz kutsuvat tällaisia klusteroituneita verkkoja, joiden karakteristisen polun pituus on lyhyt, *pieni maailma -verkoiksi* [10]. He osoittivat lisäksi, että pieni maailma -verkot eivät ainoastaan esiinny ihmisten sosiaalisissa verkostoissa. Heidän kaksi muuta tutkittavaa verkostoa, joiden täydelliset kytkentäkaaviot oli tallennettu, olivat *Caenorhabditis elegans* -madon hermoverkosto ja läntisen Yhdysvaltain sähköverkosto (ks. taulukko 1). Pieni maailma -verkkorakenne näyttää siis esiintyvän myös sekä biologisissa että ihmisten rakentamissa verkostoissa.

TAULUKKO 1. Eri verkostoille tilastollisesti lasketut klusterointikerroimet C ja $C_{V'}$, transitiivisuuskerroin C' sekä karakteristisen polun pituus L . Verkostojen täydelliset kytkentäkaaviot oli tallennettu tietokantoihin.

Verkosto	n	\bar{k}	L	L_{sat}	C	$C_{V'}$	C'	C'_{sat}
Sähköverkosto [10]	4941	2,67	18,7	12,4	0,080	-	-	0,0005
Hermoverkosto [10]	282	14	2,65	2,25	0,28	-	-	0,05
Näyttelijöitä [10]	225226	61	3,65	2,99	0,79	-	-	0,00027
Näyttelijöitä [16]	382000	78,5	-	-	0,780	0,785	0,012	0,00021*
Näyttelijöitä [15]	449913	113,43	3,48	-	0,78	-	0,20	0,00025*
Internet [15]	10697	5,98	3,31	-	0,39	-	0,035	0,00056*
Internet [16]	13164	4,33	-	-	0,311	0,458	0,012	0,00033*

Symbolien L_{sat} ja C'_{sat} arvot vastaavat satunnaisverkolle $G_{n,M}$ (tai $G_{n,p}$) laskettua karakteristisen polun pituutta $L(G_{n,M})$ ja transitiivisuuskerrointa $c'(G_{n,M})$, kun $M = \bar{k}n/2$ (tai $p = \bar{k}/(n-1)$). Tähdellä (*) merkityt arvot on itse lisättyjä arvoja.



KUVA 4.3. Vuoden 2011 Facebookin sosiaalisen verkoston [19] (a): k -asteiden solmujen keskimääräinen klusterointikerroin asteen k funktiona, $C^{(k)} = \frac{1}{n_k} \sum_{i:k_i=k} C_i$, missä $n_k = |\{i \in V_n : k_i = k\}|$. (b): Solmun i naapureiden naapurien keskimääräinen lukumäärä asteen k_i funktiona, missä solmun i ”ei-uniikkien” naapureiden lukumäärä on $\sum_{j \in N_i} (k_j - 1)$ ja ”uniikkien” lukumäärä on $|\{j \in V : d_{ij} = 2\}|$.

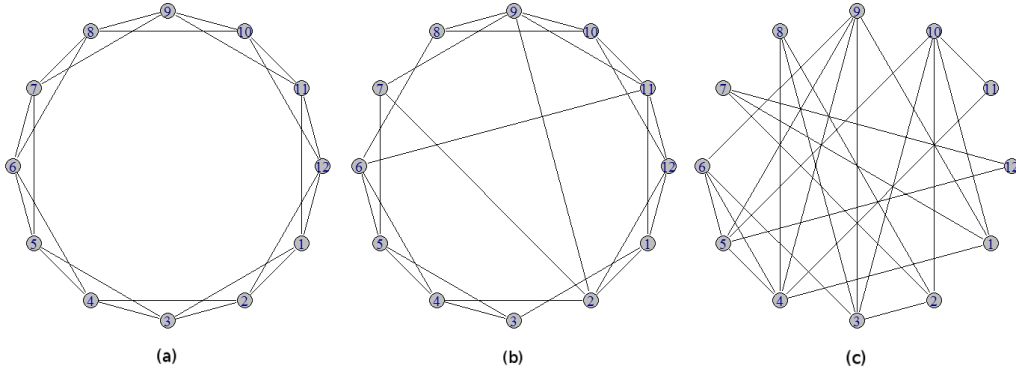
4.4. Facebookin sosiaalisen verkoston tutkimus

Vuonna 2011 Facebookin Johan Ugander, Brian Karrer, Lars Backstrom ja Cameron Marlow [19] tutkivat julkaisussaan Facebookin sosiaalista verkostoa, joka toukokuussa 2011 koostui maailmanlaajuisesti noin 721 miljoonasta aktiivisesta käyttäjästä. He tutkivat myös Facebookin sosiaalisen verkoston aliverkostoa, joka koostui noin 149 miljoonasta Yhdysvaltain väestön aktiivisesta käyttäjästä ja noin 15,9 miljardista ystävyyslinkeistä, jolloin aliverkon tiheys oli $D \approx 1,43 \cdot 10^{-6}$. Keskimäärin yhdellä Facebookin käyttäjällä Yhdysvaltain väestöstä oli $\bar{k} \approx 214$ [20] Facebook-kaveria Yhdysvaltain väestöstä ja Facebookin Yhdysvaltain väestön käyttäjäverkoston karakteristisen polun pituus oli $L_{\text{FB-Yhdysvallat}} \approx 4,3$ [19]. Facebookin globaalin käyttäjäverkoston käyttäjällä oli keskimäärin $\bar{k} \approx 190$ [20] Facebook-kaveria ja Facebookin globaalin verkoston karakteristisen polun pituus oli $L_{\text{FB-globaali}} \approx 4,7$ [19].

Verkoston klusteroituneisuuden määrittämiseksi Ugander, Karrer, Backstrom ja Marlow laskivat verkoston k -asteiden solmujen klusterointikertoimien keskiarvon $C^{(k)} = \frac{1}{n_k} \sum_{i:k_i=k} C_i$, missä k_i on solmun i aste, $n_k = |\{i \in V_n : k_i = k\}|$ ja C_i solmun i klusterointikerroin. Kuvassa 4.3(a) nähdään, että $C^{(k)}$ on iso asteesta k riippumatta, joten Facebookin sosiaalinen verkosto on pieni maailma -verkko.

Ugander, Karrer, Backstrom ja Marlow huomauttivat, että Facebookin sosiaalista verkostoa karakterisoi ilmiö, jossa käyttäjän Facebook-kavereilla oli enemmän Facebook-kavereita kuin käyttäjällä itse, kunnes käyttäjällä oli yli 700 Facebook-kaveria [19] (ks. kuva 4.3(b)). Esimerkiksi solmulle i , jonka aste oli $k_i = 100$ päti $|\{j \in V : d_{ij} = 2\}| = 27500$ [19] vaikka klusterointikerroin $C_{k=100}$ oli iso. Tämä oli enemmän kuin $k_i(k_i - 1) = 100 \cdot 99 = 9900$, joka on solmun i naapurien naapureiden lukumäärä, jos solmun i naapurien asteet ovat samat kuin solmulla i ja verkon klusterointikerroin on nolla.

Wattsin ja Strogatzin pieni maailma -verkkomalli



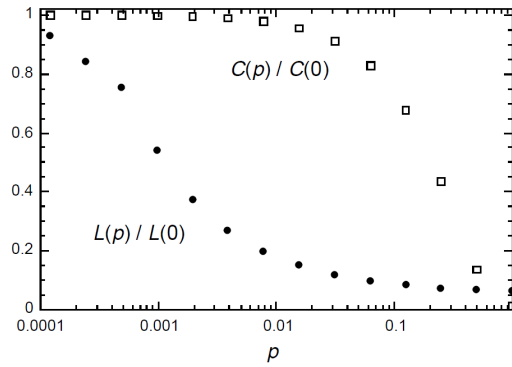
KUVA 5.1. Esimerkkejä Wattsin ja Strogatzin verkkomallin $\mathcal{G}_{n,k,p}$ generoiduista verkoista parametreilla $n = 12$, $k = 4$ ja eri parametrin p arvoilla. (a): $p = 0$, jolloin verkko on n -kokoinen k -säännöllinen rengashila. (b): $p = 0, 1$, jolloin rengashilan linkit on kytketty uudelleen todennäköisyydellä p riippumattomasti toisistaan algoritmin 1 tapaan. (c): $p = 1$.

Duncan J. Watts ja Steven H. Strogatz [10] esittävät vuoden 1998 julkaisussaan satunnaisverkkomallin $\mathcal{G}_{n,k,p}$, jolla pystytään mallintamaan pieni maailma -verkkojen rakennetta. Siitä käytetään myös nimitystä Wattsin ja Strogatzin *pieni maailma -verkkomalli* [15]. Verkkomallin verkkoja voidaan säätää säännöllisen rengashilan ja lähes satunnaisen verkon välillä kytkemällä rengashilan linkit uudelleen kytkentätodennäköisyydellä p algoritmin 1 tapaan. Kytkettäessä linkkejä uudelleen ilmaantuu satunnaisia oikoteitä solmuparien välille, jotka muuten olisivat kaukana toisistaan rengashilassa (katso kuva 5.1). Watts ja Strogatz osoittavat empiirisesti, että tietyillä parametrin p arvoilla satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ generoidulla verkolla voi olla säännöllisen rengashilan tapaan iso klusterointikerroin C samalla, kun sen karakteristisen polun pituus L on pieni, kuten satunnaisverkoissa (katso kuva 5.2).

5.1. Pieni maailma -verkkomallin kytkentämenetelmä

Wattsin ja Strogatzin satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ lähtökohtana on n -kokoinen k -säännöllinen rengashila $H_{n,k}$, missä k on parillinen (ks. kuva 5.1(a)). Seuraavaksi määritellään metriikka, joka kertoo kuinka lähellä toisiinsa rengasmuotoon asetetut solmut ovat

$$(5.1) \quad \ell_{ij} = \min\{|j - i|, n - |j - i|\}$$



KUVA 5.2. Wattsin ja Strogatzin [10] laskemat klusterointikertoimet $C(p)$ ja karakteristisen polun pituudet $L(p)$ pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ generoiduille verkoille parametrin p funktiona, missä $n = 1000$ ja $k = 10$. Tulokset ovat keskiarvoja kahdestakymmenestä generoidusta verkosta, jotka on normalisoitu täysin säännöllisen rengashilan arvoilla $C(0)$ ja $L(0)$. Karakteristisen polun pituuden $L(p)$ arvo laskee nopeasti samalla, kun klusterointikerroin $C(p)$ pysyy lähes vakiona.

kaikille $\{i, j\} \in V_n^{(2)}$. Rengashilassa $H_{n,k}$ jokainen solmu on kytketty linkillä sen k lähimpiin solmuihin renkaassa, jolloin rengashilan linkit ovat

$$(5.2) \quad E(H_{n,k}) = \{\{i, j\} \in V_n^{(2)} : 0 < \ell_{ij} \leq k/2\}.$$

Wattsin ja Strogatzin verkkomallin kytkentämenetelmässä käydään rengashilan $H_{n,k}$ jokainen linkki läpi kytkemällä linkki uudelleen todennäköisyydellä p riippumattomasti toisistaan menetelmällä, joka on kuvailtu algoritmin 1 pseudokoodissa.

Algoritmin 1 kytkentämenetelmän lähtökohtana olevaa rengashilaa $H_{n,k}$ kannattaa analysivaiheessa ajatella *suunnattuna verkkona*.

MÄÄRITELMÄ 5.1. Suunnattu verkko [5] G on pari $G = (V, E)$, missä E on joukko pareja $(i, j) \in V \times V$. Joukon E alkiota (i, j) kutsutaan *suunnatuksi linkiksi*, jossa solmu i on linkin *lähtösolmu* ja solmu j linkin *maalisolmu*. Solmun $i \in V$ *lähtöaste* [5] k_i^+ tarkoittaa niiden verkon linkkien lukumäärää, joiden lähtösolmuna on solmu i . Vastaavasti solmun $i \in V$ *tuloaste* [5] k_i^- tarkoittaa niiden linkkien lukumäärää, joiden tulosolmuna on solmu i .

Wattsin ja Strogatzin satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ lähtökohtana olevaa rengashilaa $H_{n,k}$ voidaan ajatella suunnattuna verkkona $H_{n,k/2\text{-out}} = H_{k/2\text{-out}}$, jossa jokaisesta solmusta on $k/2$ suunnattua linkkiä lähimpiin myötäpäiväisiin solmuihin. Algoritmissa 1 valitaan satunnainen solmu $i \in V_n$ ja suunnattu linkki $(i, j) \in E(H_{k/2\text{-out}})$, jolle $\ell_{ij} = 1$. Jos $k_i^+ + k_i^- < n - 1$, kytketään linkki uudelleen todennäköisyydellä p , valitsemalla linkille satunnaisesti uusi maalisolmu $j' \in V_n$. Jos verkossa on jo olemassa linkki (i, j') tai (j', i) , valitaan linkille satunnaisesti toinen maalisolmu. Samoin pätee, jos $j' = i$. Tämän jälkeen siirrytään solmun i viereiseen myötäpäivässä olevaan solmuun $i' \in V_n$. Valitaan solmun i' myötäpäiväinen linkki (i', k') , jolle $\ell_{i'k'} = 1$. Jos $k_{i'}^+ + k_{i'}^- < n - 1$, kytketään linkki uudelleen todennäköisyydellä p . Käydään tällä tavalla verkon $H_{k/2\text{-out}}$ kaikki solmut läpi myötäpäiväiseen suuntaan, kunnes yksi

Algorithm 1 Wattsin ja Strogatzin verkkomallin kytkentämenetelmä

Require: k -regular ring lattice $H_{n,k}$ (k even), rewiring probability p

```

i' ← RandomNode(1, n)
for lap ∈ {1, ..., k/2} do
  for next ∈ {0, ..., n - 1} do
    i ← i' + next
    if n < i then
      i ← i mod n
    end if
    j ← i + lap
    if n < j then
      j ← j mod n
    end if
    if RandomRealNumber(0, 1) ≤  $p$  and Degree(i) < n - 1 then
      repeat
        j' ← RandomNode(1, n)
        until not EdgeExists(i, j') and j' ≠ i
        AddEdge(i, j')
        DeleteEdge(i, j)
      end if
    end for
  end for

```

kierros on tehty. Tämän jälkeen siirrytään solmun i myötäpäiväiseen linkkiin (i, j') , jolle $\ell_{ij} = 2$. Käydään samalla tavalla verkon jokainen solmu $i' \in V_n$ läpi kytkemällä uudelleen solmun i' myötäpäiväinen linkki (i', j') , $\ell_{i'j'} = 2$, todennäköisyydellä p , jos $k_i^+ + k_i^- < n - 1$. Jatketaan samaa menetelmää siirtymällä jokaisen kierroksen jälkeen verkon $H_{k/2\text{-out}}$ myötäpäiväisiin linkkeihin, jotka kytkvät verkon yhä etäisempiä solmuja. Kun ollaan käyty myötäpäiväiset linkit $\{(i', j') \in E(H_{k/2\text{-out}}) : \ell_{i'j'} = k/2\}$ läpi, on lähtökohtaisen suunnatun verkon $H_{k/2\text{-out}}$ jokainen linkki kytketty uudelleen todennäköisyydellä p riippumattomasti toisistaan. Lopuksi muutetaan verkko suuntaamattomaksi asettamalla suunnatut linkit suuntaamattomiksi. Kyseinen suuntaamaton verkko on Wattsin ja Strogatzin satunnaisverkkomallin generoitu verkko $g \in \mathcal{G}_{n,k,p}$.

5.2. Pieni maailma -verkkomallin satunnaisverkko

MÄÄRITELMÄ 5.2. Olkoon $\mathcal{G}_{n,k,p}$ joukko algoritmin 1 kaikista mahdollisista generoiduista verkoista parametreilla n , k ja p , missä k on parillinen. Merkitään kyseistä satunnaisverkkomallia lyhyesti $\mathcal{G}_{n,k,p} = (\mathcal{G}_{n,k,p}, \mathcal{P}(\mathcal{G}_{n,k,p}), \mathbb{P}_{\mathcal{G}_{n,k,p}})$ ja sen satunnaisverkkoa $G_{n,k,p}$.

Seuraavaksi esitetään satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ ominaisuuksia, jotka pätevät satunnaisverkolle $G_{n,k,p}$ riippumattomasti parametrin p arvosta.

LAUSE 5.3. *Olkoon $\mathcal{G}_{n,k,p}$ satunnaisverkkomalli. Tällöin satunnaisverkon $G_{n,k,p}$ linkkien lukumäärä on*

$$|E(G_{n,k,p})| = kn/2,$$

missä n ja k ovat satunnaisverkon $G_{n,k,p}$ parametreja.

TODISTUS. Olkoon $p = 0$. Satunnaisverkko $G_{n,k,p=0}$ on k -säännöllinen rengashila, jonka solmujen asteiden keskiarvo on määritelmän 1.23 mukaan $\bar{k}(G_{n,k,p=0}) = k$. Seurauksen 1.27 mukaan $|E(G_{n,k,p=0})| = \bar{k}n/2 = kn/2$. Olkoon $p > 0$. Lähtökohtaisen rengashilan linkkien lukumäärä ei muutu satunnaisverkossa $G_{n,k,p}$, koska algoritmin 1 kytkentämenetelmässä linkkiä ei voida kytkeä uudelleen solmupariin, joka on jo kytketty linkillä. \square

SEURAUUS 5.4. *Olkoon $\mathcal{G}_{n,k,p}$ satunnaisverkkomalli. Satunnaisverkon $G_{n,k,p}$ parametri k määrittää solmujen asteiden keskiarvon yksikäsitteisesti*

$$\bar{k}(G_{n,k,p}) = k.$$

TODISTUS. Lauseen 5.3 mukaan $|E(G_{n,k,p})| = kn/2$, joten seurauksesta 1.26 saadaan $\bar{k}(G_{n,k,p}) = 2|E(G_{n,k,p})|/|V(G_{n,k,p})| = 2kn/(2n) = k$. \square

SEURAUUS 5.5. *Olkoon $\mathcal{G}_{n,k,p}$ satunnaisverkkomalli. Tällöin satunnaisverkon $G_{n,k,p}$ tiheys on*

$$D(G_{n,k,p}) = \frac{k}{n-1},$$

missä n ja k ovat satunnaisverkon $G_{n,k,p}$ parametreja.

TODISTUS. Seurauksesta 5.4 nähdään, että $\bar{k}(G_{n,k,p}) = k$, joten seurauksen 1.28 mukaan $D = \bar{k}/(n-1) = k/(n-1)$. \square

SEURAUUS 5.6. *Satunnaisverkkomalli $\mathcal{G}_{n,k,p}$ on harva, kun mallin parametri k on alempaa kertaluokkaa kuin parametri n , $k = o(n)$, eli*

$$\frac{k}{n} \rightarrow 0, \text{ kun } n \rightarrow \infty,$$

missä parametri $k = k(n)$ on parametrin n funktio.

TODISTUS. Satunnaisverkon $G_{n,k,p}$ linkkien lukumäärä on lauseen 5.3 mukaan $|E(G_{n,k,p})| = kn/2$, jossa parametri $k = k(n)$ voi olla vakio tai parametrin n funktio. Parametrit n ja k määrittävät satunnaisverkon $G_{n,k,p}$ linkkien lukumäärän $|E(G_{n,k,p})|$ yksikäsitteisesti. Määritelmän 1.17 mukaan satunnaisverkkomalli $\mathcal{G}_{n,k,p}$ on harva, jos $|E(G_{n,k,p})| = o(n^2)$. Lauseen 1.16 mukaan tämä on yhtäpitävä pieni o -estimaatin $|E(G_{n,k,p})| = o(\frac{1}{2}n^2)$ kanssa, jolloin $|E(G_{n,r\text{-reg}})|/(\frac{1}{2}n^2) = \frac{1}{2}kn/(\frac{1}{2}n^2) = k/n \rightarrow 0$, kun $n \rightarrow \infty$. \square

Satunnaisverkossa $G_{n,k,p}$ solmujen asteet $k_i(G_{n,k,p})$, $i \in V_n$ voivat olla eri suuruksia, kun parametri $p > 0$. Algoritmin 1 kytkentämenetelmä varmistaa kuitenkin, että solmujen asteet ovat vähintään $k/2$, jolloin satunnaisverkolla ei ole eristettyjä solmuja.

LAUSE 5.7. *Olkoon $\mathcal{G}_{n,k,p}$ satunnaisverkkomalli. Tällöin satunnaisverkon $G_{n,k,p}$ solmujen asteille pätee*

$$k_i(G_{n,k,p}) \geq k/2 \text{ kaikilla } i \in V_n,$$

missä k on satunnaisverkon $G_{n,k,p}$ parametri.

TODISTUS. Algoritmin 1 lähtökohtana olevassa suunnatussa verkossa $H_{k/2\text{-out}}$ jokaisen solmun $i \in V_n$ lähtöaste on $k_i^+ = k/2$ ja tuloaste $k_i^- = k/2$. Satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ kytkentämenetelmässä solmun i ulosmenevät linkit kytketään uudelleen todennäköisyydellä p siten, että ainoastaan linkkien maalisolmut muuttuvat. Tällöin

solmun $i \in V_n$ lähtöaste pysyy samana, mutta solmun tuloaste on satunnaismuuttuja $k_i^- \in \{0, \dots, n - k/2 - 1\}$. Kun linkkien suunnat unohdetaan, on solmun $i \in V_n$ aste $k_i(G_{n,k,p})$ lähtöasteen k_i^+ ja tuloasteen k_i^- summana vähintään $k/2$. \square

5.3. Pieni maailma -verkkomallin kytkentätodennäköisyydet

LAUSE 5.8. *Olkoon $\mathcal{G}_{n,k,p}$ satunnaisverkkomalli. Todennäköisyys, että satunnaisverkon $G_{n,k,p}$ satunnaisesti valittu solmupari $\{i, j\} \in V_n^{(2)}$ on kytketty linkillä, on*

$$\mathbb{P}(X_{ij} = 1) = \frac{k}{n-1},$$

missä n ja k ovat satunnaisverkon $G_{n,k,p}$ parametreja ja X_{ij} on indikaattori tapahtumalle, että solmupari $\{i, j\}$ on kytketty linkillä.

TODISTUS. Lauseen 5.3 mukaan satunnaisverkon $G_{n,k,p}$ linkkien lukumäärä on $|E(G_{n,k,p})| = kn/2$. Tämä on satunnaisverkon $G_{n,k,p}$ solmuparien lukumäärä, jotka on kytketty linkillä. Satunnaisverkossa $G_{n,k,p}$ on yhteensä $\binom{n}{2} = n(n-1)/2$ eri solmuparia, joten todennäköisyys millä satunnaisverkon $G_{n,k,p}$ satunnaisesti valittu solmupari $\{i, j\}$ on kytketty linkillä on $\mathbb{P}(X_{ij} = 1) = |E(G_{n,k,p})|/\binom{n}{2} = kn/(n(n-1)) = k/(n-1)$. \square

Algoritmin 1 lähtökohtana olevassa suunnatussa verkossa $H_{k/2\text{-out}}$ solmuparin $\{i, j\}$ solmuilla i ja j on yhteensä k ulosmenevää linkkiä. Linkit kytketään uudelleen todennäköisyydellä p riippumattomasti toisistaan. Lasketaan millä todennäköisyydellä yksi näistä linkeistä kytkee solmuparin $\{i, j\}$ satunnaisverkossa $G_{n,k,p}$. Merkitään muuttujalla $t \in \{1, 2, \dots, k\}$ missä järjestyksessä kyseiset linkit käydään läpi algoritmissa 1. Merkitään lisäksi tapahtumat $A_t :=$ ”linkki t kytkee solmuparin $\{i, j\}$ ” ja $B_t :=$ ”linkki t ei kytke solmuparia $\{i, j\}$ ” kaikilla $t \in \{1, 2, \dots, k\}$. Lauseista 3.16 ja 3.17 saadaan laskettua satunnaisverkon $G_{n,k,p}$ solmuparin $\{i, j\} \in V_n^{(2)}$ kytkentätodennäköisyys

(5.3)

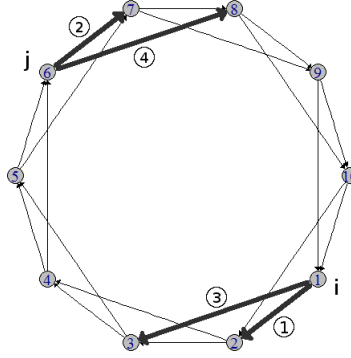
$$\begin{aligned} \mathbb{P}(X_{ij} = 1) &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \cap B_1) + \mathbb{P}(A_3 \cap B_1 \cap B_2) + \dots + \mathbb{P}(A_k \cap B_1 \cap \dots \cap B_{k-1}) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2 | B_1) \mathbb{P}(B_1) + \mathbb{P}(A_3 | B_1 \cap B_2) \mathbb{P}(B_2 | B_1) \mathbb{P}(B_1) + \dots \\ &\quad + \mathbb{P}(A_k | B_1 \cap \dots \cap B_{k-1}) \mathbb{P}(B_{k-1} | B_1 \cap \dots \cap B_{k-2}) \dots \mathbb{P}(B_2 | B_1) \mathbb{P}(B_1). \end{aligned}$$

Seuraavaksi lasketaan järjestämättömien solmuparien $\{i, j\} \in V_n^{(2)}$ kytkentätodennäköisyydet satunnaisverkossa $G_{n,k,p}$ metriikan ℓ_{ij} suhteen. Empiirisissä kokeissa [11] on havaittu, että harvan satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkolle $G_{n,k,p}$ pätee $\mathbb{P}(k_i(G_{n,k,p}) = n-1) \approx 0$, kun n on iso.

5.3.1. Kaukaisten solmujen kytkentätodennäköisyys. Harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ solmuparin $\{i, j\}$ kytkentätodennäköisyys ehdolla $\{\ell_{ij} = m\}$, $m \in \{k/2 + 1, k/2 + 1, \dots, n/2\}$, on

$$(5.4) \quad \mathbb{P}(X_{ij} = 1 | \ell_{ij} = m) \approx \frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Solmuparin $\{i, j\} \in V_n^{(2)}$ solmujen järjestyksellä ei ole väliä, joten oletetaan, että solmu i tulee ennen solmua j algoritmin 1 kytkentämenetelmässä (ks. kuva 5.3). Algoritmin 1 lähtökohtana olevassa sunna-



KUVA 5.3. Pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ lähtökohtainen suunnattu verkko $H_{n,k/2\text{-out}}$, kun $n = 10$ ja $k = 4$. Solmuparille $\{i, j\}$ pätee $\ell_{ij} = \min\{|j - i|, n - |j - i|\} > k/2$. Kuvaan on merkitty solmujen i ja j ulosmenevät linkit sekä järjestys, millä niitä kytketään uudelleen todennäköisyydellä p . Solmuparin $\{i, j\}$ kytkentätodennäköisyys harvan satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkossa $G_{n,k,p}$ ehdolla $\{\ell_{ij} = m\}$, $m \in \{k/2 + 1, k/2 + 2, \dots, n/2\}$ on

$$\mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \approx \frac{k}{n - k - 1}p - \frac{k(k - 1)}{2(n - k - 1)^2}p^2,$$

kun satunnaisverkon $G_{n,k,p}$ koko n on tarpeeksi iso.

tussa verkossa $H_{k/2\text{-out}}$, solmupari $\{i, j\}$ ei ole kytketty linkillä, kun $\ell_{ij} > k/2$, joten tapahtuman A_1 todennäköisyys ehdolla $\{\ell_{ij} = m\}$, $m \in \{k/2 + 1, k/2 + 2, \dots, n/2\}$ on

$$\mathbb{P}(A_1 \mid \ell_{ij} = m) = \frac{p}{n - k_{i,\text{hetkellinen}} - 1}$$

ja $\mathbb{P}(B_1 \mid \ell_{ij} = m) = 1 - \mathbb{P}(A_1 \mid \ell_{ij} = m)$. Liitteessä A osoitetaan, että solmun i hetkellinen aste $k_{i,\text{hetkellinen}}$ algoritmissa 1 on satunnaisuuttuja, joka riippuu muiden solmujen asteista. Tämä hankaloittaa laskuja, joten tehdään approksimaatio

$$(5.5) \quad \frac{p}{n - k_{i,\text{hetkellinen}} - 1} \approx \frac{p}{n - k - 1},$$

missä seurauksen 5.4 mukaan parametri $k = \bar{k}(G_{n,k,p})$. Kaavan (5.5) approksimaation virhe on empiirisissä kokeissa [11] havaittu pieneksi, kun harvan satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ koko n on iso. Käyttäen kaavan (5.5) approksimaatiota saadaan ehdolla $\{\ell_{ij} = m\}$, $m \in \{k/2 + 1, k/2 + 2, \dots, n/2\}$ laskettua $\mathbb{P}(A_t \mid B_1 \cap \dots \cap B_{t-1} \cap \ell_{ij} = m) \approx p/(n - k - 1)$ ja $\mathbb{P}(B_t \mid B_1 \cap \dots \cap B_{t-1} \cap \ell_{ij} = m) \approx 1 - p/(n - k - 1)$ kaikilla $t \in \{2, 3, \dots, k\}$, jolloin

$$\begin{aligned} \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) &\approx \frac{p}{n - k - 1} + \frac{p}{n - k - 1} \left(1 - \frac{p}{n - k - 1}\right) + \frac{p}{n - k - 1} \\ &\cdot \left(1 - \frac{p}{n - k - 1}\right)^2 + \dots + \frac{p}{n - k - 1} \left(1 - \frac{p}{n - k - 1}\right)^{k-1}. \end{aligned}$$

Tämä vastaa *geometrista summaa* [1] $\sum_{m=1}^k a_1 q^{m-1}$ termeillä $a_1 = p/(n - k - 1)$ ja $q = (1 - p/(n - k - 1))$. Geometrisen summan $\sum_{m=1}^k a_1 q^{m-1}$ saadaan laskettua

kaavasta [1] $S_k = \sum_{m=1}^k a_1 q^{m-1} = a_1(1 - q^k)/(1 - q)$, kun $q \neq 1$, jolloin saadaan

$$\begin{aligned} \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} > k/2) &\approx \sum_{m=1}^k \frac{p}{n-k-1} \left(1 - \frac{p}{n-k-1}\right)^{m-1} \\ &= \frac{\frac{p}{n-k-1} \left(1 - \left(1 - \frac{p}{n-k-1}\right)^k\right)}{1 - \left(1 - \frac{p}{n-k-1}\right)} = 1 - \left(1 - \frac{p}{n-k-1}\right)^k. \end{aligned}$$

Termi $(1 - p/(n-k-1))^k$ saadaan laskettua *binomilauseen* [1] avulla $(a+b)^k = \sum_{s=0}^k \binom{k}{s} a^{k-s} b^s$, jolloin

$$\begin{aligned} \left(1 - \frac{p}{n-k-1}\right)^k &= \binom{k}{0} 1^k (-1)^0 \left(\frac{p}{n-k-1}\right)^0 + \binom{k}{1} 1^{k-1} (-1)^1 \left(\frac{p}{n-k-1}\right)^1 \\ &\quad + \binom{k}{2} 1^{k-2} (-1)^2 \left(\frac{p}{n-k-1}\right)^2 + \dots + \binom{k}{k} 1^0 (-1)^k \left(\frac{p}{n-k-1}\right)^k \\ &= 1 - k \frac{p}{n-k-1} + \frac{k(k-1)}{2} \frac{p^2}{(n-k-1)^2} \\ &\quad - \frac{k(k-1)(k-2)}{6} \frac{p^3}{(n-k-1)^3} + \dots + (-1)^k \binom{k}{k} \frac{p^k}{(n-k-1)^k}. \end{aligned}$$

Harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ parametreille n ja k pätee $k^s/(n-k-1)^s \approx 0$ kaikilla $s \geq 3$, kun $n \rightarrow \infty$, jolloin

$$\left(1 - \frac{p}{n-k-1}\right)^k \approx 1 - \frac{k}{n-k-1} p + \frac{k(k-1)}{2(n-k-1)^2} p^2.$$

Todennäköisyys, millä satunnaisverkon $G_{n,k,p}$ solmupari $\{i, j\}$ on kytketty linkillä ehdolla $\{\ell_{ij} = m\}$, $m \in \{k/2+1, k/2+2, \dots, n/2\}$, on harvassa satunnaisverkkomallissa

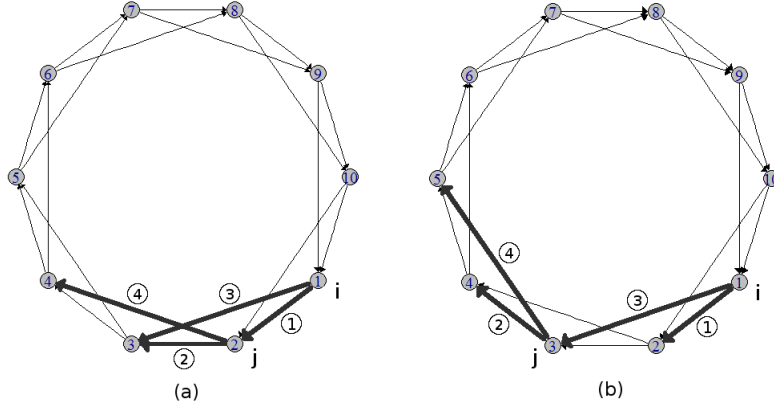
$$\begin{aligned} \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) &\approx 1 - \left(1 - \frac{k}{n-k-1} p + \frac{k(k-1)}{2(n-k-1)^2} p^2\right) \\ &= \frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$.

5.3.2. Läheisten solmujen kytkentätodennäköisyys. Satunnaisverkon $G_{n,k,p}$ solmuparin $\{i, j\}$ kytkentätodennäköisyys harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$, on

$$(5.6) \quad \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \approx (1-p) + \frac{k-2m+1}{n-k-1} p^2,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Kyseinen todennäköisyys riippuu metriikan $\ell_{ij} = m \in \{1, 2, \dots, k/2\}$ arvosta. Solmupari $\{i, j\}$ on kytketty linkillä algoritmin 1 lähtökohtana olevassa suunnatussa verkossa $H_{k/2\text{-out}}$, kun $\ell_{ij} = m \in \{1, 2, \dots, k/2\}$. Oletetaan, että kyseisen linkin lähtösolmu on i . Todennäköisyydellä $1-p$ linkkiä (i, j) ei kytketä uudelleen, jolloin $ij \in E(G_{n,k,p})$. Oletetaan, että linkki $(i, j) \in E(H_{k/2\text{-out}})$ kytketään uudelleen todennäköisyydellä p . Lasketaan millä todennäköisyydellä jokin suunnatun verkon $H_{k/2\text{-out}}$ solmujen i ja j jäljelle jääneistä



KUVA 5.4. Pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ lähtökohtainen suunnattu verkko $H_{n,k/2\text{-out}}$, kun $n = 10$ ja $k = 4$. Kuvassa on merkitty solmujen i ja j ulosmenevät linkit sekä järjestys millä niitä kytetään uudelleen todennäköisyydellä p , kun algoritmin 1 aloitussolmu $i' \in \{j+1, \dots, n\}$. (a): $\ell_{ij} = 1$. Linkkiä (i, j) kytkettäessä uudelleen voi tilalle tulla jokin jäljelle jääneistä $k - 2\ell_{ij} + 1 = 3$ merkityistä linkeistä. (b): $\ell_{ij} = 2$, joten $k - 2\ell_{ij} + 1 = 1$. Solmuparin $\{i, j\}$ kytkentätodennäköisyys harvan satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkossa $G_{n,k,p}$ ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$, on

$$\mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \approx (1 - p) + \frac{k - 2m + 1}{n - k - 1} p^2$$

kun satunnaisverkon $G_{n,k,p}$ koko n on tarpeeksi iso.

ulosmenevistä linkeistä kytkee solmuparin $\{i, j\}$ satunnaisverkossa $G_{n,k,p}$ (ks. kuva 5.4).

Olkoon $\{i, j\}$ suunnatun verkon $H_{k/2\text{-out}}$ solmupari, jolle $\ell_{ij} = m \in \{1, 2, \dots, k/2\}$. Solmuparin järjestyksellä ei ole väliä, joten oletetaan, että solmu i on solmuparia $\{i, j\}$ kytkevän linkin lähtösolmu. Algoritmin 1 kytkentämenetelmän aloitussolmusta riippuen linkki $t = 1$ on suunnatun verkon $H_{k/2\text{-out}}$ solmun i ulosmenevä linkki todennäköisyydellä $(n - m)/n$. Todennäköisyydellä m/n se on solmun j ulosmenevä linkki (ks. kuva 5.4). Oletetaan ensin, että linkki $t = 1$ on solmun i ulosmenevä linkki. Tällöin ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$ pätee $\mathbb{P}(A_t \mid \ell_{ij} = m) = 0$ ja $\mathbb{P}(B_t \mid \ell_{ij} = m) = 1$ riippumatta tapahtumista $(B_s)_{s=1}^{t-1}$ kaikilla $1 \leq t < 2m - 1$, sillä solmupari $\{i, j\}$ on jo kytketty linkillä $t = 2m - 1$ (ks. kuva 5.4). Vasta, kun $t = 2m - 1$ pätee $\mathbb{P}(A_{2m-1} \mid \ell_{ij} = m) = 1 - p$ ja $\mathbb{P}(B_{2m-1} \mid \ell_{ij} = m) = p$ riippumatta tapahtumista $(B_s)_{s=1}^{t-1}$. Kaavan (5.5) approksimaatiota käyttäen saadaan laskettua $\mathbb{P}(A_t \mid B_1 \cap \dots \cap B_{t-1} \cap \ell_{ij} = m) \approx p/(n - k - 1)$ ja $\mathbb{P}(B_t \mid B_1 \cap \dots \cap B_{t-1} \cap \ell_{ij} = m) \approx 1 - p/(n - k - 1)$ kaikilla $t \in \{2m, 2m + 1, \dots, k\}$. Kaavasta 5.3 saadaan laskettua solmuparin $\{i, j\}$ kytkentätodennäköisyys ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$

$$\begin{aligned} \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) &\approx (1 - p) + p \frac{p}{n - k - 1} + p \frac{p}{n - k - 1} \left(1 - \frac{p}{n - k - 1}\right) \\ &+ \dots + p \frac{p}{n - k - 1} \left(1 - \frac{p}{n - k - 1}\right)^{k-2m}. \end{aligned}$$

Tämä vastaa termiä $(1-p)$ lukuun ottamatta geometrista summaa $p^2/(n-k-1) \sum_{s=1}^{k-2m+1} (1-p/(n-k-1))^{s-1}$. Geometrisen summan kaavasta [1] saadaan laskettua solmuparin $\{i, j\}$ kytkentätodennäköisyys ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$

$$\begin{aligned} \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) &\approx (1-p) + \sum_{s=1}^{k-2m+1} \frac{p^2}{n-k-1} \left(1 - \frac{p}{n-k-1}\right)^{s-1} \\ &= (1-p) + \frac{\frac{p^2}{n-k-1} \left(1 - \left(1 - \frac{p}{n-k-1}\right)^{k-2m+1}\right)}{1 - \left(1 - \frac{p}{n-k-1}\right)} \\ &= (1-p) + p \left(1 - \left(1 - \frac{p}{n-k-1}\right)^{k-2m+1}\right) \\ &= 1 - p \left(1 - \frac{p}{n-k-1}\right)^{k-2m+1}. \end{aligned}$$

Termi $(1 - p/(n-k-1))^{k-2m+1}$ saadaan laskettua binomilauseen [1] avulla, $(a+b)^{k-2m+1} = \sum_{s=0}^{k-2m+1} \binom{k-2m+1}{s} a^{k-2m+1-s} b^s$, jolloin

$$\begin{aligned} \left(1 - \frac{p}{n-k-1}\right)^{k-2m+1} &= 1 - (k-2m+1) \frac{p}{n-k-1} + \binom{k-2m+1}{2} \frac{p^2}{(n-k-1)^2} \\ &\quad + \dots + \binom{k-2m+1}{k-2m+1} (-1)^{k-2m+1} \frac{p^{k-2m+1}}{(n-k-1)^{k-2m+1}}. \end{aligned}$$

Harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ parametreille n ja k pätee $k^m/(n-k-1)^m \approx 0$ kaikilla $m \geq 2$, kun $n \rightarrow \infty$, jolloin

$$\left(1 - \frac{p}{n-k-1}\right)^{k-2m+1} \approx 1 - \frac{k-2m+1}{n-k-1} p.$$

Todennäköisyys, millä satunnaisverkon $G_{n,k,p}$ solmupari $\{i, j\}$ on kytketty linkillä harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$ on

$$\mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \approx 1 - p \left(1 - \frac{k-2m+1}{n-k-1} p\right) = (1-p) + \frac{k-2m+1}{n-k-1} p^2,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$.

Oletetaan, että linkki $t = 1$ on solmun j ulosmenevä linkki. Tällöin ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$ pätee $\mathbb{P}(A_t \mid \ell_{ij} = m) = 0$ ja $\mathbb{P}(B_t \mid \ell_{ij} = m) = 1$ riippumatta tapahtumista $(B_s)_{s=1}^{t-1}$ kaikilla $1 \leq t < 2m$, sillä solmupari $\{i, j\}$ on jo kytketty linkillä $t = 2m$. Vasta, kun $t = 2m$ pätee $\mathbb{P}(A_{2m} \mid \ell_{ij} = m) = 1 - p$ ja $\mathbb{P}(B_{2m} \mid \ell_{ij} = m) = p$ riippumatta tapahtumista $(B_s)_{s=1}^{t-1}$. Vastaavasti saadaan laskettua kaavasta 5.3 solmuparin $\{i, j\}$ kytkentätodennäköisyys ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$

$$\mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \approx 1 - p(1 - p/(n-k-1))^{k-2m}.$$

Binomilauseen [1] mukaan

$$\left(1 - \frac{p}{n-k-1}\right)^{k-2m} = 1 - \frac{k-2m}{n-k-1} p + \dots + \binom{k-2m}{k-2m} (-1)^k \frac{p^k}{(n-k-1)^k},$$

jolloin

$$\mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \approx 1 - p \left(1 - \frac{k - 2m}{n - k - 1} p \right) = (1 - p) + \frac{k - 2m}{n - k - 1} p^2,$$

kaikilla $m \in \{1, 2, \dots, k/2\}$, kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Lauseen 3.17 mukaan todennäköisyys, millä satunnaisverkon $G_{n,k,p}$ solmupari $\{i, j\}$ on kytketty linkillä harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ ehdolla $\{\ell_{ij} = m\}$, $m \in \{1, 2, \dots, k/2\}$, on

$$\begin{aligned} & \mathbb{P}(X_{ij} = 1 \mid \ell_{ij} = m) \\ & \approx \frac{n - m}{n} \left((1 - p) + \frac{k - 2m + 1}{n - k - 1} p^2 \right) + \frac{m}{n} \left((1 - p) + \frac{k - 2m}{n - k - 1} p^2 \right) \\ & = (1 - p) + \frac{k - 2m}{n - k - 1} p^2 + \frac{n - m}{n(n - k - 1)} p^2 \approx (1 - p) + \frac{k - 2m + 1}{n - k - 1} p^2, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$.

5.4. Solmujen asteiden jakauma pieni maailma -verkkomallissa

Fyysikot Alain Barrat ja Martin Weigt [11] analysoivat vuoden 1999 julkaisussaan Wattsin ja Strogatzin verkkomallin ominaisuuksia tarkemmin sekä empiirisesti että teoreettisesti. He esittävät estimaatin satunnaisverkon $G_{n,k,p}$ tyypillisen solmun $i \in V_n$ asteen jakaumalle, kun satunnaisverkko $G_{n,k,p}$ on iso. Parametrilla $p > 0$ solmun i aste $k_i(G_{n,k,p})$ on satunnaismuuttuja, joka lauseen 5.7 mukaan on vähintään $k/2$. Voidaan siis kirjoittaa solmun i aste muodossa

$$k_i(G_{n,k,p}) = k/2 + X_i,$$

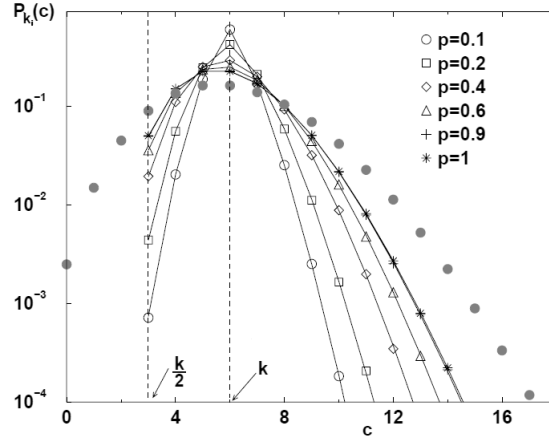
missä $X_i \in \{0, 1, \dots, n - k/2 - 1\}$ on satunnaismuuttuja. Barrat ja Weigt esittävät satunnaismuuttujaa X_i kahden muun satunnaismuuttujien summana

$$X_i = Y_i + Z_i,$$

missä $Y_i \in \{0, 1, \dots, k/2\}$ ja $Z_i = X_i - Y_i$. Satunnaismuuttuja Y_i määrittää algoritmin 1 lähtökohtana olevan suunnatun verkon $H_{k/2\text{-out}}$ solmun i sisääntulevien linkkien lukumäärän, jotka ovat pysyneet paikallaan kytkentämenetelmän aikana todennäköisyydellä $1 - p$ riippumattomasti toisistaan. Satunnaismuuttuja Y_i noudattaa siis binomijakaumaa

$$Y_i \sim \text{Bin} \left(\frac{k}{2}, 1 - p \right).$$

Barrat ja Weigt määrittelevät satunnaismuuttujan Z_i ilmaisemaan kuinka monta linkkiä on kytketty uudelleen *kohti* solmua i todennäköisyydellä p/n . Liitteessä A osoitetaan, että kyseinen todennäköisyys on approksimaatio. Barrat ja Weigt approksimoivat satunnaismuuttujan Z_i jakaumaksi $Z_i \sim \text{Poisson}(\lambda)$ parametrilla $\lambda = pk/2$, kun n on iso, jolloin Barrat ja Weigt määrittävät satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i asteen $k_i(G_{n,k,p})$ jakaumaksi



KUVA 5.5. Barratin ja Weigtin [11] esittämä todennäköisyysjakauma $P_{k_i}(c)$ satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i asteille k_i , missä $c \geq k/2$. Symbolit ovat numeerisesti simuloitujen verkkojen solmujen asteiden suhteellisia frekvenssejä parametreilla $n = 1000$ ja $k = \bar{k} = 6$. Täytetyt ympyrät näyttävät satunnaisverkon $G_{n,p}$ solmun asteen $k_i(G_{n,p}) \sim \text{Poisson}(\lambda)$ jakauman parametrilla $\lambda = \bar{k} = 6$.

$$\begin{aligned}
 (5.7) \quad \mathbb{P}(k_i(G_{n,k,p}) = c) &= \mathbb{P}(k_i(G_{n,k,p}) = k/2 + x) \\
 &= \sum_{y=0}^{\min(x, \frac{k}{2})} \mathbb{P}((Y_i = y) \cap (Z_i = x - y)) \\
 &\approx \sum_{y=0}^{\min(x, \frac{k}{2})} \binom{\frac{k}{2}}{y} (1-p)^y p^{\frac{k}{2}-y} \frac{(\frac{kp}{2})^{x-y}}{(x-y)!} e^{-\frac{kp}{2}},
 \end{aligned}$$

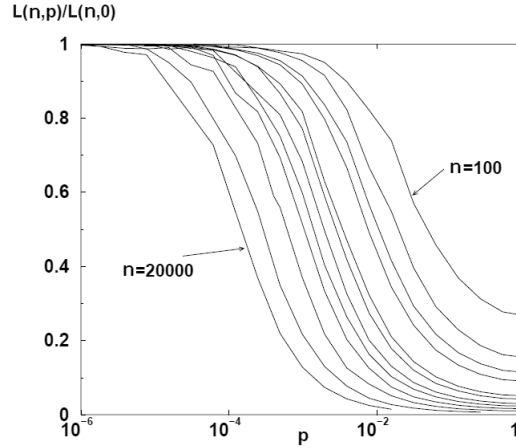
kun $c \geq k/2$ ja n on iso. Barrat ja Weigt huomauttavat, että asteen $k_i(G_{n,k,p})$ jakauma levenee kun parametrin p arvo kasvaa (ks. kuva 5.5), mutta he eivät kuitenkaan laske jakauman tunnuslukuja. Asteen $k_i(G_{n,k,p})$ jakauman odotusarvo saadaan laskettua kaavan $k_i(G_{n,k,p}) = k/2 + Y_i + Z_i$ avulla, jossa $Y_i \sim \text{Bin}(k/2, 1-p)$ ja $Z_i \sim \text{Poisson}(pk/2)$, kun n on iso. Lauseesta 3.7 saadaan laskettua odotusarvo

$$(5.8) \quad \mathbb{E}[k_i(G_{n,k,p})] = \mathbb{E}[k/2] + \mathbb{E}[Y_i] + \mathbb{E}[Z_i] \approx \frac{k}{2} + \frac{k}{2}(1-p) + \frac{k}{2}p = k,$$

joka ei riipu parametrin p arvosta. Jakauman varianssi saadaan myös laskettua kaavan $k_i(G_{n,k,p}) = k/2 + Y_i + Z_i$ avulla, sillä satunnaismuuttujat Y_i ja Z_i ovat riippumattomia, kun n on iso. Lauseesta 3.11 sekä binomijakauman että Poissonjakauman varianssien kaavoista [3] saadaan laskettua varianssille kaavan

$$(5.9) \quad \text{Var}(k_i(G_{n,k,p})) \approx \text{Var}(Y_i) + \text{Var}(Z_i) \approx \frac{k}{2}p(1-p) + \frac{k}{2}p = \frac{k}{2}p(2-p),$$

missä vakion $k/2$ varianssi on nolla [3]. Asteen $k_i(G_{n,k,p})$ varianssi kasvaa siis, kun parametrin p arvo kasvaa. Kaavasta 5.9 saadaan kuitenkin laskettua, että jopa parametrilla $p = 1$ pätee $\text{Var}(k_i(G_{n,k,p=1})) = \frac{1}{2} \text{Var}(k_i(G_{n,p'}))$, missä $G_{n,p'}$ on satunnaisverkko, jossa $k_i(G_{n,p'}) \sim \text{Poisson}(\lambda)$ parametrilla $\lambda = k$ (ks. kuva 5.5).



KUVA 5.6. Barratin ja Weigtin [11] numeerisesti simuloitujen pieni maailma -verkkomallin verkkojen karakteristisen polun pituus $L(n, p)$ parametrin p funktiona useilla eri parametrin $n \in [100, 20000]$ arvoilla, missä parametri $k = 4$. Lasketut arvot ovat keskiarvoja viidestä sadasta generoidusta verkosta ja normalisoitu täysin säännöllisen rengashilan arvoilla $L(n, 0)$. Käyrän lasku ilmenee yhä pienemmällä parametrin p arvoilla samalla kun parametrin n arvo kasvaa.

5.5. Karakteristisen polun pituus pieni maailma -verkkomallissa

D. J. Watts ja S. H. Strogatz [10] tutkivat pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ generoitujen verkkojen karakteristisen polun pituutta L ääritapauksissa $p = 0$ ja $p = 1$. Kun $p = 0$ mallin generoima verkko on yhtenäinen k -säännöllinen rengashila (ks. kuva 5.1(a)), jonka karakteristisen polun pituus L kasvaa lineaarisesti verkon koon n mukaan. A. Barrat ja M. Weigt [11] laskevat rengashilan $G_{n,k,p=0}$ karakteristisen polun pituuden tarkasti määritelmän 1.31 kaavalla

$$L(G_{n,k,p=0}) = \frac{n(n-k-2)}{2k(n-1)} \sim \frac{n}{2k}.$$

Kun $p = 1$ Watts ja Strogatz [10] arvioivat, että satunnaisverkon $G_{n,k,p=1}$ karakteristisen polun pituus on lähes sama kuin satunnaisverkossa $G_{n,M(n)}$ parametrilla $M(n) = kn/2$. Lisäksi satunnaisverkkomallin $\mathcal{G}_{n,M(n)}$ tyypillinen verkko on yhtenäinen, kun $\ln(n) = o(k)$ [6]. Barrat ja Weigt [11] osoittavat numeerisella simuloinnilla, että pieni maailma -verkkomallin $\mathcal{G}_{n,k,p=1}$ generoitujen verkkojen karakteristisen polun pituus parametrilla $p = 1$ tosiaan kasvaa, kuten

$$L(G_{n,k,p=1}) \sim \frac{\ln(n)}{\ln(k-1)}.$$

Kuvassa 5.6 Barrat ja Weigt [11] näyttävät, että karakteristisen polun pituuden L käyrän lasku riippuu myös parametrilla n . He osoittavat, että käyrän lasku ilmenee yhä pienemmällä parametrin p arvoilla samalla kun verkon koko n kasvaa.

Pieni maailma -verkkomallin klusteroituneisuus

Alain Barrat ja Martin Weigt [11] määrittävät julkaisussaan heuristisen lausekkeen pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ transitiivisuuskertoimelle, johon usein viitataan kirjallisuudessa [15] [17]

$$(6.1) \quad c'(G_{n,k,p}) = \frac{3(k-2)}{4(k-1)}(1-p)^3.$$

Kyseessä on kuitenkin heuristinen estimaatti kaavan (2.9) transitiivisuuskertoimelle $C'(G_{n,k,p}) = \frac{1}{n} \sum_{i=1}^n |\Delta_i(G_{n,k,p})| / \left(\frac{1}{n} \sum_{i=1}^n |\Upsilon_i(G_{n,k,p})| \right)$, jonka tarkkuuden Barrat ja Weigt väittävät olevan kertaluokkaa $|c'(G_{n,k,p}) - 3(k-2)/(4(k-1)) \cdot (1-p)^3| = O(1/n)$. He päättävät heuristisin perusteluin, että solmun i kolmio $\nabla \in \Delta_i(H_{n,k})$ algoritmin 1 lähtökohtaisessa rengashilassa $H_{n,k}$ on vielä solmun i kolmio $\nabla \in \Delta_i(G_{n,k,p})$ satunnaisverkossa $G_{n,k,p}$ todennäköisyydellä $(1-p)^3 + O(1/n)$, kun $p > 0$. Tästä he johtavat satunnaisverkon $G_{n,k,p}$ transitiivisuuskertoimen kaavan 2.9 avulla

$$C'(G_{n,k,p}) = \frac{\frac{1}{n} \sum_{i=1}^n |\Delta_i(G_{n,k,p})|}{\frac{1}{n} \sum_{i=1}^n |\Upsilon_i(G_{n,k,p})|} \approx \frac{|\Delta_i(H_{n,k})|(1-p)^3}{|\Upsilon_i(H_{n,k})|} = \frac{3(k-2)}{4(k-1)}(1-p)^3,$$

missä k -säännöllisen rengashilan $H_{n,k}$ solmun i kytkettyjen kolmikoiden- ja kolmioiden lukumäärä on $|\Delta_i(H_{n,k})| = 3k(k-2)/8$ ja $|\Upsilon_i(H_{n,k})| = k(k-1)/2$. Barrat ja Weigt osoittavat numeerisella simuloinnilla, että pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ generoitujen verkkojen klusterointikerroin C ja transitiivisuuskerroin C' eroavat vain hyvin vähän toisistaan, joten he antavat satunnaisverkon $G_{n,k,p}$ klusterointikerroimelle estimaatin $c(G_{n,k,p}) \approx c'(G_{n,k,p})$.

6.1. Pieni maailma -verkkomallin satunnaisverkon transitiivisuuskerroin

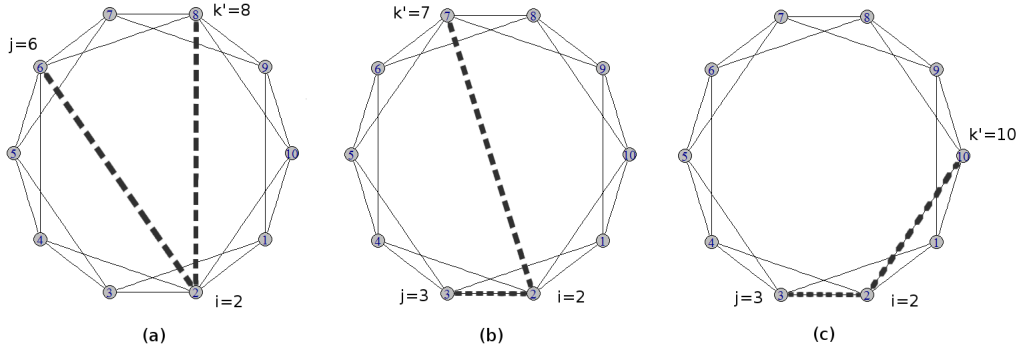
Seuraavaksi lasketaan matemaattinen lauseke satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ transitiivisuuskertoimelle parametrien n , k ja p funktiona. Määritelmän 3.38 mukaan satunnaisverkon $G_{n,k,p}$ transitiivisuuskerroin on

$$c'(G_{n,k,p}) = \frac{\mathbb{E}[|\Delta_i(G_{n,k,p})|]}{\mathbb{E}[|\Upsilon_i(G_{n,k,p})|]},$$

kun $\mathbb{E}[|\Upsilon_i(G_n)|] > 0$ ja kokoelmat $(|\Delta_i(G_{n,k,p})| : i \in V_n)$ ja $(|\Upsilon_i(G_{n,k,p})| : i \in V_n)$ muodostuvat samoin jakautuneista satunnaismuuttujista.

6.1.1. Solmun kytkettyjen kolmikoiden odotusarvo. Lauseen 3.25 mukaan satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kytkettyjen kolmikoiden lukumäärän odotusarvo saadaan laskettua asteen $k_i(G_{n,k,p})$ jakauman tunnuslukujen avulla

$$\mathbb{E}[|\Upsilon_i(G_{n,k,p})|] = \frac{1}{2} (\mathbb{E}[k_i(G_{n,k,p})]^2 - \mathbb{E}[k_i(G_{n,k,p})] + \text{Var}(k_i(G_{n,k,p}))).$$



KUVA 6.1. Esimerkkejä pieni maailma -verkkomallin satunnaisverkon $G_{n,k,p}$ solmun i mahdollisista kytketyistä kolmikoista (katkoviivat) parametreilla $n = 10$, $k = 4$ ja $p > 0$. Olkoon $\ell_{ij} = \min\{|j - i|, n - |j - i|\}$ metriikka. (a): $\{\ell_{ij} > k/2, \ell_{ik'} > k/2\}$, (b): $\{\ell_{ij} \leq k/2, \ell_{ik'} > k/2\}$ ja (c): $\{\ell_{ij} \leq k/2, \ell_{ik'} \leq k/2\}$.

Kaavassa 5.7 on esitetty asteen $k_i(G_{n,k,p})$ jakauma, kun satunnaisverkon $G_{n,k,p}$ koko on iso. Kaavojen 5.8 ja 5.9 mukaan $\mathbb{E}[k_i(G_{n,k,p})] \approx k$ ja $\text{Var}(k_i(G_{n,k,p})) \approx kp(2-p)/2$, joten isolle satunnaisverkolle $G_{n,k,p}$ saadaan laskettua

$$(6.2) \quad \mathbb{E}[|\Upsilon_i(G_{n,k,p})|] \approx \frac{k(k-1)}{2} + \frac{k}{4}p(2-p).$$

Transitiivisuuskertoimen $c'(G_{n,k,p})$ lausekkeen määrittämiseen parametrien n , k ja p funktiona lasketaan odotusarvo $\mathbb{E}[|\Upsilon_i(G_{n,k,p})|]$ myös pienemmille satunnaisverkoille $G_{n,k,p}$. Lauseen 3.23 mukaan

$$\mathbb{E}[|\Upsilon_i(G_{n,k,p})|] = \sum_{\{i,j,k'\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik'\}} = 1) = \sum_{\{i,j,k'\} \in V_i^{(3)}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1),$$

missä satunnaisverkon $G_{n,k,p}$ solmuparien kytkentätodennäköisyydet saadaan laskettua kaavoista 5.4 ja 5.6. Odotusarvon $\mathbb{E}[|\Upsilon_i(G_{n,k,p})|]$ laskemista varten tehdään ositus solmujoukon V_n järjestämättömien kolmikoiden $V_n^{(3)}$ osajoukolle $V_i^{(3)} = \{\{i', j', k'\} \in V_n^{(3)} : i \in \{i', j', k'\}\}$. Alkiossa $\{i, j, k'\} \in V_i^{(3)}$ solmujen järjestyksellä ei ole väliä, joten voidaan olettaa, että $\ell_{ij} \leq \ell_{ik'}$. Olkoon $V_i^{(3)} = \bigcup_{s=1}^3 A_s$ joukon $V_i^{(3)}$ ositus erillisiin osajoukkoihin

$$\begin{aligned} A_1 &= \{\ell_{ij} > k/2, \ell_{ik'} > k/2\}, \\ A_2 &= \{\ell_{ij} \leq k/2, \ell_{ik'} > k/2\}, \\ A_3 &= \{\ell_{ij} \leq k/2, \ell_{ik'} \leq k/2\}, \end{aligned}$$

missä k on satunnaisverkon $G_{n,k,p}$ parametri (ks. kuva 6.1). Lauseesta 3.17 saadaan nyt satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kytkettyjen kolmikoiden lukumäärän

odotusarvo muotoon

$$(6.3) \quad \begin{aligned} \mathbb{E} [|\Upsilon_i(G_{n,k,p})|] &= \sum_{\{i,j,k'\} \in V_i^{(3)}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &= \sum_{s=1}^3 \sum_{\{i,j,k'\} \in A_s} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1), \end{aligned}$$

sillä $V_i^{(3)} = \bigcup_{s=1}^3 A_s$ on joukon $V_i^{(3)}$ ositus. Ensin lasketaan kaavan (6.3) summan osa alkioille $\{i, j, k'\} \in A_1 = \{\ell_{ij} > k/2, \ell_{ik'} > k/2\}$ (ks. kuva 6.1(a)). Joukon A_1 alkioden lukumäärä on

$$|\{\{i, j, k'\} \in A_1\}| = \binom{n-k-1}{2} = \frac{(n-k-1)^2 - (n-k-1)}{2}.$$

Merkitään $A_{1,m,m'} = \{\ell_{ij} = m, \ell_{ik'} = m'\}$ kaikilla $k/2 < m < n/2, m \leq m' \leq n/2$. Liitteessä B lasketaan kaavan (5.4) avulla, että joukon $A_{1,m,m'}$ alkioille pätee

$$\begin{aligned} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) &= \mathbb{P}(X_{ij} = 1) \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) \\ &\approx \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \left(\frac{k-\frac{1}{2}}{n-k-1} p - \frac{(k-1)^2}{2(n-k-1)^2} p^2 \right) \\ &\approx \frac{k(k-\frac{1}{2})}{(n-k-1)^2} p^2 - \frac{k(k-1)(k-\frac{3}{4})}{(n-k-1)^3} p^3, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$, jolloin

$$(6.4) \quad \begin{aligned} &\sum_{\{i,j,k'\} \in A_1} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &= \sum_{m>k/2}^{n/2-1} \sum_{m' \geq m}^{n/2} \sum_{\{i,j,k'\} \in A_{1,m,m'}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &\approx \frac{(n-k-1)^2 - (n-k-1)}{2} \left(\frac{k(k-\frac{1}{2})}{(n-k-1)^2} p^2 - \frac{k(k-1)(k-\frac{3}{4})}{(n-k-1)^3} p^3 \right) \\ &\approx \frac{k(k-\frac{1}{2})}{2} \left(1 - \frac{1}{n-k-1} \right) p^2 - \frac{k(k-1)(k-\frac{3}{4})}{2(n-k-1)} p^3 \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Seuraavaksi lasketaan kaavan (6.3) summan osa alkioille $\{i, j, k'\} \in A_2 = \{\ell_{ij} \leq k/2, \ell_{ik'} > k/2\}$ (ks. kuva 6.1(b)). Joukon A_2 alkioden lukumäärä on

$$|\{\{i, j, k'\} \in A_2\}| = k(n-k-1).$$

Merkitään $A_{2,m,m'} = \{\ell_{ij} = m, \ell_{ik'} = m'\}$ kaikilla $m \in \{1, 2, \dots, k/2\}$ ja $m' \in \{k/2+1, k/2+2, \dots, n/2\}$. Liitteessä B lasketaan kaavojen (5.4) ja (5.6) avulla, että

joukon $A_{2,m,m'}$ alkiuille pätee

$$\begin{aligned}\mathbb{P}(X_{ij} = 1, X_{ik'} = 1) &= \mathbb{P}(X_{ij} = 1) \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) \\ &\approx \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \left(\frac{k-\frac{1}{2}}{n-k-1} p - \frac{(k-1)^2}{2(n-k-1)^2} p^2 \right)\end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Liitteessä C osoitetaan, että

$$(6.5) \quad \sum_{m=1}^{k/2} \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) = \frac{k}{2} (1-p) + \frac{k^2}{4(n-k-1)} p^2,$$

jolloin saadaan laskettua

(6.6)

$$\begin{aligned}& \sum_{\{i,j,k'\} \in A_2} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &= \sum_{m=1}^{k/2} \sum_{m' > k/2}^{n/2} \sum_{\{i,j,k'\} \in A_{2,m,m'}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &\approx 2(n-k-1) \sum_{m=1}^{k/2} \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \left(\frac{k-\frac{1}{2}}{n-k-1} p - \frac{(k-1)^2}{2(n-k-1)^2} p^2 \right) \\ &= \left((k-\frac{1}{2})p - \frac{(k-1)^2}{2(n-k-1)} p^2 \right) 2 \sum_{m=1}^{k/2} \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\ &= \left((k-\frac{1}{2})p - \frac{(k-1)^2}{2(n-k-1)} p^2 \right) \left(k(1-p) + \frac{k^2}{2(n-k-1)} p^2 \right) \\ &\approx k(k-\frac{1}{2})(1-p)p - \frac{k(k-1)^2}{2(n-k-1)} (1-p)p^2 + \frac{k^2(k-\frac{1}{2})}{2(n-k-1)} p^3,\end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Lopuksi lasketaan kaavan (6.3) summan osa alkiuille $\{i, j, k'\} \in A_3 = \{\ell_{ij} \leq k/2, \ell_{ik'} \leq k/2\}$ (ks. kuva 6.1(c)). Joukon A_3 alkioiden lukumäärä on

$$|\{\{i, j, k'\} \in A_3\}| = \binom{k}{2} = \frac{k(k-1)}{2}.$$

Jaetaan joukko A_3 kahteen erilliseen osajoukkoon $A_3 = A_{3a} \cup A_{3b}$, missä

$$\begin{aligned}A_{3a} &= \{1 \leq \ell_{ij} = \ell_{ik'} \leq k/2\} \text{ ja} \\ A_{3b} &= \{1 \leq \ell_{ij} < \ell_{ik'} \leq k/2\}.\end{aligned}$$

Joukon A_{3a} alkioiden lukumäärä on

$$|\{\{i, j, k'\} \in A_{3a}\}| = \frac{k}{2}.$$

Merkitään $A_{3a,m} = \{\ell_{ij} = m, \ell_{ik'} = m\}$, $m \in \{1, 2, \dots, k/2\}$. Liitteessä B lasketaan kaavan (5.6) avulla, että joukon $A_{3a,m}$ alkiolle pätee

$$\begin{aligned} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) &= \mathbb{P}(X_{ij} = 1) \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) \\ &\approx \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \left((1-p) + \frac{k-2m+\frac{1}{2}}{n-k-1} p^2 \right) \\ &= (1-p)^2 + \frac{2(k-2m+\frac{3}{4})}{n-k-1} (1-p)p^2 + \frac{(k-2m+1)(k-2m+\frac{1}{2})}{(n-k-1)^2} p^4, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Liitteessä C osoitetaan, että

$$\sum_{m=1}^{k/2} \frac{(k-2m+1)(k-2m+\frac{1}{2})}{(n-k-1)^2} p^4 = \frac{k(k^2 - \frac{3}{4}k - 1)}{6(n-k-1)^2} p^4$$

jolloin saadaan laskettua

$$\begin{aligned} &\sum_{\{i,j,k'\} \in A_{3a}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &= \sum_{m=1}^{k/2} \sum_{\{i,j,k'\} \in A_{3a,m}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &\approx \sum_{m=1}^{k/2} \left((1-p)^2 + \frac{2(k-2m+\frac{3}{4})}{n-k-1} (1-p)p^2 + \frac{(k-2m+1)(k-2m+\frac{1}{2})}{(n-k-1)^2} p^4 \right) \\ &\approx \frac{k}{2} (1-p)^2 + \frac{2k(k+\frac{3}{4})}{2(n-k-1)} (1-p)p^2 - \frac{k(k+2)}{2(n-k-1)} (1-p)p^2 \\ &= \frac{k}{2} (1-p)^2 + \frac{k(k-\frac{1}{2})}{2(n-k-1)} (1-p)p^2, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Joukon A_{3b} alkioiden lukumäärä on

$$|\{\{i,j,k'\} \in A_{3b}\}| = \frac{k(k-1)}{2} - \frac{k}{2} = \frac{k(k-2)}{2}.$$

Merkitään $A_{3b,m,m'} = \{\ell_{ij} = m, \ell_{ik'} = m'\}$ kaikilla $1 \leq m < m' \leq k/2$. Liitteessä B lasketaan kaavan (5.6) avulla, että joukon $A_{3b,m,m'}$ alkiolle pätee

$$\begin{aligned} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) &= \mathbb{P}(X_{ij} = 1) \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) \\ &\approx \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \left((1-p) + \frac{k-2m'+\frac{1}{2}}{n-k-1} p^2 \right) \\ &= (1-p)^2 + \frac{2(k+\frac{3}{4} - (m+m'))}{n-k-1} (1-p)p^2 + \frac{(k-2m+1)(k-2m'+\frac{1}{2})}{(n-k-1)^2} p^4, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Liitteessä C osoitetaan, että

$$4 \sum_{m=1}^{k/2-1} \sum_{m'>m}^{k/2} (m+m') = \frac{k(k-2)(k+2)}{4}$$

ja

$$4 \sum_{m=1}^{k/2-1} \sum_{m'>m}^{k/2} \frac{(k-2m+1)(k-2m'+\frac{1}{2})}{(n-k-1)^2} p^4 = \frac{k(k-2)(k^2-2k-2)}{8(n-k-1)^2} p^4,$$

missä

$$4 \sum_{m=1}^{k/2-1} \sum_{m'>m}^{k/2} 1 = \frac{k(k-2)}{2},$$

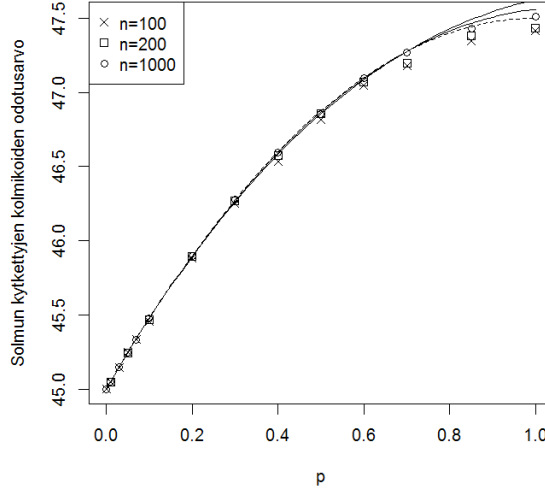
jolloin saadaan laskettua

$$\begin{aligned} & \sum_{\{i,j,k'\} \in A_{3b}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &= \sum_{m=1}^{k/2-1} \sum_{m'>m}^{k/2} \sum_{\{i,j,k'\} \in A_{3b,m,m'}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &\approx 4 \sum_{m=1}^{k/2-1} \sum_{m'>m}^{k/2} \left((1-p)^2 + \frac{2(k+\frac{3}{4}-(m+m'))}{n-k-1} (1-p)p^2 \right) \\ &= \frac{k(k-2)}{2} (1-p)^2 + \frac{k(k-2)(k+\frac{3}{4})}{n-k-1} (1-p)p^2 - \frac{k(k-2)(k+2)}{2(n-k-1)} (1-p)p^2 \\ &= \frac{k(k-2)}{2} (1-p)^2 + \frac{k(k-2)(k-\frac{1}{2})}{2(n-k-1)} (1-p)p^2, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Nyt saadaan laskettua kaavan (6.3) summan osa alkioille $\{i, j, k'\} \in A_3 = A_{3a} \cup A_{3b}$,

$$\begin{aligned} (6.7) \quad & \sum_{\{i,j,k'\} \in A_3} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\ &\approx \left(\frac{k}{2} (1-p)^2 + \frac{k(k-\frac{1}{2})}{2(n-k-1)} (1-p)p^2 \right) \\ &+ \left(\frac{k(k-2)}{2} (1-p)^2 + \frac{k(k-2)(k-\frac{1}{2})}{2(n-k-1)} (1-p)p^2 \right) \\ &= \frac{k(k-1)}{2} (1-p)^2 + \frac{k(k-\frac{1}{2})(k-1)}{2(n-k-1)} (1-p)p^2, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys on $D(G_{n,k,p}) \rightarrow 0$. Satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kytkettyjen kolmikoiden lukumäärän odotusarvo saadaan laskettua kaavoista



KUVA 6.2. Satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kytkettyjen kolmikoiden lukumäärän odotusarvo $\mathbb{E}[\Upsilon_i(G_{n,k,p})]$ kaavasta (6.8) parametrin p funktiona (käyrät), kun $k = 10$ ja $n = 100, 200, \infty$ (katkoviiva). Ristit ($n = 100$), neliöt ($n = 200$) ja ympyrät ($n = 1000$) ovat numeerisen simuloinnin keskiarvoja sadasta generoidusta verkosta.

(6.4), (6.6) ja (6.7),

$$\begin{aligned}
 (6.8) \quad \mathbb{E}[\Upsilon_i(G_{n,k,p})] &= \sum_{s=1}^3 \sum_{\{i,j,k'\} \in A_s} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1) \\
 &\approx \left(\frac{k(k-\frac{1}{2})}{2} \left(1 - \frac{1}{n-k-1} \right) p^2 - \frac{k(k-1)(k-\frac{3}{4})}{2(n-k-1)} p^3 \right) \\
 &+ \left(k(k-\frac{1}{2})(1-p)p - \frac{k(k-1)^2}{2(n-k-1)}(1-p)p^2 + \frac{k^2(k-\frac{1}{2})}{2(n-k-1)} p^3 \right) \\
 &+ \left(\frac{k(k-1)}{2}(1-p)^2 + \frac{k(k-\frac{1}{2})(k-1)}{2(n-k-1)}(1-p)p^2 \right) \\
 &= \frac{k(k-1)}{2} + \frac{k}{4}p(2-p) + \frac{k^2(3p-2) - kp}{8(n-k-1)}p^2,
 \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ tiheydelle pätee seurausten 5.5 ja 5.6 mukaan $D(G_{n,k,p}) = k/(n-1) \rightarrow 0$, kun $n \rightarrow \infty$, joten odotusarvo $\mathbb{E}[\Upsilon_i(G_{n,k,p})]$ on asymptoottisesti

$$(6.9) \quad \mathbb{E}[\Upsilon_i(G_{n,k,p})] \rightarrow \frac{k(k-1)}{2} + \frac{k}{4}p(2-p), \text{ kun } n \rightarrow \infty.$$

Kuvasta 6.2 nähdään, että tyypillisen solmun i kytkettyjen kolmikoiden odotusarvo ei juuri riipu satunnaisverkon koosta n . Kaavojen (6.8) ja (6.9) suhteellinen erotus on

$$\frac{\mathbb{E}[\Upsilon_i(G_{n,k,p})] - \mathbb{E}[\Upsilon_i(G_{n=\infty,k,p})]}{\mathbb{E}[\Upsilon_i(G_{n=\infty,k,p})]} = \frac{(k^2(3p-2) - kp)p^2}{2(n-k-1)(2k^2 - k + kp(2-p))} \approx 0,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$, joten voidaan estimoida odotusarvo $\mathbb{E}[|\Upsilon_i(G_{n,k,p})|]$ kaavalla (6.9). Tämä on sama kuin kaavassa 6.2 laskettu odotusarvo

$$\mathbb{E}[|\Upsilon_i(G_{n,k,p})|] \approx \frac{k(k-1)}{2} + \frac{1}{2} \text{Var}(k_i(G_{n,k,p})) = \frac{k(k-1)}{2} + \frac{k}{4}p(2-p).$$

Tyypillisen solmun i kytkettyjen kolmikoiden odotusarvo parametrilla $p = 0$ on $\mathbb{E}[|\Upsilon_i(G_{n,k,p=0})|] = k(k-1)/2$, joka on riippumaton parametrilla n . Parametrin p arvon kasvaessa odotusarvo $\mathbb{E}[|\Upsilon_i(G_{n,k,p})|]$ kasvaa lineaarisesti asteen $k_i(G_{n,k,p})$ varianssin kanssa. Parametrilla $p = 1$ odotusarvo $\mathbb{E}[|\Upsilon_i(G_{n,k,p=1})|]$ on kuitenkin pienempi kuin vastaavan kokoisessa satunnaisverkossa $G_{n,p'}$, jossa $k_i(G_{n,p'}) \sim \text{Poisson}(\lambda)$ parametrilla $\lambda = \bar{k} = k$. Lauseen 3.25 mukaan $\mathbb{E}[|\Upsilon_i(G_{n,p'})|] = k(k-1)/2 + k/2$, joten

$$(6.10) \quad \mathbb{E}[|\Upsilon_i(G_{n,k,p=1})|] \approx \frac{k^2}{2} - \frac{k}{4} < \frac{k^2}{2} = \mathbb{E}[|\Upsilon_i(G_{n,p'})|].$$

Odotusarvojen $\mathbb{E}[|\Upsilon_i(G_{n,k,p=1})|]$ ja $\mathbb{E}[|\Upsilon_i(G_{n,p'})|]$ suhteellinen erotus menee kuitenkin nolnaan, kun parametri k kasvaa isoksi. Solmun i kytkettyjen kolmikoiden odotusarvo kaavasta (6.9) eroaa selvästi A. Barratin ja M. Weigtin [11] lausekkeesta (6.1) käytetystä estimaatista $\frac{1}{n} \sum_{i=1}^n |\Upsilon_i(G_{n,k,p})| \approx |\Upsilon_i(H_{n,k})| = k(k-1)/2$, joka ei kasva parametrin p funktiona.

6.1.2. Solmun kolmioiden odotusarvo. Satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kolmioiden lukumäärän odotusarvo on seurauksen 3.29 mukaan

$$\mathbb{E}[|\Delta(G_n)|] = \sum_{\{i,j,k'\} \in V_n^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik',jk'\}} = 1) = \sum_{\{i,j,k'\} \in V_i^{(3)}} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1).$$

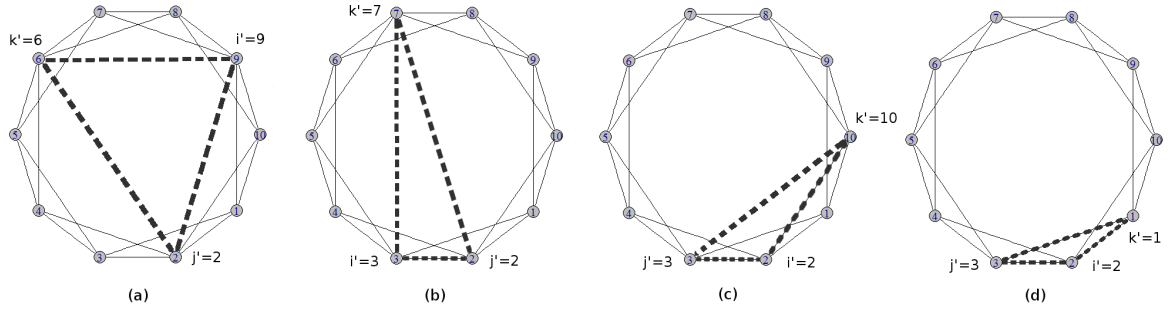
Odotusarvon $\mathbb{E}[|\Delta_i(G_{n,k,p})|]$ laskemista varten tehdään ositus solmujoukon V_n järjestämättömien kolmikoiden $V_n^{(3)}$ osajoukolle $V_i^{(3)} = \{\{i',j',k'\} \in V_n^{(3)} : i \in \{i',j',k'\}\}$. Tapahtuman $\{X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1\}$ todennäköisyys on riippumaton siitä, minkä solmun $s \in \{i',j',k'\}$ joukkoa $V_s^{(3)}$ ollaan tarkistamassa, joten voidaan olettaa, että $l_{i'j'} \leq l_{i'k'} \leq l_{j'k'}$ (ks. kuva 6.3). Olkoon $V_i^{(3)} = \bigcup_{s=1}^4 B_s$ joukon $V_i^{(3)}$ ositus erillisiin osajoukkoihin

$$\begin{aligned} B_1 &= \{l_{i'j'} > k/2, l_{i'k'} > k/2, l_{j'k'} > k/2\}, \\ B_2 &= \{l_{i'j'} \leq k/2, l_{i'k'} > k/2, l_{j'k'} > k/2\}, \\ B_3 &= \{l_{i'j'} \leq k/2, l_{i'k'} \leq k/2, l_{j'k'} > k/2\}, \\ B_4 &= \{l_{i'j'} \leq k/2, l_{i'k'} \leq k/2, l_{j'k'} \leq k/2\}, \end{aligned}$$

missä k on satunnaisverkon $G_{n,k,p}$ parametri (ks. kuva 6.3). Lauseen 3.17 avulla saadaan satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kolmioiden lukumäärän odotusarvo muotoon

$$(6.11) \quad \begin{aligned} \mathbb{E}[|\Delta_i(G_{n,k,p})|] &= \sum_{\{i',j',k'\} \in V_i^{(3)}} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ &= \sum_{s=1}^4 \sum_{\{i',j',k'\} \in B_s} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1), \end{aligned}$$

sillä $V_i^{(3)} = \bigcup_{s=1}^4 B_s$ on joukon $V_i^{(3)}$ ositus. Ensinnäkin lasketaan kaavan (6.11) summan



KUVA 6.3. Esimerkkejä pieni maailma -verkkomallin satunnaisverkon $G_{n,k,p}$ solmun $i = 2$ mahdollisista muodostuvista kolmioista (katkoviivat) parametreilla $n = 10$, $k = 4$ ja $p > 0$. Olkoon $\ell_{ij} = \min\{|j - i|, n - |j - i|\}$ metriikka. Solmujen $\{i', j', k'\} \in V_i^{(3)}$ järjestyksellä ei ole väliä, joten oletetaan, että $\ell_{i'j'} \leq \ell_{i'k'} \leq \ell_{j'k'}$. (a): $\{\ell_{i'j'} > k/2, \ell_{i'k'} > k/2, \ell_{j'k'} > k/2\}$, (b): $\{\ell_{i'j'} \leq k/2, \ell_{i'k'} > k/2, \ell_{j'k'} > k/2\}$, (c): $\{\ell_{i'j'} \leq k/2, \ell_{i'k'} \leq k/2, \ell_{j'k'} > k/2\}$ ja (d): $\{\ell_{i'j'} \leq k/2, \ell_{i'k'} \leq k/2, \ell_{j'k'} \leq k/2\}$.

osa alkiolle $\{i, j, k'\} \in B_1 = \{\ell_{i'j'} > k/2, \ell_{i'k'} > k/2, \ell_{j'k'} > k/2\}$ (ks. kuva 6.3(a)). Joukon B_1 alkioiden lukumäärä on

$$\begin{aligned} |\{\{i', j', k'\} \in B_1\}| &= \sum_{m=1}^{(n-k-1)-(k/2+1)} m = \sum_{m=1}^{n-3k/2-2} m = \frac{(n - \frac{3k}{2} - 2)(n - \frac{3k}{2} - 1)}{2} \\ &= \frac{(n - k - 1)^2 - (n - k - 1) - k(n - \frac{5}{4}k - \frac{1}{4})}{2}, \end{aligned}$$

missä $n - 5k/4 - 1/4 < n - k - 1$, kun $k > 3$. Merkitään $B_{1,m,m'} = \{\ell_{i'j'} = m, \ell_{i'k'} = m', \ell_{j'k'} = m+m'\}$ kaikilla $m' \geq m > k/2$ siten, että $\min\{m+m', n-(m+m')\} \leq n/2$. Liitteessä B osoitetaan, että joukon $B_{1,m,m'}$ alkiolle pätee

$$\begin{aligned} &\mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ &\approx \frac{k(k-1)(k-\frac{1}{4})}{(n-k-1)^3} p^3 - \frac{k^2}{4(n-k-1)^3} p^3 - \frac{3k(12k^3 - 34k^2 + 27k - 3)}{8(n-k-1)^4} p^4, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Tämä on riippumaton arvoista $m' \geq m > k/2$, jolloin

$$\begin{aligned} (6.12) \quad &\sum_{\{i', j', k'\} \in B_1} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ &\approx \left(\frac{(n-k-1)^2 - (n-k-1) - k(n - \frac{5}{4}k - \frac{1}{4})}{2} \right) \\ &\cdot \left(\frac{k(k-1)(k-\frac{1}{4})}{(n-k-1)^3} p^3 - \frac{k^2}{4(n-k-1)^3} p^3 - \frac{3k(12k^3 - 34k^2 + 27k - 3)}{8(n-k-1)^4} p^4 \right) \\ &\approx \frac{k((k-1)(4k-1) - k)}{8(n-k-1)} p^3 = \frac{k(4k^2 - 6k + 1)}{8(n-k-1)} p^3, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Joukon B_2 alkioiden lukumäärä on

$$|\{\{i', j', k'\} \in B_2\}| = 3 \sum_{m=1}^{k/2} (n - k - 1 - m) = \frac{3k(n - k - 1)}{2} - \frac{3k(k + 2)}{8}.$$

Merkitään $B_{2,m,m'} = \{\ell_{i'j'} = m, \ell_{i'k'} = m', \ell_{j'k'} = m + m'\}$ kaikilla $1 \leq m \leq k/2 < m' \leq n/2 - m$. Liitteessä B osoitetaan, että joukon $B_{2,m,m'}$ alkiolle pätee

$$\begin{aligned} & \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\ & \approx \left(\frac{(k-1)(k-\frac{1}{4})}{(n-k-1)^2} p^2 - \frac{k}{4(n-k-1)^2} p^2 - \frac{k^3 - \frac{13}{4}k^2 + 3k - \frac{1}{2}}{(n-k-1)^3} p^3 \right) \\ & \cdot \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right), \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Tämä on riippumaton arvosta $m' > k/2$, joten kaavan (6.11) summan osa alkiolle $\{i', j', k'\} \in B_2$ on

$$\begin{aligned} (6.13) \quad & \sum_{\{i', j', k'\} \in B_2} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ & = \sum_{m=1}^{k/2} \sum_{m' > k/2}^{n/2-m} \sum_{\{i', j', k'\} \in B_{2,m,m'}} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ & \approx 3 \sum_{m=1}^{k/2} (n - k - 1 - m) \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\ & \cdot \left(\frac{(k-1)(k-\frac{1}{4})}{(n-k-1)^2} p^2 - \frac{k}{4(n-k-1)^2} p^2 - \frac{k^3 - \frac{13}{4}k^2 + 3k - \frac{1}{2}}{(n-k-1)^3} p^3 \right) \\ & = 3 \left(\frac{k(n-k-1)}{2} (1-p) - \frac{k(k+2)}{8} (1-p) + \frac{k^2}{4} p^2 - \frac{k(k+1)(k+2)}{24(n-k-1)} p^2 \right) \\ & \cdot \left(\frac{(k-1)(k-\frac{1}{4})}{(n-k-1)^2} p^2 - \frac{k}{4(n-k-1)^2} p^2 - \frac{k^3 - \frac{13}{4}k^2 + 3k - \frac{1}{2}}{(n-k-1)^3} p^3 \right) \\ & \approx \frac{3k((k-1)(4k-1) - k)}{8(n-k-1)} (1-p)p^2 = \frac{3k(4k^2 - 6k + 1)}{8(n-k-1)} (1-p)p^2, \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys on $D(G_{n,k,p}) \rightarrow 0$. Joukon B_3 alkioiden lukumäärä on

$$|\{\{i', j', k'\} \in B_3\}| = 3 \sum_{m=1}^{k/2} m = 3 \frac{\frac{k}{2}(\frac{k}{2} + 1)}{2} = \frac{3k(k+2)}{8}.$$

Merkitään $B_{3,m,m'} = \{\ell_{i'j'} = m, \ell_{i'k'} = m', \ell_{j'k'} = m+m'\}$ kaikilla $1 \leq m \leq m' \leq k/2$, joille $m + m' > k/2$. Liitteessä B osoitetaan, että joukon $B_{3,m,m'}$ alkiolle pätee

$$\mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \approx \frac{k-1}{n-k-1} (1-p)^2 p,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys on $D(G_{n,k,p}) \rightarrow 0$. Tämä on riippumaton arvoista $1 \leq m \leq m' \leq k/2$, joten kaavan (6.11) summan osa alkioille $\{i', j', k'\} \in B_3$ on

$$(6.14) \quad \sum_{\{i', j', k'\} \in B_3} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\ \approx \frac{3k(k+2)}{8} \left(\frac{k-1}{n-k-1} (1-p)^2 p \right) = \frac{3k(k-1)(k+2)}{8(n-k-1)} (1-p)^2 p,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Joukon B_4 alkioden lukumäärä on

$$|\{\{i', j', k'\} \in B_4\}| = 3 \sum_{s=1}^{k/2-1} \sum_{t=1}^{k/2-s} 1 = \frac{3k(k-2)}{8}.$$

Merkitään $B_{4,s,t} = \{\ell_{i'j'} = \min\{s, t\}, \ell_{i'k'} = \max\{s, t\}, \ell_{j'k'} = s + t\}$ kaikilla $s, t \in \{1, 2, \dots, k/2\}$, joille $s + t \leq k/2$. Liitteessä B osoitetaan, että joukon $B_{4,s,t}$ alkioille pätee

$$\mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \approx (1-p)^3 + \frac{3(k + \frac{1}{4}) - 4(s+t)}{n-k-1} (1-p)^2 p^2,$$

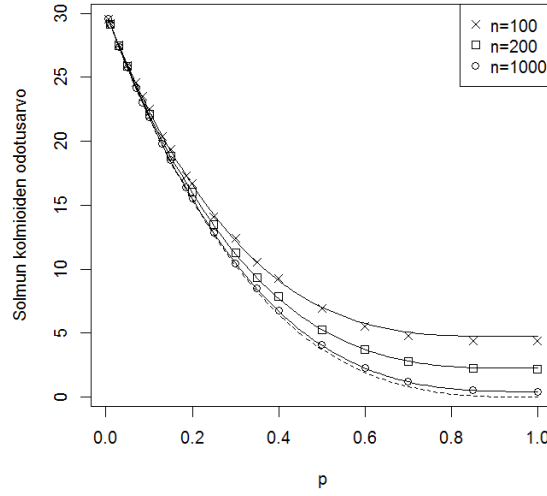
kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Liitteessä C osoitetaan, että

$$\sum_{s=1}^{k/2-1} \sum_{t=1}^{k/2-s} (s+t) = \frac{k(k+2)(k-2)}{24},$$

joten kaavan (6.11) summan osa alkioille $\{i', j', k'\} \in B_4$ on

$$(6.15) \quad \sum_{\{i', j', k'\} \in B_4} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ = 3 \sum_{s=1}^{k/2-1} \sum_{t=1}^{k/2-s} \sum_{\{i', j', k'\} \in B_{4,s,t}} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1) \\ \approx 3 \sum_{s=1}^{k/2-1} \sum_{t=1}^{k/2-s} \left((1-p)^3 + \frac{3(k + \frac{1}{4}) - 4(s+t)}{n-k-1} (1-p)^2 p^2 \right) \\ = \frac{3k(k-2)}{8} (1-p)^3 + \frac{9k(k-2)(k + \frac{1}{4})}{8(n-k-1)} (1-p)^2 p^2 - \frac{k(k+2)(k-2)}{2(n-k-1)} (1-p)^2 p^2 \\ = \frac{3k(k-2)}{8} (1-p)^3 + \frac{k(k-2)(9(k + \frac{1}{4}) - 4(k+2))}{8(n-k-1)} (1-p)^2 p^2 \\ = \frac{3k(k-2)}{8} (1-p)^3 + \frac{k(k-2)(5k - \frac{23}{4})}{8(n-k-1)} (1-p)^2 p^2,$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kolmioiden lukumäärän odotusarvo saadaan estimoitua kaavoista (6.12),



KUVA 6.4. Satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kolmioiden lukumäärän odotusarvo $\mathbb{E} [|\Delta_i(G_{n,k,p})|]$ kaavasta (6.16) parametrin p funktiona (käyrät), kun $k = 10$ ja $n = 100, 200, 1000, \infty$ (katkovii-va). Ristit ($n = 100$), neliöt ($n = 200$) ja ympyrät ($n = 1000$) ovat numeerisen simuloinnin keskiarvoja sadasta generoiduista verkoista.

(6.13), (6.14) ja (6.15),

$$\begin{aligned}
 (6.16) \quad \mathbb{E} [|\Delta_i(G_{n,k,p})|] &= \sum_{s=1}^4 \sum_{\{i,j,k'\} \in B_s} \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\
 &\approx \left(\frac{k(4k^2 - 6k + 1)}{8(n-k-1)} p^3 \right) + \left(\frac{3k(4k^2 - 6k + 1)}{8(n-k-1)} (1-p)p^2 \right) \\
 &+ \left(\frac{3k(k-1)(k+2)}{8(n-k-1)} (1-p)^2 p \right) \\
 &+ \left(\frac{3k(k-2)}{8} (1-p)^3 + \frac{k(k-2)(5k - \frac{23}{4})}{8(n-k-1)} (1-p)^2 p^2 \right) \\
 &\approx \frac{3k(k-2)}{8} (1-p)^3 + \frac{4k^3 - 6k^2 + k}{8(n-k-1)} p^2 (3-2p) \\
 &+ \frac{k^3(3+5p) + k^2(3 - \frac{63}{4}p) + k(\frac{23}{2}p - 6)}{8(n-k-1)} (1-p)^2 p,
 \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ tiheydelle pätee $D(G_{n,k,p}) = k/(n-1) \rightarrow 0$, kun $n \rightarrow \infty$, joten odotusarvo $\mathbb{E} [|\Delta_i(G_{n,k,p})|]$ on asymptoottisesti

$$(6.17) \quad \mathbb{E} [|\Delta_i(G_{n,k,p})|] \rightarrow \frac{3k(k-2)}{8} (1-p)^3, \text{ kun } n \rightarrow \infty.$$

Satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i kolmioiden lukumäärän odotusarvo parametrilla $p = 0$ on $\mathbb{E} [|\Delta_i(G_{n,k,p=0})|] = 3k(k-2)/8$, joka on riippumaton parametrilla n . Parametrin p arvon kasvaessa odotusarvo pienenee siten, että odotusarvon $\mathbb{E} [|\Delta_i(G_{n,k,p})|]$ käyrän jyrkkyys riippuu tiheydestä $D(G_{n,k,p}) = k/(n-1)$ (ks. kuva

6.4). Satunnaisverkon $G_{n,k,p}$ tiheys määrää myös käyrän päätepisteen, kun $p = 1$, joka on selkeästi nolasta eroava, kun tiheys $D(G_{n,k,p})$ on suhteellisen iso. Parametrilla $p = 1$ se on silti hieman pienempi kuin täysin satunnaisessa verkossa

$$(6.18) \quad \mathbb{E} [|\Delta_i(G_{n,k,p=1})|] \approx \frac{4k^3 - 6k^2 + k}{8(n - k - 1)} < \frac{k^3}{2(n - 1)} \approx \mathbb{E} [|\Delta_i(G_{n,p'})|],$$

missä $G_{n,p'}$ on samankokoinen satunnaisverkko parametrilla $p' = k/(n - 1)$, jolloin $\mathbb{E} [k_i(G_{n,p'})] = k$. Odotusarvojen $\mathbb{E} [|\Delta_i(G_{n,p'})|]$ ja $\mathbb{E} [|\Delta_i(G_{n,k,p=1})|]$ erotus menee kuitenkin noltaan, kun $D(G_{n,k,p}) = k/(n - 1) \rightarrow 0$. Solmun i kolmioiden lukumäärän odotusarvo on asymptoottisesti sama kuin A. Barratin ja M. Weigtin [11] lausekkeessa (6.1) käytetty estimaatti $\frac{1}{n} \sum_{i=1}^n |\Delta_i(G_{n,k,p})| \approx |\Delta_i(H_{n,k})|(1 - p)^3 = 3k(k - 2)/8 \cdot (1 - p)^3$, joka pienenee, kuten funktio $f(p) = (1 - p)^3$.

6.1.3. Satunnaisverkon transitiivisuuserroin. Wattsin ja Strogatzin pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ transitiivisuuserroin $c'(G_{n,k,p})$ saadaan määritelmän 3.38 mukaan laskettua kaavasta $c'(G_{n,k,p}) = \mathbb{E} [|\Delta_i(G_{n,k,p})|] / \mathbb{E} [|\Upsilon_i(G_{n,k,p})|]$. Kaavoista 6.9 ja 6.16 saadaan laskettua

$$(6.19) \quad c'(G_{n,k,p}) \approx \frac{3(k - 2)(1 - p)^3}{4(k - 1) + 2p(2 - p)} + \frac{4k^2 - 6k + 1}{(n - k - 1)(4(k - 1) + 2p(2 - p))} p^2(3 - 2p) \\ + \frac{k^2(3 + 5p) + k(3 - \frac{63}{4}p) - 6 + \frac{23}{2}p}{(n - k - 1)(4(k - 1) + 2p(2 - p))} (1 - p)^2 p,$$

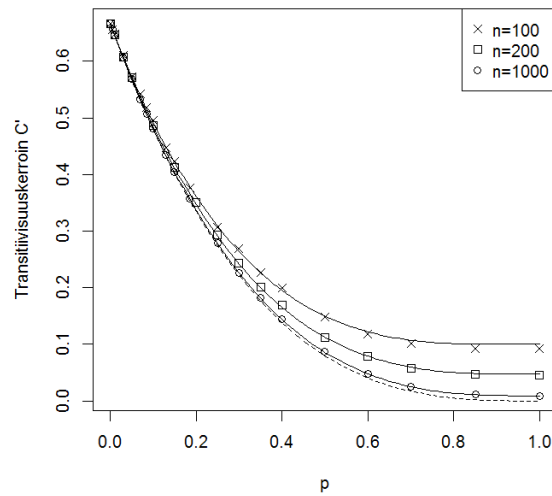
kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ tiheydelle pätee $D(G_{n,k,p}) = k/(n - 1) \rightarrow 0$, kun $n \rightarrow \infty$, joten satunnaisverkon $G_{n,k,p}$ transitiivisuuserroin on asymptoottisesti

$$(6.20) \quad c'(G_{n,k,p}) \rightarrow \frac{3(k - 2)(1 - p)^3}{4(k - 1) + 2p(2 - p)}, \text{ kun } n \rightarrow \infty.$$

Satunnaisverkon $G_{n,k,p}$ transitiivisuuserroin parametrilla $p = 0$ on $c'(G_{n,k,p}) = 3(k - 2)/4(k - 1)$, joka on riippumaton parametrilla n . Parametrin p arvon kasvaessa transitiivisuuserroin $c'(G_{n,k,p})$ pienenee siten, että käyrän jyrkkyys riippuu tiheydestä $D(G_{n,k,p}) = k/(n - 1)$ (ks. kuva 6.5). Satunnaisverkon $G_{n,k,p}$ tiheys määrää myös käyrän päätepisteen, kun $p = 1$, joka kaavan (6.19) mukaan on

$$c'(G_{n,k,p=1}) \approx \frac{k}{n - k - 1} \left(1 - \frac{1}{k - \frac{1}{2}}\right) \sim \frac{k}{n - 1} = c'(G_{n,p'}),$$

missä $G_{n,p'}$ on samankokoinen satunnaisverkko parametrilla $p' = k/(n - 1)$, jolloin $\mathbb{E} [k_i(G_{n,p'})] = k$. Tämä eroaa A. Barratin ja M. Weigtin [11] heuristisesta lausekkeesta $c'(G_{n,k,p}) = 3(k - 2)/(4(k - 1)) \cdot (1 - p)^3$ kaavasta (6.1), jossa $c'(G_{n,k,p}) = 0$, kun $p = 1$. Barrat ja Weigt väittävät, että heidän esittämänsä lausekkeen tarkkuus $|c'(G_{n,k,p}) - 3(k - 2)/(4(k - 1)) \cdot (1 - p)^3|$ on kertaluokkaa $O(1/n)$. Kaava (6.20) osoittaa kuitenkin, että asymptoottiselle erotukselle pätee $|c'(G_{n,k,p}) - 3(k - 2)/(4(k - 1)) \cdot (1 - p)^3| \neq O(1/n)$, kun parametri k on pieni.



KUVA 6.5. Satunnaisverkon $G_{n,k,p}$ transitiivisuuskertoimen $c'(G_{n,k,p})$ kaavasta (6.19) parametrin p funktiona (käyrät), kun $k = 10$ ja $n = 100, 200, 1000, \infty$ (katkoviiva). Ristit ($n = 100$), neliöt ($n = 200$) ja ympyrät ($n = 1000$) ovat numeerisen simuloinnin keskiarvoja sadasta generoiduista verkoista.

6.2. Pieni maailma -verkkomallin satunnaisverkon klusterointikerroin

Seuraavaksi yritetään laskea matemaattinen lauseke satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ satunnaisverkon $G_{n,k,p}$ klusterointikertoimelle $c(G_{n,k,p})$ parametrin n, k ja p funktiona. Määritelmästä 3.30 ja lauseesta 3.31 saadaan laskettua satunnaisverkon $G_{n,k,p}$ klusterointikerroin kaavasta

$$c(G_{n,k,p}) = \mathbb{E}[C_i(G_n)] = \sum_{k': \mathbb{P}(k_i(G_n)=k') > 0} \mathbb{E}[C_i(G_n) \mid k_i(G_n) = k'] \mathbb{P}(k_i(G_n) = k').$$

Kaavan (5.7) todennäköisyysjakaumasta nähdään, että se tuottaa monimutkaisen lausekkeen satunnaisverkon $G_{n,k,p}$ klusterointikertoimelle $c(G_{n,k,p})$. Alain Barrat ja Martin Weigt [11] osoittavat numeerisella simuloinnilla, että pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ generoitujen verkkojen klusterointikerroin ja transitiivisuuskertoimen eivät juuri eroa toisistaan. Kuvassa 6.6 osoitetaan, että klusterointikertoimen $c(G_{n,k,p})$ ero kaavan (6.19) transitiivisuuskertoimesta on pieni, joten käytetään kaavaa (6.19) estimoimaan klusterointikertoimen $c(G_{n,k,p})$ lauseketta

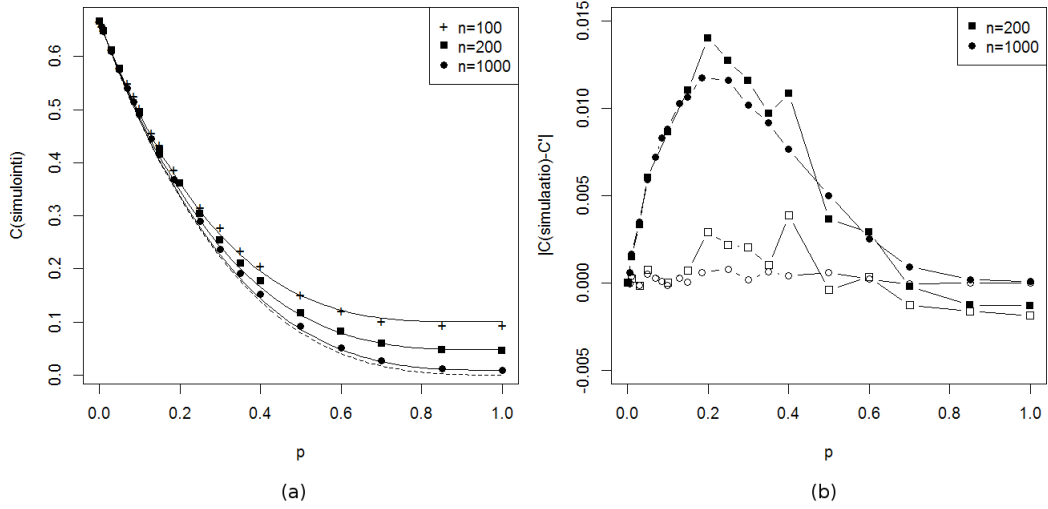
$$(6.21) \quad c(G_{n,k,p}) \approx c'(G_{n,k,p}).$$

On ehkä mielekkäämpää tutkia miksi satunnaisverkon $G_{n,k,p}$ klusterointikerroin ja transitiivisuuskertoimen ovat lähes samoja, kuin yrittää määrittää klusterointikertoimen $c(G_{n,k,p})$ monimutkainen lauseke. Ensin esitetään lause, joka kertoo klusterointikertoimen $c(G_{n,k,p})$ ja $c_{V'}(G_{n,k,p})$ yhteyden satunnaisverkossa $G_{n,k,p}$.

LAUSE 6.1. *Olkoon $\mathcal{G}_{n,k,p}$ satunnaisverkkomalli parametrilla $k \geq 4$. Tällöin satunnaisverkon $G_{n,k,p}$ klusterointikertoimelle pätee*

$$c(G_{n,k,p}) = c_{V'}(G_{n,k,p}),$$

missä $c_{V'}(G_{n,k,p}) = \mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p}) \geq 2]$.



KUVA 6.6. (a): Satunnaisverkon $G_{n,k,p}$ klusterointikerroin c laskettuna numeerisella simuloinnilla parametrin p eri arvoilla, kun $k = 10$ ja $n = 100$ (ristit), 200 (neliöt) ja 1000 (ympyrät). Arvot ovat keskiarvoja sadasta generoidusta verkosta. Kuvaan on lisätty transitiivisuuskertoimen $c'(G_{n,k,p})$ vastaavat käyrät kaavasta (6.19). (b): Numeerisella simuloinnilla lasketut klusterointikertoimien C (tummat symbolit) ja transitiivisuuskertoimien C' (vaaleat symbolit) erot kaavasta (6.19).

TODISTUS. Olkoon parametri $k \geq 4$. Lauseen 5.7 mukaan satunnaisverkolle $G_{n,k,p}$ pätee $k_i(G_{n,k,p}) \geq k/2 \geq 2$ kaikilla $i \in V_n$, joten $\mathbb{P}(k_i(G_{n,k,p}) \geq 2) = 1$. Lauseesta 3.37 saadaan nyt laskettua $c(G_{n,k,p}) = \mathbb{P}(k_i(G_{n,k,p}) \geq 2)c_{V'}(G_{n,k,p}) = c_{V'}(G_{n,k,p})$. \square

Seuraavaksi osoitetaan, että satunnaisverkon $G_{n,k,p}$ klusterointikerroin $c(G_{n,k,p})$ on sama kuin sen transitiivisuukerroin $c'(G_{n,k,p})$, kun parametri $p = 0$.

LAUSE 6.2. *Olkoon $G_{n,k,p}$ satunnaisverkko, jossa parametri $p = 0$. Tällöin*

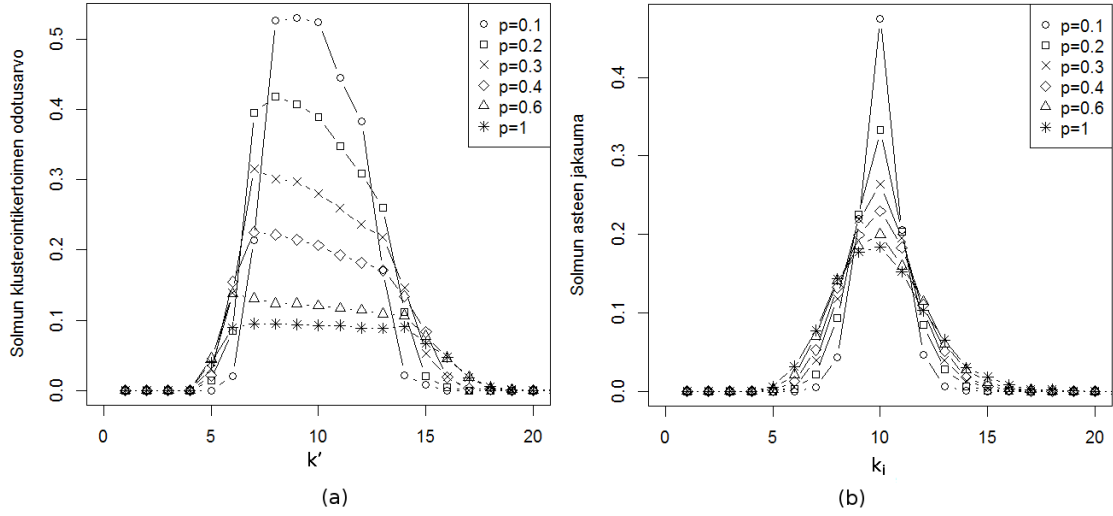
$$c(G_{n,k,p=0}) = c'(G_{n,k,p=0}).$$

TODISTUS. Satunnaisverkko $G_{n,k,p}$ on parametrilla $p = 0$ k -säännöllinen rengashila, joten lauseen 2.21 mukaan $c(G_{n,k,p=0}) = c'(G_{n,k,p=0})$. \square

Kun parametri $p > 0$, saadaan lauseella 3.40 osoitettua, että $c(G_{n,k,p}) = c'(G_{n,k,p})$, jos satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i klusterointikertoimen ehdollinen odotusarvo $\mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p}) = k']$ on riippumaton asteen $k_i(G_{n,k,p})$ arvosta $k' \geq k/2$. Seuraavaksi osoitetaan, että näin käy parametrilla $p = 1$.

Olkoon $G_{n,k,p}$ satunnaisverkko parametrilla $p = 1$, jossa tiheys $D(G_{n,k,p}) = k/(n-1)$ on pieni. Tällöin kaavojen (5.4) ja (5.6) mukaan $\mathbb{P}(X_{ij} = 1) \approx k/(n-k-1)$ kaikilla $\{i, j\} \in V_n^{(2)}$. Lemman 3.34 mukaan satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i klusterointikertoimen ehdollinen odotusarvo ehdolla $\{k_i(G_{n,k,p}) = k' \geq k/2\}$ on

$$\mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p}) = k'] = \frac{1}{\binom{k'}{2}} \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1 \mid k_i(G_{n,k,p}) = k'),$$



KUVA 6.7. (a): Pieni maailma -verkkomallin k' -asteiden solmujen keskimääräinen klusterointikerroin asteen k' funktiona $C^{(k')} = \frac{1}{n_{k'}} \sum_{i:k_i=k'} C_i$, jossa $n_{k'} = |\{i \in V_n : k_i = k'\}|$. Symbolit ovat numeerisen simuloinnin keskiarvoja sadasta generoidusta verkosta, parametrilla $n = 100$, $k = 10$ ja parametrin p eri arvoilla. (b): Asteen k_i vastaavat jakaumat.

kun $\mathbb{P}(k_i(G_{n,k,p}) = k') > 0$. Indikaattorin $\mathbb{1}_{\{ij,ik,jk\}}$ määritelmän mukaan $\mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1 \mid k_i(G_{n,k,p}) = k') = \mathbb{P}(X_{ij} = 1, X_{ik} = 1, X_{jk} = 1 \mid k_i(G_{n,k,p}) = k')$, missä tapahtuma $\{X_{jk} = 1\}$ on (lähes) ehdollisesti riippumaton tapahtumasta $\{X_{ij} = 1, X_{ik} = 1\}$ ehdolla $\{k_i(G_{n,k,p}) = k'\}$. Tapahtuma $\{X_{jk} = 1\}$ on lisäksi riippumaton tapahtumasta $\{k_i(G_{n,k,p}) = k'\}$, joten määritelmistä 3.13 ja 3.14 saadaan $\mathbb{P}(\mathbb{1}_{\{ij,ik,jk\}} = 1 \mid k_i = k') \approx \mathbb{P}(X_{jk} = 1) \mathbb{P}(X_{ij} = 1, X_{ik} = 1 \mid k_i(G_{n,k,p}) = k')$. Ehdollinen todennäköisyys $\mathbb{P}(X_{ij} = 1, X_{ik} = 1 \mid k_i(G_{n,k,p}) = k')$ saadaan laskettua *hypergeometrisen jakauman* [2] avulla, $\mathbb{P}(X_{ij} = 1, X_{ik} = 1 \mid k_i(G_{n,k,p}) = k') \approx k'(k' - 1)/((n - 1)(n - 2))$, kun $k' \geq k/2$. Lemmasta 3.34 saadaan nyt laskettua satunnaisverkon $G_{n,k,p}$ tyypillisen solmun i klusterointikertoimen ehdollinen odotusarvo ehdolla $\{k_i(G_{n,k,p}) = k' \geq k/2\}$,

$$\begin{aligned} \mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p}) = k'] &\approx \frac{1}{\binom{k'}{2}} \sum_{\{i,j,k\} \in V_i^{(3)}} \mathbb{P}(X_{jk} = 1) \mathbb{P}(X_{ij} = 1, X_{ik} = 1 \mid k_i = k') \\ &\approx \frac{2}{k'(k' - 1)} \sum_{\{i,j,k\} \in V_i^{(3)}} \frac{k}{n - k - 1} \frac{k'(k' - 1)}{(n - 1)(n - 2)} = \frac{k}{n - k - 1}, \end{aligned}$$

joka on riippumaton asteen $k_i(G_{n,k,p})$ arvosta $k' \geq k/2$. Tällöin lauseen 3.40 seuraus on, että parametrilla $p = 1$ pätee

$$c(G_{n,k,p=1}) \approx c'(G_{n,k,p=1}).$$

Ehdollista todennäköisyyttä $\mathbb{P}(X_{ij} = 1, X_{ik} = 1 \mid k_i(G_{n,k,p}) = k')$ ei voida laskea hypergeometrisella jakaumalla, kun parametri p on selkeästi pienempi kuin yksi. Tällöin kaavoista (5.4) ja (5.6) nähdään, että solmuparien kytkentätodennäköisyydet

eroavat paljon toisistaan. Kyseisessä tilanteessa voidaan käyttää *moniulotteista Fisherin epäkeskeistä hypergeometrista jakaumaa* [21]. Tämä tuottaa kuitenkin monimutkaisia kaavoja, joten tutkitaan ehdollista odotusarvoa $\mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p}) = k']$ numeerisella simuloinnilla.

Kuvassa 6.7 nähdään k' -asteiden solmujen keskimääräinen klusterointikerroin asteen k' funktiona, $C^{(k')} = \frac{1}{n_{k'}} \sum_{i:k_i(G_{n,k,p})=k'} C_i$, jossa $n_{k'} = |\{i \in V_n : k_i(G_{n,k,p}) = k'\}|$. Tämä poikkeaa hyvin paljon asteen $k_i(G_{n,k,p})$ jakaumasta, sillä klusterointikerroin $C^{(k')}$ kasvaa, kun asteen $k_i(G_{n,k,p})$ arvo k' pienenee. Klusterointikerroin $C^{(k')}$ vähenee lähes lineaarisesti asteen $k_i(G_{n,k,p})$ odotusarvon k ympäristössä, kun $p > 0$. Lemman 2.20 ja lauseen 2.21 mukaan verkon klusterointikerroin C painottaa solmujen klusterointikertoimia C_i yhtä paljon, kun taas transitiivisuuskerroin C' painottaa enemmän sellaisia solmujen $i \in V$ klusterointikertoimia C_i , joiden aste k_i on iso. Tästä seuraa, että satunnaisverkon $G_{n,k,p}$ klusterointikerroinille pätee

$$c(G_{n,k,p}) > c'(G_{n,k,p}),$$

kun parametri $p > 0$. Asteen $k_i(G_{n,k,p})$ jakauman terävästä huipusta ja pienestä varianssista johtuen satunnaisverkon $G_{n,k,p}$ klusterointikerroin $c(G_{n,k,p}) = \mathbb{E}[C_i(G_{n,k,p})] = \mathbb{E}[\mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p})]]$ painottaa eniten asteen $k_i(G_{n,k,p})$ odotusarvon k klusterointikerrointa C_k . Satunnaisverkon $G_{n,k,p}$ klusterointikerroin on siis vain vähän isompi kuin sen transitiivisuuskerroin, kuten nähdään kuvassa 6.6(b).

Johtopäätös

Tämän tutkielman alkupuolella on määritelty symmetrisesti jakautuneen satunnaisverkon G_n eri klusterointikertoimet $c(G_n) = \mathbb{E}[C_i(G_n)]$ ja $c_{V'}(G_n) = \mathbb{E}[C_i(G_n) \mid k_i(G_n) \geq 2]$, joiden ero jää usein huomioimatta tieteellisissä julkaisuissa. Satunnaisverkon G_n solmun i asteen $k_i(G_n)$ jakaumasta saadaan laskettua, että $c(G_n) = \mathbb{P}(k_i(G_n) \geq 2)_{c_{V'}(G_n)}$. Yleisen verkon G transitiivisuuskertoimen yleisimmin käytetty tilastollinen kaava on $C'(G) = 3|\Delta(G)|/|\Upsilon(G)|$, josta tieteellisissä julkaisuissa ei ole tehty selkeää määritelmää satunnaisverkolle. Transitiivisuuskerroin voidaan kuitenkin kirjoittaa yhtäpitävässä muodossa $C'(G) = \frac{1}{n} \sum_{i=1}^n |\Delta_i| / (\frac{1}{n} \sum_{i=1}^n |\Upsilon_i|)$, joten tässä tutkielmassa symmetrisesti jakautuneen satunnaisverkon G_n transitiivisuuskerroin on määritelty kaavalla $c'(G_n) = \mathbb{E}[|\Delta_i(G_n)|] / \mathbb{E}[|\Upsilon_i(G_n)|] = 3\mathbb{E}[|\Delta(G_n)|] / \mathbb{E}[|\Upsilon(G_n)|]$.

Transitiivisuuskerroin C' on tieteellisissä julkaisuissa osoitettu eri klusterointikertoimeksi kuin C tai $C_{V'}$. Transitiivisuuskerroin painottaa enemmän sellaisten solmujen klusterointikertoimia C_i , joiden aste on iso. Tässä tutkielmassa on todistettu, että symmetrisesti jakautuneessa satunnaisverkossa $c_{V'}(G_n) = c'(G_n)$, kun satunnaisverkon solmuilla on sama aste, kuten satunnaisverkossa $G_{n,r\text{-reg}}$. Lisäksi on todistettu, että $c_{V'}(G_n) = c'(G_n)$, kun symmetrisesti jakautuneen satunnaisverkon solmun i klusterointikertoimen ehdollinen odotusarvo $\mathbb{E}[C_i(G_n) \mid k_i(G_n) = k']$ on riippumaton asteen $k_i(G_n)$ arvosta $k' \geq 2$, kuten satunnaisverkossa $G_{n,p}$.

Tämän tutkielman loppupuolella on tutkittu Wattsin ja Strogatzin satunnaisverkkomallin $\mathcal{G}_{n,k,p}$ klusteroituneisuutta. Satunnaisverkon $G_{n,k,p}$ transitiivisuuskertoimelle $c'(G_{n,k,p})$ on laskettu matemaattinen lauseke parametrien n , k ja p funktiona. Lauseke osoittaa, että harvassa satunnaisverkkomallissa $\mathcal{G}_{n,k,p}$ transitiivisuuskerroin $c'(G_{n,k,p}) \rightarrow 3(k-2)(1-p)^3 / (4(k-1) + 2p(2-p))$, kun $n \rightarrow \infty$. Pienemmässä satunnaisverkossa $G_{n,k,p}$ transitiivisuuskertoimen käyrä vähenee parametrin p funktiona kohti täysin satunnaisen verkon arvoa $c'(G_{n,k,p=1}) \approx k / (n-1) > 0$.

Wattsin ja Strogatzin satunnaisverkon $G_{n,k,p}$ klusterointikertoimen $c(G_{n,k,p})$ lausekkeen määrittäminen on tuottanut vaikeuksia, sillä laskuissa esiintyy mahdollisia todennäköisyyksiä sekä asteen $k_i(G_{n,k,p})$ monimutkaisen jakauman todennäköisyyksiä. Tässä tutkielmassa on kuitenkin osoitettu, että $c(G_{n,k,p}) = c'(G_{n,k,p})$ parametrilla $p = 0$ ja $c(G_{n,k,p}) \approx c'(G_{n,k,p})$ parametrilla $p = 1$. Numeerisen simuloinnin avulla on lisäksi näytetty, että ehdollinen odotusarvo $\mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p}) = k']$ asteen $k_i(G_{n,k,p})$ arvon k' funktiona on lähes vähenevä funktio, kun $p > 0$, joten $c(G_{n,k,p}) > c'(G_{n,k,p})$. Klusterointi- ja transitiivisuuskertoimen erotus on kuitenkin havaittu hyvin pieneksi.

Wattsin ja Strogatzin satunnaisverkkomallille on tullut kilpailevia satunnaisverkkomalleja, joissa asteen jakauma on tietyille verkostoille realistisempi. Tässä tutkielmassa on kuitenkin osoitettu, että Wattsin ja Strogatzin satunnaisverkon $G_{n,k,p}$ satunnaisuuttujan $\mathbb{E}[C_i(G_{n,k,p}) \mid k_i(G_{n,k,p})]$ jakauma muistuttaa asteen $k_i(G_{n,k,p})$ odotusarvon ympäristössä jakaumaa, jota tavataan esimerkiksi sosiaalisissa verkostoissa.

LIITE A

Pieni maailma -verkkomallille tehty approksimaatio

Pieni maailma -verkkomallin kytkentämenetelmässä solmujen asteet voivat vaihdella menetelmän aikana. Tämä vaikuttaa satunnaisverkon $G_{n,k,p}$ solmuparien kytkentätodennäköisyyksiin. Merkitään satunnaismuuttujalla $k_{i,m}$ solmun $i \in V_n$ senhetkinen aste, kun algoritmin 1 kytkentämenetelmä on solmun i ulosmenevän linkin (i, j) kohdalla, jolle $\ell_{ij} = m$. Satunnaismuuttujan $k_{i,m}$ arvo on summa solmun i lähtöasteesta $k_i^+ = k/2$ ja senhetkisestä tuloasteesta $k_{i,m}^-$, kun algoritmin 1 kytkentämenetelmä on solmun i ulosmenevän linkin (i, j) kohdalla, jolle $\ell_{ij} = m$. Tällöin todennäköisyys, jolla linkki (i, j) kytketään uuteen satunnaisesti valittuun maalisolmuun j' on

$$(A.1) \quad \frac{p}{n - k_{i,m} - 1}.$$

Kyseinen todennäköisyys riippuu satunnaismuuttujan $k_{i,m}$ arvosta, joka taas riippuu solmun i järjestysluvusta $t(i) \in \{1, 2, \dots, n\}$ algoritmin 1 kytkentämenetelmässä. Solmun i järjestysluku $t(i)$ on satunnaismuuttuja, sillä se määrittyy algoritmin 1 satunnaisesti valitusta aloitussolmusta $i' \in V_n$. Satunnaismuuttuja $k_{i,m=1} = k_{i,1}$ voidaan kirjoittaa muodossa

$$k_{i,1} = k_i^+ + k_{i,1}^- = k/2 + X_{i,1} + Y_{i,1},$$

jossa $X_{i,1}$ on indikaattorisatunnaismuuttuja tapahtumalle, jossa solmun i sisäänmenevä linkki (j, i) , jolle $\ell_{ij} = 1$, pysyy paikallaan todennäköisyydellä $1 - p$. Tällöin satunnaismuuttuja $X_{i,1}$ noudattaa jakaumaa

$$X_{i,1} \sim \text{Bernoulli}(1 - p).$$

Satunnaismuuttuja $Y_{i,1}$ ilmaisee linkkien lukumäärän, jotka on kytketty uudelleen *kohti* maalisolmua i . Kyseiset linkit ovat joukon $\{j \in V_n : t(j) < t(i) - k/2\}$ lähtösolmujen linkkejä (j, k') , joille $\ell_{jk'} = 1$. Satunnaismuuttuja $Y_{i,1}$ saadaan kirjoitettua muotoon

$$Y_{i,1} = \sum_{s=1}^{t(i)-k/2-1} Z_{i,1}^{(s)}.$$

$Z_{i,1}^{(s)}$ on indikaattorisatunnaismuuttuja tapahtumalle, jossa lähtösolmun $j \in V_n$ linkki (j, k') , jolle $\ell_{jk'} = 1$, kytketään uudelleen maalisolmuun i , kun $t(j) = s$. Tällöin satunnaismuuttujan $Z_{i,1}^{(s)}$, $s \in \{1, 2, \dots, t(i) - k/2 - 1\}$, jakauma on

$$Z_{i,1}^{(s)} \sim \text{Bernoulli}\left(\frac{p}{n - k_{j,1} - 1}\right).$$

Satunnaismuuttuja $k_{j,1}$ voidaan jälleen kirjoittaa muodossa $k_{j,1} = k/2 + X_{j,1} + Y_{j,1}$ ja niin edelleen. Kaavan (A.1) kytkentätodennäköisyyden määrittäminen tuottaa hankalan rekursiivisen laskun, joten tehdään sille approksimaatio

$$(A.2) \quad \frac{p}{n - k_{i,m} - 1} \approx \frac{p}{n - k - 1},$$

jossa seurauksen 5.4 ja kaavan 5.8 mukaan parametri $k = \bar{k}(G_{n,k,p}) \approx \mathbb{E}[k_i(G_{n,k,p})]$. Satunnaismuuttujalle $k_{i,1}$, kirjoitettuna muodossa $k_{i,1} = k/2 + X_{i,1} + Y_{i,1}$, pätee nyt $X_{i,1} \sim \text{Bernoulli}(1 - p)$ ja

$$Y_{i,1} \sim \text{Bin} \left(t(i) - k/2 - 1, \frac{p}{n - k - 1} \right),$$

sillä $Y_{i,1} = \sum_{s=1}^{t(i)-k/2-1} Z_{i,1}^{(s)}$ on riippumattomien Bernoullijakautuneiden satunnaismuuttujien $Z_{i,1}^{(s)} \sim \text{Bernoulli}(p/(n - k - 1))$ summana binomijakautunut.

Alain Barrat ja Martin Weigt [11] ovat empiirisissä kokeissaan havainneet, että kaavan (A.1) virhe on pieni, kun satunnaisverkon $G_{n,k,p}$ koko n on iso. Lisäksi empiiristen kokeiden perusteella harvalle satunnaisverkkomallille $\mathcal{G}_{n,k,p}$ pätee $\mathbb{P}(k_{i,m} = n - 1) \approx 0$, kun satunnaisverkon $G_{n,k,p}$ koko n on iso.

LIITE B

Linkkien ehdolliset todennäköisyydet pieni maailma -verkkomallissa

Satunnaisverkon $G_{n,k,p}$ solmuparin $\{i, k'\} \in V_n^{(2)}$ ehdollinen kytkentätodennäköisyys $\mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1, \ell_{ij} = m, \ell_{ik'} = m')$ riippuu metriikan ℓ arvojen lisäksi siitä, kuinka monta ulosmenevää linkkiä solmuparin $\{i, k'\}$ solmuilla on käytettävissä, ehdolla $\{X_{ij} = 1\}$. Kolmikon $\{i, j, k'\} \in V_i^{(3)}$ solmujen järjestyksellä ei ole väliä, joten voidaan olettaa, että $\ell_{ij} \leq \ell_{ik'}$. Merkitään $A_{m,m'} = \{\ell_{ij} = m, \ell_{ik'} = m'\}$ kaikilla $m \leq m'$ ja tapahtuma $B_{ij}^{(s)} :=$ ”linkki ij on solmun $s \in \{i, j\}$ ulosmenevä linkki”. Tällöin joukon $A_{m,m'}$ alkiolle pätee $\mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1, B_{ij}^{(j)}) = \mathbb{P}(X_{ik'} = 1)$. Tapahtumassa $\{X_{ij} = 1, B_{ij}^{(i)}\}$ solmuparia $\{i, j\}$ kytkevä linkki on solmun i ulosmenevä linkki. Solmuparilla $\{i, k'\}$ on tällöin käytössään yksi ulosmenevä linkki vähemmän, joten joukon $A_{m,m'}$ alkiolle pätee $\mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1, B_{ij}^{(i)}) < \mathbb{P}(X_{ik'} = 1)$. Alkiolle $\{i, j, k'\} \in A_{m,m'}$ saadaan laskettua solmuparin $\{i, k'\}$ ehdollinen kytkentätodennäköisyys ehdolla $\{X_{ij} = 1\}$ lauseesta 3.17,

$$\begin{aligned}
 (B.1) \quad & \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) \\
 &= \sum_{s \in \{i, j\}} \mathbb{P}(B_{ij}^{(s)} \mid X_{ij} = 1) \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1, B_{ij}^{(s)}) \\
 &= \frac{1}{2} \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1, B_{ij}^{(i)}) + \frac{1}{2} \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1, B_{ij}^{(j)}),
 \end{aligned}$$

sillä pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ lähtökohtaisen rengashilan $H_{n,k}$ symmetrian perusteella $\mathbb{P}(B_{ij}^{(s)} \mid X_{ij} = 1) = 1/2$ kaikilla $s \in \{i, j\}$. Kaavan (B.1) ehdollinen todennäköisyys alkiolle $\{i, j, k'\} \in A_{m,m'}$, joille pätee $1 \leq \ell_{ij} = m \leq \ell_{ik'} = m' \in \{k/2 + 1, k/2 + 1, \dots, n/2\}$, on kaavan (5.4) mukaan

$$\begin{aligned}
 & \mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) \\
 & \approx \frac{1}{2} \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 + \frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \\
 & = \frac{k - \frac{1}{2}}{n-k-1} p - \frac{(k-1)^2}{2(n-k-1)^2} p^2,
 \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Kaavan (B.1) ehdollinen todennäköisyys alkiolle $\{i, j, k'\} \in A_{m,m'}$, joille pätee $1 \leq \ell_{ij} = m \leq \ell_{ik'} = m' \in \{1, 2, \dots, k/2\}$,

on kaavan (5.6) mukaan

$$\begin{aligned}\mathbb{P}(X_{ik'} = 1 \mid X_{ij} = 1) &\approx \frac{1}{2} \left((1-p) + \frac{(k-1) - 2m + 1}{n-k-1} p^2 + (1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\ &= (1-p) + \frac{k-2m+\frac{1}{2}}{n-k-1} p^2,\end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$.

Alkion $\{i, j, k'\} \in A_{m,m'}$ solmuparin $\{j, k'\}$ ehdollinen kytkentätodennäköisyys

$$\mathbb{P}(X_{j'k'} = 1 \mid X_{i'j'} = 1, X_{i'k'} = 1)$$

riippuu siitä, kuinka monta ulosmenevää linkkiä solmuparin $\{j', k'\}$ solmuilla on käytettävissään, ehdolla $\{X_{i'j'} = 1, X_{i'k'} = 1\}$. Kolmikön $\{i', j', k'\} \in V_i^{(3)}$ solmujen järjestyksellä ei ole väliä, joten voidaan olettaa, että $\ell_{i'j'} \leq \ell_{i'k'} \leq \ell_{j'k'}$, jolloin joukon $A_{m,m'} = \{\ell_{i'j'} = m, \ell_{i'k'} = m'\}$ alkioille pätee $\ell_{j'k'} = \min\{m+m', n-(m+m')\} \leq n/2$. Lauseesta 3.15 saadaan joukon $A_{m,m'}$ alkioille laskettua

$$\mathbb{P}(X_{j'k'} = 1 \mid X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(i')}, B_{i'k'}^{(i')}) = \mathbb{P}(X_{j'k'} = 1).$$

Tapahtumissa $\{X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(j')}, B_{i'k'}^{(i')}\}$ ja $\{X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(i')}, B_{i'k'}^{(k')}\}$ solmuparin $\{j', k'\}$ solmuilla on käytössään yksi ulosmenevä linkki vähemmän ja tapahtumassa $\{X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(j')}, B_{i'k'}^{(k')}\}$ kaksi ulosmenevää linkkiä vähemmän. Lauseista 3.16 ja 3.17 saadaan joukon $A_{m,m'}$ alkioille laskettua

$$\begin{aligned}(\text{B.2}) \quad &\mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\ &= \sum_{s \in \{i', j'\}} \sum_{t \in \{i', k'\}} \mathbb{P}(X_{i'j'} = 1, X_{i'k'} = 1, X_{j'k'} = 1, B_{i'j'}^{(s)}, B_{i'k'}^{(t)}) \\ &= \sum_{s \in \{i', j'\}} \sum_{t \in \{i', k'\}} \mathbb{P}(X_{j'k'} = 1 \mid X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(s)}, B_{i'k'}^{(t)}) \\ &\quad \cdot \mathbb{P}(B_{i'k'}^{(t)} \mid X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(s)}) \mathbb{P}(X_{i'k'} = 1 \mid X_{i'j'} = 1, B_{i'j'}^{(s)}) \\ &\quad \cdot \mathbb{P}(B_{i'j'}^{(s)} \mid X_{i'j'} = 1) \mathbb{P}(X_{i'j'} = 1) \\ &= \frac{1}{4} \sum_{s \in \{i', j'\}} \sum_{t \in \{i', k'\}} \mathbb{P}(X_{j'k'} = 1 \mid X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(s)}, B_{i'k'}^{(t)}) \\ &\quad \cdot \mathbb{P}(X_{i'k'} = 1 \mid X_{i'j'} = 1, B_{i'j'}^{(s)}) \mathbb{P}(X_{i'j'} = 1),\end{aligned}$$

sillä pieni maailma -verkkomallin $\mathcal{G}_{n,k,p}$ lähtökohtaisen rengashilan $H_{n,k}$ symmetrian perusteella $\mathbb{P}(B_{i'j'}^{(s)} \mid X_{i'j'} = 1) = 1/2$ kaikilla $s \in \{i', j'\}$ ja $\mathbb{P}(B_{i'k'}^{(t)} \mid X_{i'j'} = 1, X_{i'k'} = 1, B_{i'j'}^{(s)}) = 1/2$ kaikilla $t \in \{i', k'\}$. Joukon $A_{m,m'}$ alkioille, joille $m' \geq m > k/2$ siten, että $\min\{m+m', n-(m+m')\} \leq n/2$, saadaan kaavoista (5.4) ja (B.2) laskettua

todennäköisyys

$$\begin{aligned}
& \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\
&= \frac{1}{4} \sum_{s \in \{i', j'\}} \sum_{t \in \{i', k'\}} \mathbb{P}(X_{j'k'} = 1 \mid X_{i'j'} = 1, X_{j'k'} = 1, B_{i'j'}^{(s)}, B_{i'k'}^{(t)}) \\
&\cdot \mathbb{P}(X_{i'k'} = 1 \mid X_{i'j'} = 1, B_{i'j'}^{(s)}) \mathbb{P}(X_{i'j'} = 1) \\
&\approx \frac{1}{4} \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 \right) \\
&\cdot \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) + \frac{1}{4} \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 \right) \\
&\cdot \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 \right) \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \\
&+ \frac{1}{4} \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 \right) \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \\
&\cdot \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) + \frac{1}{4} \left(\frac{k-2}{n-k-1} p - \frac{(k-2)(k-3)}{2(n-k-1)^2} p^2 \right) \\
&\cdot \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \\
&\approx \frac{1}{4} \left(\frac{k^2(k-1)}{(n-k-1)^3} p^3 - \frac{3k^2(k-1)(k-\frac{4}{3})}{2(n-k-1)^4} p^4 \right) \\
&+ \frac{1}{4} \left(\frac{k(k-1)^2}{(n-k-1)^3} p^3 - \frac{3k(k-1)^3}{2(n-k-1)^4} p^4 \right) \\
&+ \frac{1}{4} \left(\frac{k^2(k-1)}{(n-k-1)^3} p^3 - \frac{3k^2(k-1)(k-\frac{4}{3})}{2(n-k-1)^4} p^4 \right) \\
&+ \frac{1}{4} \left(\frac{k^2(k-1)}{(n-k-1)^3} p^3 - \frac{k^2}{(n-k-1)^3} p^3 - \frac{3k^2(k-2)(k-\frac{5}{3})}{2(n-k-1)^4} p^4 \right) \\
&= \frac{k(k-1)(k+(k-1)+k+k)}{4(n-k-1)^3} p^3 - \frac{k^2}{4(n-k-1)^3} p^3 - \frac{3k(12k^3-34k^2+27k-3)}{8(n-k-1)^4} p^4 \\
&= \frac{k(k-1)(k-\frac{1}{4})}{(n-k-1)^3} p^3 - \frac{k^2}{4(n-k-1)^3} p^3 - \frac{3k(12k^3-34k^2+27k-3)}{8(n-k-1)^4} p^4,
\end{aligned}$$

kun satunnaisverkossa $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$, sillä $k^s/(n-k-1)^s \approx 0$ kaikilla $s \geq 5$, kun tiheys $D(G_{n,k,p}) \rightarrow 0$.

Joukon $A_{m,m'}$ alkiolle, joille pätee $1 \leq m \leq k/2 < m' \leq n/2 - m$, saadaan kaavoista (5.4), (5.6) ja (B.2) laskettua

$$\begin{aligned}
 & \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\
 &= \frac{1}{4} \sum_{s \in \{i', j'\}} \sum_{t \in \{i', k'\}} \mathbb{P}(X_{j't'} = 1 \mid X_{i'j'} = 1, X_{j'k'} = 1, B_{i'j'}^{(s)}, B_{i'k'}^{(t)}) \\
 & \cdot \mathbb{P}(X_{i'k'} = 1 \mid X_{i'j'} = 1, B_{i'j'}^{(s)}) \mathbb{P}(X_{i'j'} = 1) \\
 & \approx \frac{1}{4} \left(\frac{k(k-1)}{(n-k-1)^2} p^2 - \frac{k(k-1)(k-\frac{3}{2})}{(n-k-1)^3} p^3 \right) \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\
 & + \frac{1}{4} \left(\frac{(k-1)^2}{(n-k-1)^2} p^2 - \frac{(k-1)^2(k-2)}{(n-k-1)^3} p^3 \right) \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\
 & + \frac{1}{4} \left(\frac{k(k-1)}{(n-k-1)^2} p^2 - \frac{k(k-1)(k-\frac{3}{2})}{(n-k-1)^3} p^3 \right) \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\
 & + \frac{1}{4} \left(\frac{k(k-2)}{(n-k-1)^2} p^2 - \frac{k(k-2)^2}{(n-k-1)^3} p^3 \right) \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right) \\
 & = \left(\frac{(k-1)(k-\frac{1}{4})}{(n-k-1)^2} p^2 - \frac{k}{4(n-k-1)^2} p^2 - \frac{k^3 - \frac{13}{4}k^2 + 3k - \frac{1}{2}}{(n-k-1)^3} p^3 \right) \\
 & \cdot \left((1-p) + \frac{k-2m+1}{n-k-1} p^2 \right),
 \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Joukon $A_{m,m'}$ alkiolle, joille pätee $1 \leq m \leq m' \leq k/2$ siten, että $\ell_{jk'} = m + m' > k/2$, saadaan kaavoista (5.4), (5.6) ja (B.2) laskettua

$$\begin{aligned}
 & \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\
 & \approx \frac{1}{4} \left(\frac{k}{n-k-1} p - \frac{k(k-1)}{2(n-k-1)^2} p^2 \right) \left((1-p)^2 + \frac{2k-2(m+m')+1}{n-k-1} p^2 \right) \\
 & + \frac{1}{4} \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 \right) \left((1-p)^2 + \frac{2k-2(m+m')+1}{n-k-1} p^2 \right) \\
 & + \frac{1}{4} \left(\frac{k-1}{n-k-1} p - \frac{(k-1)(k-2)}{2(n-k-1)^2} p^2 \right) \left((1-p)^2 + \frac{2k-2(m+m')+2}{n-k-1} p^2 \right) \\
 & + \frac{1}{4} \left(\frac{k-2}{n-k-1} p - \frac{(k-2)(k-3)}{2(n-k-1)^2} p^2 \right) \left((1-p)^2 + \frac{2k-2(m+m')+2}{n-k-1} p^2 \right) \\
 & \approx \frac{k-1}{n-k-1} (1-p)^2 p,
 \end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$. Joukon $A_{m,m'}$ alkiolle, joille pätee $1 \leq m \leq m' \leq k/2 - m$, jolloin $\ell_{j'k'} = m + m' \leq k/2$, saadaan kaavoista (5.6) ja (B.2)

laskettua

$$\begin{aligned}
& \mathbb{P}(X_{ij} = 1, X_{ik'} = 1, X_{jk'} = 1) \\
& \approx \frac{1}{4} \left((1-p) + \frac{k - 2(m + m') + 1}{n - k - 1} p^2 \right) \left((1-p)^2 + \frac{2k - 2(m + m') + 1}{n - k - 1} (1-p)p^2 \right) \\
& + \frac{1}{4} \left((1-p) + \frac{k - 2(m + m')}{n - k - 1} p^2 \right) \left((1-p)^2 + \frac{2k - 2(m + m') + 1}{n - k - 1} (1-p)p^2 \right) \\
& + \frac{1}{4} \left((1-p) + \frac{k - 2(m + m')}{n - k - 1} p^2 \right) \left((1-p)^2 + \frac{2k - 2(m + m') + 2}{n - k - 1} (1-p)p^2 \right) \\
& + \frac{1}{4} \left((1-p) + \frac{k - 2(m + m'3) - 1}{n - k - 1} p^2 \right) \left((1-p)^2 + \frac{2k - 2(m + m') + 2}{n - k - 1} (1-p)p^2 \right) \\
& \approx \frac{1}{2} \left((1-p)^3 + \frac{3k - 4(m + m') + 2}{n - k - 1} (1-p)^2 p^2 \right) \\
& + \frac{1}{2} \left((1-p)^3 + \frac{3k - 4(m + m') + 1}{n - k - 1} (1-p)^2 p^2 \right) \\
& \approx (1-p)^3 + \frac{3k - 4(m + m') + \frac{3}{4}}{n - k - 1} (1-p)^2 p^2 \\
& = (1-p)^3 + \frac{3(k + \frac{1}{4}) - 4(m + m')}{n - k - 1} (1-p)^2 p^2,
\end{aligned}$$

kun satunnaisverkon $G_{n,k,p}$ tiheys $D(G_{n,k,p}) \rightarrow 0$.

LIITE C

Summia

Kaavasta $\sum_{m=1}^N m = N(N+1)/2$ [1] saadaan laskettua

$$\begin{aligned} \sum_{m=1}^{k/2} \frac{(k-2m+1)}{n-k-1} p^2 &= \frac{k(k+1)}{2(n-k-1)} p^2 - \frac{2p^2}{n-k-1} \sum_{m=1}^{k/2} m \\ &= \frac{k(k+1)}{2(n-k-1)} p^2 - \frac{k(k+2)}{4(n-k-1)} p^2 = \frac{k^2}{4(n-k-1)} p^2, \end{aligned}$$

joten

$$\sum_{m=1}^{k/2} \left((1-p) + \frac{(k-2m+1)}{n-k-1} p^2 \right) = \frac{k}{2} (1-p) + \frac{k^2}{4(n-k-1)} p^2.$$

Kaavoista $\sum_{m=1}^N m = N(N+1)/2$ [1] ja $\sum_{m=1}^N m^2 = N(N+1)(2N+1)/6$ [1] saadaan laskettua

$$\begin{aligned} \sum_{m=1}^{k/2} \frac{(k-2m+1)(k-2m+\frac{1}{2})}{(n-k-1)^2} p^4 &= \sum_{m=1}^{k/2} \frac{(k+1)(k+\frac{1}{2}) - 4(k+\frac{3}{4})m + 4m^2}{(n-k-1)^2} p^4 \\ &= \frac{k(k+1)(k+\frac{1}{2})}{2(n-k-1)^2} p^4 - \frac{k(k+2)(k+\frac{3}{4})}{2(n-k-1)^2} p^4 + \frac{k(k+1)(k+2)}{6(n-k-1)^2} p^4 \\ &= \frac{k(k+1)(4k+\frac{7}{2})}{6(n-k-1)^2} p^4 - \frac{k(k+2)(k+\frac{3}{4})}{2(n-k-1)^2} p^4 = \frac{k(k^2 - \frac{3}{4}k - 1)}{6(n-k-1)^2} p^4 \end{aligned}$$

ja

$$\begin{aligned} &\sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} \frac{(k-2s+1)(k-2t+\frac{1}{2})}{(n-k-1)^2} p^4 \\ &= \sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} \frac{(k+1)(k+\frac{1}{2}) - 2k(s+t) - s - 2t + 4st}{(n-k-1)^2} p^4 \\ &= \frac{k(k-2)(k+1)(k+\frac{1}{2})}{8(n-k-1)^2} - \frac{k^2(k-2)(k+2)}{8(n-k-1)^2} p^4 - \frac{5k(k-2)(k+2)}{48(n-k-1)^2} p^4 \\ &\quad + \frac{k(k+2)(k-2)(k+\frac{4}{3})}{32(n-k-1)^2} p^4 \\ &= \frac{k(k-2)(k+1)(k+\frac{1}{2})}{8(n-k-1)^2} - \frac{k(k-2)(k+2)(3k+2)}{32(n-k-1)^2} p^4 = \frac{k(k-2)(k^2 - 2k - 2)}{32(n-k-1)^2} p^4, \end{aligned}$$

jossa kaavasta $\sum_{m=1}^N m = N(N+1)/2$ [1] saadaan laskettua

$$\sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} 1 = \sum_{s=1}^{k/2-1} \binom{k}{2} - s = \frac{k(k-2)}{4} - \frac{k(k-2)}{8} = \frac{k(k-2)}{8}.$$

Kaavojen $\sum_{m=1}^N m = N(N+1)/2$ ja $\sum_{m=1}^N m^2 = N(N+1)(2N+1)/6$ [1] mukaan

$$\begin{aligned} \sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} s &= \sum_{s=1}^{k/2-1} s \binom{k}{2} - s = \frac{k}{2} \sum_{s=1}^{k/2-1} s - \sum_{s=1}^{k/2-1} s^2 \\ &= \frac{k^2(k-2)}{16} - \frac{k(k-1)(k-2)}{24} = \frac{k(k-2)(3k-2k+2)}{48} = \frac{k(k-2)(k+2)}{48} \end{aligned}$$

ja

$$\begin{aligned} \sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} t &= \sum_{t=2}^{k/2} \sum_{s=1}^{t-1} t = \sum_{t=2}^{k/2} t(t-1) = \sum_{s=1}^{k/2-1} s(s+1) = \sum_{s=1}^{k/2-1} s^2 + \sum_{s=1}^{k/2-1} s \\ &= \frac{k(k-1)(k-2)}{24} + \frac{k(k-2)}{8} = \frac{k(k-2)(k+2)}{24}, \end{aligned}$$

joten $\sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} (s+t) = k(k-2)(k+2)/48 + k(k-2)(k+2)/24 = k(k-2)(k+2)/16$.

Kaavojen $\sum_{m=1}^N m^2 = N(N+1)(2N+1)/6$ [1] ja $\sum_{m=1}^N m^3 = N^2(N+1)^2/4$ [1] mukaan

$$\begin{aligned} \sum_{s=1}^{k/2-1} \sum_{t>s}^{k/2} st &= \sum_{t=2}^{k/2} t \sum_{s=1}^{t-1} s = \sum_{t=2}^{k/2} t \left(\frac{t(t-1)}{2} \right) = \frac{1}{2} \sum_{t=2}^{k/2} t^3 - \frac{1}{2} \sum_{t=2}^{k/2} t^2 \\ &= \frac{1}{2} \left(\frac{(\frac{k}{2})^2(\frac{k}{2}+1)^2}{4} - 1 \right) - \frac{1}{2} \left(\frac{\frac{k}{2}(\frac{k}{2}+1)(k+1)}{6} - 1 \right) \\ &= \frac{1}{2} \left(\frac{k^2(k+2)^2}{64} - \frac{k(k+1)(k+2)}{24} \right) = \frac{k(k+2)(k-2)(k+\frac{4}{3})}{128}. \end{aligned}$$

Kaavojen $\sum_{m=1}^N m = N(N+1)/2$ ja $\sum_{m=1}^N m^2 = N(N+1)(2N+1)/6$ [1] mukaan

$$\begin{aligned} \sum_{s=1}^{k/2-1} \sum_{t=1}^{k/2-s} s &= \sum_{s=1}^{k/2-1} \left(\frac{k}{2}s - s^2 \right) = \frac{k^2(k-2)}{16} - \frac{k(k-1)(k-2)}{24} \\ &= \frac{k(k-2)(3k-2(k-1))}{48} = \frac{k(k-2)(k+2)}{48} \end{aligned}$$

ja

$$\begin{aligned} \sum_{s=1}^{k/2-1} \sum_{t=1}^{k/2-s} t &= \sum_{s=1}^{k/2-1} \frac{(\frac{k}{2}-s)(\frac{k}{2}-s+1)}{2} = \sum_{s=1}^{k/2-1} \left(\frac{k(k+2)}{8} - \frac{k+1}{2}s + \frac{1}{2}s^2 \right) \\ &= \frac{k(k-2)(k+2)}{16} - \frac{k(k-2)(k+1)}{16} + \frac{k(k-1)(k-2)}{48} \\ &= \frac{k(k-2)}{16} + \frac{k(k-1)(k-2)}{48} = \frac{k(k-2)(3+(k-1))}{48} = \frac{k(k-2)(k+2)}{48}. \end{aligned}$$

LIITE D

Käytetty koodi

```
# WattsStrogatzRewiring.R
# Syöte:
# n: solmujen lkm, k: solmujen asteiden keskiarvo,
# p: todennäköisyys, jolla linkki kytketään uudelleen
# Tulos:
# G: Watts ja Strogatzin pieni maailma -verkkomallin verkko
# Funktiota kutsutaan komennoilla:
# source('WattsStrogatzRewiring.R')
# G <- WattsStrogatzRewiring(n,k,p)
library(igraph)
library(Matrix)

WattsStrogatzRewiring <- function (n,k,p) {
  # k-saannollinen rengashila, jossa on n solmua
  G <- graph.lattice(length=n, dim=1, nei=k/2, circular=TRUE)
  A <- get.adjacency(G, sparse=TRUE) # naapuruusmatriisi
  r <- runif(n*k/2)
  h <- 0
  for ( lap in 1:(k/2) ){
    for ( i in 1:n ){
      h <- h+1
      if ( r[h] <= p && sum(A[i,]) < n-1 ){
        j <- i+lap
        if ( n < j ){ j <- j %% n }
        repeat{
          j2 <- sample(1:n,1)
          if ( i != j2 && A[i,j2] == 0 ){
            A[i,j2] <- 1
            A[j2,i] <- 1
            break
          }
        }
      }
      A[i,j] <- 0
      A[j,i] <- 0
    }
  }
}
```

LIITE E

Merkintöjä

<i>Merkintä</i>	<i>Selitys</i>
G	Verkko
V	Solmujen joukko
E	Linkkien joukko
$ \cdot $	Joukon alkioiden lukumäärä
n	Verkon koko
V_n	Numeroilla merkityn n -kokoisen verkon solmujen joukko
$V_n^{(2)}$	Solmujoukon V_n järjestämättömät parit
$V_i^{(2)}$	Solmujoukon V_n järjestämättömät parit, jotka sisältävät solmun i
$V_n^{(3)}$	Solmujoukon V_n järjestämättömät kolmikot
$V_i^{(3)}$	Solmujoukon V_n järjestämättömät kolmikot, jotka sisältävät solmun i
\mathcal{G}^n	Solmujoukolla V_n määriteltyjen yleisten verkkojen kokoelma
Ω	Perusjoukko
ω	Alkeistapaus
\mathcal{F}	Perusjoukon osajoukkojen σ -algebra
(Ω, \mathcal{F})	Mitallinen avaruus
\mathcal{P}	Potenssijoukko
P	Todennäköisyysfunktio
\mathbb{P}	Todennäköisyysmitta
$(\Omega, \mathcal{F}, \mathbb{P})$	Todennäköisyysavaruus
X, Y, Z	Satunnaismuuttujia
P_X	Satunnaismuuttujan jakauma
Ω_n	Solmujoukolla V_n määritelty satunnaisverkkomalli
G_n	Satunnaisverkkomallin Ω_n satunnaisverkko
P_{G_n}	Satunnaisverkon jakauma
$\mathcal{G}_{n,M}$	Satunnaisverkkomalli
$\mathcal{G}_{n,p}$	Satunnaisverkkomalli
$\mathcal{G}_{n,r\text{-reg}}$	Satunnaisverkkomalli
$\mathcal{G}_{n,k,p}$	Pieni maailma -satunnaisverkkomalli
$G_{n,M}$	Satunnaisverkkomallin $\mathcal{G}_{n,M}$ satunnaisverkko
$G_{n,p}$	Satunnaisverkkomallin $\mathcal{G}_{n,p}$ satunnaisverkko
$G_{n,k,p}$	Pieni maailma -satunnaisverkkomallin satunnaisverkko
$H_{n,k}$	k -säännöllinen rengashila
$D(G)$	Verkon tiheys
o	Pieni o -estimaatti
o_P	Stokastinen pieni o -estimaatti
$X_n \xrightarrow{m.v.} X$	Suppeneminen melkein varmasti

$X_n \xrightarrow{P} X$	Suppeneminen stokastisesti
d_{ij}	Solmujen i ja j välinen etäisyys
$L(G)$	Verkon karakteristisen polun pituus
$\text{diam}(G)$	Verkon halkaisija
N_i	Solmun i naapurusto
k_i	Solmun i aste
\bar{k}	Solmujen asteiden keskiarvo
Υ	Kytetty kolmikko
$\Upsilon(G)$	Verkon kytkettyjen kolmikoiden joukko
Υ_i	Keskisolmun i kytkettyjen kolmikoiden joukko
∇	Kolmio
$\Delta(G)$	Verkon kolmioiden joukko
Δ_i	Solmun i sisältävien kolmioiden joukko
C_i	Solmun i klusterointikerroin
$C(G)$	Verkon klusterointikerroin
V'	Verkon solmujoukko $\{i \in V : k_i \geq 2\}$
$C_{V'}(G)$	Verkon klusterointikerroin laskettuna solmujoukolle V'
$C'(G)$	Verkon transitiivisuuskerroin
$C_w(G)$	Verkon painotettu klusterointikerroin
$w : V \rightarrow \mathbb{R}_+$	Solmujoukon V painokerroinfunktio
$C^{(k)}$	k -asteiden solmujen keskimääräinen klusterointikerroin
$c(G_n)$	Satunnaisverkon klusterointikerroin
$c_{V'}(G_n)$	Satunnaisverkon klusterointikerroin laskettuna solmujoukolle V'
$c'(G_n)$	Satunnaisverkon transitiivisuuskerroin
l_{ij}	Rengashilan $H_{n,k}$ solmujoukon metriikka $l_{ij} = \min\{ j - i , n - j - i \}$
k_i^+	Suunnatun verkon solmun i lähtöaste
k_i^-	Suunnatun verkon solmun i tuloaste
$H_{n,k/2\text{-out}}$	Suunnattu verkko, jossa jokaisen solmun lähtöaste on $k/2$
X_{ij}	Indikaattorisatunnaismuuttuja
$\mathbb{1}_{\{ij,ik\}}$	Indikaattorisatunnaismuuttuja
$\mathbb{1}_{\{ij,ik,jk\}}$	Indikaattorisatunnaismuuttuja.

Kirjallisuutta

- [1] RICHARD COURANT ja FRITZ JOHN: *Introduction to Calculus and Analysis, Volume I*. Springer, 1999.
- [2] JEAN JACOD ja PHILIP PROTTER: *Probability Essentials*. Second printing, Springer, 2004.
- [3] ALLAN GUT: *Probability: A Graduate Course*. Springer, 2005.
- [4] LASSE LESKELÄ: *Stokastiikan perusteet*. Elektroninen dokumentti osoitteessa <http://math.aalto.fi/~lleskela/teaching/MATA280-2013/lectures/Muistio.pdf> (luettu 31.01.2014).
- [5] PERTTI KOIVISTO ja RIITTA NIEMISTÖ: *Graafiteoriaa*. Elektroninen dokumentti osoitteessa <http://www.sis.uta.fi/matematiikka/graafteoraa/graafteoraa.pdf>, 2001 (luettu 11.12.2013).
- [6] BÉLA BOLLOBÁS: *Random Graphs*. Academic press (London), 1985.
- [7] BÉLA BOLLOBÁS ja OLIVER M. RIORDAN: *Sparse graphs: metrics and random models*. arXiv:0812.2656v3 [math.PR], 2010.
- [8] ITHIEL DE SOLA POOL ja MANFRED KOCHEN: *Contacts and Influence*. Social Networks, 1:5-51, 1978/79.
- [9] SANLEY MILRAM: *The small-world problem*. Psychology Today, 1.1:61-67, 1967.
- [10] DUNCAN J. WATTS ja STEVEN H. STROGATZ: *Collective dynamics of 'small-world'-networks*. Nature, 393:440-442, 1998.
- [11] ALAIN BARRAT ja MARTIN WEIGT: *On the properties of small-world network models*. The European Physical Journal B, 13:547-560, 2000.
- [12] BÉLA BOLLOBÁS ja OLIVER M. RIORDAN: *Mathematical results on scale-free random graphs*. Handbook of Graphs and Networks: From the Genome to the Internet (eds S. Bornholdt and H. G. Schuster), Wiley-VCH, 2002.
- [13] MARK E. J. NEWMAN, DUNCAN J. WATTS and STEVEN H. STROGATZ: *Random graph models of social networks*. Proceeding of the National Academy of Science of the United States of America, 99:2566-2572, 2002.
- [14] MARK E. J. NEWMAN: *Models of the Small World: A Review*. J. Stat. Phys. 101, 819-841, 2000.
- [15] MARK E. J. NEWMAN: *The structure and function of complex networks*. SIAM Review 45:167-256, 2003.
- [16] THOMAS SCHANK ja DOROTHEA WAGNER: *Approximating clustering coefficient and transitivity*. Journal of Graph Algorithms and Applications, 9.2:265-275, 2005.
- [17] MATTHIAS GROSSGLAUSER ja PATRICK THIRAN: *Networks out of Control: Models and Methods for Random Networks*. Elektroninen dokumentti osoitteessa <http://www.stat.berkeley.edu/~aldous/Networks/nooc2006.pdf> (luettu 27.05.2014).
- [18] REMCO VAN DER HOFSTAD: *Random Graphs and Complex Networks. Vol. I*. Elektroninen dokumentti osoitteessa <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf> (luettu 9.06.2014).
- [19] JOHAN UGANDER, BRIAN KARRER, LARS BACKSTROM, ja CAMERON MARLOW: *The Anatomy of the Facebook Socal Graph*. arXiv:1111.4503v1 [cs.SI], 2011.
- [20] LARS BACKSTROM, PAOLA BOLDI, MARCO ROSA, JOHAN UGANDER ja SEBASTIANO VIGNA: *Four Degrees of Separation*. arXiv:1111.4570v3 [cs.SI], 2012.
- [21] AGNER FOG: *Sampling Methods for Wallenius' and Fisher's Noncentral Hypergeometric Distributions*. Communications in Statistics: Simulation and Computation, Vol. 37, no. 2, 241-257, 2008.
- [22] U.S. DEPARTMENT OF COMMERCE: *1990 Census of Population and Housing, Population and Housing Unit Counts United States*. Elektroninen dokumentti osoitteessa <http://www.census.gov/prod/cen1990/cph2/cph-2-1-1.pdf>.