

Satunnaiskävelyt ja niiden kasvattamat puut

Joonas Laukka

Perustieteiden korkeakoulu

Kandidaatintyö
Espoo 1.5.2018

Vastuuopettaja

TkT Riikka Korte

Työn ohjaaja

Prof. Lasse Leskelä



**Aalto-yliopisto
Perustieteiden
korkeakoulu**

Copyright © 2018 Joonas Laukka



Tekijä Joonas Laukka

Työn nimi Satunnaiskävelyt ja niiden kasvattamat puut

Koulutusohjelma Teknillinen fysiikka ja matematiikka

Pääaine Matematiikka ja systeemitieteen

Pääaineen koodi SCI3029

Vastuuopettaja TkT Riikka Korte

Työn ohjaaja Prof. Lasse Leskelä

Päivämäärä 1.5.2018

Sivumäärä 20+2

Kieli Suomi

Tiivistelmä

TODO: real networks -> preferential attachment Yksi kiinnostava kasvumalli on *No Restart Random Walk* (NRRW), jossa satunnaiskävelijä kasvattaa suuntaamatonta verkkoa kulkissaan siinä. Kävelijä lähtee alkusolmusta, kulkee s :n kaaren läpi satunnaiseen solmuun, ja sen naapuriksi lisätään uusi solmu. Uuden solmun asteluku on aina yksi. Tätä prosessia jatketaan satunnaiskävelijän viimeisimmästä tilasta, ja prosessin toistaminen kasvattaa satunnaisen puurakenteen. Mallin ominaisuudet määräät siis askelparametri, s .

NRRW-malli mukailee reaalimaailman verkoille tyypillistä suosivaa kiinnittymistä. Sen rakentaman verkon ominaisuudet riippuvat kuitenkin voimakkaasti askelparametrista. Malli ei generoi skaalautumattomia verkkuja, mutta se demonstroi kiinnostavan yhteyden satunnaisverkon astelujen jakauman ja sitä kasvattavan satunnaiskulun palautuvuuden välillä. Tämän yhteyden ymmärtäminen on askel kohti entistä realistisempia verkkojen kasvumalleja. Tässä kandidaatintyössä keskityn tarkastelemaan parillisen askelparametrin NRRW-malleja. Muodostan NRRW-prosessille erään stokastisen esityksen, todistan sen oikeaksi ja hyödynnän sitä parillisen askelparametrin NRRW-mallin ominaisuuksien tutkimiseen.

Avainsanat Satunnaiskävely, puu, verkkو, NRRW

Author Joonas Laukka

Title Random Walks and the Trees they Grow

Degree programme Engineering Physics and Mathematics

Major Mathematics and System Analysis**Code of major** SCI3029

Supervisor DSc (Tech.) Riikka Korte

Advisor Prof. Lasse Leskelä

Date 1.5.2018

Number of pages 20+2

Language Finnish

Abstract

The modern world is full of networks from the Internet to social circles. Modeling the growth of graphs has caught the attention of many researchers during the last decade. One such growth model is a *No Restart Random Walk* (NRRW). In an NRRW, a random walker grows an undirected graph while it is traversing it. The walker starts from the root vertex and crosses s edges to end up in a random destination vertex. Then, a new vertex with degree 1 is added as its neighbour. This process is repeated starting from the destination vertex and consequently, a random tree is generated. The properties of the model are unambiguously determined by the step parameters, s .

The NRRW model conforms to many properties that are common in real life graphs such as preferential attachment. The random walk process and the random graph process are mutually dependent. Examining these dependent processes can be simplified by proving that they are equivalent with a simpler stochastic process. I will show that the NRRW model is stochastically equivalent to a model seeded by a sequence of independent, uniformly distributed random numbers. Finally, I will use this alternative model to prove a theorem about the recurrence of the vertices in an NRRW generated graph.

Keywords Random walk, graph, tree, NRRW

Esipuhe

Otaniemi, 15.9.2017

Joonas Laukka

Sisällysluettelo

Tiivistelmä	3
Tiivistelmä (englanniksi)	4
Esipuhe	5
Sisällysluettelo	6
Symbolit ja lyhenteet	7
1 Johdanto	8
2 Malli	10
2.1 Yleiskatsaus	10
2.2 Satunnaisverkko ja -kävely	10
2.2.1 Satunnaisverkko	11
2.2.2 Satunnaiskävely	12
2.3 Kokonaisuus	12
2.3.1 Tilajoukko	12
2.3.2 Siirtymätodennäköisyydet	13
3 Stokastinen esitys	14
3.1 Yleistä teoriaa	14
3.2 Prosessin esitys	15
4 Parillinen prosessi	17
4.1 Astelukujen jakauma	18
4.2 Palautuvuus	18
5 Yhteenvetö	19
Viitteet	20
A Esimerkki liitteestä	21
B Toinen esimerkki liitteestä	22

Symbolit ja lyhenteet

Symbolit

\mathbb{N}_0 ei-negatiiviset kokonaisluvut $(0, 1, 2, \dots)$

Operaattorit

$\mathbf{1}\{X\}$ indikaattorifunktio ehdolla X
 $\mathbb{P}[X]$ tapahtuman X todennäköisyys

Lyhenteet

NRRW No Restart Random Walk
i.i.d. riippumaton ja identtisesti jakautunut

1 Johdanto

Nykymaailma on täynnä erilaisia verkostoja Internetistä sosiaalisiin piireihin ja niiden syvällinen ymmärtäminen on tärkeämpää kuin koskaan. Verkkoteoria on matematiikan ala, joka pyrkii ymmärtämään solmuista ja niitää yhdistävistä kaarista rakentuvia abstrakteja verkkooja. Koska verkostoja tarkastellaan riittävän abstraktilla tasolla, verkkoteorian tulokset pätevät sovelluskohteesta riippumatta. Sama looginen rakenne, verkko, voi kuvata niin kaupunkien välisiä tieyhteyksiä kuin nettisivujen välisiä linkkejäkin, ja verkkoteorian tulokset ovat aina yhtä oleellisia. Vain verkon tulkinta vaihtuu.

Yksi tärkeä askel kohti reaalimaailman verkkojen ymmärtämistä on erilaisten kasvumallien luominen. Kun tiedämme, miten ja miksi verket kehittyvät sellaisiksi kuin ne ovat, pystymme simuloimaan niitä ja ennustamaan niiden kehitystä. Kuluneina vuosikymmeninä onkin kehitetty lukuisia erilaisia verkkojen kasvumalleja. Niiden generoimien verkkojen topologiassa on kiinnitetty huomioita moniin erilaisiin suureisiin, mutta yksi kiinnostavimmista ominaisuuksista on solmujen astelukujen jakauma. Empiiriset kokeet ovat osoittaneet, että monet suuret ja kiinnostavat verket, kuten Internet, ovat skaalautumattomia (eng. scale-free) [1]. Skaalautumattoman verkon astelukujen jakauman määrittää jokin potenssilaki ($\mathbb{P}[k] = k^\gamma$). Yksinkertaisstaen tämä tarkoittaa sitä, että ääriarvot, eli sellaiset solmut, joilla on hyvin korkea astelu, ovat skaalautumattomissa verkoissa verrattain yleisiä. [2]

Monet tosielämän verket syntyvät näennäisesti hyvinkin toisistaan poikkeavista prosesseista, mutta niille on kuitenkin tyypillistä skaalautumaton rakenne. Tämä viittaa siihen, että kyseisten verkkojen kasvu pohjautuu samankaltaisiin stokastisiin mekanismiin. Näyttää siltä, että skaalautumattomia verkkooja syntyy erityisesti sellaisista prosesseista, jotka noudattavat kasvaessaan suosivaa kiinnitymistä (eng. preferential attachment) [3]. Suosivalla kiinnitymisellä tarkoitetaan satunnaisverkkojen yhteydessä sitä, että uusia solmuja liitetään todennäköisimmin sellaisiin solmuihin, joilla on jo ennestään korkea astelu. Koska suosiva kiinnityminen näyttäisi olevan tosielämässä esiintyvien skaalautumattomien verkkojen kasvun taustalla, on siihen perustuvien kasvuprosessien kehittäminen kiinnittänyt tutkijoiden huomion kuluneen vuosikymmenen aikana.

Yksi suosivaa kiinnitymistä ilmentävä kasvumalli on *No Restart Random Walk* (NRRW), jossa satunnaiskävelijä kasvattaa suuntaamatonta verkkoa kulkissaan siinä. Nimensä mukaisesti NRRW-prosessin satunnaiskävely ei ala koskaan alusta, vaan jatkuu katkeamatta verkon kasvaessa. Prosessin ominaisuudet määräätaskelparametri, s , ja se kehittyy seuraavanlaisesti:

0. Aluksi verkossa on yksi solmu, juuri, josta on kaari itseensä. Satunnaiskävelyn alkutila on juurisolmu.
1. Satunnaiskävely ottaa s askelta verkossa.
2. Verkkoon lisätään uusi solmu ja se liitetään satunnaiskävelyn senhetkiseen sijaintiin.
3. Palataan kohtaan 1.

Prosessin simulointi ja matemaattinen analyysi osoittaa, että parametrin s parillisuus

vaikuttaa voimakkaasti satunnaiskävelyn ja satunnaisverkon luonteisiin:

Askelparametri	Satunnaiskävely	Satunnaisverkko
$s = 1$	Väistyvä	Sisäsolmujen astelukujen jakauma on geometrisen jakauman ylhäältä rajoittama.
$s = 2n$	Palautuva	Sisäsolmujen astelukujen jakauma on potentiaalin alhaalta rajoittama.

NRRW-prosessin luonnetta voidaan tarkastella joko satunnaisen verkon tai satunnaiskävelyn näkökulmasta. Satunnaiskävelyn palautuvuus ja sen kasvattaman verkon topologia ovat kuitenkin kietoutuneet selvästi toisiinsa. [4]

Koska NRRW-prosessin kasvattamat verkot ilmentävät sekä skaalautumatonta että skaalautuvalaa sisäsolmujen rakennetta, ovat ne epäilemättä kiinnostavia perinteisen verkkoanalyysin näkökulmasta. Osuvimmat sovelluskohteet löytynevät kuitenkin sellaisista systeemeistä, joissa satunnaiskulun ja verkon kasvun yhteyks on prosessin sisäänrakennettu, kiistaton ominaisuus. Yksi tällainen systeemi esiintyy uudenlaisen kryptovalutan, IOTA:n, toteutuksessa. Toisin kuin perinteiset kryptovaluutat, IOTA ei perustu lohkoketjuteknologiaan. Lohkoketjun sijaan transaktiot tallennetaan suuntaamattomaan, syklittömään verkoon. Kun uusi transaktio saapuu, on sen vahvistettava kaksi aiempaa transaktiota liittymällä niihin transaktiopuussa. Koska transaktioiden vahvistusmekanismi ei perustu louhimiseen (eng. mining), ei transaktioihin liity myöskään transaktiokustannuksia. Tämä on huomattava etu verrattuna perinteisiin kryptovaluuttoihin, sillä se mahdollistaa muun muassa mikromaksujen maksamisen. [5]

IOTA:n transaktiopuun stabiilius ja turvallisuus ovat kuitenkin monimutkaisia kysymyksiä. Puuta mallinnetaan suuntaamattomia ja syklittömiä verkkoja rakentavalla stokastisella prosessilla, Tanglella, jossa uusien solmujen liitoskohdat valitaan satunnaiskulun avulla [6]. NRRW-prosessi on pohjimmiltaan satunnaiskulku satunnaisessa verkossa, ja siten sen ymmärtäminen on osa matemaattista kivijalkaa, jonka päälle Tanglen kaltaiset tietotekniset sovellukset rakennetaan.

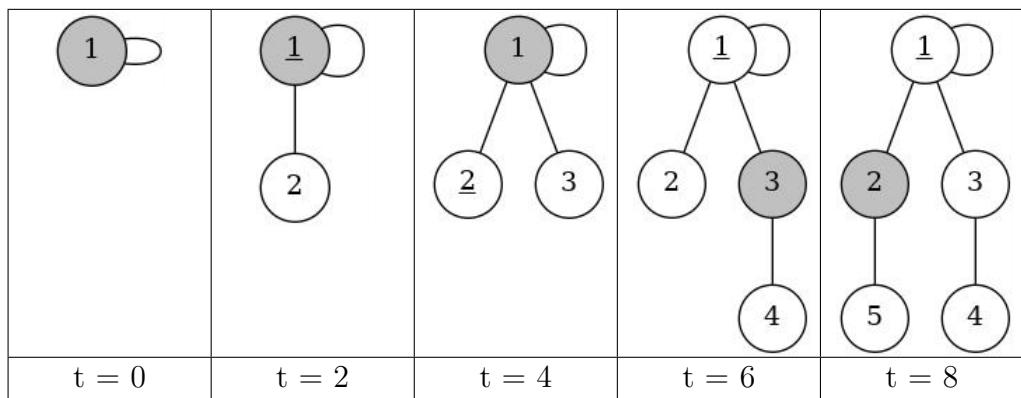
Tässä kandidaatintyössä perehdyn sellaisiin NRRW-prosessin ominaisuuksiin, joita on käsitellyt Giulio Iacobellin, Daniel Figueiredon ja Giovanni Neglian paperissa "Transient and Slim versus Recurrent and Fat: Random Walks and the Trees they Grow". Keskityn erityisesti parillisen askelparametrin prosessiin, sillä sen generoimien verkkojen sisäsolmut muodostavat monien kasvumallien sovelluksien kannalta kiinnostavan skaalautumattoman verkon. Formuloin NRRW-prosessille stokastisen esityksen, todistan sen oikeaksi ja hyödynnän sitä parillisen askelparametrin ominaisuuksien tarkasteluun.

2 Malli

2.1 Yleiskatsaus

NRRW-prosessi on kahden toisistaan riippuvan stokastisen prosessin pari, $(\mathcal{G}_t^s, \mathcal{W}_t)_{t \geq 0}$. molemmat prosessit ovat kytkettyneet vahasti toisiinsa, mutta niitä voidaan jaksottain tarkastella myös erillisinä. Taulukko 1 esittää yhden mahdollisen NRRW-prosessin realisaation askelparametrilla $s = 2$ ja parillisilla ajanhetkillä.

Taulukko 1: Yksi mahdollinen NRRW-prosessin realisaatio askelparametrilla $s = 2$. Satunnaiskävely on harmaan värin osoittamassa tilassa ja uusi solmu on juuri lisätty sen naapuriksi. Satunnaiskävely siirtyi nykyiseen tilaan alleviivatun solmun kautta.

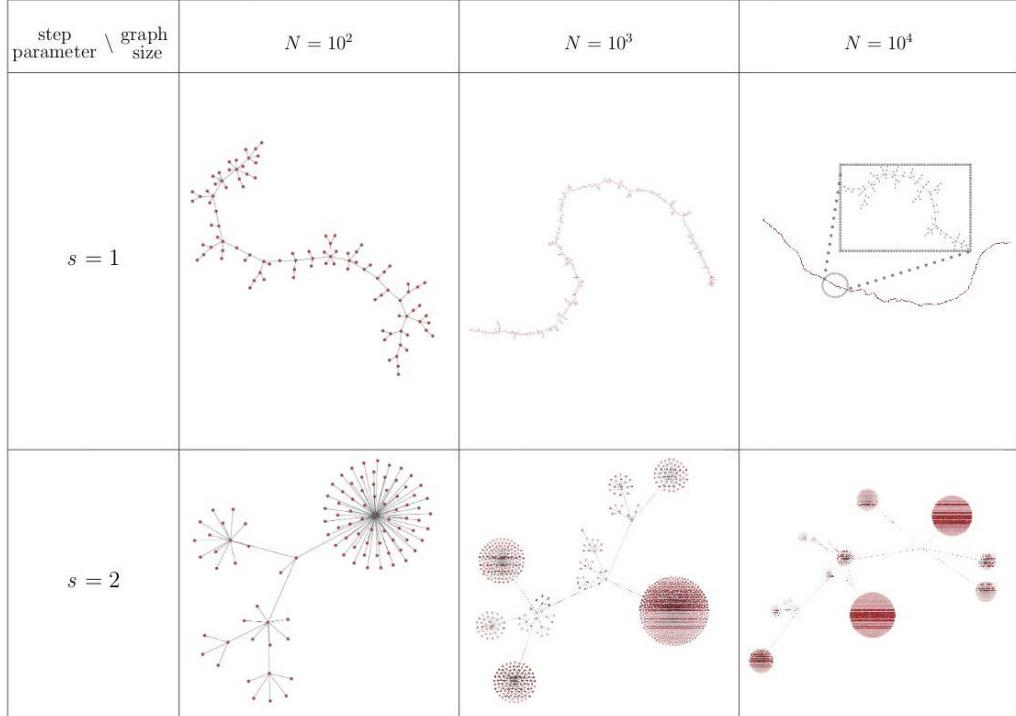


Askelparametri vaikuttaa huomattavasti siihen, millaisia verkkoja NRRW-prosessi kasvattaa. Tämä ilmenee hyvin tietokonesimuloiduista realisaatioista kuvassa 1. Parametrilla $s = 1$ jokaiseen satunnaiskävelyn vierailemaan tilaan lisätään uusi solmu. Tämä johtaa korkeisiin puihin, joissa solmut jakutuvat tasaisesti parillisille ja parittomille tasoille. Satunnaiskävelyn kaikki tilat ovat väistyviä.

Sen sijaan parillinen askelparametri ($s \in 2\mathbb{N}_0$) johtaa hyvin toisenlaisiin verkkoihin. Simulaatiotulokset viittaavat solmujen kasaantuvan suurissa määrin joidenkin yksittäisten solmujen jälkeläisiksi. Kyseessä on niin sanottu suoivan kiinnittymisen ilmiö, jossa jotain mitattavaa suuretta kertyy sellaisille joukon alkioille, joilla on jo ennestään paljon kyseistä suuretta. Satunnaisverkossa tämä ilmenee siten, että uusia solmuja liitetään todennäköisimmin sellaisiin verkon solmuihin, joilla on jo ennestään korkea asteluku. Kyseinen ilmiö näyttää olevan monien todellisten verkkojen, kuten Internetin, skaalautumattoman rakenteen takana [3].

2.2 Satunnaisverkko ja -kävely

NRRW-prosessin ominaisuudet ovat seurausta satunnaiskulun ja sen kasvattaman satunnaisen verkon välisestä vuorovaikutuksesta. Malli on monimutkainen ja yksinkertaisuuden vuoksi on ensin syytä tarkastella satunnaisverkkoa ja satunnaiskulkua



Kuva 1: NRRW-prosessin realisaatioita tietokonesimulaatioissa. Kuva on kopioitu aiemmasta tutkimuksesta (Iacobelli et al.) [1].

erillisinä prosesseina. Tämä mahdollistaa erillisiin malleihin liittyvän terminologian ja yksityiskohtien esittelyn ennen NRRW-prosessin kokonaisuuteen paneutumista.

2.2.1 Satunnaisverkko

Olkoon $(\mathcal{G}_t^s)_{t \in \mathbb{N}_0}$ satunnaisverkko, johon lisätään uusi solmu s aikayksikön välein. Satunnaisprosessin tilajoukko on kaikkien suuntaamattomien, yhtenäisten ja syklittömien verkkojen eli puiden joukko.

Määritelmä 1. *Suuntaamaton verkko G on pari (V, E) , missä joukko V on verkon solmujen joukko ($v_i \in V$) ja E on suuntaamattomien kaarien joukko. Kaarien joukon alkiot ovat muotoa $\{v_i, v_j\}$. Kaari $\{v_i, v_j\}$ yhdistää solmut v_i ja v_j toisiinsa.*

Kasvavan verkon käsittelyn helpottamiseksi satunnaisverkon $(\mathcal{G}_t^s)_{t \in \mathbb{N}_0}$ solmut indeksoidaan kokonaisluvuilla $(0, 1, 2, \dots)$ lisäysjärjestyksen mukaan. Juurisolmu on siis v_0 ja hetkellä $t = sn$, $n \in \mathbb{N}_0$ lisätään solmu v_n . Jos satunnaisverkkoa $(\mathcal{G}_t^s)_{t \in \mathbb{N}_0}$ tarjotaan itsenäisenä prosessina sen siirtymätodennäköisyydet pois nykytilasta ovat nollasta poikkeavia vain ajanhetkillä $t = sn$, $n \in \mathbb{N}_0$. Ne määritävät satunnaiskävelyn $(\mathcal{W}_t)_{t \in \mathbb{N}_0}$ siirtymätodennäköisyydet hetkellä $t = sn$:

$$\begin{aligned} & \mathbb{P} [\mathcal{G}_{t+s}^s = (V_t \cup \{v_{j+1}\}, E_t \cup \{(v_i, v_{j+1})\}) \mid \mathcal{G}_t^s = (V_j, E_j), \mathcal{W}_t = v_j] \\ &= \mathbb{P} [\mathcal{W}_{t+s} = v_i \mid \mathcal{W}_t = v_n, G_t = (V_t, E_t)] \end{aligned} \quad (1)$$

2.2.2 Satunnaiskävely

Vastaavasti satunnaiskävelyprosessin $(\mathcal{W}_t)_{t \in \mathbb{N}_0}$ tilajoukon ja siirtymätodennäköisyydet määräät sen kasvattama satunnaisverkko. Sekä tilajoukko että siirtymätodennäköisyydet muuttuvat aina, kun verkoon lisätään uusi solmu. Lisäyshetkien välillä tilajoukko on kuitenkin vakio ja satunnaiskävelyprosessi voidaan silloin nähdä tavallisena äärellisen tilajoukon Markov-ketjuna. Jos satunnaiskävely on solmussa v_i hetkellä t , määrätyy seuraava tila satunnaisesti valitsemalla yksi solmesta v_i lähevä kaari ja siirtymällä sen osoittamaan solmuun. Todennäköisyysmassa on tasajakautunut kaikkien solmun v_i kaarien kesken.

Koska NRRW-prosessi rakentaa puita, ei verkossa ole muita syklejä kuin juuren kaari itseensä. Verkkoteoriassa kaaret solmesta itseensä lasketaan kahdesti. Merkitään solmun v_i solmuun v_j yhdistävien kaarien määräät verkossa G_i

$$\psi_{G_i}(v_i, v_j) = \begin{cases} \mathbf{1}\{\{v_i, v_j\} \in E_i\} & \text{jos } v_i \neq v_j \\ 2 \cdot \mathbf{1}\{v_i = v_0\} & \text{muulloin} \end{cases}$$

missä E_i on verkon G_i kaarien joukko. Nyt todennäköisyys siirtyä solmesta v_i solmuun v_j on yksinkertaisesti

$$\mathbb{P}[\mathcal{W}_{t+1} = v_j \mid \mathcal{W}_t = v_i, \mathcal{G}_t^s = G_i] = \frac{\psi_{G_i}(v_i, v_j)}{d_{G_i}(v_i)} \quad (2)$$

sillä oletuksella, että satunnaiskävely tapahtuu verkossa G_i . Kaavassa $d_{G_i}(v_i)$ on solmun v_i asteluku verkossa G_i . Toisin sanoen

$$d_{G_i}(v_i) = \sum_{w \in V_i} \psi_{G_i}(v_i, w) \quad (3)$$

eli kaikkien solmuun v_i yhdistyvien kaarien lukumäärä.

2.3 Kokonaisuus

Satunnaisverkon ja satunnaiskävelyn erillinen tarkastelu paljastaa, että näistä jälkimmäinen on NRRW-prosessin todellinen satunnaisuuden lähde. Satunnaiskävely muuttaa liikkeillään sen omaa tilajoukkoa eli satunnaisverkkoa \mathcal{G}_t^s . Kun verkko muuttuu, myös satunnaiskävelyn siirtymätodennäköisyydet muuttuvat. Satunnaisverkko sen sijaan kasvaa satunnaiskävelyn realisaation ohjaamana. Kokonaisuutena NRRW-prosessi on siis ajassa epähomogeeninen äärettömän tilajoukon Markov-prosessi. Proseszin epähomogenisuus johtuu siitä, että satunnaisverkko voi muuttua vain s :llä jaollisilla ajanhetkillä.

2.3.1 Tilajoukko

Olkoon $V = \{v_1, v_2, \dots\}$ kaikkien solmujen numeroituvasti ääretön joukko. Jos n :n solmun joukkoa merkitään $V_n = \{v_1, v_2, \dots, v_n\} \subset V$ niin n :n solmun suuntaamattonien verkkojen joukko on

$$\mathbb{G}_n = \{(V_n, E) \mid E \subset V_n^2\}$$

missä $V_n^2 = \{e \subset V_n \mid |e| = 2\}$ on solmujoukon V_n järjestämättömien parien joukko. NRRW-prosessiin kuuluu myös satunnaiskävely, jonka tilajoukko n :n solmun verkkoa kulkissa sen solmujen joukko V_n . Selvästi n :n solmun NRRW-prosessin tilajoukko on $S_n = V_n \times \mathbb{G}_n$.

Kaikkien solmujen joukkoja merkitään numeroituvasti äärettömällä joukolla V . Vastaavasti kaikkien verkkojen joukkoon voidaan merkitä äärettömällä unionilla

$$\mathbb{G} = \bigcup_{n=1}^{\infty} \mathbb{G}_n \quad (4)$$

Koska satunnaisverkko kasvaa NRRW-prosessin edetessä, täytyy sen tilajoukon sisältää kaikki eri solmulukumäärien mahdollistamat tilat. NRRW-prosessin tilajoukko onkin siis

$$S = V \times \mathbb{G} \quad (5)$$

2.3.2 Siirtymätodennäköisyydet

NRRW-prosessin siirtymätodennäköisyydet ovat ajassa epähomogeenisia, sillä satunnaiskävelyn realisaatio kasvattaa sen omaa tilajoukkoa, eli NRRW-prosessin verkkoa vain kun $t \bmod s = 0$. Siirtymätodennäköisyyksiä on siten syytä tarkastella kahdessa eri tapauksessa ajanhetken jaollisuuden mukaan.

Merkitään NRRW-prosessia hetkellä t yksinkertaisesti $(\mathcal{W}_t, \mathcal{G}_t^s)$. Jos $v, v' \in V$ ja $G, G' \in \mathbb{G}$, niin NRRW-prosessin siirtymätodennäköisyydet saavat esityksen

$$P((v', G'), (v, G)) = \mathbb{P}[(\mathcal{W}_{t+1}, \mathcal{G}_{t+1}^s) = (v, G) \mid (\mathcal{W}_t, \mathcal{G}_t^s) = (v', G')] \quad (6)$$

$$= \mathbb{P}[\mathcal{W}_{t+1} = v, \mathcal{G}_{t+1}^s = G \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \quad (7)$$

Kun $t + 1 \bmod s \neq 0$, ei prosessin verkkoon lisätä uusia solmuja. Satunnaisverkko \mathcal{G}_{t+1}^s on siis varmasti sama kuin sen edeltäjä \mathcal{G}_t^s . Siten $\mathbb{P}[\mathcal{G}_{t+1}^s = G \mid \mathcal{G}_t^s = G] = 1$ riippumattomasti satunnaiskävelyn tilasta. Selvästi tapahtuman komplementti ($\mathcal{G}_{t+1}^s \neq G$) on mahdoton. Lausekkeen 7 mukaisesti

$$\begin{aligned} P((v', G'), (v, G)) &= \mathbb{P}[\mathcal{W}_{t+1} = v, \mathcal{G}_{t+1}^s = G \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \\ &= \mathbb{P}[\mathcal{G}_{t+1}^s = G \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \\ &\quad \times \mathbb{P}[\mathcal{W}_{t+1} = v \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G', \mathcal{G}_{t+1}^s = G] \\ &= \mathbb{P}[\mathcal{G}_{t+1}^s = G \mid \mathcal{G}_t^s = G'] \mathbb{P}[\mathcal{W}_{t+1} = v \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \\ &= \mathbb{1}\{G = G'\} \mathbb{P}[\mathcal{W}_{t+1} = v \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \\ &= \mathbb{1}\{G_i = G_j\} \frac{\psi^{(t)}(v_i, v_j)}{d_t(v_i)} \end{aligned}$$

missä viimeinen yhtäsuuruus pohjautuu kaavan 2 lausekkeeseen. **Kun** $t + 1 \bmod s = 0$, satunnaisverkko \mathcal{G}_{t+1}^s määräytyy satunnaiskävelyn \mathcal{W}_{t+1} realisaatiosta.

$$\begin{aligned} P((v', G'), (v, G)) &= \mathbb{P}[\mathcal{W}_{t+1} = v, \mathcal{G}_{t+1}^s = G \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \\ &= \mathbb{P}[\mathcal{W}_{t+1} = v \mid \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \\ &\quad \times \mathbb{P}[\mathcal{G}_{t+1}^s = G \mid \mathcal{W}_{t+1} = v, \mathcal{W}_t = v', \mathcal{G}_t^s = G'] \end{aligned}$$

Uusi solmu lisätään nyt siihen verkon G' solmuun, johon satunnaiskävely siirtyy ajanhetkellä $t + 1$. Selvästi mahdollisia satunnaisverkon \mathcal{G}_{t+1}^s realisaatioita ovat vain sellaiset verkot, jotka on saatu liittämällä verkkoon G' yksi uusi solmu, $v^+ = v_{\lfloor \frac{t+1}{s} \rfloor}$. Kun verkkojen G ja G' merkitään verkon määritelmän (1) mukaisina solmujen ja kaarien pareina (V, E) ja (V', E') ,

$$\begin{aligned} P((v', G'), (v, G)) &= \mathbb{P}[\mathcal{W}_{t+1} = v \mid \mathcal{W}_t = v', \mathcal{G}_t^s = (V', E')] \\ &\quad \times \mathbb{P}[\mathcal{G}_{t+1}^s = (V, E) \mid \mathcal{W}_{t+1} = v, \mathcal{W}_t = v', \mathcal{G}_t^s = (V', E')] \\ &= \mathbf{1}\{(V, E) = (V' \cup \{v^+\}, E' \cup \{\{v, v^+\}\})\} \\ &\quad \times \mathbb{P}[\mathcal{W}_{t+1} = v' \mid \mathcal{W}_t = v', \mathcal{G}_t^s = (V', E')] \\ &= \mathbf{1}\{G = (V' \cup \{v^+\}, E' \cup \{\{v, v^+\}\})\} \frac{\psi_{G'}(v, v')}{d_{G'}(v')} \end{aligned}$$

missä viimeinen yhtäsuuruus perustuu kaavan 2 lausekkeeseen.

3 Stokastinen esitys

TODO: Yleistä...

3.1 Yleistä teoriaa

Olkoon $(\omega_t)_{t \geq 0}$ jono riippumattomia välin $[0, 1]$ tasajakautuneita satunnaislukuja. Toisin sanoen $w_t \sim \text{Tas}(0, 1)$. Tällaista satunnaislukujen sarjaa voidaan käyttää monimutkaisemman stokastisen prosessin algoritmiseen simulointiin [3]. Simulointi rakennetaan etsimällä deterministinen funktio $\phi : S \times [0, 1] \rightarrow S$, joka liittää stokastisen prosessin tilan ja välin $[0, 1]$ reaaliluvun prosessin seuraavaan tilaan. Funktion ϕ tulee toteuttaa yhtälö

$$\mathbb{P}[\phi(x, \omega_t) = y] = \int_0^1 \mathbf{1}\{\phi(x, w) = y\} dw \quad (8)$$

$$= \mathbb{P}[X_{t+1} = y \mid X_t = x] = P_{x,y} \quad \forall x, y \in S \quad (9)$$

Yhtälön 9 toteuttavaa funktiota ϕ ja satunnaislukujonoa $(\omega_t)_{t \geq 0}$ kutsutaan siirtymä-matriisiin P *stokastiseksi esityksekseksi*. Jos stokastisen prosessin alkutila on $X_0 = x_0$, sitä voidaan simuloida sen stokastisen esityksen avulla rekursiivisesti:

$$X_{t+1} = \phi(X_t, \omega_t) \quad (10)$$

Mikäli stokastinen prosessi on epähomogeinen ajassa, vaaditaan stokastiseen esitykseen jono deterministisiä funktioita $(\phi^{(t)})_{t \geq 0}$, jotka ovat muotoa $\phi^{(t)} : S \times [0, 1] \rightarrow S \quad \forall t \geq 0$. Funktioiden $\phi^{(t)}$ tulee toteuttaa yhtälöä 9 vastaava ehto $\forall t \geq 0$:

$$\mathbb{P}[\phi^{(t)}(x, \omega_t) = y] = \mathbb{P}[X_{t+1} = y \mid X_t = x] = P_{x,y}^{(t)} \quad \forall x, y \in S \quad (11)$$

Prosessin simulointi tehdään, kuten homogeenisessa tilanteessa, mutta rekursio-funktio valitaan funktiojonosta $(\phi^{(t)})_{t \geq 0}$ järjestyksessä:

$$X_{t+1} = \phi^{(t)}(X_t, \omega_t) \quad (12)$$

3.2 Prosessin esitys

Edellä havaittiin, että NRRW-prosessi on epähomogeeninen ajassa. Sen stokastinen esitys on siis muotoa $(\phi_t, \omega_t)_{t \geq 0}$, missä funktiot ϕ_t toteuttavat ehdon 11. Muodostetaan sopiva jono deterministisiä funktioita NRRW-prosessille. Prosessin tilat ovat muotoa (v, G) , missä $G = (V, E)$ on verkko ja $v \in V$ on verkon G solmu. Jakamalla rekursiofunktiot satunnaiskävely- ja satunnaisverkko-osiin saadaan niille korkeatasoinen muoto

$$\phi^{(t)}((v, G), w) = (\lambda((v, G), w), \mu^{(t)}(\lambda((v, G), w), G)) \quad (13)$$

missä λ on satunnaiskävelyn $(\mathcal{W}_t)_{t \geq 0}$ stokastisen esityksen rekursiofunktio parametriksi annetussa verkossa G ja μ päivittää verkkoa G satunnaiskävelyn realisaation mukaisesti. Satunnaiskävelyn rekursiofunktio

$$\lambda((v_i, G), w) = \begin{cases} v_1 & w \in [0, \rho(G)_{i,1}] \\ v_2 & w \in [\rho(G)_{i,1}, \rho(G)_{i,1} + \rho(G)_{i,2}] \\ \vdots & \\ v_{|V|} & w \in [\rho(G)_{i,|V|-1}, \rho(G)_{i,|V|-1} + \rho(G)_{i,|V|}] \end{cases} \quad (14)$$

missä ρ on funktio verkosta sen siirtymämatriisiin ja $|V|$ on verkon G solmujen lukumäärä. Hyödyntäen määritelmän 2.2.2 funktiota solmuja yhdistävien kaarien määälle, saadaan kuvaukselle ρ alkioittainen määritelmä

$$\rho(G)_{i,j} = \frac{\psi_G(i, j)}{d_G(i)}. \quad (15)$$

Tämä seuraa suoraan kaavasta 2. Lopuksi Stokastisen esityksen verkkoa päivittää funktio

$$\mu^{(t)}(v, (V, E)) = \begin{cases} (V \cup \{v_{t/s}\}, E \cup \{\{v, v_{t/s}\}\}) & \text{kun } t \bmod s = 0 \\ (V, E) & \text{muulloin} \end{cases} \quad (16)$$

Se siis lisää verkon solmuun v uuden solmun $v_{t/s}$ aina kun aika t on jaollinen askelparametrillä s . Muulloin verkkoa ei muuteta.

Teoreema 1. Olkoon $(\omega_t)_{t \geq 0}$ jono riippumattomia välin $[0, 1]$ tasajakautuneita satunnaislukuja. Kaavojen 13-16 määrittelemä jono deterministisiä funktioita $(\phi^t)_{t \geq 0}$ muodostaa yhdessä satunnaislukujonon $(\omega_t)_{t \geq 0}$ kanssa NRRW-prosessin stokastisen esityksen.

Todistus. Esitän, että stokastisen esityksen $(\phi^{(t)}, \omega_t)_{t \geq 0}$ simuloima stokastinen prosessi on todellakin NRRW-prosessi askelparametrillä s . NRRW-prosessin siirtymätodennäköisyydet on laskettu kappaleessa 2.3.1. Prosessin käyttäytyminen ajanhetkellä t riippuu siitä, onko t jaollinen askelparametrilla s . Osoitetaan stokastisen esitykseen yhdenmukaisuus NRRW-prosessin kanssa molemmissa tapauksissa. Olkoon $X_t = (\mathcal{W}_t, \mathcal{G}_t^s)$ NRRW-prosessin tila hetkellä t .

On näytettävä kaavan 11 mukaisesti, että

$$\mathbb{P} [\phi^{(t)}(x, \omega_t) = y] = \mathbb{P} [X_{t+1} = y \mid X_t = x] \quad (17)$$

kaikilla $x, y \in S$ ja $t \in \mathbb{N}$, missä S on kaavan 5 määrittelemä NRRW-prosessin tilajoukko. Sijoittamalla aiemmin määritelty stokastinen esitys yhtälön vasempaan puoleen saadaan

$$\begin{aligned} \mathbb{P} [\phi^{(t)}(x, \omega_t) = y] &= \int_0^1 \mathbb{1} \{ \phi^{(t)}(x, w) = y \} dw \\ &= \int_0^1 \mathbb{1} \{ (\lambda((v_x, G_x), w), \mu^{(t)}(\lambda((v_x, G_x), w), G_x)) = (v_y, G_y) \} dw \\ &= \int_0^1 \mathbb{1} \{ \lambda((v_x, G_x), w) = v_y \} \mathbb{1} \{ \mu^{(t)}(\lambda((v_x, G_x), w), G_x) = G_y \} dw \end{aligned}$$

Kun $t + 1 \bmod s \neq 0$

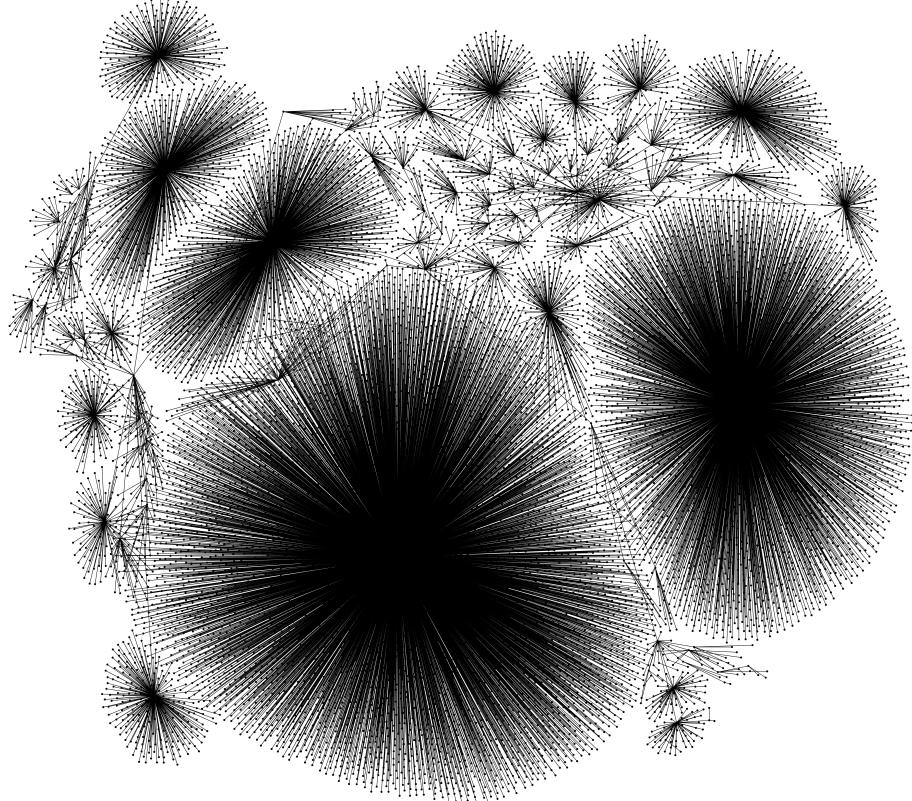
$$\begin{aligned} \mathbb{P} [\phi^{(t)}(x, \omega_t) = y] &= \int_0^1 \mathbb{1} \{ \lambda((v_x, G_x), w) = v_y \} \mathbb{1} \{ \mu^{(t)}(\lambda((v_x, G_x), w), G_x) = G_y \} dw \\ &= \mathbb{1} \{ G_x = G_y \} \int_0^1 \mathbb{1} \{ \lambda((v_x, G_x), w) = v_y \} dw \\ &= \mathbb{1} \{ G_x = G_y \} \rho(G_x)_{x,y} = \mathbb{1} \{ G_x = G_y \} \frac{\psi_{G_x}(v_x, v_y)}{d_{G_x}(v_x)} \\ &= \mathbb{P} [X_{t+1} = y \mid X_t = x] \end{aligned}$$

Kun $t + 1 \bmod s = 0$

$$\begin{aligned} \mathbb{P} [\phi^{(t)}(x, \omega_t) = y] &= \int_0^1 \mathbb{1} \{ \lambda((v_x, G_x), w) = v_y \} \mathbb{1} \{ \mu^{(t)}(\lambda((v_x, G_x), w), G_x) = G_y \} dw \\ &= \mathbb{1} \{ G_y = (V_x \cup \{v_y\}, E_x \cup \{\{v_x, v_y\}\}) \} \int_0^1 \mathbb{1} \{ \lambda((v_x, G_x), w) = v_y \} dw \\ &= \mathbb{1} \{ G_y = (V_x \cup \{v_y\}, E_x \cup \{\{v_x, v_y\}\}) \} \rho(G_x)_{x,y} \\ &= \mathbb{1} \{ G_y = (V_x \cup \{v_y\}, E_x \cup \{\{v_x, v_y\}\}) \} \frac{\psi^{(t)}(v_i, v_j)}{d_t(v_i)} \\ &= \mathbb{P} [X_{t+1} = y \mid X_t = x] \end{aligned}$$

Selvästi rakentamani stokastinen esitys on ekvivalentti NRRW-prosessin kanssa.

□



Kuva 2: Tietokonesimuloitu NRRW-prosessi parillisella askelparametrilla $s = 4$.

4 Parillinen prosessi

NRRW-prosessin käyttäytymiseen vaikuttaa voimakkaasti askelparametrin s parillisuus. Parillisen parametrin prosessissa satunnaiskävely $(\mathcal{W}_t)_{t \in \mathbb{N}}$ on palautuva eli se vierailee kaikissa sen tilajoukon solmuissa äärettömän monta kertaa lähes varmasti. Sen sijaan ainakin askelparametrin $s = 1$ satunnaiskulku on väistyvä. Askelparametrin parillisuus vaikuttaa yhtä radikaalisti satunnaisverkkojen ominaisuuksiin. Parillisen askelparametrin satunnaisverkon $(\mathcal{G}_t^s)_{t \in \mathbb{N}}$ sisäsolmujen astelukujen jakausma on potenssilain alhaalta rajoittama, kun taas askelparametrin $s = 1$ solmujen astelukujen jakausma on geometrisen jakauksen ylhäältä rajoittama.

Erityisesti vain parillisen askelparametrin NRRW-prosessissa kaksi peräkkäistä uutta solmua voidaan lisätä samaan verkon solmuun. Parittoman askelparametrin prosessissa tämä ei ole mahdollista muualla kuin puun juuressa, sillä satunnaiskävely siirtyy s :ssä askeleessa puun parittomalta tasolta parilliselle ja parilliselta tasolta parittomalle. Parillisen askelparametrin prosessissa satunnaiskävelyn pariteetti muuttuu vain, jos satunnaiskävely kulkee puun juuren silmukan läpi.

Uuden, peräkkäisen solmun lisääminen samaan solmuun parillisessa NRRW-verkossa on paitsi mahdollista niin myös sitä todennäköisempää mitä suurempi kyseisen solmun asteluku on. Koska uusia solmuja lisätään ennen kaikkea saman pariteetin tasolle puussa, jäävät uudet solmut lehdiksi (*eng. leaf*), kunnes prosessin pariteetti vaihtuu. Lehtisolmuista voi siirtyä vain takaisin lähtöpisteeseen, joten

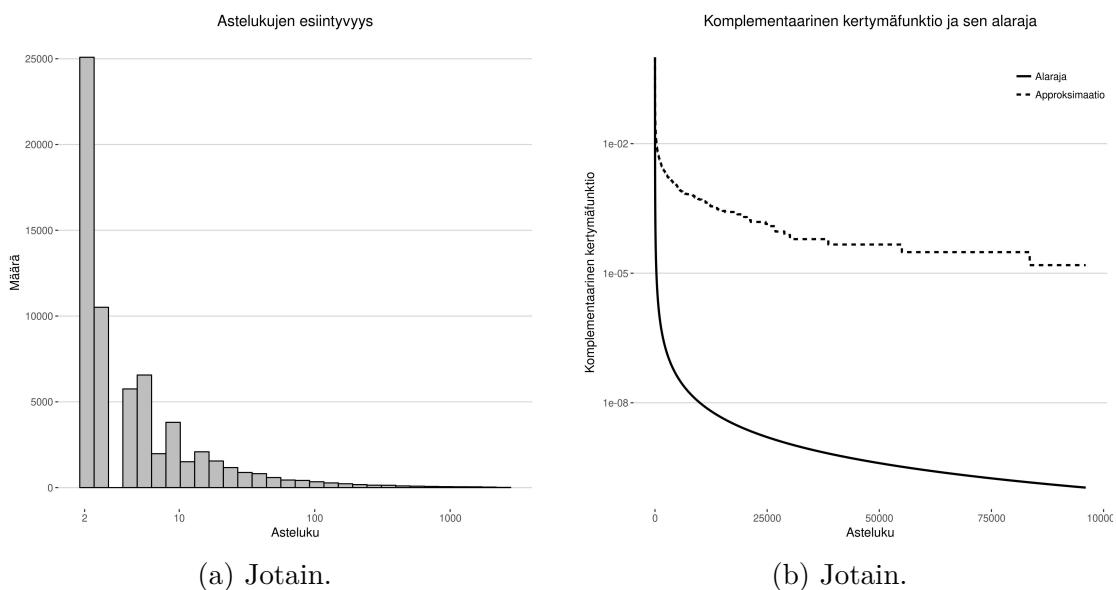
todennäköisyys palata johonkin satunnaisverkon solmuun suurenee, kun solmussa vieraillaan ja sen asteluku kasvaa. Kuvaaja 2 havainnollistaa, kuinka lehtisolmuja kasaantuu tähtimaiseen muodostelmaan joidenkin solmujen ympärille.

Tämä parillisen prosessin ominaisuus liittyy vahasti suoivaan kiinnittymiseen (*eng. preferential attachment*), sillä uusia solmuja kertyy erityisen todennäköisesti sellaisille verkon solmuille, joilla on jo ennestään paljon naapureita. Suosiva kiinnittyminen on tosielämän verkkorakenteiden kannalta kiinnostava ominaisuus. Käsite on yhteydessä sosialogiseen Matteus-efektiin, joka tarkoittaa etujen kasaantumista niille, joilla niitä on jo ennestään. Efektiin ydinajatus tiivistetään usein lausahdukseen: "Rikkaat rikastuu ja köyhät köyhtyy".

TODO: Preferential attachment in IOTA?

Keskitynkin tässä kandidaatintyössä parillisen NRRW-prosessin ominaisuuksia käsittelemiseen.

4.1 Astelukujen jakauma



(a) Jotain.

(b) Jotain.

Teoreema 2. Olkoon NRRW-prosessin askelparametri s parillinen. Oletetaan, että T_j on se ajanhetki, jolloin solmuun j lisätään sen ensimmäinen naapuri. Tällöin kaikilla $t \geq T_j$ ja $k \in \{1, \dots, \lfloor \frac{t-T_j}{s} \rfloor + 1\}$ pätee, että

$$\mathbb{P}[d_t(j) \geq k+1 \mid T_j < \infty] \geq k^{-\frac{s}{2}}, \quad \text{jos } j \neq \text{root}$$

$$\mathbb{P}[d_t(j) \geq k+2] \geq \left(\frac{k(k+1)}{2}\right)^{-\frac{s}{2}}, \quad \text{jos } j = \text{root}$$

4.2 Palautuvuus

Teoreema 3. Olkoon i solmu ja $\{i, j\}$ kaari NRRW-prosessin verkossa. Jos i on palautuva, niin satunnaiskävely $(\mathcal{W}_t)_{t \in \mathbb{N}}$ kulkee kaaren $\{i, j\}$ läpi äärettömän monta

kertaa lähes varmasti.

Todistus. TODO: INSERT PROOF HERE! □

Teoreema 4.

Todistus. □

5 Yhteenveton

Viitteet

- [1] Clauset, A., Shalizi, C. R. ja Newman, M. E. Power-law distributions in empirical data. Department of Computer Science, University of New Mexico, Albuquerque, USA, 2009.
- [2] Barabási, A. Linkit - verkostojen uusi teoria. K. Helsinki: Terra Gognita Oy. 2002. ISBN 952-5202-66-6.
- [3] Barabási, A., Albert, R. Emergence of Scaling in Random Networks. Department of Physics, University of Notre-Dame, Notre-Dame, France, 1999.
- [4] Iacobelli, G., Figueiredo, D. R., Neglia, G. Transient and Slim versus Recurrent and Fat: Random Walks and the Trees they Grow. Istituto de Matemática, Federal University of Rio de Janeiro, Brazil, 2017.
- [5] Popov, S. The Tangle. Viitattu 3.5.2018. Saatavissa: https://iota.org/IOTA_Whitepaper.pdf.
- [6] Popov, S., Saa, O., Finardi, P. Equilibria in the Tangle. Department of Statistics, Institute of Mathematics, Statistics and Scientific Computation, University of Campinas, Brazil, 2017.

A Esimerkki liitteestä

Liitteet eivät ole opinnäytteen kannalta välittämättömiä ja opinnäytteen tekijän on kirjoittamaan ryhtyessään hyvä ajatella pärjäävänsä ilman liitteitä. Kokemattomat kirjoittajat, jotka ovat huolissaan tekstiosan pituudesta, paisuttavat turhan helposti liitteitä pitääkseen tekstiosan pituuden annetuissa rajoissa. Tällä tavalla ei synny hyvää opinnäytettä.

Liite on itsenäinen kokonaisuus, vaikka se täydentääkin tekstiosaa. Liite ei siten ole pelkkä listaus, kuva tai taulukko, vaan liitteessä selitetään aina sisällön laatu ja tarkoitus.

Liitteeseen voi laittaa esimerkiksi listauksia. Alla on listausesimerkki tämän liitteen luomisesta.

```
\clearpage
\appendix
\addcontentsline{toc}{section}{Liite A}
\section*{Liite A}
...
\thispagestyle{empty}
...
tekstiä
...
\clearpage
```

Kaavojen numerointi muodostaa liitteissä oman kokonaisuutensa:

$$d \wedge A = F, \tag{A1}$$

$$d \wedge F = 0. \tag{A2}$$

B Toinen esimerkki liitteestä

Liitteissä voi myös olla kuvia, jotka eivät sovi leipätekstin joukkoon: Liitteiden taulu-

Kuva B1: Kuvateksti, jossa on liitteen numeroointi

koiden numeroointi on kuvien ja kaavojen kaltainen: Kaavojen numeroointi muodostaa

Taulukko B1: Taulukon kuvateksti.

9.00–9.55	Käytettävyyystestauksen tiedotustilaisuus (osanottajat ovat saaneet sähköpostitse val- mistautumistehtävät, joten tiedotustilai- suus voidaan pitää lyhyenä).
9.55–10.00	Testausalueelle siirtyminen

liitteissä oman kokonaisuutensa:

$$T_{ik} = -pg_{ik} + wu_i u_k + \tau_{ik}, \quad (\text{B1})$$

$$n_i = nu_i + v_i. \quad (\text{B2})$$