

hyväksymispäivä arvosana

arvostelija

## **Aine**

Ola Länsman

Helsinki 12.3.2017

HELSINGIN YLIOPISTO  
Tietojenkäsittelytieteen laitos

# 1 Johdanto

Käytännön kokeilla sosiaalisista verkoista muodostettu Pieni maailma-ilmiö on vaikuttanut myös tietojenkäsittelytieteeseen. Ilmiötä on yleistetty useampiin ulottuvuuksiin sekä sen tärkeimpiä ominaisuuksia kuten klusteroitumista tutkittu. Ilmiön perusteella on tehty useita polunetsintä-algoritmeja jotka käyttävät ainoastaan lokaalia tietoa verkon solmuista ja kaarista.

Tutkielma pyrkii selittämään Pieni maailma-ilmiön verkkoteoriaa ja muutaman yleisen polunetsintä-algoritmin. Näiden jälkeen kiinnostuksen kohteeksi nousevat käytännön sovellukset joissa näitä käytetään.

## 2 Pieni maailma

### 2.1 Määritelmä

Pieni maailma-ilmiö tarkoittaa seuraavan ehdon toteumista verkossa:

**Lemma 1** *Yksittäiset solmut voivat lähettää viestin verkon kaaria pitkin muihin verkon solmuihin lyhyitä polkuja pitkin käyttämällä ainoastaan paikallista tietoa.*

Paikallisella tiedolla tarkoitetaan globaalin tiedon puuttumisella. Jokaisessa solmussa valitaan viestin seuraava vastaanottaja käyttäen hyväksi ainoastaan tietoa viestiä lähettävän solmun kaarista. [DHLS06]

Pieni maailma-verkot (PM) yleensä täyttävät myös seuraavat ehdot:

1. *ne ovat harvoja* [BBS11]: 1 täyttyy implisiittisesti, mikäli lähtösolmusta löytyy kaari kohteeseen. Tällaiset verkot eivät ole mielenkiintoisia ilmiön kannalta.
2. *solmujen välillä esiintyy lyhyitä polkuja* [BBS11]: 1 aiheuttaa tämän suoraan.
3. *ne ovat ryhmittyneitä ja niillä on pieni halkaisija* [BBS11]: Intuitiivisesti ehto 3 tarkoittaa, että mikäli solmu  $u$  ja  $v$  ovat lähellä toisiaan niin niiden välillä on todennäköisesti kaari. Yleisimmissä polunetsintä-algoritmeissa viesti lähetetään aina mahdollisimman lähelle kohdetta ehdosta 3 johtuen.

Ehdon 1 on usein täytyttävä, jotta verkko olisi mielenkiintoinen. Ääriesimerkkinä voi ajatella verkkoa, jossa jokaisesta solmusta on kaari kaikkiin muihin solmuihin. Ehto 2 on suoraa seurausta lemmasta 1.

## 2.2 Kuinka muodostetaan pieni maailma

### 2.2.1 Kleinbergin malli

PM-verkkoja voidaan muodostaa monella tavalla. Aloitamme muodostamalla mallin, joka tehtiin kuvaamaan verkkoa ihmisten sosiaalisista suhteista.

Mallinnamme ihmisiä solmuina. Tämä joukko solmuja muodostaa  $n \times n$  ruudukon, missä

$$V = \{(i, j) : i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, n\}\}.$$

Olkoon ruudukkoetäisyys  $d((i, j), (k, l)) = |k - i| + |l - j|$ . Solmulla  $u$  on paikallinen kontakti solmun  $v$  kanssa, jos  $d(u, v) \leq p$  jollain vakiolla  $p \geq 1$ . Solmulla  $u$  on myös vakiomäärä,  $q \geq 0$ , etäkontakteja. Etäontaktit muodostetaan satunnaisesti todennäköisyysfunktiolla joka riippuu vakioista  $r \geq 0$  ja etäisyydestä  $d(u, v)$ . Tarkemmin, etäkontakti muodostuu solmusta  $u$  solmuun  $v$  todennäköisyydellä

$$[d(u, v)]^{-r}.$$

Tämä tapa muodostaa PM-verkko voidaan tulkita myös geometrisesti. Solmulla  $u$  on kaari jokaiseen tarpeeksi lähellä olevaan solmuun. Näiden yhteyksien lisäksi solmulla  $u$  on kaaria kauempana ruudukossa. Jos vakio  $r = 0$ , niin solmujen etäkontaktit ovat jakautuneet tasaisesti ruudukolle. Vakion  $r$  kasvaessa etäkontaktit ovat lähempänä ja lähempänä solmua itseään.

[Kle00]

## 3 Polunetsintä

### 3.1 Strategiat

#### 3.1.1 Ahne reititys

Ensimmäisenä tutkimme normaalia ahnetta algoritmia lyhyen polun etsintään. Ahneissa algoritmeissa viestiä kuljettava solmu lähettää viestin aina lähimpänä kohdetta olevalle naapurilleen. Esimerkiksi Kleinbergin esittämässä ahneessa algoritmissa viestiä kuljettava solmu tietää

1. kaikkien solmujen paikalliset kontaktit
2. kohteen  $y$  sijainti ruudukossa
3. ja kaikkien viestiä kuljettaneiden solmujen etäkontaktit ja sijainnit

.

### 3.1.2 Epäsuora ahne reititys

Epäsuora ahne reititys toimii kuten ahne reititys. Poikkeuksena, viestiä kuljettavalla solmulla  $u$  on tiedossa myös solmun  $v$  etäkontaktit jos solmulle pätee  $d(u, v) \leq q$  jollain etäisyysfunktioilla  $d$  ja vakiolla  $q$ . Tällöin viestiä kuljettavalla solmulla on mahdollisuus lähettää viesti jonkin itseään lähellä olevan solmun kautta.

### 3.1.3 Naapurien naapurit

Ahneella naapurien naapurit-algoritmi toimii samalla periaatte kuin epäsuora ahne reititys. Lähellä olevien solmujen sijaan viestiä kuljettavalla solmulla on tiedossa omien naapureidensa kontaktit.

## 4 Käytännön sovellukset

### 4.1 Peer-to-peer verkot

### 4.2 Luonnollisen kielen tiivistäminen

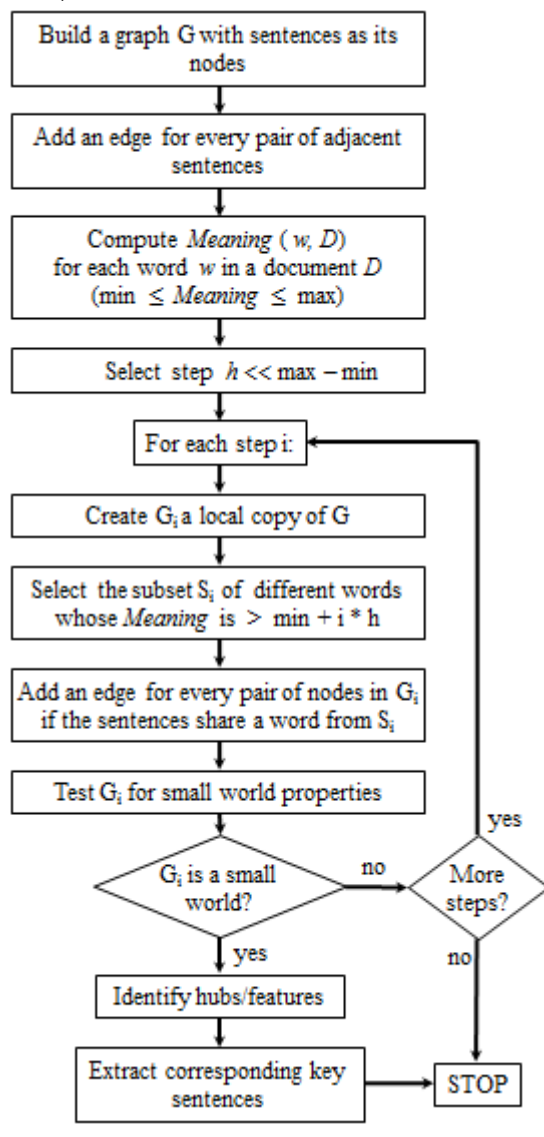
Luonnollisten kielten tekstien tiivistäminen on hyödyllistä nykymaailmassa elektronisesti käsillä olevan tiedon kasvaessa. Tekstiä voi tiivistää valikoiden ydinkohdat tai esittäen nämä lyhyesti uusin sanoin. Seuraavaksi esittelemme automaattisen valikoivan tiivistämiskeinon, joka käyttää hyväkseen pienten maailmojen topologiaa. Rakennamme tekstin virkkeiden verkon ja poimimme virkkeistä ne, jotka edesauttavat verkkoa eniten olemaan pieni maailma.

Määrittelemme verkon  $G = (V, E)$ , jossa pisteet  $V$  ovat lauseita ja kaaret  $E$  kuvaavat virkkeiden välisiä suhteita. Virkkeellä  $L$  on lähikontakti virkkeen  $L'$  kanssa, jos virkkeet  $L$  ja  $L'$  ovat peräkkäisiä. Etäkontaktien muodostamiseen tarvitsemme keinon määrittää virkkeiden yhteyden toisiinsa. Tätä varten rakennamme tekstin tärkeimmistä sanoista joukon  $MeaningfulSet(e)$ . Etäkontakti kahden virkkeen välille muodostuu vain, jos kummassakin virkkeessä esiintyy jokin joukon  $MeaningfulSet(e)$  sanoista.

Joukon  $MeaningfulSet(e)$  sanojen määrä suhteessa joukkoon kaikista tekstissä esiintyvistä sanoista vaikuttaa verkon  $G$  kaarien määrän ja täten myös tiivistelmän pituuteen ja olennaisuuteen. Jos joukko  $MeaningfulSet(e)$  on liian suuri, verkko  $G$  ei näytä enää pieneltä maailmalta vaan sattumanvaraisesti muodostetulta verkolta. Kuitenkin joukon  $MeaningfulSet(e)$  ollessa liian pieni verkko  $G$  näyttää molempiin suuntiin linkitetyltä listalta josta löytyy muutama poikkeus. Tiivistysmenetelmän toimintaperiaatteen kannalta tällöin on suuresti merkitystä, miten tämä joukko valitaan. Tätä menetelmää emme esitä tässä tutkielmassa.

[BBS11]

Kuva 1 (Automaattinen tekstin tiivistäminen)



//

## Lähteet

- BBS11 Balinsky, H., Balinsky, A. ja Simske, S. J., Automatic text summarization and small-world networks. *Proceedings of the 11th ACM Symposium on Document Engineering, DocEng '11*, New York, NY, USA, 2011, ACM, sivut 175–184, URL <http://doi.acm.org/10.1145/2034691.2034731>.
- DHLS06 Duchon, P., Hanusse, N., Lebhar, E. ja Schabanel, N., Could any graph be turned into a small-world? *Theoretical Computer Science*, 355,1(2006), sivut 96 – 103. URL <http://www.sciencedirect.com/science/article/pii/S0304397505009187>.

- Kle00 Kleinberg, J., The small-world phenomenon: An algorithmic perspective. *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, STOC '00, New York, NY, USA, 2000, ACM, si-vut 163–170, URL <http://doi.acm.org.libproxy.helsinki.fi/10.1145/335305.335325>.