

1. Johdanto

Harjoitustyössä analysoidaan Rovira ja Virgilin yliopiston henkilökunnan sähköpostiliikenteestä kerättyä aineistoa. Aineistoa voidaan kuvata verkostona, jossa yliopiston henkilöt ovat solmuja, joiden välille piirretään kaari, jos henkilöiden välillä on kulkenut yksikin sähköpostiviesti.

2. Adjasenssimatriisiksi Matlab-ohjelmistossa

Aineisto saatiin EDG-tiedostossa. Ladattaessa tiedosto, päästiin käsiksi 3x5451-kokoisen matriisin, jossa ensimmäinen ja toinen kolumni vastaa verkoston solmujen indeksejä. Matlab-ohjelmiston min- ja max-toimintojen mukaan indeksit on numeroitu välille [1,1133]. Ensimmäisen ja toisen kolumnin osoittamien indeksien solmujen välillä on kaari, jonka paino on kerrottu kolmannessa kolumnissa. Kaikki kaaret ovat painoltaan 1, joten verkosto on painottomaton.

Tarkempaa analysointia varten muunnetaan tiedosto adjasenssimatriisiksi. Muodostetaan adjasenssimatriisi seuraavalla ohjelmakoodilla:

```
A=sparse(zeros(1133,1133))
for y = 1:5451
    rivi = N(y,1);
    kolumni = N(y,2);
    A(rivi, kolumni) = N(y,3);
    A(kolumni, rivi) = N(y,3);
end
```

3. Verkoston perusominaisuudet

Koska solmut on numeroitu järjestyksessä kaikki numerot käyttäen, saadaan koko verkoston solmujen lukumäärä helposti selville MatLabin max-toiminnon avulla. Solmuja on siis $N=1133$. Kaarten lukumäärä on myös heti alussa jo selvillä. Alkuperäisessä tiedostossa on lueteltu kaikki yhteydet, joten verkoston **kaarten lukumäärä** on sama kuin alkuperäisen matriisin rivien määrä eli $|E|=5451$.

Verkoston **kaaritiheys** saadaan jakamalla kaarten lukumäärä mahdollisten kaarten lukumäärällä [kaava 1].

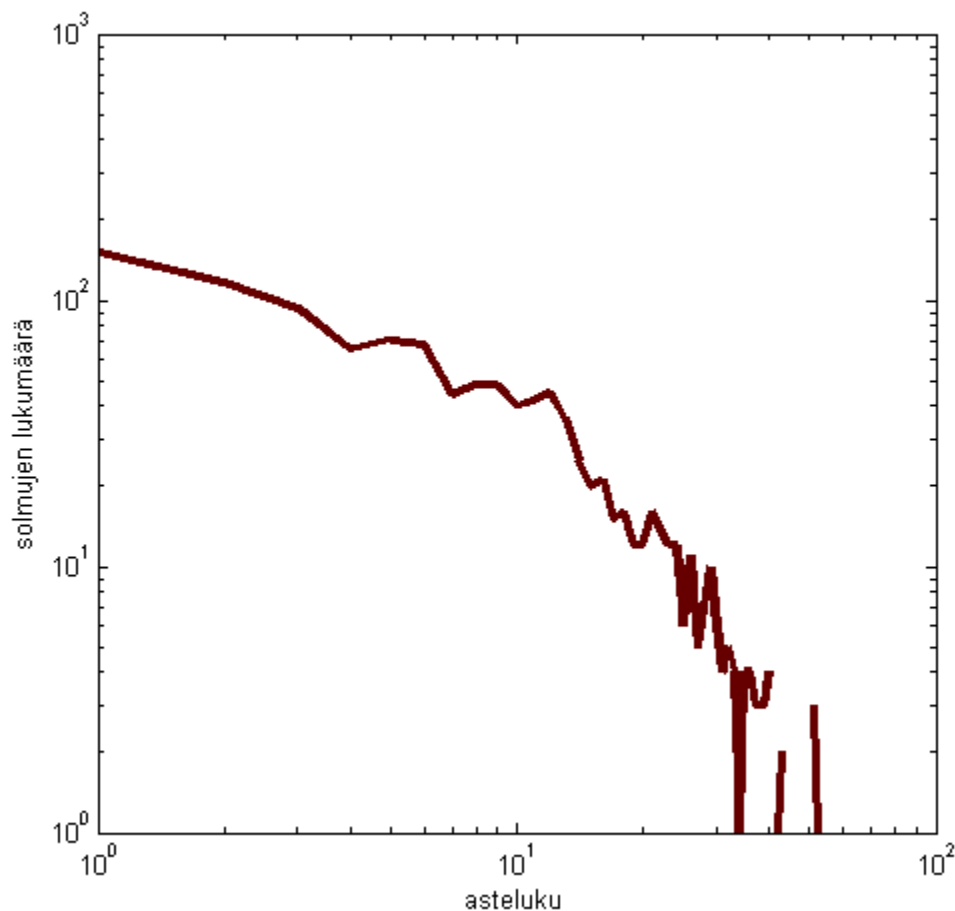
$$\rho = \frac{2|E|}{N(N-1)} = \frac{2 * 5451}{1133(1133-1)} \approx 0,0085 \quad \text{kaava 1}$$

Luku on juuri sosiaalisille verkostoille tyypillisen pieni. Edes yliopistomaailmassa ei kaikki voi tuntea kaikkia.

4. Astejakauma

Verkoston astejakaumaa kuvaava astevektori saadaan komennolla $k = \text{sum}(A, 1)$. Astevektorin avulla saadaan selville verkoston solmujen **keskiaste** komennolla $\text{mean}(k)$, joka antaa vastaukseksi **n. 9,6**. Suurin ja pienin asteluku saadaan selville min- ja max-komentojen avulla. **Minimiaste on 1** ja **maksimiaste 71**.

Verkoston astejakaumaa voidaan kuvata log-log-asteikolla [Kuvaaja 1] käyttämällä Matlabin plottoolityökalua. X-akselilla on merkittynä verkoston kaikki eri asteluvut ja y-akselilla kunkin asteluvun omaavien solmujen lukumäärä.



Kuvaaja 1: verkoston astejakauma

Kuvaajasta huomataan odotetusti, että suurimmalla osalla solmuista eli henkilöistä on suhteellisen vähän kontakteja. Tämä on aivan luonnollista, koska kukaan ei ehdi todellisuudessa olemaan kaikkien kanssa yhteyksissä.

5. Klusterointikerroin

Klusterointikerroin kuvaa, miten tiheässä lähinaapureiden väliset linkit ovat. Klusterointikerrointa kuvaava osuus saadaan, kun tietyn solmun lähinaapureiden väliset linkit jaetaan lähinaapureiden mahdollisten linkkien lukumäärällä. Verkoston keskimääräinen klusterointikerroin saatiin selville seuraavanlaisella ohjelmakoodilla:

```

klusterijakaumaRatkaisu = 0;
% klusterijakaumaRatkaisu-matriisi:
% 1.sarakkeeseen kavereiden lkm,
% 2.sarakkeeseen yhteisten kavereiden lkm,
% 3.sarakkeeseen kavereiden välisten yhteyksien max-lkm.

for tyontekijaNro = 1:1133
    % nollataan kyseisen solmun "kaveritaulukko"
    kaverit = 0;
    kavereidenLkm = k(tyontekijaNro);
    klusterijakaumaRatkaisu(tyontekijaNro,1) = kavereidenLkm; %#ok<AGROW>

    % jos tarkasteltavan solmun aste on 1, talletetaan tulokset heti
    if (kavereidenLkm == 1)
        klusterijakaumaRatkaisu(tyontekijaNro,2) = 0; %#ok<AGROW>
        klusterijakaumaRatkaisu(tyontekijaNro,3) = 0; %#ok<AGROW>

    else
        % kerätään tarkasteltavan solmun "kaveritaulukko"
        x = 1;
        for mahdollinenKaveriNro = 1:1133
            if (A(tyontekijaNro, mahdollinenKaveriNro) == 1)
                && (tyontekijaNro ~= mahdollinenKaveriNro)
                    kaverit(x, 1) = mahdollinenKaveriNro; %#ok<AGROW>
                    x = x + 1;
            end
        end

        % lasketaan solmun kaverit, joilla on keskenään yhteys
        yhteisiaKavereita = 0;
        for i = 1: (length(kaverit) - 1)
            for j = (i+1):length(kaverit)
                if (A(kaverit(i, 1), kaverit(j, 1)) == 1) && (i ~= j)
                    yhteisiaKavereita = yhteisiaKavereita + 1;
                end
            end
        end

        klusterijakaumaRatkaisu(tyontekijaNro, 2) = yhteisiaKavereita;
        %#ok<AGROW>
    end
end

```

```

klusterijakaumaRatkaisu(tyontekijaNro,3) = (kavereidenLkm *
      (kavereidenLkm - 1)) / 2;
      %#ok<AGROW>
      % talletetaan osuus 4. kolumniin
if klusterijakaumaRatkaisu(tyontekijaNro,3) ~= 0
      klusterijakaumaRatkaisu(tyontekijaNro,4) =
klusterijakaumaRatkaisu(tyontekijaNro,2) /
      klusterijakaumaRatkaisu(tyontekijaNro,3); %#ok<AGROW>
end
end
      % tulostetaan saatu, kokoverkoston keskimääräinen klusterikerroin
kKlusterointikerroin = mean(klusterijakaumaRatkaisu(:,4)) %#ok<NOPTS>

```

Ohjelmakoodini tarjoaa **keskimääräiseksi klusterointikertoimeksi 0,2202**, mikä tuntuu järkevältä, koska kyseessä on sosiaalinen verkosto. Sosiaalisessa verkostossa on yleistä, että henkilön kaverit tuntevat myös toisensa.

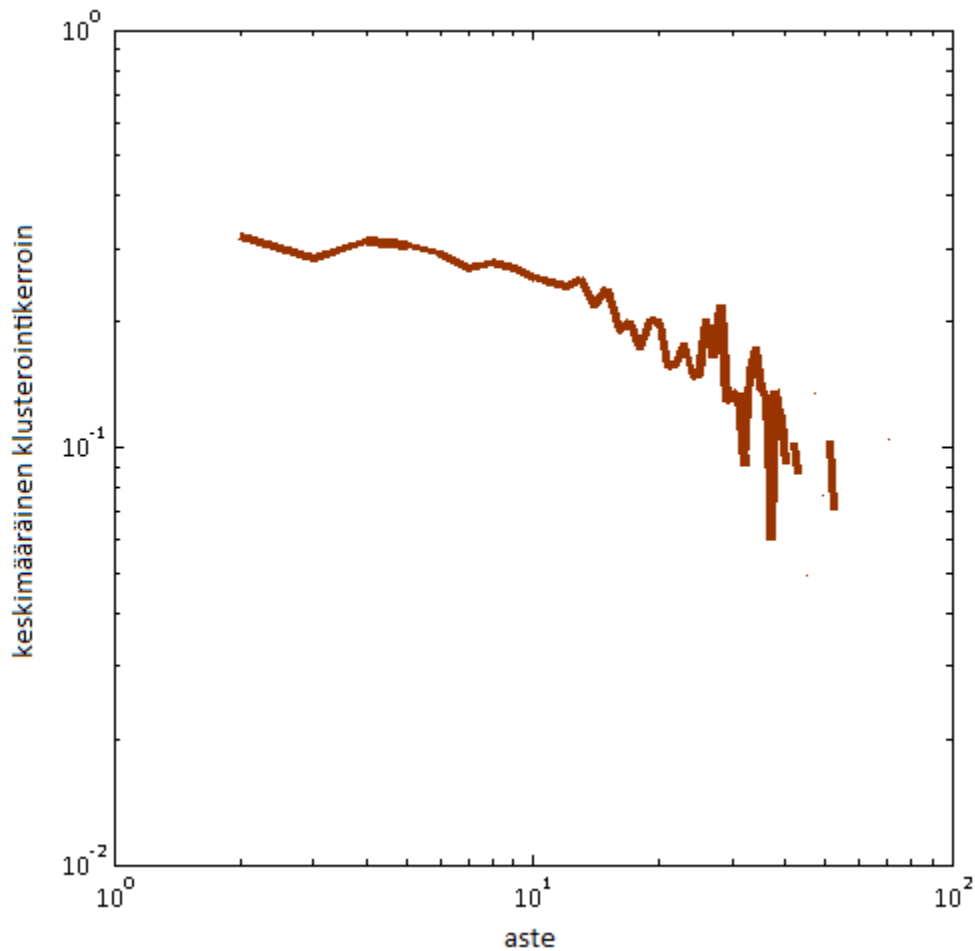
Keskimääräisestä klusterointikertoimesta asteen funktiona esittävään kuvaajaan [Kuvaaja 2] tarvittava data saatiin helposti kerättyä edellisestä kohdasta saatua tulosta hyväksikäyttäen. Seuraava ohjelma kerää edellisen kohdan ratkaisumatriisista kuvaajaa varten tarvittavat tiedot:

```

klustJakauma = 0;
for aste = 1:71
      summa = 0;
      lkm = 0;
      for i = 1:1133
            if klusterijakaumaRatkaisu(i,1) == aste
                  summa = summa + ratkaisu(i,4);
                  lkm = lkm + 1;
            end
      end

      if lkm == 0
            klustJakauma(aste,1) = 0; %#ok<AGROW>
      else
            klustJakauma(aste,1) = summa / lkm; %#ok<AGROW>
      end
end
end

```



Kuvaaja 2: keskimääräinen klusterointikerroin asteen funktiona

Kuvaajasta 2 huomataan, miten korkeimman asteen hubisolmujen kaikki naapurit eivät tunne toisiaan yhtä hyvin kuin pienemmän asteen solmujen naapurit.

6. Keskimääräinen naapuriaste

Keskimääräinen naapuriaste saadaan aineistosta laskemalla jokaiselle solmulle naapureiden asteiden keskiarvo, jonka jälkeen keskiarvoistetaan solmujen asteiden yli [kaava 2].

$$k_{nn}(k) = \frac{1}{N(k)} \sum_{i, k_i=k} \left[\frac{1}{k} \sum_{j \in [i]} k_j \right]$$

kaava 2

Keskimääräisen naapuriasteen kuvaajan [kuvaaja 3] tarvitsemat tiedot saatiin seuraavalla ohjelmakoodilla:

```

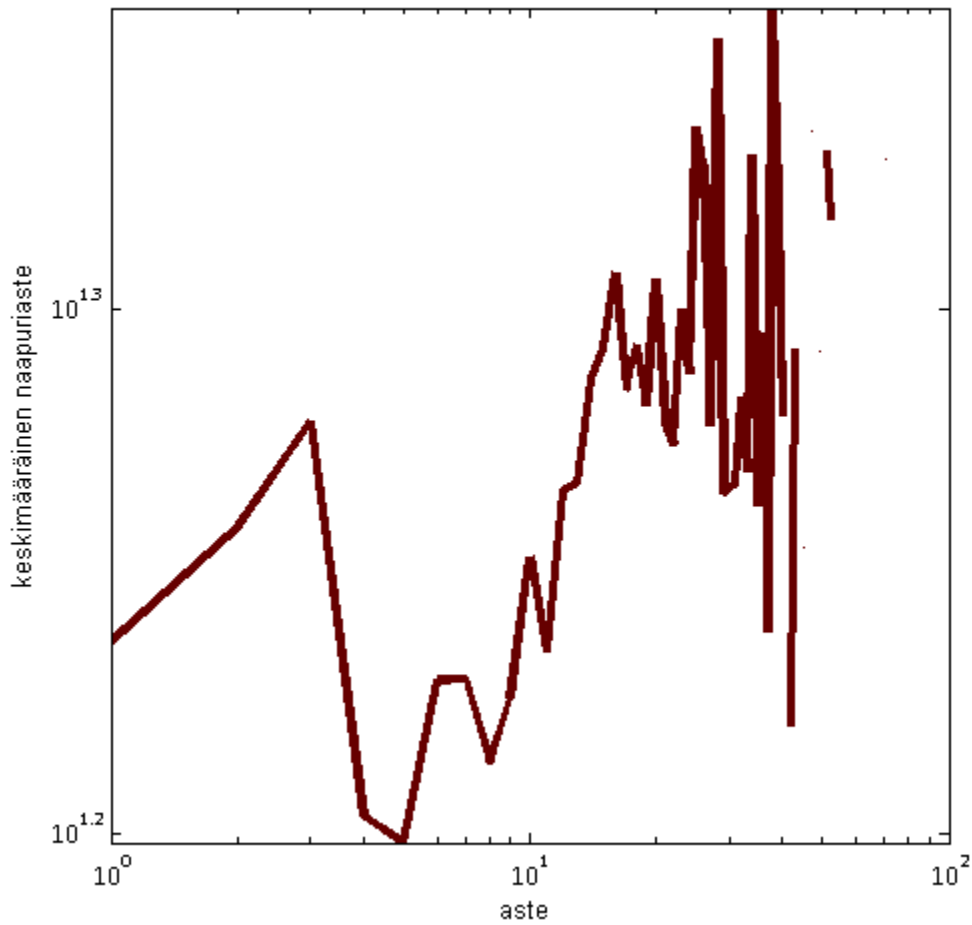
kNaapuriAsteet = 0;
for tyontekijaNro = 1:1133
    % nollataan kyseisen solmun "kavereiden asteet"-taulukko, johon
    % kerätään käsiteltävän solmun kaikkien kavereiden asteluvut
    solmunKavereidenAsteet = 0;

    % kerätään tarkasteltavan solmun "kavereiden asteet"-taulukko
    x = 1;
    for mahdollinenKaveriNro = 1:1133
        if (A(tyontekijaNro, mahdollinenKaveriNro) == 1) && (tyontekijaNro
            ~= mahdollinenKaveriNro)
            solmunKavereidenAsteet(x, 1) = k(mahdollinenKaveriNro);
            %#ok<AGROW>
            x = x + 1;
        end
    end
    % talletetaan asteiden keskiarvo
    kNaapuriAsteet(tyontekijaNro, 1) = mean(solmunKavereidenAsteet);
    %#ok<AGROW>
end

% kerätään tiedot asteiden mukaan keskiarvoistettuina.
% ohjelma käyttää hyväkseen edellisen ohjelman luomaa
% klusterijakaumaRatkaisu-matriisia
asteenSuhteenKNaapuriaste = 0;
for aste = 1:71
    summa = 0;
    lkm = 0;
    for i = 1:1133
        if klusterijakaumaRatkaisu(i,1) == aste
            summa = summa + kNaapuriAsteet(i,1);
            lkm = lkm + 1;
        end
    end

    if lkm == 0
        asteenSuhteenKNaapuriaste(aste,1) = 0; %#ok<AGROW>
    else
        asteenSuhteenKNaapuriaste(aste,1) = summa / lkm; %#ok<AGROW>
    end
end
end

```



Kuvaaja 3: keskimääräinen naapuriaste asteen funktiona

Kuvaajasta 3 voidaan miltei tehdä johtopäätös, että $k_{nn}(k)$ on kasvava. Tämä tarkoittaa sitä, että käsittelemämme verkosto on muiden sosiaalisten verkostojen tapaan assortatiivinen, eli korkean asteen solmut viihtyvät toistensa seurassa. Käytännössä siis ahkerat sähköpostinlähettäjät käyvät sähköpostinvaihtoa muiden ahkerien lähettelijöiden kanssa.

7. Yhteenveto

Sähköpostilokeista saadut tunnusluvut vastasivat yllättävän hyvin ennako-odotuksiani. Kaaritiheys on pieni, astejakaumaa kuvaava käyrä on laskeva, keskimääräinen klusterikerroin on suhteellisen suuri ja näiden lisäksi verkosto on assortatiivinen. Nämä kaikki ominaisuudet ovat sosiaalisille verkostoille tyypillisiä.