Verkot ja vierusmatriisi

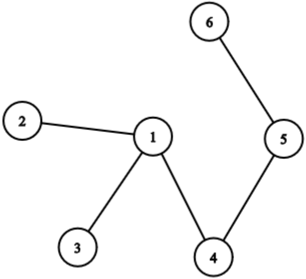
Aleksis Koski

Aalto-yliopisto
Perustieteiden
korkeakoulu

Suomen
Akatemia

Verkot

Matematiikassa **verkko** tai **graafi** voi kuvata mitä tahansa verkostomaista rakennetta. Verkko koostuu kahdesta osasta: **solmut** ja niiden väliset **linkit**.



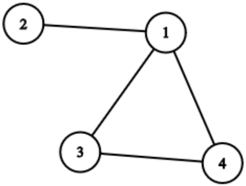
Solmuja: 6 kpl

Linkkejä: 5 kpl

Englanniksi verkko on **graph**, solmu on **vertex**, ja linkki on **edge**.

Vierusmatriisi

Verkkoa voidaan myös kuvata **vierusmatriisilla**. Jos verkossa on n solmua, niin vierusmatriisi on $n \times n$-matriisi, jossa rivin i sarakkeen j alkio on 1 tai 0 sen perusteella, onko verkon solmun i ja solmun j välillä linkki vai ei.



	1	2	3	4
1	0	**1**	**1**	**1**
2	**1**	0	0	0
3	**1**	0	0	**1**
4	**1**	0	**1**	0

Vierusmatriisi on aina symmetrinen matriisi, ja sen diagonaalilla alkiot ovat nollia.

Vierusmatriisi

Vierusmatriisia tutkimalla voidaan selvittää ominaisuuksia sitä esittävästä verkosta.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Yllä oleva matriisi esittää erään verkon vierusmatriisia.

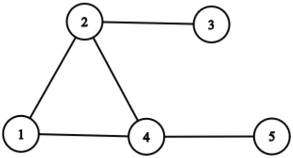
Kuinka monta solmua ja linkkiä verkossa on? Mikä on suurin määrä linkkejä, jota yhdestä solmusta lähtee tässä verkossa?

Vierusmatriisi

Kertomalla vierusmatriisin pystyvektorilla, jossa on pelkkiä ykkösiä, saadaan selville jokaisen solmun **asteluku**, eli kuinka monta linkkiä siitä lähtee.

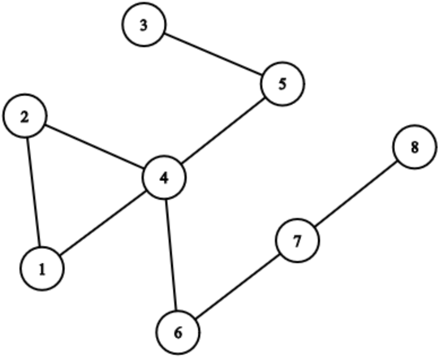
$$Av = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 3 \\ 1 \end{bmatrix}$$

Kyseinen verkko:



Kävelyt

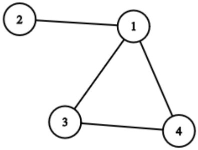
Verkossa **kävely** tarkoittaa mitä tahansa jonoa solmuja, missä kahden perättäisen solmun välillä on linkki.



Kuvan verkossa kävely voisi olla vaikkapa 1,2,4,5,4,6,7,8. Kävely saa käydä saman solmun läpi enemmän kuin kerran.

Kävelyt ja vierusmatriisi

Tutkitaan, mitä vierusmatriisin toinen potenssi kertoo.



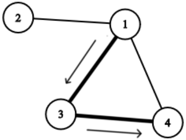
$$A = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Lasketaan...

$$A^2 = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$$

Kävelyt ja vierusmatriisi

Toinen potenssi A^2 kertoo, kuinka monta kahden askeleen kävelyä on jokaisen solmuparin välillä.



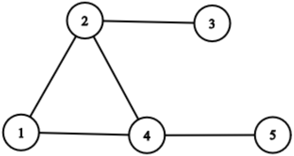
Esimerkiksi solmusta 1 pääsee solmuun 4 tässä verkossa kahdella askeleella vain yhdellä tavalla - solmun 3 kautta. Tämä näkyy kertolaskussa seuraavasti:

$$A^2 = \underbrace{\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}}_{\substack{\text{Ykkösestä pääsee} \\ \text{kolmoseen.}}} \underbrace{\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}}_{\substack{\text{Kolmosesta pääsee} \\ \text{neloseen.}}} = \underbrace{\begin{bmatrix} 3 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}}_{\substack{\text{Tämä tuottaa reitin} \\ \text{ykkösestä neloseen.}}}$$

Matriisin A^2 alkio kohdassa $(1, 1)$ taas on 3, sillä solmusta 1 pääsee takaisin solmuun 1 kolmella eri kahden askeleen kävelyllä.

Kävelyt ja vierusmatriisi

Vastaavasti matriisi A^k kertoo jokaisella luvulla k, montako kävelyä solmujen välillä löytyy, joissa askelia on tasan k kappaletta.



Lasketaan, montako kahdeksan askeleen kävelyä tästä verkosta löytyy:

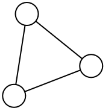
$$A^8 = \begin{bmatrix} 194 & 217 & 97 & 217 & 97 \\ 217 & 271 & 98 & 237 & 119 \\ 97 & 98 & 55 & 119 & 42 \\ 217 & 237 & 119 & 217 & 98 \\ 97 & 119 & 42 & 98 & 55 \end{bmatrix}$$

Miksi kolmosrivin luvut ovat pienempiä kuin kakkosrivin? Miksi vitosrivin luvut ovat samannäköisiä kuin kolmosrivillä?

Kävelyt ja vierusmatriisi

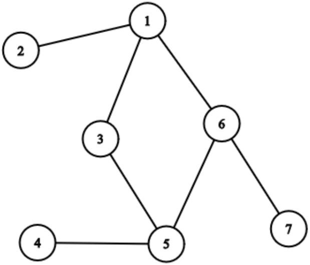
Tehtävä. Erään verkon vierusmatriisin kolmas potenssi on

$$A^3 = \begin{bmatrix} 0 & 4 & 0 & 1 & 2 & 1 \\ 4 & 2 & 6 & 5 & 1 & 7 \\ 0 & 6 & 0 & 2 & 4 & 1 \\ 1 & 5 & 2 & 2 & 2 & 4 \\ 2 & 1 & 4 & 2 & 0 & 5 \\ 1 & 7 & 1 & 4 & 5 & 2 \end{bmatrix}.$$

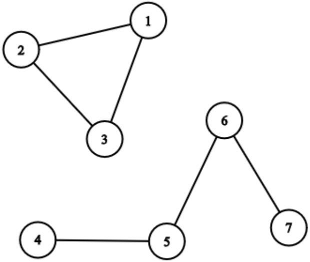
Montako kolmiota  verkosta löytyy?

Yhtenäisyys

Verkko on yhtenäinen, jos sen jokaisen kahden eri solmun välille löytyy kävely (tai polku).



Yhtenäinen



Ei yhtenäinen

Miten yhtenäisyys selviää vierusmatriisin avulla?

Yhtenäisyys

Vastaus: Verkko on yhtenäinen, jos sen vierusmatriisin potensseilla $A, A^2, A^3, \ldots$ jokaista kohtaa (i, j) vastaava matriisin alkio poikkeaa jossain vaiheessa nollasta.

Nimittäin jos kohdassa (i, j) matriisin A^k alkio on positiivinen, niin tällöin on löytynyt k askeleen kävely solmusta i solmuun j.

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \; A^2 = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \; A^3 = \begin{bmatrix} 0 & 3 & 0 & 2 \\ 3 & 0 & 2 & 0 \\ 0 & 2 & 0 & 1 \\ 2 & 0 & 1 & 0 \end{bmatrix}$$

Huom! Jos verkossa on n solmua, riittää tarkistaa vain potenssit $A, A^2, \ldots, A^{n-1}$. Tämä siksi, että lyhin mahdollinen yhteys käy aina korkeintaan läpi jokaisen solmun tasan kerran.

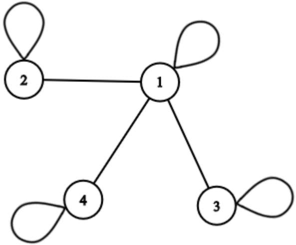
Yhtenäisyys

Tämän perusteella yhtenäisyys selviää myös laskemalla yhteen matriisisumma

$$A + A^2 + A^3 + \cdots + A^{n-1}$$

ja tarkistamalla, että tämän matriisin jokainen (ei-diagonaalinen) alkio on positiivinen. Mitä tämän matriisin alkiot esittävät?

Vähän näppärämpi tapa:
Hyväksytään myös sellaiset kävelyt, joissa joka askeleella voimme päättää myös pysyä paikallaan solmussa.



Yhtenäisyys

Lisätään vierusmatriisiin A samankokoinen identtinen matriisi:

$$\mathrm{I} + A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Nyt matriisin $(\mathrm{I} + A)^k$ alkiot kuvaavat, kuinka monta tapaa on kävellä solmusta toiseen kun paikallaan pysyminen on myös sallittu.

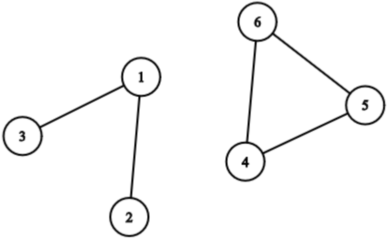
Tällöin verkon yhtenäisyys voidaan tarkistaa matriisin $(\mathrm{I} + A)^{n-1}$ alkioista. Jos jokainen alkio on positiivinen, verkko on yhtenäinen. Tietokoneella laskettaessa kannattaa selvittää vain matriisit

$$(\mathrm{I} + A)^2, \quad (\mathrm{I} + A)^4, \quad (\mathrm{I} + A)^8, \quad \ldots$$

Yhtenäisyys

Edellä esitelty tapa selvittää verkon yhtenäisyys ei kuitenkaan ole kovin käytännöllinen. Nopeampi algoritmi olisi käydä verkko läpi solmu kerrallaan yhdestä solmusta alkaen.

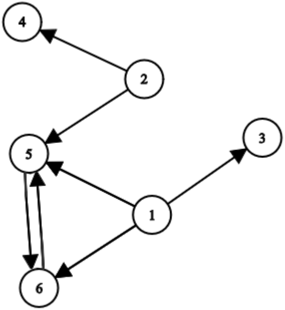
Epäyhtenäisen verkon vierusmatriisin rivit voi myös aina järjestää uudelleen niin, että se menee "komponenttimuotoon":



$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Vierusmatriisi yleisemmin

Vierusmatriisin voi muodostaa myös yleisemmille verkoille, kuten suunnatuille verkoille, joissa linkki saattaa mennä vain yhteen suuntaan:



$$A = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Googlen hakualgoritmi

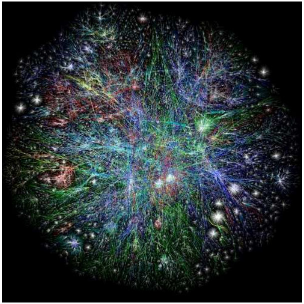
Mitä tapahtuu tarkalleen, kun käyttäjä laittaa syötteen hakukoneeseen?

Tutustumme seuraavaksi Googlen haussa käytettyyn algoritmiin nimeltä PageRank.

PageRank

Nimensä mukaisesti verkkosivut muodostavat yhden modernin ajan tärkeimmistä verkoista. Verkkosivut ovat tämän verkon solmuja ja niiden väliset linkit ovat, no, linkkejä.

Koko internetin käsittävä verkko on massiivinen suunnattu verkko, ja kahden solmun välillä voi mennä useampia linkkejä.



PageRankin tehtävä on selvittää, mitkä sivut ovat tärkeimpiä ja näytetään hakutuloksissa ensin.

PageRank

Verkkosivun "tärkeys" perustuu PageRank-algoritmissa kysymykseen:

Miten todennäköisesti nettiä satunnaisesti selaava käyttäjä päätyy tietylle sivulle?

Toimintaa mallinnetaan seuraavasti:

1. Käyttäjä aloittaa satunnaiselta sivulta.
2. Todennäköisyydellä d käyttäjä valitsee tältä sivulta satunnaisen linkin eteenpäin.
3. Todennäköisyydellä $1 - d$ käyttäjä kyllästyy ja hyppää satunnaiseen osoitteeseen.

Tyypillisesti $d \approx 0.85$.

PageRank

Tähän voidaan hyödyntää matriisilaskentaa:

1. Muodostetaan ensin internetin vierusmatriisi A.
2. Normalisoidaan A uudeksi matriisiksi M niin, että joka sarakkeen alkioiden summa on 1. Tällöin sarake j kuvaa todennäköisyyksiä hypätä sivulta j eteenpäin.
3. Jos vektori
$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_N \end{bmatrix}$$
kuvaa todennäköisyyksiä olla eri verkkosivuilla tietyllä hetkellä, niin seuraavalla klikkauksella uusi todennäköisyysvektori löytyy laskulla
$$\mathbf{r}_{\text{new}} = dM\mathbf{r} + \frac{1-d}{N} v_N,$$
missä v_N on $N \times 1$ vektori täynnä ykkösiä.

PageRank

Osoittautuu, että kaavan

$$\mathbf{r}_{\text{new}} = dM\mathbf{r} + \frac{1-d}{N} v_N$$

iterointi suppenee varsin järkevässä vauhdissa. Suunnilleen 52 iteraatiota riittää hyvään tarkkuuteen.

Lopullinen iteraatiosta löytynyt vektori kertoo komponenteissaan, kuinka todennäköisesti ollaan eri verkkosivuilla.

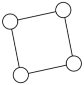
Hakua tehtäessä Googlen riittää tarkistaa, millä verkkosivuilla hakusanat esiintyvät, ja järjestää hakutulokset edellisen vektorin avulla järjestykseen.

Järjestykseen vaikuttaa myös muita asioita, kuten millä tavalla hakusanat esiintyvät sivulla, millainen sivu on kyseessä, yms.

Tehtävä lopuksi

Erään verkon vierusmatriisin toinen ja neljäs potenssi ovat

$$A^2 = \begin{bmatrix} 3 & 0 & 2 & 0 & 2 & 0 \\ 0 & 2 & 0 & 2 & 0 & 1 \\ 2 & 0 & 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 3 & 0 & 2 \\ 2 & 0 & 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 & 0 & 2 \end{bmatrix}, \quad A^4 = \begin{bmatrix} 17 & 0 & 12 & 0 & 12 & 0 \\ 0 & 9 & 0 & 12 & 0 & 8 \\ 12 & 0 & 9 & 0 & 8 & 0 \\ 0 & 12 & 0 & 17 & 0 & 12 \\ 12 & 0 & 8 & 0 & 9 & 0 \\ 0 & 8 & 0 & 12 & 0 & 9 \end{bmatrix}$$

montako neljän solmun sykliä  verkosta löytyy?