Googles sidrankning - linjär algebra värt en förmögenhet

Stefan Lemurell

Outline

1. Sökmotorer
2. Matematik
   - Grafteori
   - Linjär algebra
3. Googles sidrankning (PageRank)

Målet

Utifrån användarens sökord lista de mest relevanta webbsidorna. Dessutom i en ordning som motsvarar graden av relevans för användaren.

Förutsättningar

En första förutsättning är att man har en (alltid) aktuell databas med nödvändig information om “alla” sidor på internet.

Förutsättningar

En första förutsättning är att man har en (alltid) aktuell databas med nödvändig information om “alla” sidor på internet.
Intressantare (ur matematiksynpunkt) är den efterföljande frågeställningen:

Förutsättningar

En första förutsättning är att man har en (alltid) aktuell databas med nödvändig information om "alla" sidor på internet.
Intressantare (ur matematiksynpunkt) är den efterföljande frågeställningen:

Hur kan man vaska fram guldkornen ur denna "sörja" av miljarder webbsidor.

Nätet är stort

- Sökmotorer indexerar ungefär 50 miljarder sidor (2012).

Nätet är stort

- Sökmotorer indexerar ungefär 50 miljarder sidor (2012).
- I stort sett varje sökning ger tusentals eller miljontals träffar.

Nätet är stort

- Sökmotorer indexerar ungefär 50 miljarder sidor (2012).
- I stort sett varje sökning ger tusentals eller miljontals träffar.
- Hur ska man lyckas presentera de sidor som är mest intressanta?

Före 1998

Då byggde sökresultaten enbart på innehållet i varje sida.

- Sökte efter sökorden i titel, nyckelord, adress och i texten.

Före 1998

Då byggde sökresultaten enbart på innehållet i varje sida.

- Sökte efter sökorden i titel, nyckelord, adress och i texten.
- Rangordningen mellan sidor byggde bara på hur viktiga sökorden var för olika sidor. Inte hur "viktiga" sidorna själva var.

Före 1998

Då byggde sökresultaten enbart på innehållet i varje sida.

- Sökte efter sökorden i titel, nyckelord, adress och i texten.
- Rangordningen mellan sidor byggde bara på hur viktiga sökorden var för olika sidor. Inte hur "viktiga" sidorna själva var.
- Man fick vara klurig för att hitta en sökning som sållade ut de sidor man var intresserad av.

Före 1998

Då byggde sökresultaten enbart på innehållet i varje sida.

- Sökte efter sökorden i titel, nyckelord, adress och i texten.
- Rangordningen mellan sidor byggde bara på hur viktiga sökorden var för olika sidor. Inte hur “viktiga” sidorna själva var.
- Man fick vara klurig för att hitta en sökning som sållade ut de sidor man var intresserad av.
- Mot slutet började nätet bli så stort att sökmotorerna var i det närmaste värdelösa.

Revolutionen 1998

- Sökmotorn Google kom i beta-version 1998 och året efter släpptes en officiell version.

Revolutionen 1998

- Sökmotorn Google kom i beta-version 1998 och året efter släpptes en officiell version.
- Plötsligt fanns det en sökmotor där sidorna du var intresserad av låg högt upp på första sidan och inte (i bästa fall) någonstans på sidan 8.

Revolutionen 1998

- Sökmotorn Google kom i beta-version 1998 och året efter släpptes en officiell version.
- Plötsligt fanns det en sökmotor där sidorna du var intresserad av låg högt upp på första sidan och inte (i bästa fall) någonstans på sidan 8.
- Kändes som ett mirakel!

Inget mirakel ...

... utan smart användning av ca 100 år gammal matematik.

Idén i korthet

- Utnyttja den struktur som länkarna mellan sidorna ger.

Idén i korthet

- Utnyttja den struktur som länkarna mellan sidorna ger.
- Modellera detta med en riktad graf.

Idén i korthet

- Utnyttja den struktur som länkarna mellan sidorna ger.
- Modellera detta med en riktad graf.
- Använda en matris M som svarar mot slumpvandring på denna graf.

Idén i korthet

- Utnyttja den struktur som länkarna mellan sidorna ger.
- Modellera detta med en riktad graf.
- Använda en matris M som svarar mot slumpvandring på denna graf.
- Perron-Frobenius sats: M^t har egenvärdet 1 och dess egenvektor $\mathbf{v}$ har bara positiva element.

Idén i korthet

- Utnyttja den struktur som länkarna mellan sidorna ger.
- Modellera detta med en riktad graf.
- Använda en matris M som svarar mot slumpvandring på denna graf.
- Perron-Frobenius sats: M^t har egenvärdet 1 och dess egenvektor $\mathbf{v}$ har bara positiva element.
- Vektorn $\mathbf{v}$ ger en rankning av alla sidor.

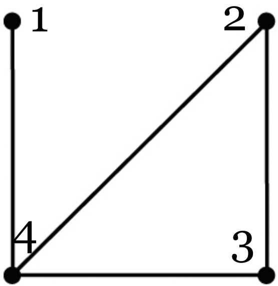
Outline

Outline

1. Sökmotorer
2. Matematik
   - Grafteori
   - Linjär algebra
3. Googles sidrankning (PageRank)

Graf

En graf består av en mängd V med *noder* och en mängd E med *kanter*. Bildligt kan man tänka sig noderna som punkter och en kant som en kurva mellan två noder.



Graf som modell

Grafer kan användas som modell för många skiftande fenomen såsom

Graf som modell

Grafer kan användas som modell för många skiftande fenomen såsom

- Nätverk

Graf som modell

Grafer kan användas som modell för många skiftande fenomen såsom

- Nätverk
- Transportförbindelser

Graf som modell

Grafer kan användas som modell för många skiftande fenomen såsom

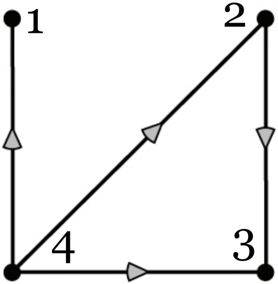
- Nätverk
- Transportförbindelser
- Datastrukturer

Exempel på sociala nätverk

Låt V vara mängden av alla personer som publicerat minst en artikel i en matematisk tidskrift. Det finns en kant mellan två personer i V om de publicerat en artikel ihop.
Eller, låt V vara alla personer som har ett konto på LinkedIn och det finns en kant mellan två personer om de har "anknyntning".

Riktad graf

I en riktad graf är kanterna "enkelriktade".



Riktad graf som modell

Riktade grafer kan användas som modell för

Riktad graf som modell

Riktade grafer kan användas som modell för

- Vägnät. Noderna är “korsningar” och kanter är vägstumpar (som ju kan vara enkelriktade) mellan korsningar.

Riktad graf som modell

Riktade grafer kan användas som modell för

- Vägnät. Noderna är "korsningar" och kanter är vägstumpar (som ju kan vara enkelriktade) mellan korsningar.
- Internet. Noderna består av sidor och kanterna är länkar mellan sidor.

Vägar och avstånd

En *väg* i en graf är en följd av kanter (som ligger “efter varandra”) och avståndet mellan två noder är längden på den kortaste vägen.

Erdös-talet

Paul Erdös är den mest produktive (räknat i antal publicerade artiklar) matematikern genom tiderna och han skrev minst en artikel tillsammans med totalt 511 andra matematiker. Det finns alltså 511 kanter från honom till andra noder i grafen där V är mängden av alla personer som publicerat minst en artikel i en matematisk tidskrift och en kant mellan två personer i V om de publicerat en artikel ihop.

Erdös-talet

Paul Erdös är den mest produktive (räknat i antal publicerade artiklar) matematikern genom tiderna och han skrev minst en artikel tillsammans med totalt 511 andra matematiker. Det finns alltså 511 kanter från honom till andra noder i grafen där V är mängden av alla personer som publicerat minst en artikel i en matematisk tidskrift och en kant mellan två personer i V om de publicerat en artikel ihop.
En persons *Erdös-tal* är avståndet från personen till Erdös i denna grafen.

Vägar på internet

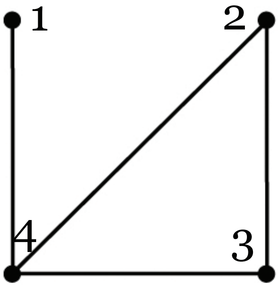
En väg från en sida A till en annan sida B i den riktade grafen som representerar internet är ett sätt att klicka sig från A till B.
Avståndet mellan A och B är minsta antalet klick från A till B.

Sammanhängande

Man säger att en graf är *sammanhängande* om det finns väg mellan varje par av noder.

Slumpvandring på graf

En slumpvandring på en graf är, precis som det låter, att man tar sig från nod till nod via kanter som väljs slumpmässigt.



Outline

Egenvärden och egenvektorer

Om M är en *kvadratisk* matris så säger man att $\mathbf{v}$ är en *egenvektor* till M med *egenvärdet* λ om $M\mathbf{v} = \lambda\mathbf{v}$.

Egenvärden och egenvektorer

Om M är en *kvadratisk* matris så säger man att $\mathbf{v}$ är en *egenvektor* till M med *egenvärdet* λ om $M\mathbf{v} = \lambda\mathbf{v}$.
Med andra ord så ändrar inte M riktningen på $\mathbf{v}$ (möjligen motsatt riktning om λ är negativt).

Egenvärden och egenvektorer

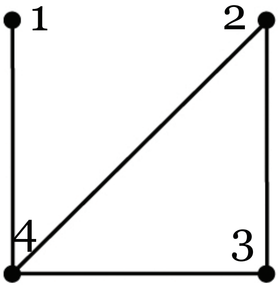
Om M är en *kvadratisk* matris så säger man att $\mathbf{v}$ är en *egenvektor* till M med *egenvärdet* λ om $M\mathbf{v} = \lambda\mathbf{v}$.
Med andra ord så ändrar inte M riktningen på $\mathbf{v}$ (möjligen motsatt riktning om λ är negativt).
Exempelvis har en identitetsmatris I alla vektorer som egenvektor med egenvärdet 1 och en matris som svarar mot rotation kring origo i planet saknar (reell) egenvektor.

Grannmatris

För en graf med n noder är *grannmatrisen* den $n \times n$-matris A vars element definieras genom

$$a_{ij} = \begin{cases} 1 & \text{om kant mellan nod i och nod j (från nod i till nod j),} \\ 0 & \text{annars.} \end{cases}$$

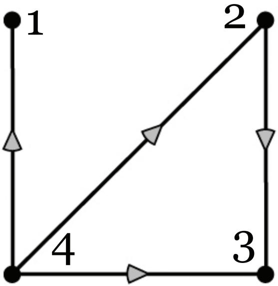
Exempel på grannmatris



har grannmatrisen

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

Exempel på grannmatris



har grannmatrisen

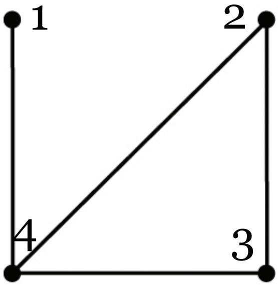
$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Övergångsmatris

För slumpvandringen på en graf med n noder är *övergångsmatrisen* den $n \times n$-matris A vars element definieras genom

$$a_{ij} = \text{sannolikheten att man går till nod j om man är i nod i.}$$

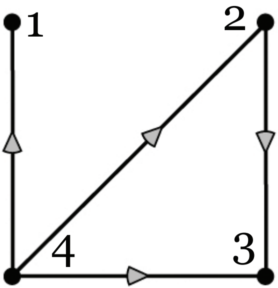
Exempel på övergångsmatris



har övergångsmatrisen

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1/2 & 0 & 1/2 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

Exempel på övergångsmatris



har övergångsmatrisen

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}$$

Perron-Frobenius sats

Sats som bevisades i 2 steg av Oskar Perron (1907) och Ferdinand Georg Frobenius (1912).
I detta sammanhanget säger satsen:
Om A är övergångsmatrisen för en slumpvandring på en sammanhängande graf, så gäller för transponatet A^t att denna matris har egenvärdet 1 med en unik egenvektor $\mathbf{v}$. Dessutom gäller att alla andra egenvärden till A^t har absolutbeloppet mindre än 1 och alla element i $\mathbf{v}$ är positiva.

Outline

1. Sökmotorer
2. Matematik
   - Grafteori
   - Linjär algebra
3. Googles sidrankning (PageRank)

Ranka sidor efter viktighetsgrad

- Man brukar länka till sidor som är viktiga. Ju fler som länkar till en sida, desto viktigare är den.

Ranka sidor efter viktighetsgrad

- Man brukar länka till sidor som är viktiga. Ju fler som länkar till en sida, desto viktigare är den.
- Länkar från en sida som bara länkar till några få utvalda är viktigare än länkar från en sida som länkar till många sidor.

Ranka sidor efter viktighetsgrad

- Man brukar länka till sidor som är viktiga. Ju fler som länkar till en sida, desto viktigare är den.
- Länkar från en sida som bara länkar till några få utvalda är viktigare än länkar från en sida som länkar till många sidor.
- Viktiga sidor länkar till viktiga sidor. (Det är bättre att ha länkar från viktiga sidor än från mindre viktiga sidor.)

En formel för vikten, version 1

Låt $v(P_i)$ vara vikten för sidan P_i. Vi vill lägga ihop antalet länkar till den sida P_i vi ska räkna ut vikt(ighet)en för. Om B_i är mängden av sidor som länkar till P_i och $\#(P_j \to P_i)$ antalet länkar från sidan P_j till sidan P_i får vi

$$v(P_i) = \sum_{P_j \in B_i} \#(P_j \to P_i).$$

På detta sätt blir det väldigt lätt att skapa sidor med hög vikt. Dessa vikter svarar mot att välja godtycklig länk på nätet och följa den.

En formel för vikten, version 2

Antag att samtliga sidor har vikt 1 att fördela till de sidor de länkar till. Då kommer varje länk från sidan P_j med ℓ_j länkar bara ge vikten $1/\ell_j$ till den sida länken går till. Vi får

$$v(P_i) = \sum_{P_j \in B_i} \frac{\#(P_j \to P_i)}{\ell_j}.$$

Något svårare att skapa sidor med hög vikt, men med hjälp av många sidor går det. Svarar mot att välja godtycklig sida på nätet och följa slumpmässig kant från den sidan.

En formel för vikten, version 3

Vi låter nu varje sida få sin egen vikt att fördela till de andra sidorna. Om P_j har ℓ_j länkar, så ger varje länk vikten $v(P_j)/\ell_j$ till den sida den leder till. Om B_i är mängden av sidor som länkar till P_i får vi då

$$v(P_i) = \sum_{P_j \in B_i} \frac{v(P_j)\#(P_j \to P_i)}{\ell_j}.$$

Svarar mot att välja en godtycklig sida och sedan upprepade gånger följa en slumpmässig kant. En *slumpvandring* på internet!
Nu går det inte längre att själv ge hög vikt åt egna sidor. Men vi får ett annat problem.

En formel för vikten, version 3

Vi låter nu varje sida få sin egen vikt att fördela till de andra sidorna. Om P_j har ℓ_j länkar, så ger varje länk vikten $v(P_j)/\ell_j$ till den sida den leder till. Om B_i är mängden av sidor som länkar till P_i får vi då

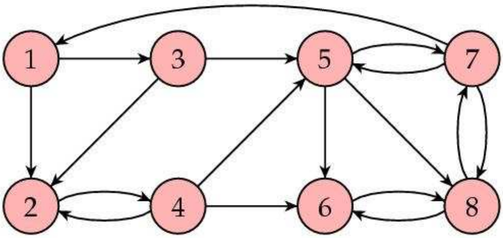
$$v(P_i) = \sum_{P_j \in B_i} \frac{v(P_j)\#(P_j \to P_i)}{\ell_j}.$$

Svarar mot att välja en godtycklig sida och sedan upprepade gånger följa en slumpmässig kant. En *slumpvandring* på internet!

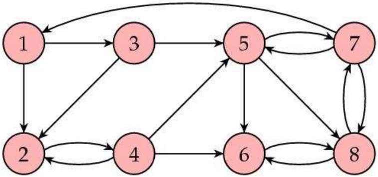
Nu går det inte längre att själv ge hög vikt åt egna sidor. Men vi får ett annat problem.

Vi måste räkna ut alla vikter samtidigt!

Nätet som en graf



. . . och som (transponatet av) en övergångsmatris

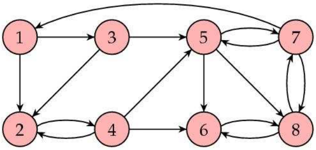


$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

Viktformel som matrismultiplikation

Viktformeln $v(P_i) = \sum_{P_j \in B_i} \frac{v(P_j)}{\ell_j}$ anger, till exempel för P_2, att

$$v(P_2) = \frac{v(P_1)}{2} + \frac{v(P_3)}{2} + \frac{v(P_4)}{3}.$$



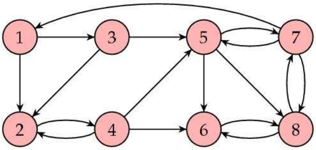
$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

Viktformel som matrismultiplikation

Viktformeln $v(P_i) = \sum_{P_j \in B_i} \frac{v(P_j)}{\ell_j}$ anger, till exempel för P_2, att

$$v(P_2) = \frac{v(P_1)}{2} + \frac{v(P_3)}{2} + \frac{v(P_4)}{3}.$$

Det är precis så som matrismultiplikation fungerar. Om vektorn v består av vikterna $(v(P_1), v(P_2), \ldots, v(P_8))$, så får vi som resultat $H\mathbf{v} =$



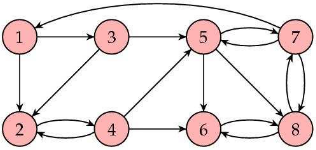
$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

Viktformel som matrismultiplikation

Viktformeln $v(P_i) = \sum_{P_j \in B_i} \frac{v(P_j)}{\ell_j}$ anger, till exempel för P_2, att

$$v(P_2) = \frac{v(P_1)}{2} + \frac{v(P_3)}{2} + \frac{v(P_4)}{3}.$$

Det är precis så som matrismultiplikation fungerar. Om vektorn v består av vikterna $(v(P_1), v(P_2), \ldots, v(P_8))$, så får vi som resultat $H\mathbf{v} = \mathbf{v}$.



$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

Att räkna ut alla vikterna samtidigt

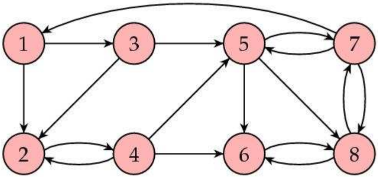
Det är vektorn $\mathbf{v}$ som innehåller vikterna. Vi vill hitta $\mathbf{v}$ så att $H\mathbf{v} = \mathbf{v}$. Med andra ord ska $\mathbf{v}$ vara en egenvektor med egenvärdet 1.

Att räkna ut alla vikterna samtidigt

Det är vektorn $\mathbf{v}$ som innehåller vikterna. Vi vill hitta $\mathbf{v}$ så att $H\mathbf{v} = \mathbf{v}$. Med andra ord ska $\mathbf{v}$ vara en egenvektor med egenvärdet 1. Det är lätt att beräkna egenvektorer till mycket små matriser: Man definierar ett polynom av samma grad som matrisens storlek, med koefficienter från matrisen, och hittar sedan nollställen till polynomet. För Google blir det ett polynom av grad 50 miljarder. Huga!

En speciell matris

Matrisen H är speciell på flera sätt.

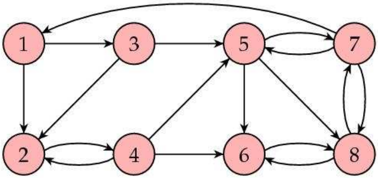


$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

En speciell matris

Matrisen H är speciell på flera sätt.

- Den har inga negativa värden.

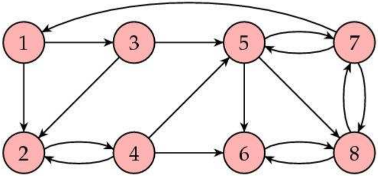


$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

En speciell matris

Matrisen H är speciell på flera sätt.

- Den har inga negativa värden.
- Kolumnerna summerar till 1.



$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

Perron-Frobenius sats

För en matris H med icke-negativa element vars kolumner summerar till 1 gäller att

Perron-Frobenius sats

För en matris H med icke-negativa element vars kolumner summerar till 1 gäller att

- den har egenvärdet 1 med unik egenvektor;

Perron-Frobenius sats

För en matris H med icke-negativa element vars kolumner summerar till 1 gäller att

- den har egenvärdet 1 med unik egenvektor;
- alla andra egenvärden λ uppfyller $|\lambda| < 1$;

Perron-Frobenius sats

För en matris H med icke-negativa element vars kolumner summerar till 1 gäller att

- den har egenvärdet 1 med unik egenvektor;
- alla andra egenvärden λ uppfyller $|\lambda| < 1$;
- egenvektorn till egenvärdet 1 har bara positiva element.

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$H\mathbf{v} = H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n)$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned} H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\ &= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \end{aligned}$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned} H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\ &= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \\ &= a_1(H\mathbf{v}_1) + a_2(H\mathbf{v}_2) + a_3(H\mathbf{v}_3) + \ldots + a_n(H\mathbf{v}_n) \end{aligned}$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned} H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\ &= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \\ &= a_1(H\mathbf{v}_1) + a_2(H\mathbf{v}_2) + a_3(H\mathbf{v}_3) + \ldots + a_n(H\mathbf{v}_n) \\ &= a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + a_3\lambda_3\mathbf{v}_3 + \ldots + a_n\lambda_n\mathbf{v}_n \end{aligned}$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned}
H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\
&= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \\
&= a_1(H\mathbf{v}_1) + a_2(H\mathbf{v}_2) + a_3(H\mathbf{v}_3) + \ldots + a_n(H\mathbf{v}_n) \\
&= a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + a_3\lambda_3\mathbf{v}_3 + \ldots + a_n\lambda_n\mathbf{v}_n \\
H^2\mathbf{v} &= a_1\lambda_1^2\mathbf{v}_1 + a_2\lambda_2^2\mathbf{v}_2 + a_3\lambda_3^2\mathbf{v}_3 + \ldots + a_n\lambda_n^2\mathbf{v}_n
\end{aligned}$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned}
H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\
&= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \\
&= a_1(H\mathbf{v}_1) + a_2(H\mathbf{v}_2) + a_3(H\mathbf{v}_3) + \ldots + a_n(H\mathbf{v}_n) \\
&= a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + a_3\lambda_3\mathbf{v}_3 + \ldots + a_n\lambda_n\mathbf{v}_n \\
H^2\mathbf{v} &= a_1\lambda_1^2\mathbf{v}_1 + a_2\lambda_2^2\mathbf{v}_2 + a_3\lambda_3^2\mathbf{v}_3 + \ldots + a_n\lambda_n^2\mathbf{v}_n \\
H^3\mathbf{v} &= a_1\lambda_1^3\mathbf{v}_1 + a_2\lambda_2^3\mathbf{v}_2 + a_3\lambda_3^3\mathbf{v}_3 + \ldots + a_n\lambda_n^3\mathbf{v}_n
\end{aligned}$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned}
H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\
&= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \\
&= a_1(H\mathbf{v}_1) + a_2(H\mathbf{v}_2) + a_3(H\mathbf{v}_3) + \ldots + a_n(H\mathbf{v}_n) \\
&= a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + a_3\lambda_3\mathbf{v}_3 + \ldots + a_n\lambda_n\mathbf{v}_n \\
H^2\mathbf{v} &= a_1\lambda_1^2\mathbf{v}_1 + a_2\lambda_2^2\mathbf{v}_2 + a_3\lambda_3^2\mathbf{v}_3 + \ldots + a_n\lambda_n^2\mathbf{v}_n \\
H^3\mathbf{v} &= a_1\lambda_1^3\mathbf{v}_1 + a_2\lambda_2^3\mathbf{v}_2 + a_3\lambda_3^3\mathbf{v}_3 + \ldots + a_n\lambda_n^3\mathbf{v}_n \\
H^4\mathbf{v} &= a_1\lambda_1^4\mathbf{v}_1 + a_2\lambda_2^4\mathbf{v}_2 + a_3\lambda_3^4\mathbf{v}_3 + \ldots + a_n\lambda_n^4\mathbf{v}_n.
\end{aligned}$$

Iterera fram viktvektorn

Antag att varje vektor kan skrivas som en summa av egenvektorer:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n.$$

Räknereglerna $H(\mathbf{v}_1 + \mathbf{v}_2) = H\mathbf{v}_1 + H\mathbf{v}_2$ och $H(a\mathbf{v}) = a(H\mathbf{v})$ ger då

$$\begin{aligned}
H\mathbf{v} &= H(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \ldots + a_n\mathbf{v}_n) \\
&= H(a_1\mathbf{v}_1) + H(a_2\mathbf{v}_2) + H(a_3\mathbf{v}_3) + \ldots + H(a_n\mathbf{v}_n) \\
&= a_1(H\mathbf{v}_1) + a_2(H\mathbf{v}_2) + a_3(H\mathbf{v}_3) + \ldots + a_n(H\mathbf{v}_n) \\
&= a_1\lambda_1\mathbf{v}_1 + a_2\lambda_2\mathbf{v}_2 + a_3\lambda_3\mathbf{v}_3 + \ldots + a_n\lambda_n\mathbf{v}_n \\
H^2\mathbf{v} &= a_1\lambda_1^2\mathbf{v}_1 + a_2\lambda_2^2\mathbf{v}_2 + a_3\lambda_3^2\mathbf{v}_3 + \ldots + a_n\lambda_n^2\mathbf{v}_n \\
H^3\mathbf{v} &= a_1\lambda_1^3\mathbf{v}_1 + a_2\lambda_2^3\mathbf{v}_2 + a_3\lambda_3^3\mathbf{v}_3 + \ldots + a_n\lambda_n^3\mathbf{v}_n \\
H^4\mathbf{v} &= a_1\lambda_1^4\mathbf{v}_1 + a_2\lambda_2^4\mathbf{v}_2 + a_3\lambda_3^4\mathbf{v}_3 + \ldots + a_n\lambda_n^4\mathbf{v}_n.
\end{aligned}$$

Om $\lambda_1 = 1$ och $|\lambda_i| < 1$ för $i \geq 2$ så går $H^j\mathbf{v}$ mot $a_1\mathbf{v}_1$.

Ger egenvektorn en ranking

Eftersom egenvektorn bara innehåller positiva element blir det en ranking.

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/2 & 0 & 1/2 & 1/3 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1/3 & 1 & 1/3 & 0 \end{bmatrix}$$

$$I = \begin{bmatrix} 0.0600 \\ 0.0675 \\ 0.0300 \\ 0.0675 \\ 0.0975 \\ 0.2025 \\ 0.1800 \\ 0.2950 \end{bmatrix}$$

En slutgiltig formel för vikten

I själva verket så inför man en dämpningsfaktor för att ta hand om det faktum att nätet inte är sammanhängande och att vissa sidor saknar utgående länkar. Dämpningsfaktorn kan man tänka på som sannolikheten att en person väljer en länk på sidan och inte väljer en sida på måfå. Om vi betecknar dämpningsfaktorn med d och totala antalet sidor men N så blir den slutgiltiga formeln

$$v(P_i) = \frac{1-d}{N} + d \sum_{P_j \in B_i} \frac{v(P_j)\#(P_j \to P_i)}{\ell_j}.$$

Beräkningen med hjälp av denna formel blir i stort sett samma som utan dämpningsfaktor.