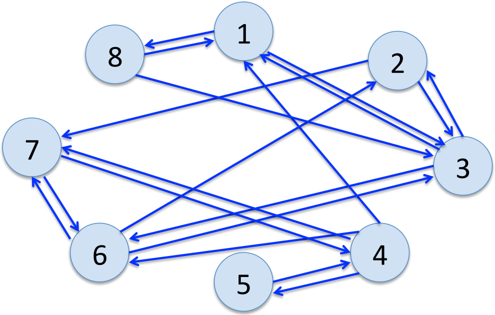
# Rangordning av internetsidor - ett egenvärdesproblem för positiva matriser

Ett litet nätverk med 8 noder och ett antal länkar mellan noderna:

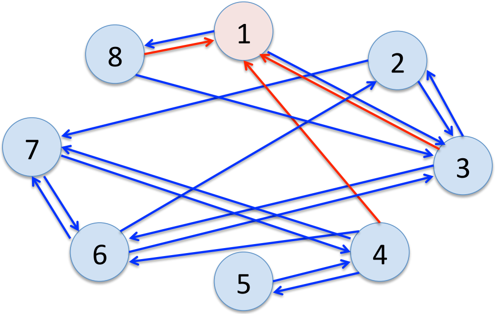



Hur kan vi rangordna noder? Vilken är "viktigast"?

Googles modell: En nods värde beror på värdena på alla noder som länkar till den. Värdet som en nod bidrar med till andra noder genom utgående länkar delas lika mellan dessa.
Låt $v_1, \dots, v_8$ vara rankingvärdena för nod 1 till 8. Noderna 3, 4 och 8 länkar till nod 1. Noderna 2, 4 och 8 har 3, 4 resp. 2 utgående länkar.

Då är modellen $v_1 = k(\frac{1}{3}v_3 + \frac{1}{4}v_4 + \frac{1}{2}v_8)$ ,

där $k$ är en proportionalitetskonstant.



Om vi skriver upp ekvationerna för alla 8 noderna får vi

$$
\begin{cases}
v_1 = & k\left(\frac{1}{3}v_3 + \frac{1}{4}v_4 + \frac{1}{2}v_8\right) \\
v_2 = & k\left(\frac{1}{3}v_3 + \frac{1}{3}v_6\right) \\
v_3 = & k\left(\frac{1}{2}v_1 + \frac{1}{2}v_2 + \frac{1}{3}v_6 + \frac{1}{2}v_8\right) \\
v_4 = & k\left(v_5 + \frac{1}{2}v_7\right) \\
v_5 = & k \cdot \frac{1}{4}v_4 \\
v_6 = & k\left(\frac{1}{3}v_3 + \frac{1}{4}v_4 + \frac{1}{2}v_7\right) \\
v_7 = & k\left(\frac{1}{2}v_2 + \frac{1}{4}v_4 + \frac{1}{3}v_6\right) \\
v_8 = & k \cdot \frac{1}{2}v_1
\end{cases}
$$

Ett system med 8 ekvationer och 9 obekanta (även $k$ !). Sätt $k = 1/\lambda$ och skriv på matrisform.

$$\begin{pmatrix} 0 & 0 & 1/3 & 1/4 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 1/3 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 1/4 & 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/4 & 0 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{pmatrix}$$

Vi har fått ett egenvärdesproblem!

Obs att kolonnsumman alltid är 1 (utdelade värden från en nod).

Kalla matrisen $A$ och vektorn $\bar{v}$, då har vi $A\bar{v} = \lambda\bar{v}$.

Matrisen $A$ kallas *icke-negativ* då alla element är $\geq 0$.

En *positiv* matris har alla element $> 0$ (inte samma som *positivt definit* matris).

Positiva matriser är både analytiskt och numeriskt något bättre än icke-negativa att handskas med. Därför modifierar Google matrisen A lite genom att byta den mot $C = 0.85 \cdot A + 0.15 \cdot B$ där

$$B = \begin{pmatrix} 1/8 & \dots & 1/8 \\ \vdots & & \vdots \\ 1/8 & \dots & 1/8 \end{pmatrix}$$

Detta kan också ses som att man tar viss hänsyn till slumpsurfaren som inte bara klickar på länkar. Vi får för vårt nätverk $C\bar{v} = \lambda\bar{v}$ med

$$C \simeq \begin{pmatrix} 0.019 & 0.019 & 0.302 & 0.231 & 0.019 & 0.019 & 0.019 & 0.444 \\ 0.019 & 0.019 & 0.302 & 0.019 & 0.019 & 0.302 & 0.019 & 0.019 \\ 0.444 & 0.444 & 0.019 & 0.019 & 0.019 & 0.302 & 0.019 & 0.444 \\ 0.019 & 0.019 & 0.019 & 0.019 & 0.869 & 0.019 & 0.444 & 0.019 \\ 0.019 & 0.019 & 0.019 & 0.231 & 0.019 & 0.019 & 0.019 & 0.019 \\ 0.019 & 0.019 & 0.302 & 0.231 & 0.019 & 0.019 & 0.444 & 0.019 \\ 0.019 & 0.444 & 0.019 & 0.231 & 0.019 & 0.302 & 0.019 & 0.019 \\ 0.444 & 0.019 & 0.019 & 0.019 & 0.019 & 0.019 & 0.019 & 0.019 \end{pmatrix}$$

Oba att kolonnsummorna i $C$ fortfarande alla är 1.

Lös $\det(C - \lambda I) = 0$ (åttondegradsekvation) numeriskt. Ger 4 reella egenvärden $\lambda_1 = 1$, $\lambda_2 \approx 0.444$, $\lambda_3 \approx 0.425$, $\lambda_4 \approx -0.425$ och 4 komplexa $\lambda_{5,6} \approx -0.489 \pm 0.089i$, $\lambda_{7,8} \approx -0.158 \pm 0.056i$.
Egenvektorerna till de reella är

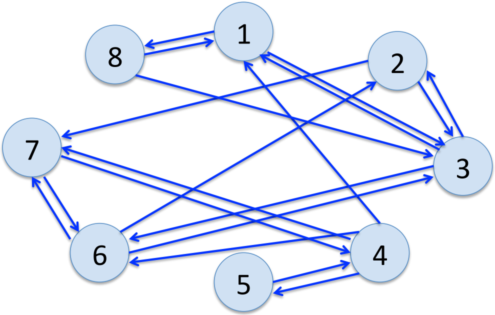
$$\bar{u}_1 \approx \begin{pmatrix} 0.35 \\ 0.33 \\ 0.55 \\ 0.31 \\ 0.11 \\ 0.43 \\ 0.37 \\ 0.20 \end{pmatrix}, \quad \bar{u}_2 \approx \begin{pmatrix} 0.14 \\ 0.31 \\ 0.53 \\ -0.69 \\ -0.33 \\ -0.04 \\ -0.06 \\ 0.13 \end{pmatrix}, \quad \bar{u}_3 \approx \begin{pmatrix} 0.09 \\ 0.35 \\ 0.52 \\ -0.69 \\ -0.35 \\ 0.00 \\ 0.00 \\ 0.09 \end{pmatrix}, \quad \bar{u}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

Vilket $\lambda_j$ och $\bar{u}_j$ ska vi ta? Alla noders värde bör vara $\geq 0$. Då måste $\bar{v} = \bar{u}_1$ och nodernas värden är $v_1 \approx 0.35$, $v_1 \approx 0.33$, $v_1 \approx 0.55$, $v_1 \approx 0.31$, $v_1 \approx 0.11$, $v_1 \approx 0.43$ $v_1 \approx 0.37$ och $v_1 \approx 0.20$. Nod 3 har högst värde och rangordningen ges av

$v_3 > v_6 > v_7 > v_1 > v_2 > v_4 > v_8 > v_5$

Obs också att $\lambda_1 > |\lambda_j|$ för $j = 2, \ldots, 8$.

Rangordning: $v_3 > v_6 > v_7 > v_1 > v_2 > v_4 > v_8 > v_5$




Verkar rimligt?

Fungerar denna metod alltid? Svar: Ja!

Perron-Frobenius sats för positiva $N \times N$ - matriser säger

SATS: Om alla element i $C$ är $> 0$ har $C$ ett positivt egenvärde $\lambda_1$ med $\lambda_1 > |\lambda_j|$, $j = 2, \ldots, N$, alla element i egenvektorn till $\lambda_1$ är $> 0$ och ingen annan egenvektor med denna egenskap finns.

Detta passar perfekt i modellen ovan, vi kommer alltid garanterat få precis en egenvektor med enbart positiva komponenter.

Obs: $C^t \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ (ty radsummorna i $C^t$ är 1) $\Rightarrow \lambda_1 = 1$ för $C^t$ ty egenvektorn $(1, \ldots, 1)^t$ har enbart positiva element. Då $C$ och $C^t$ har samma egenvärden är $\lambda_1 = 1$ även för $C$.

För hela internet bildar vi $N \times N$ matrisen ($N \sim 10^{10}, 10^{11}$, antal sidor på internet) $C = 0.85 \cdot \text{länkmatrisen} + 0.15 \cdot \begin{pmatrix} 1/N & \dots & 1/N \\ \vdots & & \vdots \\ 1/N & \dots & 1/N \end{pmatrix}$.

Vi vet att $\lambda_1 = 1$ är till beloppet största egenvärde och motsvarande egenvektor $\bar{v} \in \mathbb{R}^N$ har enbart positiva komponenter. Dessa komponenter är alla sidors rankingvärden. Vid en sökning rangordnas alla sidor som innehåller det sökta ordet efter dessa rankingvärden.

Egenvektorn till $\lambda_1 = 1$ fås ur $C\bar{v} = \bar{v} \Leftrightarrow (C - I)\bar{v} = \bar{0}$ som är det system som måste lösas. Man behöver alltså inte leta egenvärden till $C$ och enbart egenvektorn till $\lambda_1 = 1$ behövs. Systemet är den största matematiska beräkningen som utförs regelbundet. I samband med linjära ekvationssystem i början av kursen ställde vi upp systemet $(C - I)\bar{v} = \bar{0}$ utan att som nu kunna förklara hur det uppkommit.

Mer om positiva matriser kan man lära sig i linjär algebra överkurs.