

MA1487 Matematisk modellering

Projekt

Henrik Fredriksson
henrik.fredriksson@bth.se

18 december 2018

1 Upplägg

Syftet med projektet är att implementera Googles PageRank algoritm på ett fiktivt webbnätverk.

Inlämningen sker via Canvas senast 17/1 i form av länk till webbsida/webbsidor eller fristående applikationer.

2 Projektspecifikation

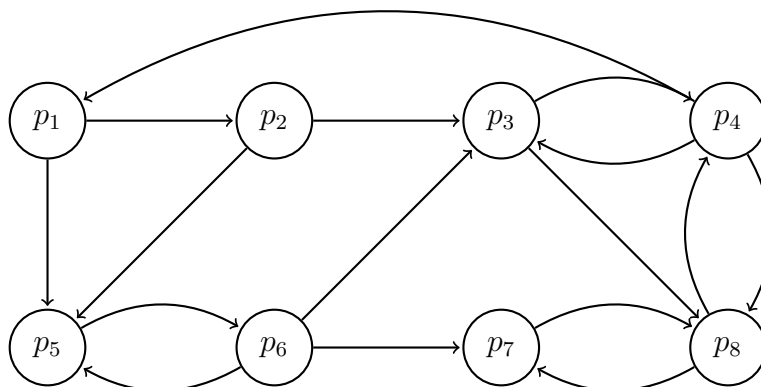
Projektet består av två uppgifter och får lösas med valfritt programmeringsspråk, samt en redovisningstext. Skriv ett allmänt stycke om hur projektet gick att genomföra. Redovisningstexten lämnas in via Canvas i samband med inlämning av resten av projektet.

Externa bibliotek får endast användas för att rita ut nätverket eller genomföra beräkningar matriser och vektorer, ej beräkna PageRank direkt.

3 Uppgifter

3.1 Uppgift 1

I Figur visas länkstrukturen hos 8 webbsidor



- (1) Bestäm hyperlink-matrisen H .
- (2) Bestäm PageRank för samtliga sidor.

3.2 Uppgift 2

Skapa ett valfritt nätverk med webbsidor med minst 30 webbsidor och minst 50 länkar. Inga sidor får länka till sig själv eller länka mer än en gång till samma sida. Varje sida har minst ett nyckelord kopplat till sig som innebär att när man som användare anger en del eller hela nyckelordet i söksträngen så ska sidan dyka upp i sökresultaten.

Nätverket ska återges antingen grafiskt eller med en sk *grannmatris*. En grannmatris A har en matriselement a_{ij} där

$$a_{ij} = \begin{cases} 1 & \text{om sida } p_i \text{ länkar till sida } p_j, \\ 0 & \text{annars.} \end{cases}$$

Tips är att använda grannmatrisen för att beskriva länkstrukturen hos sidorna.

- Användaren ska kunna mata in en söksträng
- Sökresultaten ska sorteras med fallande PageRank.

- PageRanken ska beräknas för “dangling nodes” och för nätverk som inte är sammanhängande.

Som stoppkriterium för potensberäkningarna av Hyperlink-matrisen så kan man göra på följande sätt.

1. Välj en tolerans $\varepsilon > 0$.
2. Sätt $k = 1$
3. Beräkna $r(k) = M^k r$ där M är den modifierade hyperlink-matrisen och r är PageRank-vektorn.
4. Avbryt om $|r_i(k) - r_i(k-1)| < \varepsilon$ för alla vektorelement i , annars sätt $k = k + 1$ och gå till Steg 3.

4 Extrauppgift

Slumpa fram ett nätverk med användarangivet antal webbsidor och länkar. Inga sidor får länka till sig själv eller länka mer än en gång till samma sida.