MATH 308

Archer Yang, McGill University

April 2, 2024

Homework 4

- This homework is due on **April 12 at 11:11pm**.
- You can work in groups of **up to 4 people**.
- Datasets can be downloaded from MyCourses.
- Late submission policy:
  - ≤ 1 hour, no penalty
  - > 1 hour and ≤ 2 days, 30% penalty
  - > 2 days, not accepted
- Please submit your solutions as pdf files via Crowdmark. Make sure to upload a separate pdf file containing your entire code with proper comments about the purpose of each code chunk.

1 PageRank

Ex. BrokenRank as a random walk

Given webpages numbered $1, \cdots, N$. The PageRank of webpage j is based on its linking webpages (webpages i that link to j).

Let

$$w_{ij} = \begin{cases} 1 & \text{if webpage } i \text{ links to webpage } j \text{ (written } i \to j) \\ 0 & \text{otherwise} \end{cases}$$

Also let

$$d_i = \sum_{k=1}^{N} w_{ik}$$

be **the total number of webpages** that i links to.

Now consider a specific random walk, with the vertices as webpages, and with transition matrix

$$[\boldsymbol{P}]_{ij} = p_{ij} = \frac{w_{ij}}{d_i}$$

The random walk is specified as

$$p_{ij} = P(\text{go from } i \text{ to } j) = \begin{cases} \frac{1}{d_i} & \text{if } i \text{ links to } j \\ 0 & \text{otherwise} \end{cases}$$

Show that $\sum_{j=1}^{N} p_{ij} = 1$.

Ex. PageRank as a random walk

The random walk for the PageRank case can be described as that at each time step, the random surfer has two options:

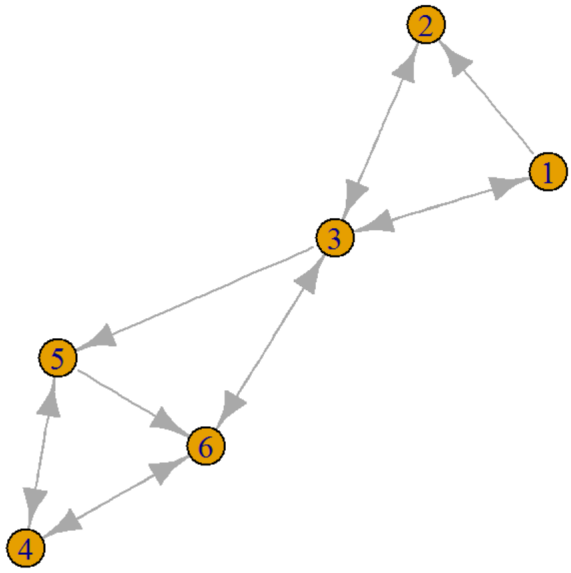
- Option 1: with prob. $\frac{1-\eta}{N} + \frac{\eta}{d_i}$, follow a link at random
- Option 2: with prob. $\frac{1-\eta}{N}$, jump to an unlinked random page

$$p_{ij} = P(\text{go from } i \text{ to } j) = \begin{cases} (1-\eta)\frac{1}{N} + \eta\frac{1}{d_i} & \text{if } i \text{ links to } j \\ (1-\eta)\frac{1}{N} & \text{otherwise} \end{cases}$$

Show that $\sum_{j=1}^{N} p_{ij} = 1$.

Ex. PageRank computation

Given the graph G below



Each vertex represent a website. Let

$$w_{ij} = \begin{cases} 1 & \text{if webpage } i \text{ links to webpage } j \text{ (written } i \to j\text{)} \\ 0 & \text{otherwise} \end{cases}$$

- Compute the weight matrix

$$\boldsymbol{W} = [w_{ij}]_{i,j \in V}$$

- Compute the degree matrix

$$\boldsymbol{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & d_N \end{pmatrix}$$

  where $d_i = \sum_{k=1}^{N} w_{ik}$

- Set $\eta = 0.85$, compute

$$\boldsymbol{P} = \frac{1-\eta}{N}\boldsymbol{E} + \eta \boldsymbol{D}^{-1}\boldsymbol{W}$$

- In R or Python, using eigendecomposition, compute PageRank score vector $\boldsymbol{\pi}$

$$\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{P}$$

- In R or Python, using update formula

$$\boldsymbol{\pi}^{(t)} = \boldsymbol{\pi}^{(t-1)} \boldsymbol{P}$$

  to compute the PageRank score vector $\boldsymbol{\pi}$