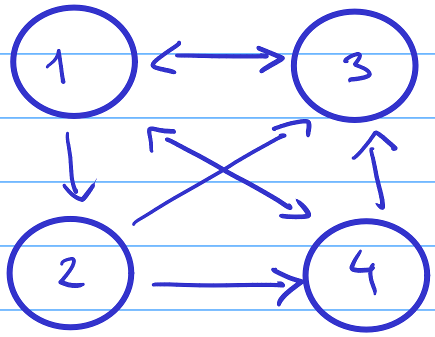
PageRank algorithm

Motivation



Graph: 4 nodes

adjacency matrix

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

Q. which node is the most important?

- "outdegree": 1 has 3
- "indegree" / backlinks: 3 has 3

Markov chain

Define the transition matrix: $P_{ij} = \dfrac{Adj(i,j)}{\sum_{k=1}^{n} Adj(i,k)}$

↳ row normalized

ex:

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

P is a stochastic matrix:

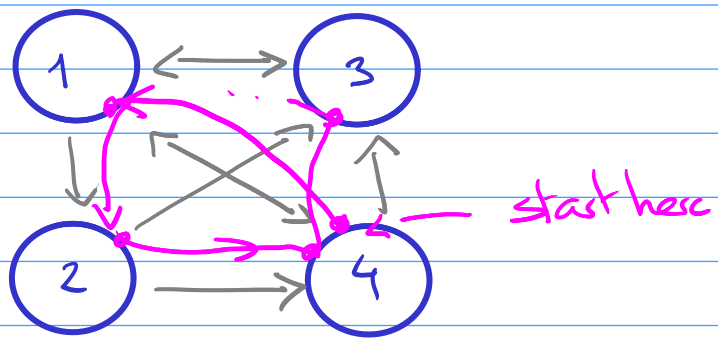
- $P_{ij} \geq 0$
- $\sum_{j=1}^{n} P_{ij} = 1$ for all i

We can define a Markov chain: X_n "jumping" on the graph

with $\mathbb{P}(X_{n+1} = j \mid X_n = i) = P_{ij}$

$\Rightarrow$ X jumping from node i to node j

ex: $x_1 = 4,\ x_2 = 1,\ x_3 = 2,\ x_4 = 4,\ x_5 = 3, \dots$



<u>Remark</u>: The paper uses the notation:

$$A = P^T \quad (\text{Transpose of } P)$$

$\Rightarrow$ A is column-stochastic matrix.

<u>Idea</u>: The most important node is the one that "appears" the most.

$\Rightarrow$ run code

<u>Q</u>: is there a fast way to find the "proportion"?

Equilibrium distribution

Denote q_n the probability density vector of X_n:

$$q_n(i) = \mathbb{P}(X_n = i)$$

ex: if $X_1 = 4$ always, Then $q_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

Then X_2 is 1 or 3 with probability $\frac{1}{2}$, thus:

$$q_2 = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{bmatrix}$$

How to get from q_1 to q_2? $\quad q_2 = P^T q_1$.

More generally: $\boxed{q_{n+1} = P^T q_n}$ $\rightsquigarrow$

extra

$\approx$ "forward Kolmogorov eq."
$\partial_t \rho = L^* \rho$

"backward Kolmogorov eq"
$\partial_t u = L u$

If the dynamics converges $q_n \to q_\infty$,
Then we have a fixed point:

$$\boxed{q_\infty = P^T q_\infty} \Rightarrow q_\infty \text{ eigenvector of } P^T \ (=A)$$

associated with $\lambda = 1$.

$\Leftrightarrow$ run code

ex: $q_\infty = \begin{bmatrix} .39 \\ .13 \\ .29 \\ .19 \end{bmatrix} \leftarrow$

Linear algebra

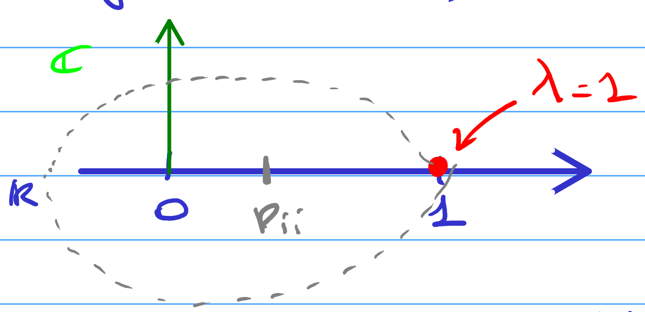
Questions: 1. Is q_∞ unique? i.e. $\lambda = 1$ has multiplicity 1?

Perron Frobenius 2. How to compute it for very large matrix?

spectrum of P (or P^T)

Since P stochastic matrix

$$P\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix} = 1 \cdot \begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}$$



Gershgorin's Thm: $\lambda \in \bigcup_i B\left(p_{ii}, \sum_{j \neq i} |p_{ij}|\right)$

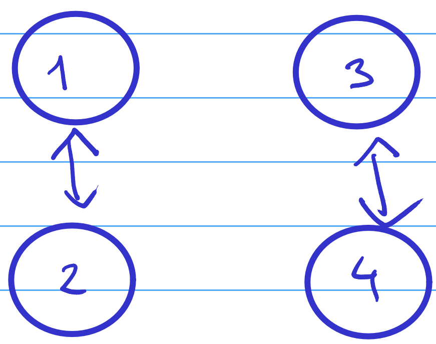
$$P = \begin{bmatrix} \cdots & p_{ii} & \cdots \\ & & \end{bmatrix}$$

Thus: $Re(\lambda) \leq 1$.

$\Rightarrow \lambda = 1$ is the "largest" eigenvalue.

Problem: it can have multiplicity > 1 if the graph is "disconnected" (i.e. P "reducible")

ex



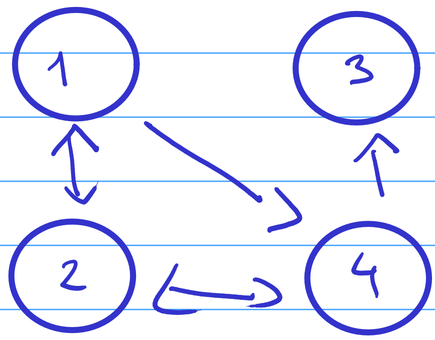
$$Adj = \begin{bmatrix} \boxed{\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \boxed{\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}} \end{bmatrix}$$

Two eigenvectors associated with $\lambda = 1$:

$$u_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad u_2 = \begin{pmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \end{pmatrix}.$$

Another problem: "absorbing states"

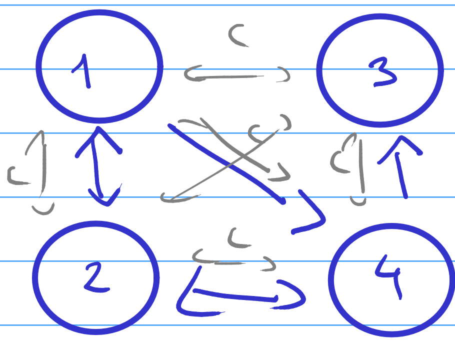
$\Leftrightarrow$ dangling node



← cannot escape (3)

Thus: $q_\infty = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$

To avoid these 2 problems, add "ε" connexion everywhere



$$P^\varepsilon = (1-\varepsilon)P + \varepsilon \begin{bmatrix} 1/n & \cdots & 1/n \\ 1/n & \cdots & 1/n \end{bmatrix}$$

$\Rightarrow$ stochastic
+ "irreducible"
+ a-periodic

} Perron Frobenius $\exists!\, q_\infty$ s.t.

$$q_\infty = P^{\varepsilon T} q_\infty$$

with $q_\infty > 0$ (and $|q_\infty| = 1$)

How to compute q_∞?

$$q_{n+1} = P q_n \quad \rightsquigarrow \quad q_n \xrightarrow[n\to\infty]{} q_\infty.$$

exponentially

(given by the spectral gap

"$\lambda_1 - \mathrm{Re}(\lambda_2)$"

$\overset{\shortparallel}{1}$

↳ link with Fiedler number

↳ Algebraic connectivity

↳ run code.