Assignment: PageRank

Sebastian Danckwardt, Nils Ekström, Reza Rezvan, Ivan Wely

October 3, 2022

1. Example network. Find the example internet with for webpages:
   http://mschauer.github.io/page1.html
   http://mschauer.github.io/page2.html
   http://mschauer.github.io/page3.html
   http://mschauer.github.io/page4.html
   In this exercise, there is no damping ($\lambda = 0$.)

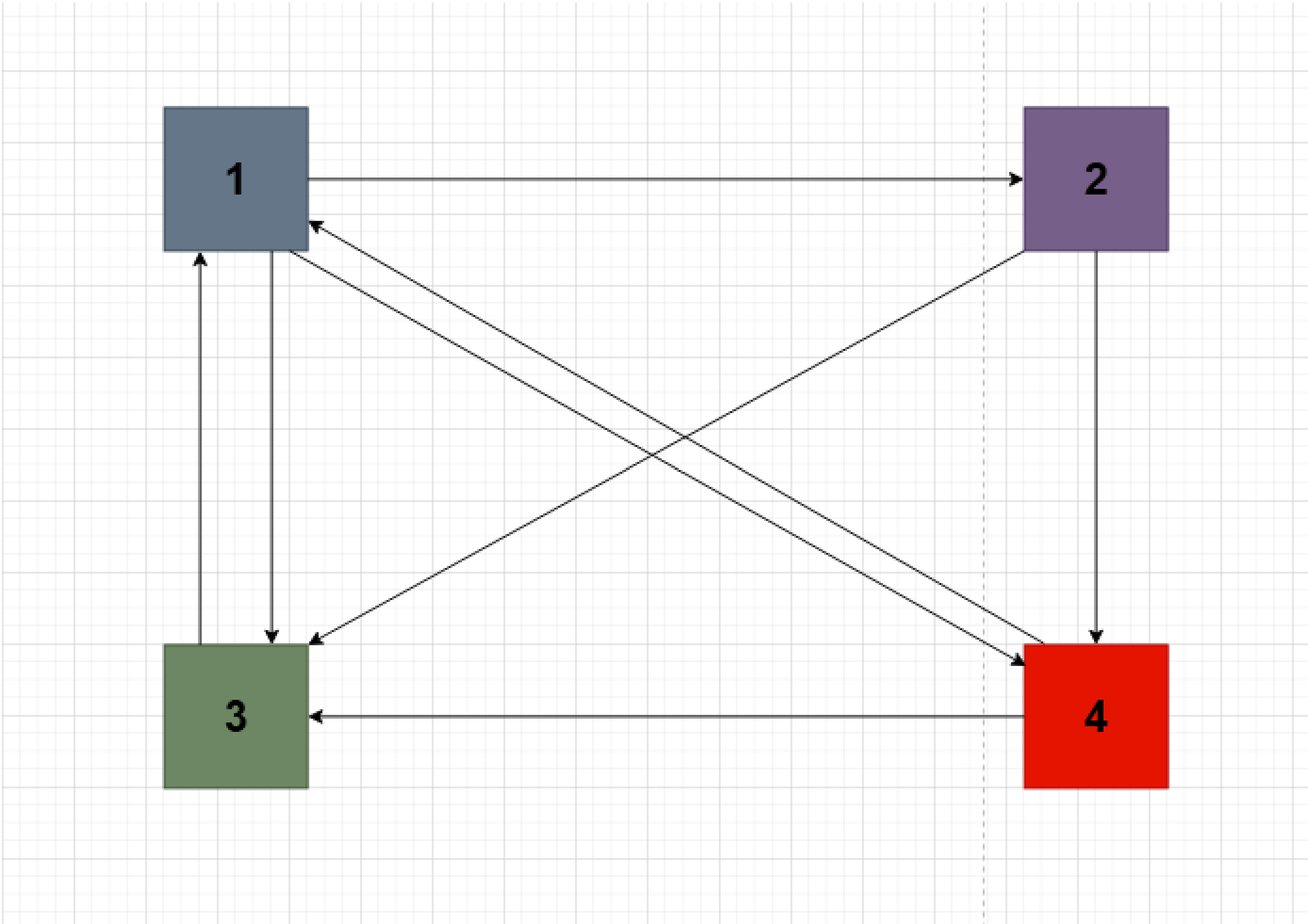
   (a) Use a die (throw again if 5 or 6) to choose one of the 4 web-pages in the network. Record your choice. Proceed by clicking on a randomly chosen link on the current webpage to get a new one. Throw a die to choose a random link with uniform probability. Record the pages you have visited. Record 15 pages. Which page have you visited most?

Throw 1	2, Page2
Throw 2	3, Page4
Throw 3	1, Page1
Throw 4	1, Page2
Throw 5	1, Page3
Throw 6	1, Page1
Throw 7	1, Page2
Throw 8	1, Page3
Throw 9	1, Page1
Throw 10	3, Page4
Throw 11	2, Page3
Throw 12	1, Page1
Throw 13	2, Page3
Throw 13	1, Page1
Throw 14	1, Page2
Throw 15	2, Page4

Table 1: Dice rolls

   In this case we have visited Page1 5 times, Page2 4 times, Page3, 4 times and Page4 3 times.

   (b) Draw a picture with arrows indicating links between the four webpages. Find the adjacency matrix A^{ex}.



The adjaceny matrix:

$$A^{ex} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

(c) Write the PageRank matrix P^{ex} for the example network assuming $\lambda = 0$.

$$P^{ex} = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

(d) Show that the rows of P^{ex} in the example sum to 1.

$$P_{sum_1} = \left(0 + \frac{1}{3} + \frac{1}{3} + \frac{1}{3}\right) = 1$$

$$P_{sum_2} = \left(0 + 0 + \frac{1}{2} + \frac{1}{2}\right) = 1$$

$$P_{sum_3} = \left(1 + 0 + 0 + 0\right) = 1$$

$$P_{sum_4} = \left(\frac{1}{2} + 0 + \frac{1}{2} + 0\right) = 1$$

(e) Describe the Markov process with transition matrix P^{ex} in words.

The transition matrix

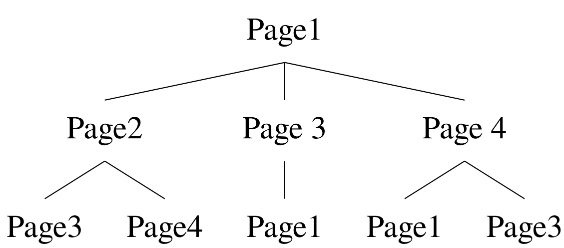
$$P^{ex} = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

can be described as a Markov process, since all rows sum up to one and is memoryless. We can describe it as: If your in state 1 (Page1/Row 1) then we have 3 choices to make (the 3 non-zero columns/entries), if we make choices at random then there's a $\frac{1}{3}$ chance to go to either state 2 3 or 4. Same logic applies to the other rows.

(f) Compute: What is the probability that a surfer starting on page 1 ends on page 2 after two clicks?

$P(\text{state will be } s_2 \text{ after two steps} \mid \text{initial state is } s_1) = 0$

If we draw a tree we can see all possible paths we can take. We can clearly see that there's no path from 1 to 2 that takes 2 clicks. If we first click on page 2 we can only end up on page 3 or 4 at the second click. If we first click on page 3 then we can only get to page 1 in the second click. Lastly, if we first click on page 4, we can only get to page 1 or 3 in the second click.



(g) Write a program which follows a random surfer in the web P^{ex}. Generate three surf history of a random surfer starting in 1.

Look at PageRank.jl

Each history consists for 5 walks and starts in State 1, (Page1)

Walk 1	Page4
Walk 2	Page3
Walk 3	Page1
Walk 4	Page2
Walk 5	Page3

Table 2: First walk

Walk 1	Page2
Walk 2	Page4
Walk 3	Page1
Walk 4	Page2
Walk 5	Page3

Table 3: Second walk

Walk 1	Page3
Walk 2	Page1
Walk 3	Page4
Walk 4	Page1
Walk 5	Page3

Table 4: Third walk

(h) Compute the probability of a random surfer to be in each Webpage after 2, after 10 steps if the starting distribution was $[\frac{1}{4}\frac{1}{4}\frac{1}{4};\frac{1}{4}]$

If the starting vector is $u^{(0)} = [\frac{1}{4}\frac{1}{4}\frac{1}{4};\frac{1}{4}]$ then the vector after 2 steps is $u^{(2)} = u^{(0)} * P^{ex^2}$

$$P^{ex^2} = \begin{pmatrix} .5 & 0.0 & .33 & .16 \\ .75 & 0.0 & .25 & 0.0 \\ 0.0 & .33 & .33 & .33 \\ .5 & .16 & .16 & .16 \end{pmatrix}$$

$u^{(2)} = \begin{pmatrix} .4375 & 0.125 & .27083 & .16 \end{pmatrix}$

Same logic applies to $u^{(10)}$

$u^{(10)} = u^{(0)} * P^{ex^{10}}$

$u^{(10)} = \begin{pmatrix} .3875 & 0.1288 & .2902 & .1933 \end{pmatrix}$

(i) Does q = qP^{ex} have a solution?

Update this answer

Yes since the P^{ex} is regular after some ”time”, it will have a solution to $q = qP^{ex}$

2. Markov chains.

(a) Show that P defined in (2) is a Markov matrix.

A (Right) Markov, or Stochastic matrix, is a matrix which:

- Is a square matrix (n x n)
- Only has non-negative entries
- The rows of the matrix represent the probability of the next state, therefor sum up to 1

In our case P fulfills all these requirments

(b) Explain how $\lambda > 0$ is related to "getting bored" in the quote above

If $\lambda > 0$ then we also get the probability to randomly selecting a page which isn't connected from our current page, therefore, "getting bored" is at random we choose a new page that isn't connected from our current page.

(c) Argue that if $\lambda > 0$, there is a solution vector q. Give an interpretation of q. Why is q a good measure for Website importance? (Evidently it was, Google won the search engine market.)

If $\lambda > 0$ the Matrix is still regular and seen as a Markov matrix, then there's a q such that $q = qP$ that solves this system. q will show us what page will be visited the most in the long run probability.

3. PageRank.

(a) Write a program to compute q for the example depending on λ.
Look at PageRank.jl

(b) Write a program which takes an adjacency (link) matrix A and λ computes the PageRank vector q.
Look at PageRank.jl

(c) Optional: Use this on a data set of your choice.
Look at PageRank.jl