

DISCUSSION WEEK 10

Part 5 Problem Set

This content was created for Information Retrieval Fall 2021 at UC Irvine by Brooke Kelsey Ryan. These slides, accompanying recordings, and any other course materials are protected by U.S. copyright law and may not be reproduced, distributed, or displayed without the express written consent of the author.
© Brooke Kelsey Ryan 2021






DISCUSSION 10

Itinerary

- Info about the final exam
- Part 5 problem solving
 - Mathematical-type problems between Quiz 4 and end of the course

Deadlines

- **December 3:** Assignment 3
- **Sometime finals week:**
Group walkthrough for Assignment 3
 - Calendly links will be posted soon on EdStem
- **December 10, 8-10am:**
Final exam!



Find these slides & recordings on Canvas / Pages / Discussion Resources

FINAL ALLOWED ITEMS

- Laptop
- Pocket Calculator
- Scratch paper
- Course book
- Lecture slides
- Notes



Exam Material

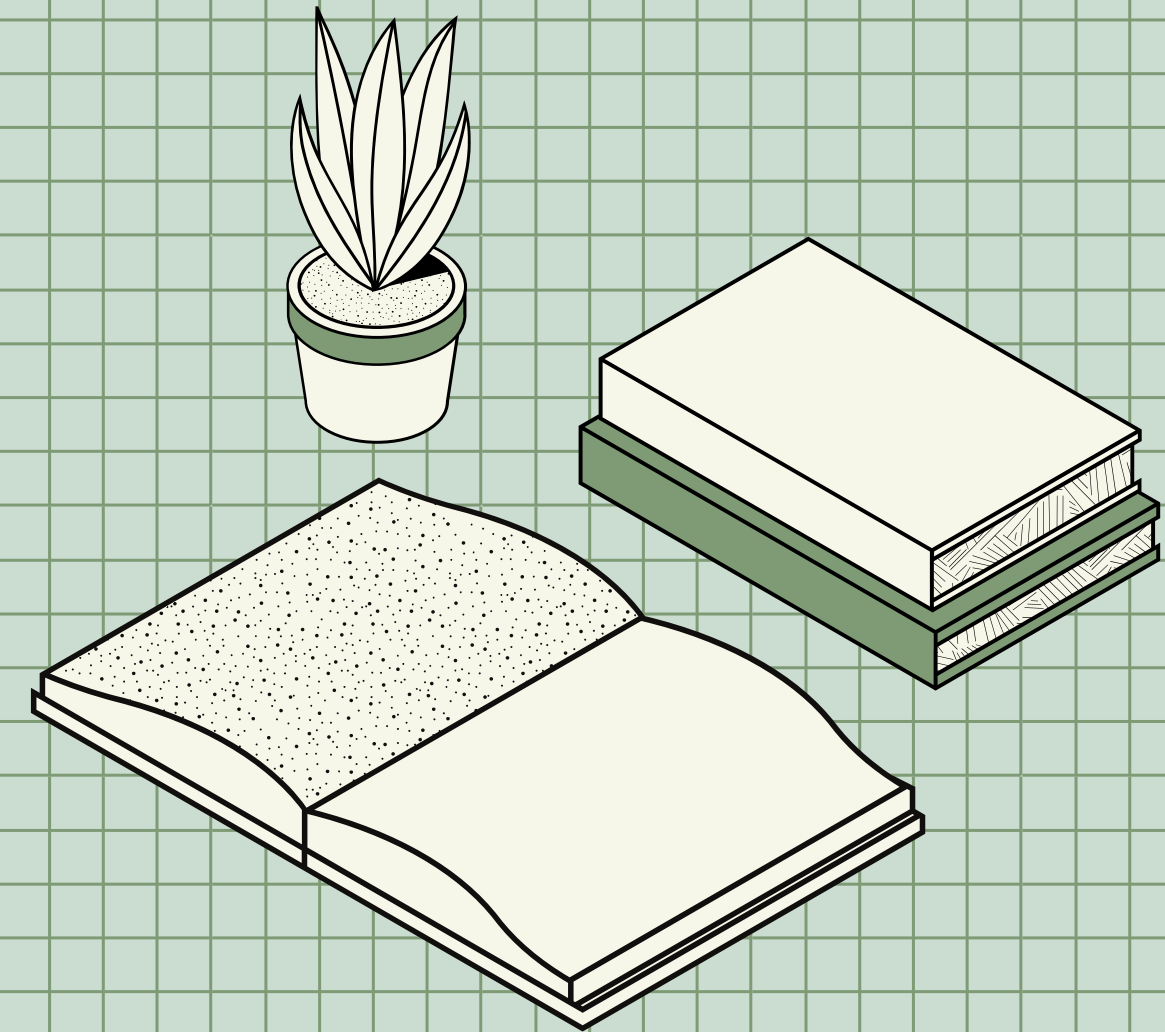
"Part 5" material encompasses problems that are after Quiz 4 material through the end of the course.

I will only be covering Part 5 today, but your Final Quiz covers approximately:

40% Quiz 1 - 4 Material (4 questions each quiz)

60% Part 5 Material

Protip: Combine all lecture slides to 1 PDF for easy searching during the exam!



I have recordings and slides available for all past quizzes! Look at Canvas / Pages / Discussion Resources



Part 5 Review

Here is an overview of the kinds of problems we will be working through today. They **will appear** on the final!



Page Rank



Precision & Recall



Authority

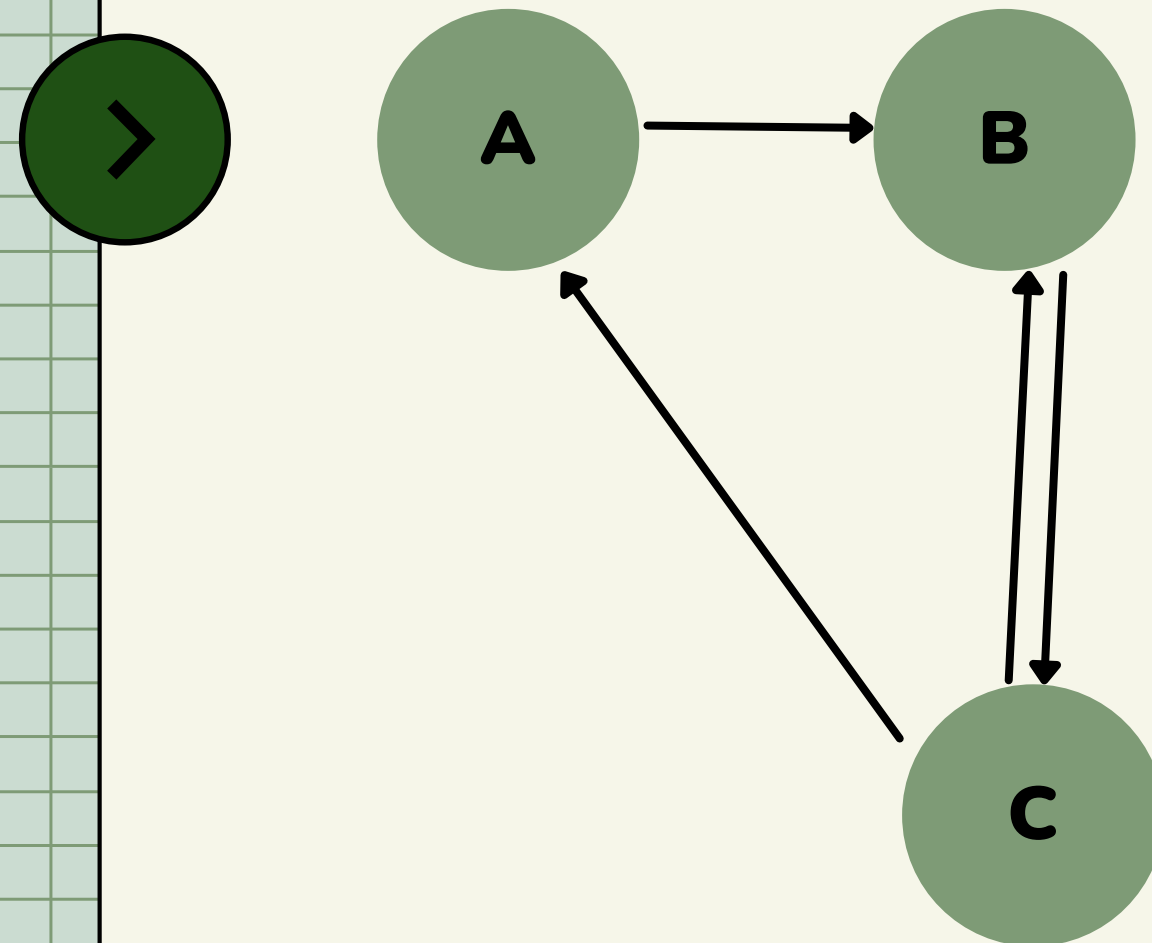


Hubs

Page Rank

What is the Page Rank of node A after 2 iterations?

$$PR(A) = (1-d) + d \sum (PR(T_i)/C(T_i))$$



$$PR(A) = (1-d) + d \sum (PR(T_i)/C(T_i))$$

A is a page

d is a damping factor (usually 0.85)

T1...Tn are pages that link to A

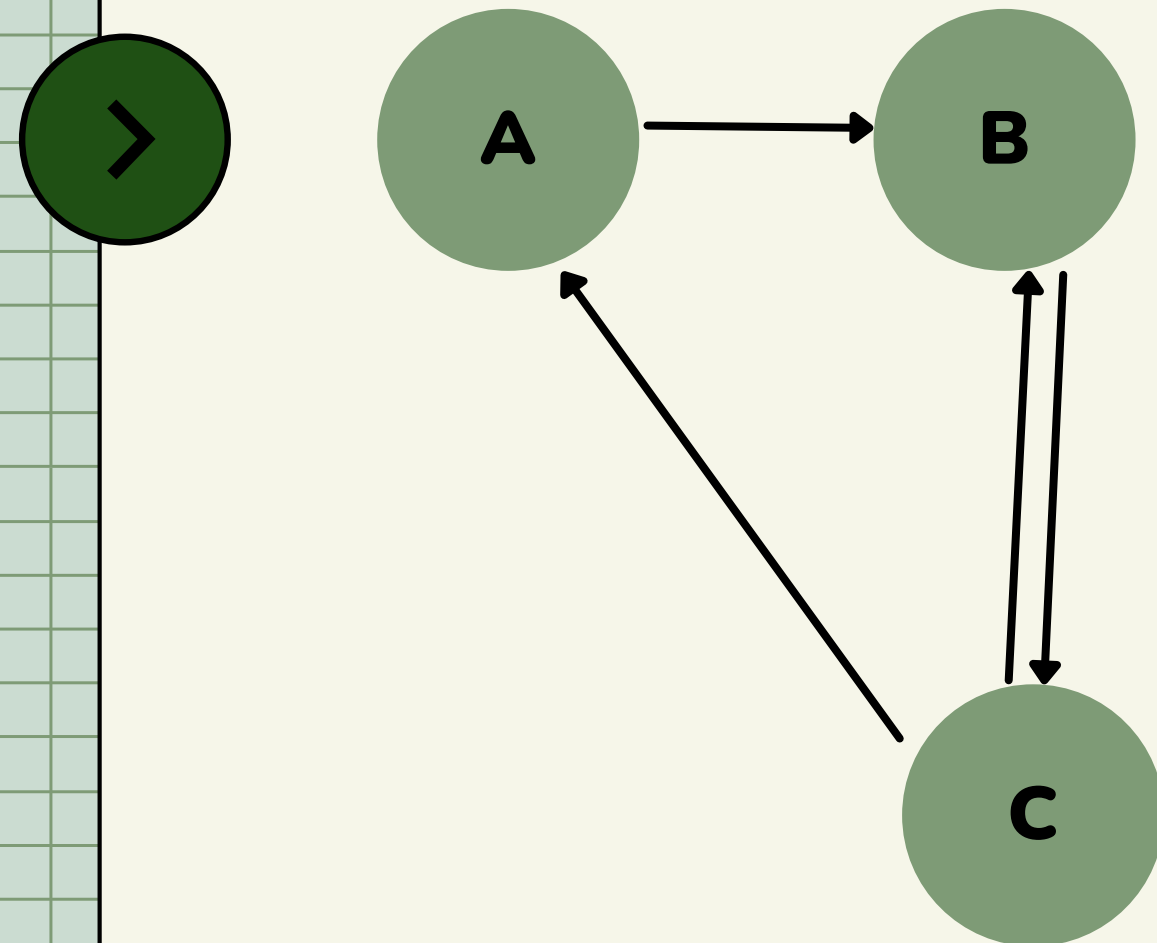
PR(Ti) is the PageRank of Ti

C(Ti) is the number of **outgoing** links from Ti

Page Rank

What is the Page Rank of node A after 2 iterations?

$$PR(A) = (1-d) + d \sum (PR(T_i)/C(T_i))$$



$$PR(A) = 0.15 + 0.85 * PR(C)/2$$

$$PR(B) = 0.15 + 0.85 * (PR(A)/1 + PR(C)/2)$$

$$PR(C) = 0.15 + 0.85 * (PR(B)/1)$$

How do we start?

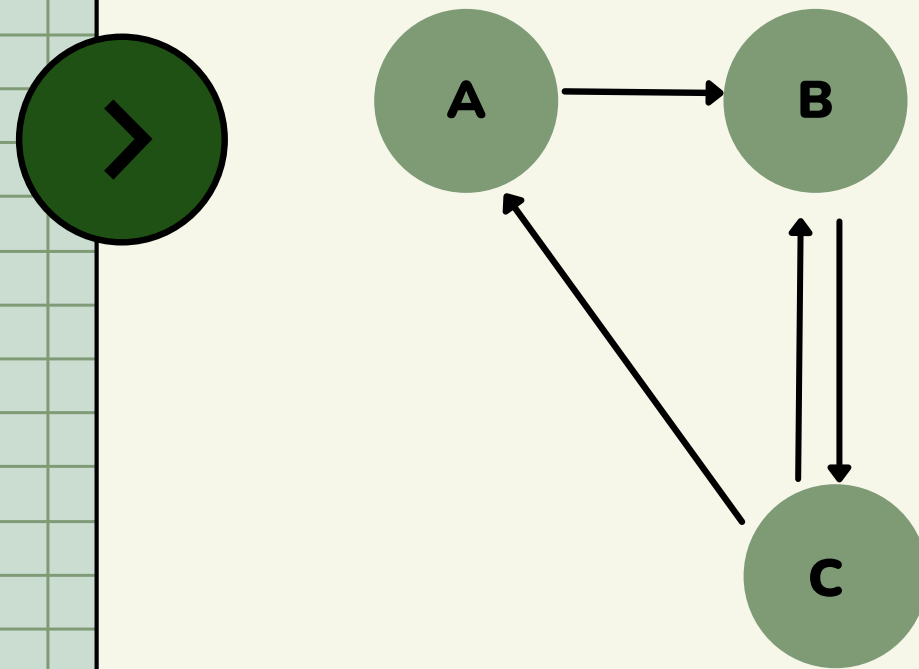
Guess $PR(p) = 1$ for starters

Then, iterate

Page Rank

What is the Page Rank of node A after 2 iterations?

$$PR(A) = (1-d) + d \sum (PR(T_i)/C(T_i))$$



$$\begin{aligned} PR(A) &= 0.15 + 0.85 * PR(C)/2 \\ PR(B) &= 0.15 + 0.85 * (PR(A)/1 + PR(C)/2) \\ PR(C) &= 0.15 + 0.85 * (PR(B)/1) \end{aligned}$$

Iteration 1

$$\begin{aligned} PR(A) &= 0.15 + 0.85 * 1/2 = 0.575 \\ PR(B) &= 0.15 + 0.85 * (1/1 + 1/2) = 1.425 \\ PR(C) &= 0.15 + 0.85 * (1/1) = 1 \end{aligned}$$

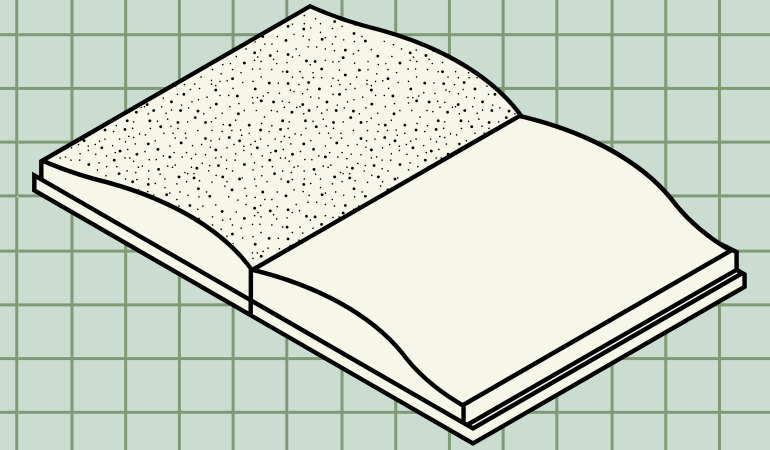
Iteration 2

$$\begin{aligned} PR(A) &= 0.15 + 0.85 * 1/2 = 0.575 \\ PR(B) &= 0.15 + 0.85 * (0.575 + 1/2) = 1.06375 \\ PR(C) &= 0.15 + 0.85 * 1.425 = 1.36125 \end{aligned}$$



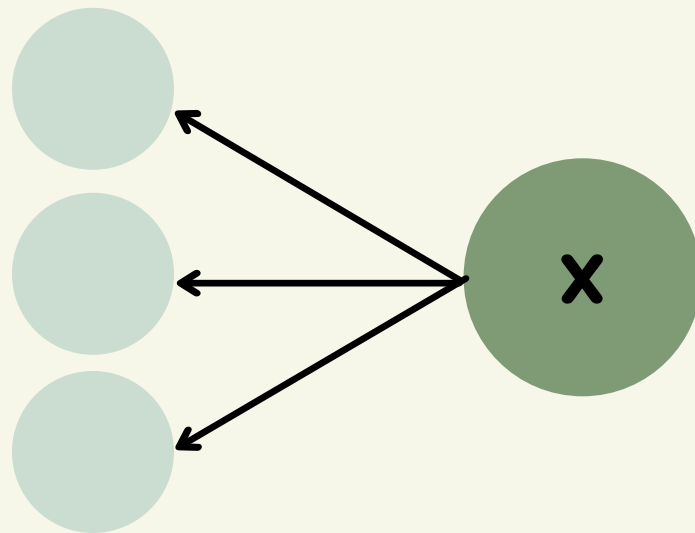
ITERATIVE UPDATE

Repeat the following updates, for all x :



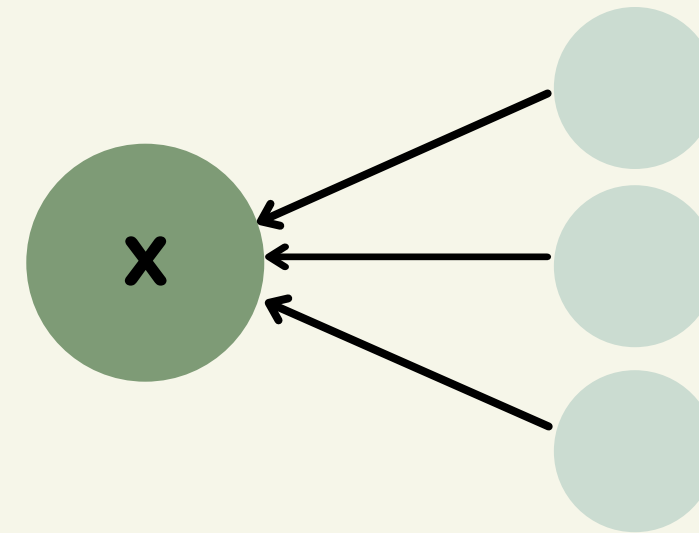
Hub

$$h(x) \leftarrow \sum_{x \mapsto y} a(y)$$



Authority

$$a(x) \leftarrow \sum_{y \mapsto x} h(y)$$

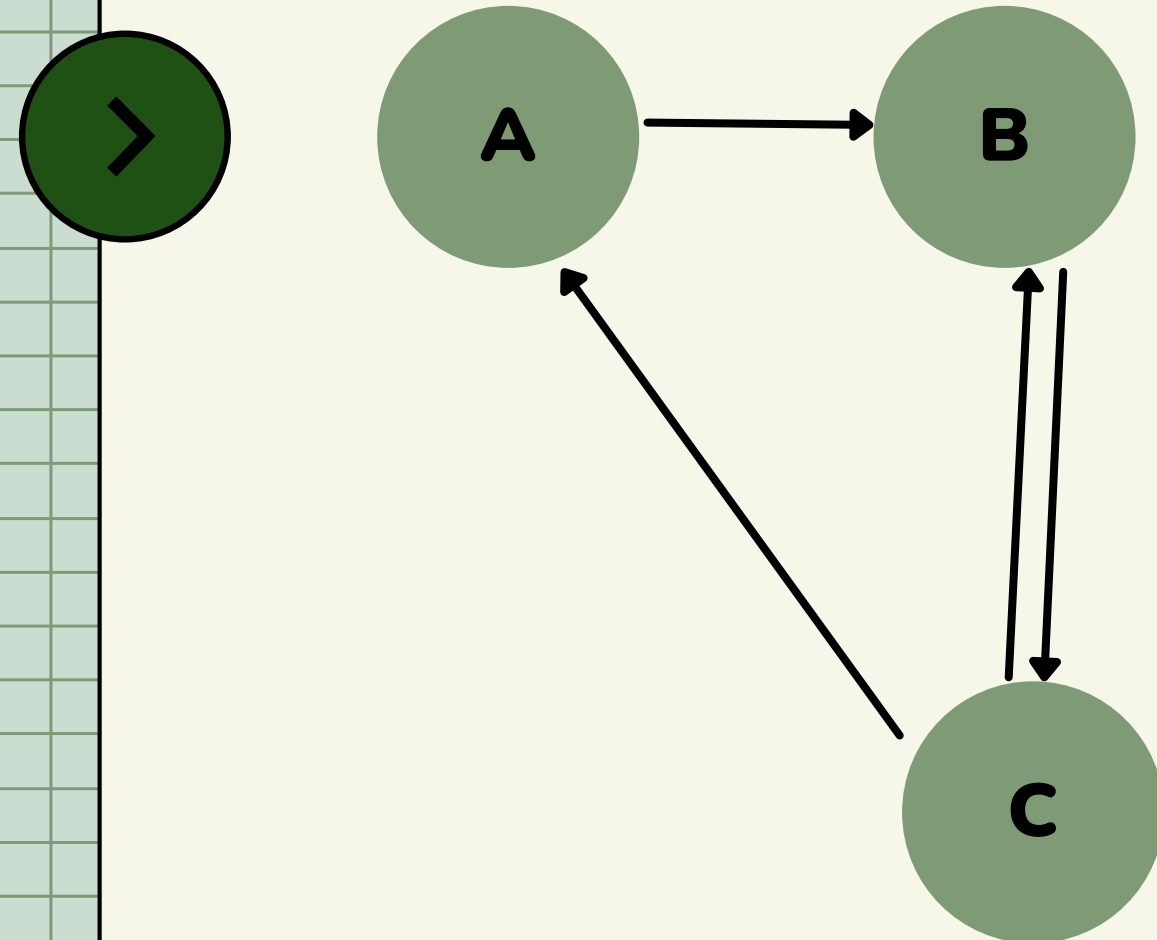


Q1: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for **node A** in the very beginning of the calculation of those scores?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$

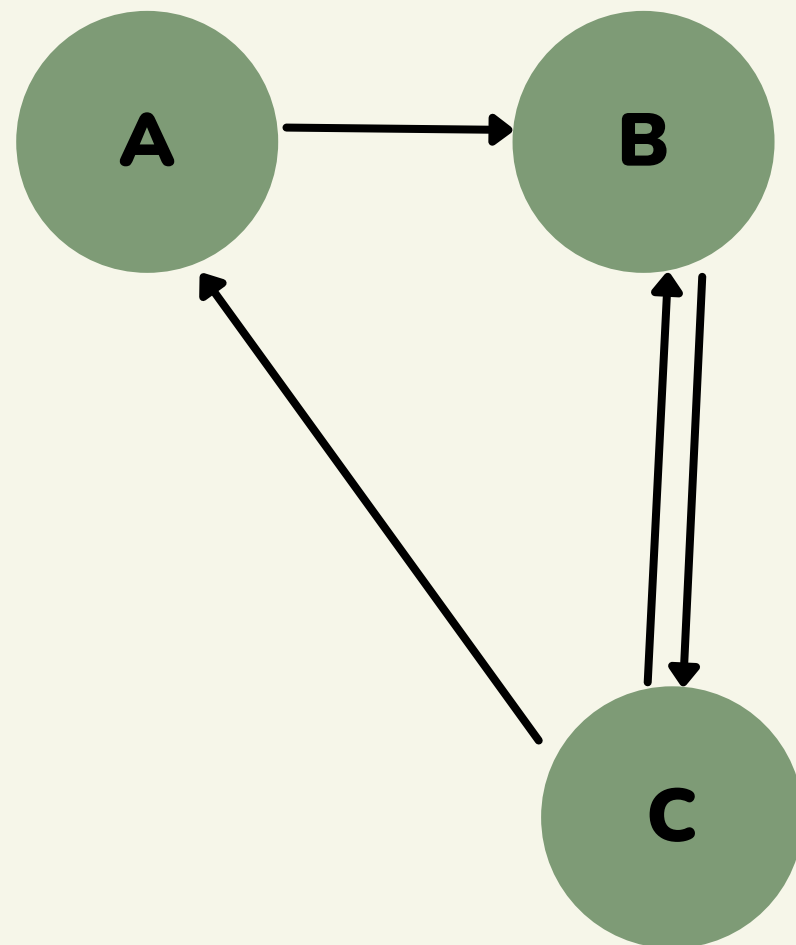
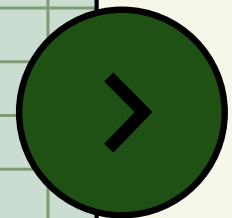


Q1: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for **node A** in the very beginning of the calculation of those scores?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$



Where to start?

Guess $h(p) = a(p) = 1$ for starters

Iteration 0

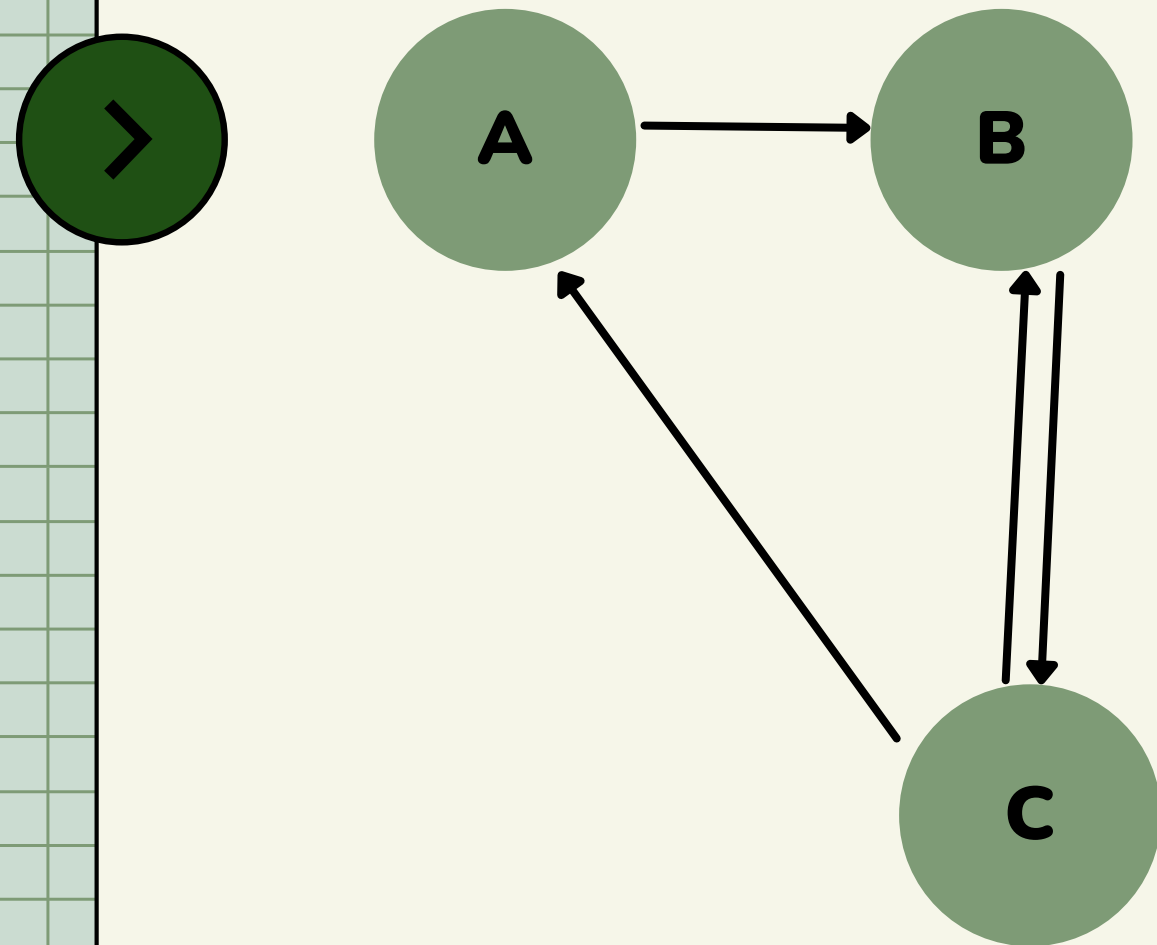
= 1

Q2: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for node B after 3 iterations?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$

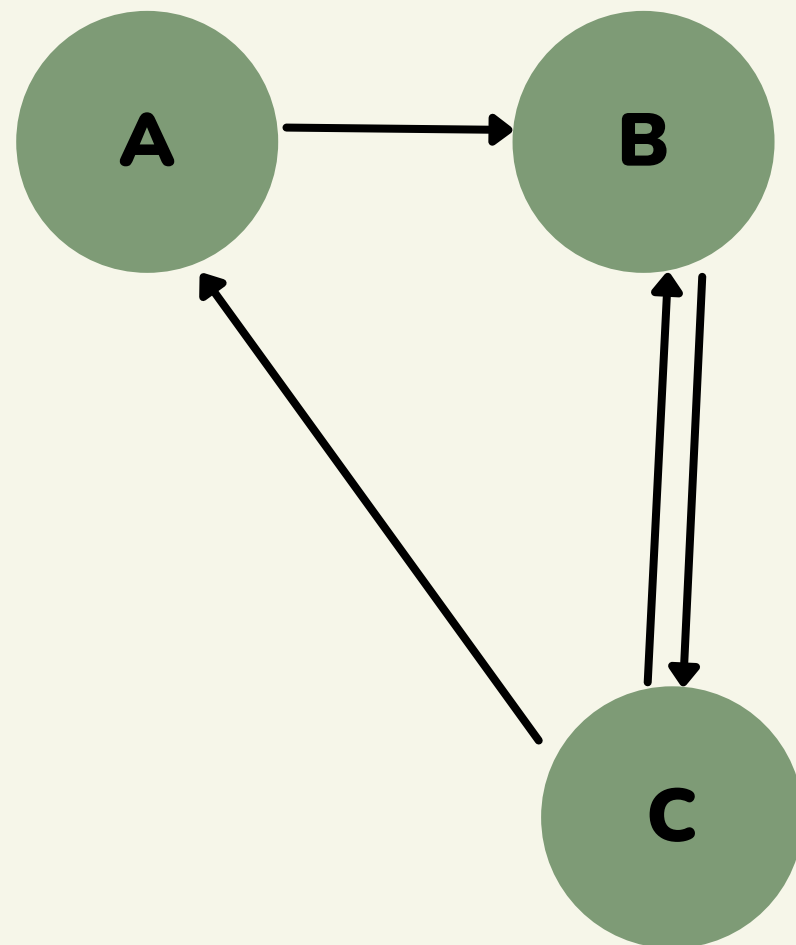
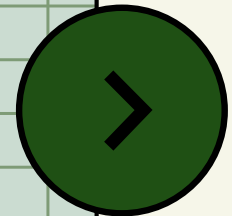


Q2: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for node B after 3 iterations?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$



Iteration 0

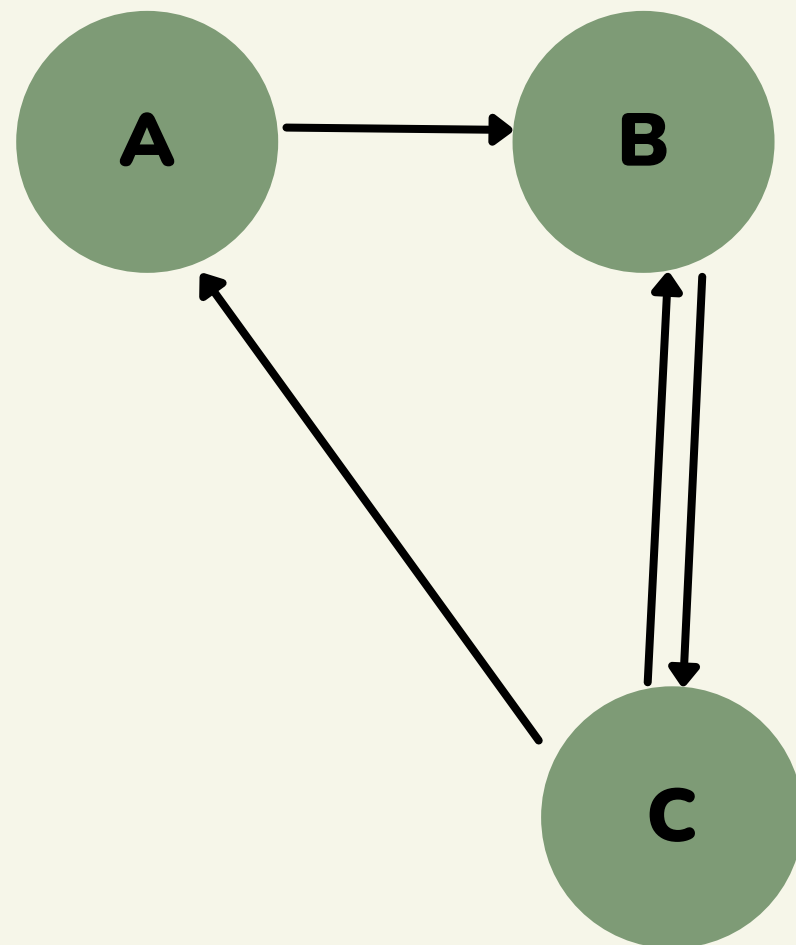
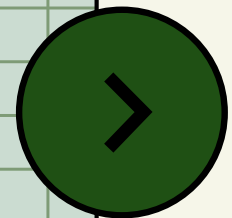
	H(P)	A(P)
A	1	1
B	1	1
C	1	1

Q2: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for node **B** after 3 iterations?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$



Iteration 1

	H(P)	A(P)
A	1	1
B	1	2
C	2	1

$$h1(A) = a0(B)$$

$$a1(A) = h0(C)$$

$$h1(B) = a0(C)$$

$$a1(B) = h0(A) + h0(C)$$

$$h1(C) = a0(A) + a0(B)$$

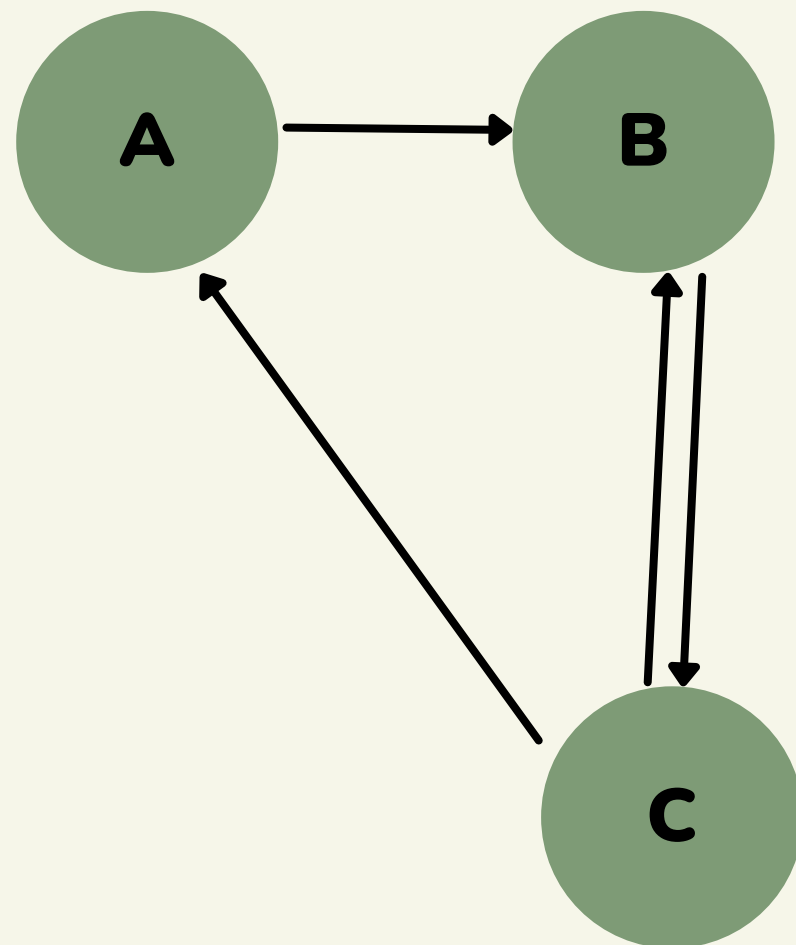
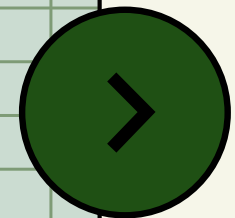
$$a1(C) = h0(B)$$

Q2: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for node **B** after 3 iterations?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$



Iteration 2

	H(P)	A(P)
A	2	2
B	1	3
C	3	1

$$h2(A) = a1(B)$$

$$a2(A) = h1(C)$$

$$h2(B) = a1(C)$$

$$a2(B) = h1(A) + h1(C)$$

$$h2(C) = a1(A) + a1(B)$$

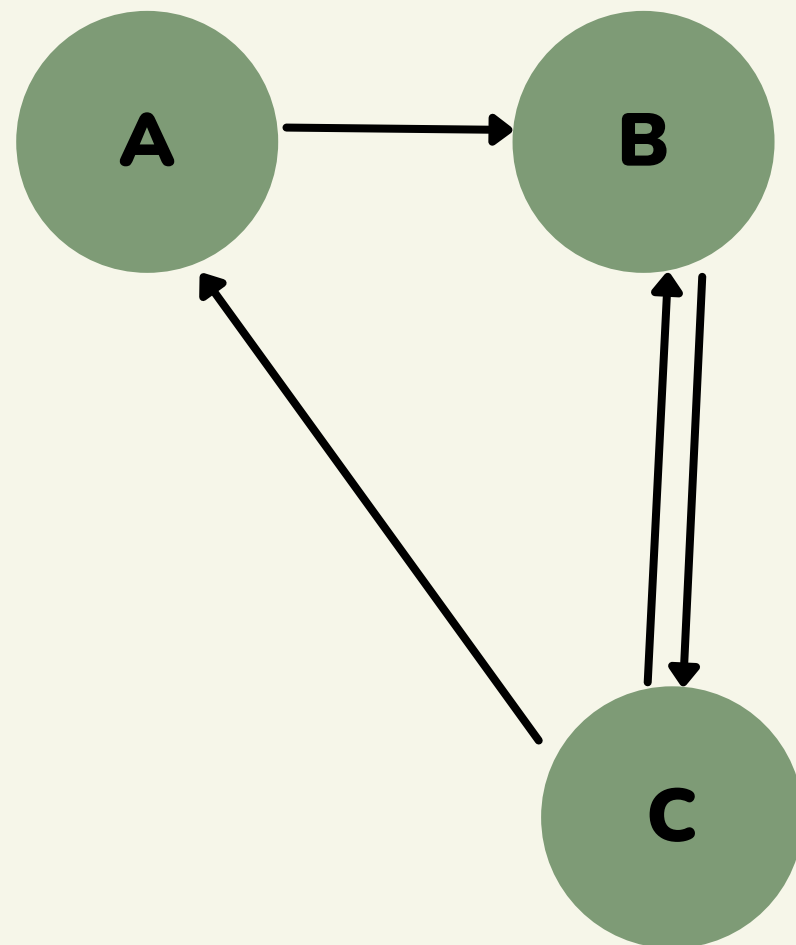
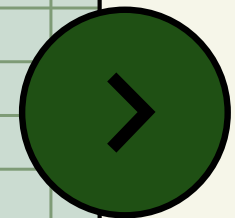
$$a2(C) = h1(B)$$

Q2: Hubness and Authoritativeness

What are the authoritativeness and hubness scores for node **B** after 3 iterations?

$$a(x) = \sum_{y \rightarrow x} h(y)$$

$$h(x) = \sum_{x \rightarrow y} a(y)$$



Iteration 3

	H(P)	A(P)
A	3	3
B	1	5
C	5	1

$$h_3(A) = a_2(B)$$

$$a_3(A) = h_2(C)$$

$$h_3(B) = a_2(C)$$

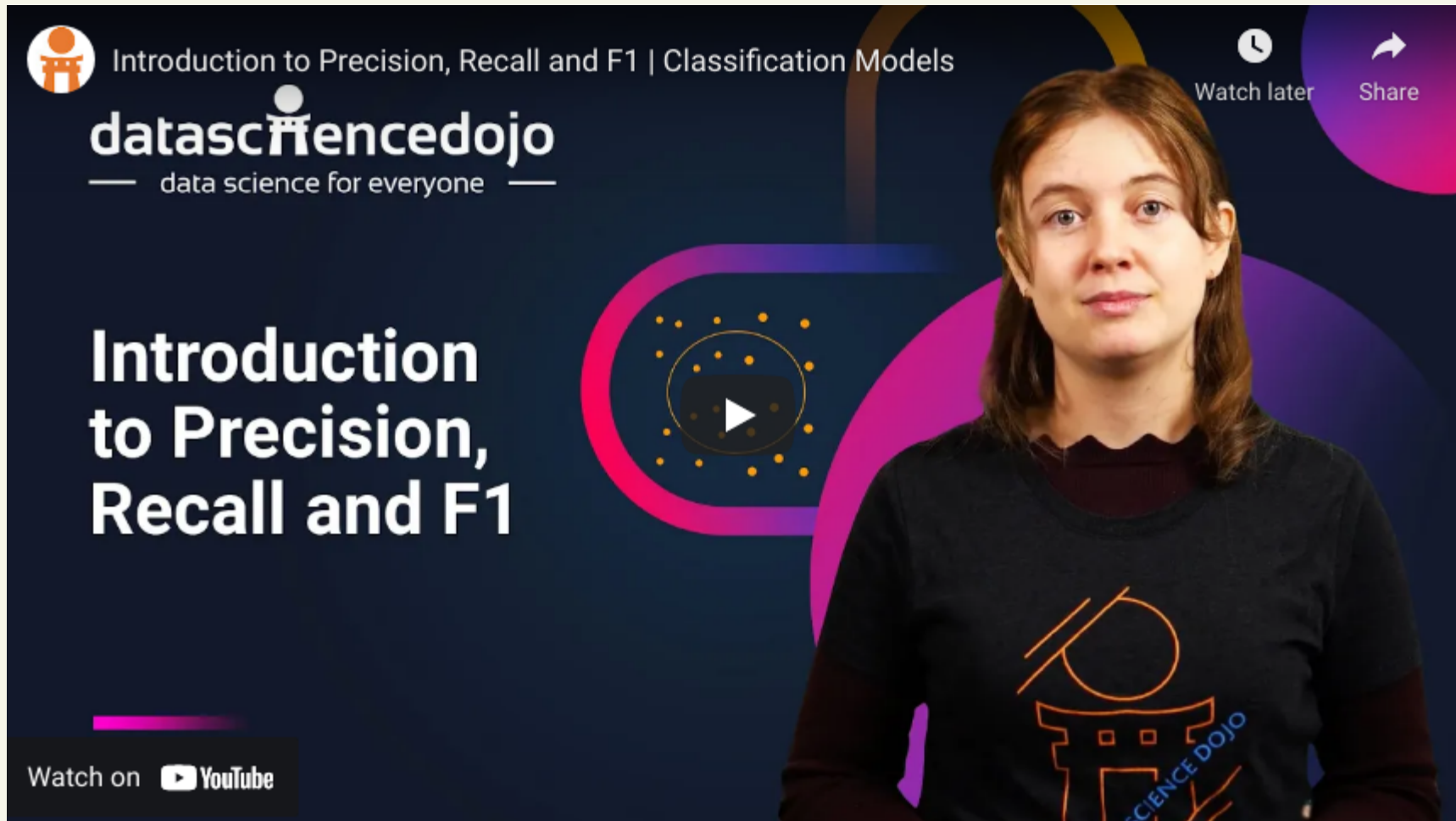
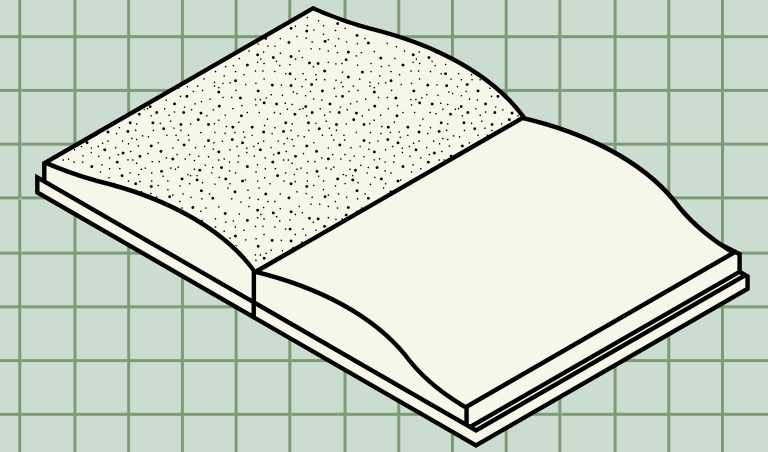
$$a_3(B) = h_2(A) + h_2(C)$$

$$h_3(C) = a_2(A) + a_2(B)$$

$$a_3(C) = h_2(B)$$



PRECISION AND RECALL



Precision

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}}$$

Recall

$$\text{Recall} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}}$$

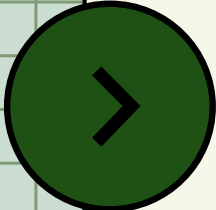
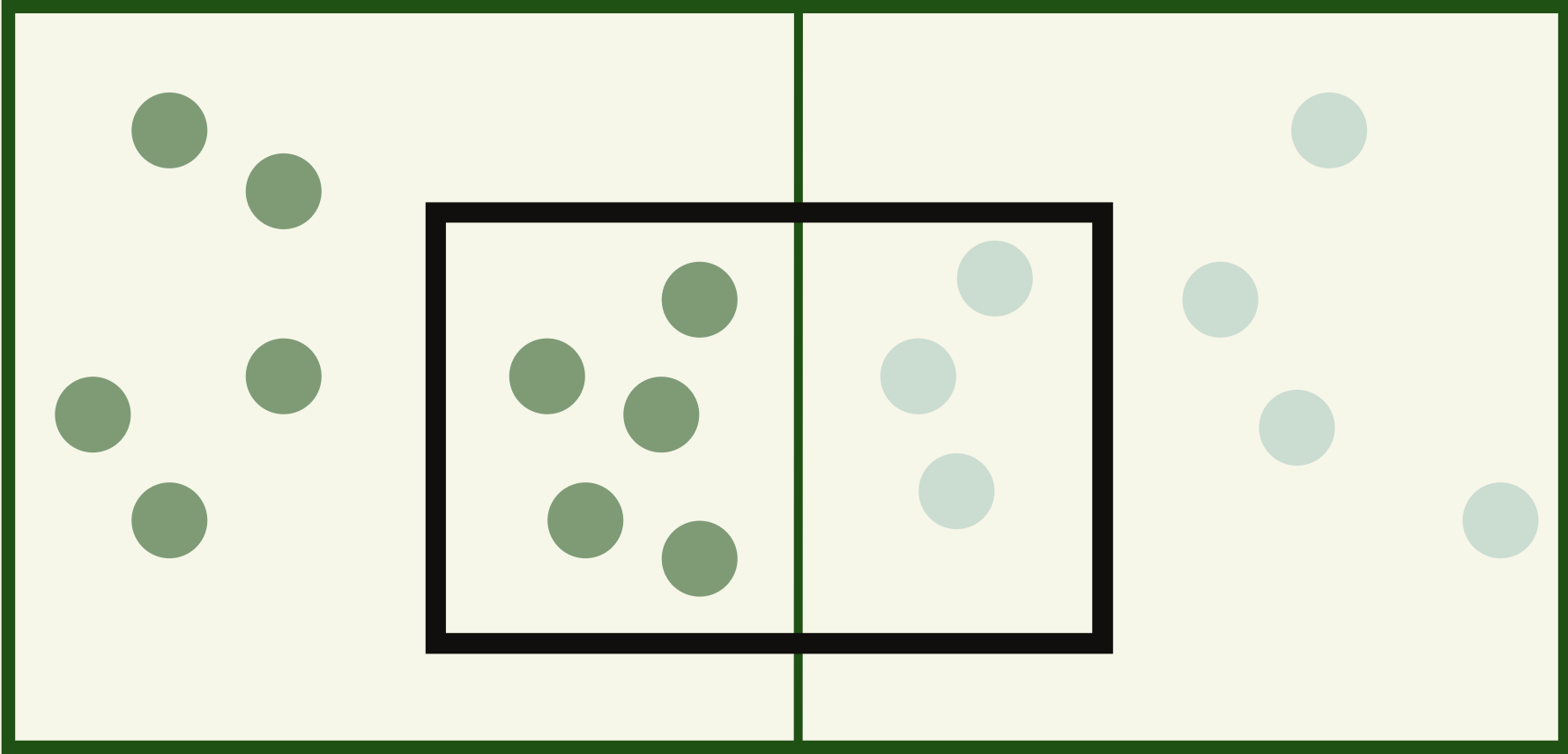
Precision & Recall

Consider the following picture depicting documents (dots) in a collection. The rectangle in the middle represents the documents retrieved for a given query.

What is the **precision** and **recall** of the retrieval scheme for this query?

Relevant

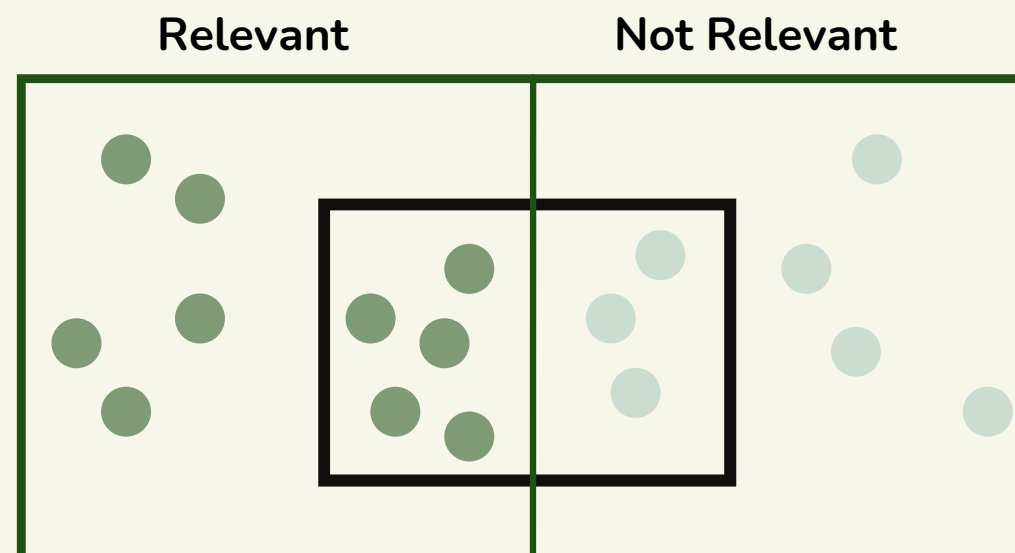
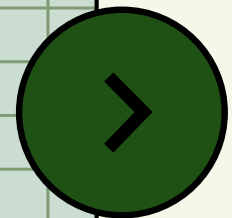
Not Relevant



Precision & Recall

Consider the following picture depicting documents (dots) in a collection. The rectangle in the middle represents the documents retrieved for a given query.

What is the **precision** and **recall** of the retrieval scheme for this query?



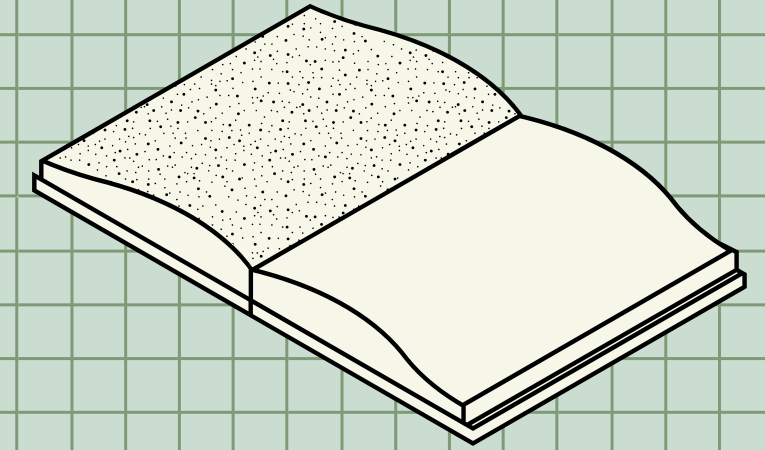
$$\begin{aligned}\text{Precision} &= \text{TruePositives} / (\text{TruePositives} + \text{FalsePositives}) \\ &= 5 / (5 + 3) \\ &= \mathbf{0.625}\end{aligned}$$

$$\begin{aligned}\text{Recall} &= \text{TruePositives} / (\text{TruePositives} + \text{FalseNegatives}) \\ &= 5 / (5 + 5) \\ &= \mathbf{0.5}\end{aligned}$$



OTHER PROBLEMS

Make sure you also study these Part 5 problems for the Final Quiz!

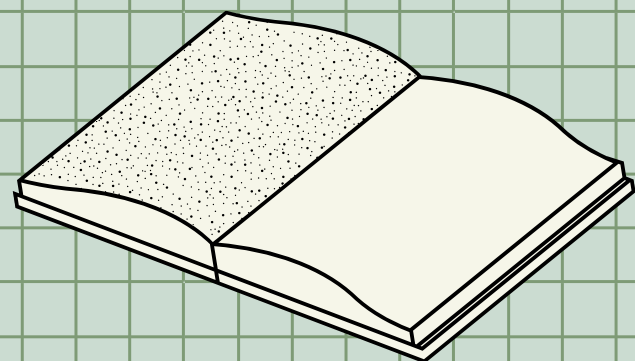


Mean Average Precision

Example available in the lecture slides.

NDCG

Example available in lecture slides from December 1 class.



THANK YOU!

Thank you so much for a great quarter everyone. Don't forget to participate in the **Course Survey** for extra credit!

