Centrality and Other Measures

CS 579 Online Social Network Analysis

Dr. Cindy Hood
9/23/25

Homework Assignments

- HW #3 – Network Metrics
  - Assigned soon
  - Good prep for Exam 1
- HW #4 – Chicago Community Areas + Census Data
  - You may work in groups up to 4 students (no exceptions) on this hw
  - Assigned later this week
- Please contact TAs with questions on hw grading

Exams and Final Project Poster Presentation

- Exam 1 – Oct 9 in class
- Exam 2 – Dec 2 in class

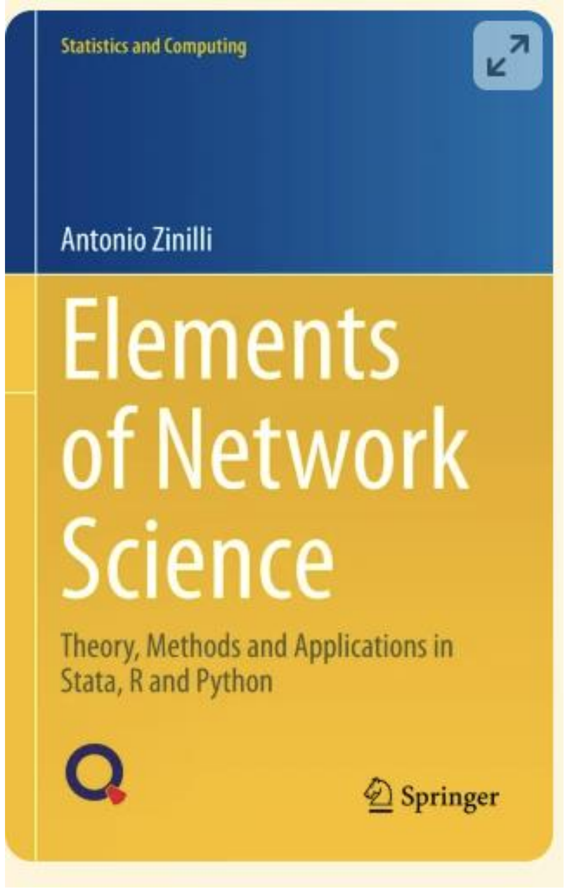
- Final Project Poster Session – Dec 4 in class

- Online students (sections 2 and 3) will have remote options

Teaching Assistants

- Siva Krishna Golla
  - sgolla2@hawk.illinoistech.edu
  - Mondays 2-3pm on zoom
- Khush Dhiren Patel
  - kpatel210@hawk.illinoistech.edu
  - Wednesdays 11-12 online
- Aswith Sama
  - asama@hawk.illinoistech.edu
  - Thursdays 3-4pm on zoom

Reference



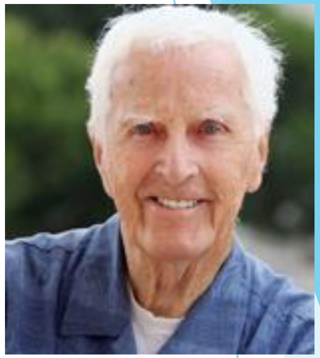
https://link.springer.com/book/10.1007/978-3-031-84712-7

Recall - Degree Centrality

- Total number of connections a vertex has
  - Degree of vertex
    - Total number of edges connected to a vertex
- Directed network
  - In-degree
  - Out-degree
- Can be considered a popularity measure
  - Is it a good popularity measure?

Recall - Betweenness Centrality

Another way of looking at centrality is by considering how important nodes are in connecting other nodes



Linton Freeman

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$$

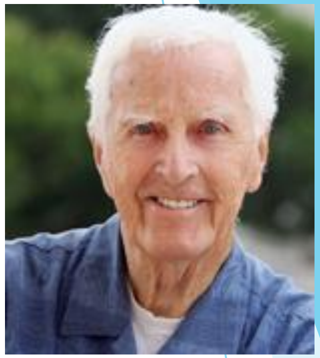
σ_{st} The number of shortest paths from vertex s to t – *a.k.a.* information pathways

$\sigma_{st}(v_i)$ The number of **shortest paths** from s to t that pass through v_i

Recall - Closeness Centrality

- The intuition is that influential/central nodes can quickly reach other nodes

- These nodes should have a smaller average shortest path length to others


Linton Freeman

Closeness centrality: $C_c(v_i) = \frac{1}{\bar{l}_{v_i}}$

$$\bar{l}_{v_i} = \frac{1}{n-1} \sum_{v_j \neq v_i} l_{i,j}$$

Page Rank Centrality

- Measure of global centrality
  - Uses entire network to assess the significance of single node
- Assesses node’s relevance based on importance of nodes that link to it
- Iterative process that estimates a node’s importance in the network
  - Calculate centrality of a node based on importance of neighbors

Page Rank Calculation (Simple)

- To begin with, each node is equally important:

$$PR_i^0 = \frac{1}{n}, \forall i$$

n = number of nodes/pages

- Each node distributes its centrality equally to the nodes it links to (outgoing)
- Page Rank of a node is calculated by adding the Page Rank fractions of the vertices that have edges to it from the previous iteration

$$PR_i^{t} = \sum_{j=1}^{n} a_{ji} \frac{PR_j^{t-1}}{d_j^{out}}, \forall t \in \{1,k\}$$

t = iteration

where d_j^{out} is the number of hyperlinks on the node/page j.

Page Rank Calculation

- The first iteration is

$$PR_i^1 = \frac{1}{n}\sum_{j=1}^{n} a_{ji}\,\frac{1}{d_j^{out}}$$

- Iterations continue until Page Rank values converge
  - When the difference between PR^{t+1} and PR^t is very small and is converging to 0

Page Rank Example

Fig. 3.7 Graph example

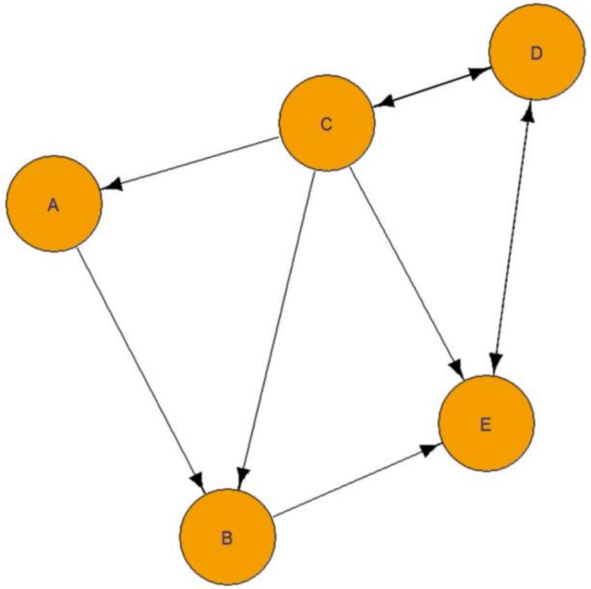
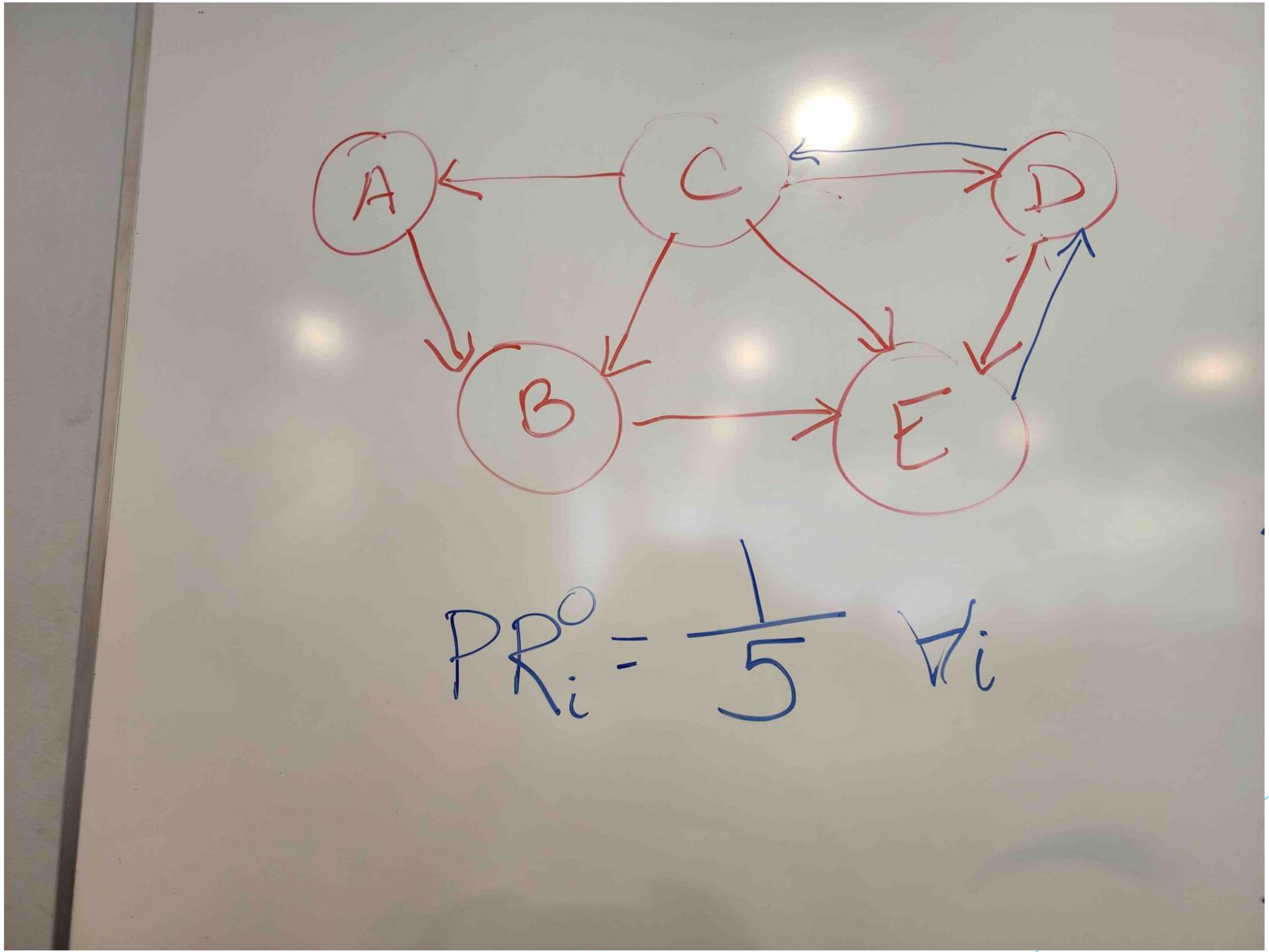
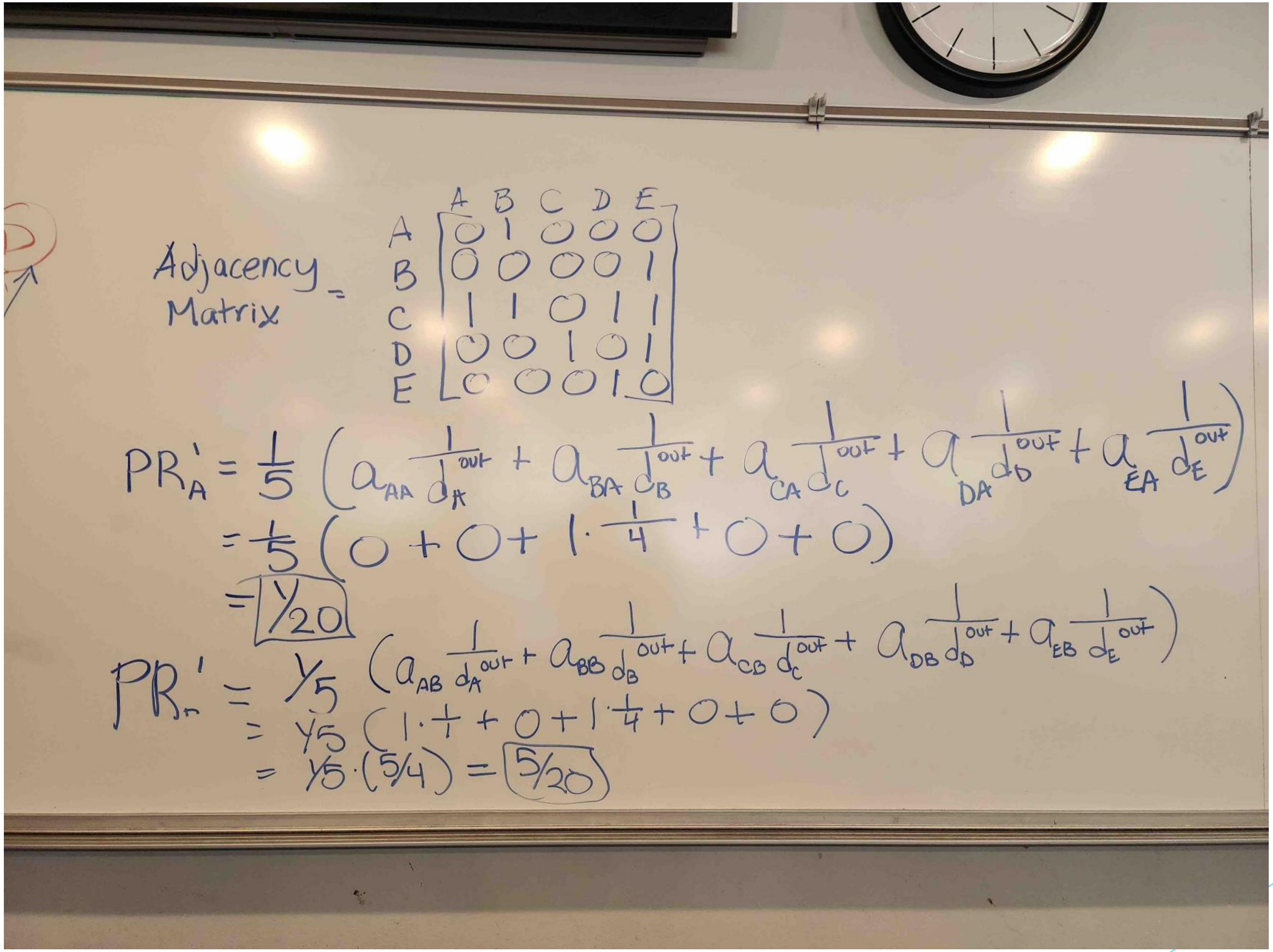


Table 3.2 Example of PageRank computation (second step)

Nodes	Iteration 1	Iteration 2	Iteration 3	Final rank
A	1/5	1/20	1/40	5
B	1/5	5/20	3/40	4
C	1/5	1/10	5/40	3
D	1/5	5/20	15/40	2
E	1/5	7/20	16/40	1


A
C
D
B
E
$PR_i^0 = \frac{1}{5} \; \forall i$


Adjacency Matrix =
$$\begin{array}{c|ccccc} & A & B & C & D & E \\ \hline A & 0 & 1 & 0 & 0 & 0 \\ B & 0 & 0 & 0 & 0 & 1 \\ C & 1 & 1 & 0 & 1 & 1 \\ D & 0 & 0 & 1 & 0 & 1 \\ E & 0 & 0 & 0 & 1 & 0 \end{array}$$
$$PR_A' = \frac{1}{5}\left(a_{AA}\frac{1}{d_A^{out}} + a_{BA}\frac{1}{d_B^{out}} + a_{CA}\frac{1}{d_C^{out}} + a_{DA}\frac{1}{d_D^{out}} + a_{EA}\frac{1}{d_E^{out}}\right)$$
$$= \frac{1}{5}\left(0 + 0 + 1\cdot\frac{1}{4} + 0 + 0\right)$$
$$= \boxed{1/20}$$
$$PR_B' = 1/5\left(a_{AB}\frac{1}{d_A^{out}} + a_{BB}\frac{1}{d_B^{out}} + a_{CB}\frac{1}{d_C^{out}} + a_{DB}\frac{1}{d_D^{out}} + a_{EB}\frac{1}{d_E^{out}}\right)$$
$$= 1/5\left(1\cdot\frac{1}{1} + 0 + 1\cdot\frac{1}{4} + 0 + 0\right)$$
$$= 1/5\cdot(5/4) = \boxed{5/20}$$

$$PR_C' = \frac{1}{5}\left(a_{AC}\frac{1}{d_A^{out}} + a_{BC}\frac{1}{d_B^{out}} + a_{CC}\frac{1}{d_C^{out}} + a_{DC}\frac{1}{d_D^{out}} + a_{EC}\frac{1}{d_E^{out}}\right)$$

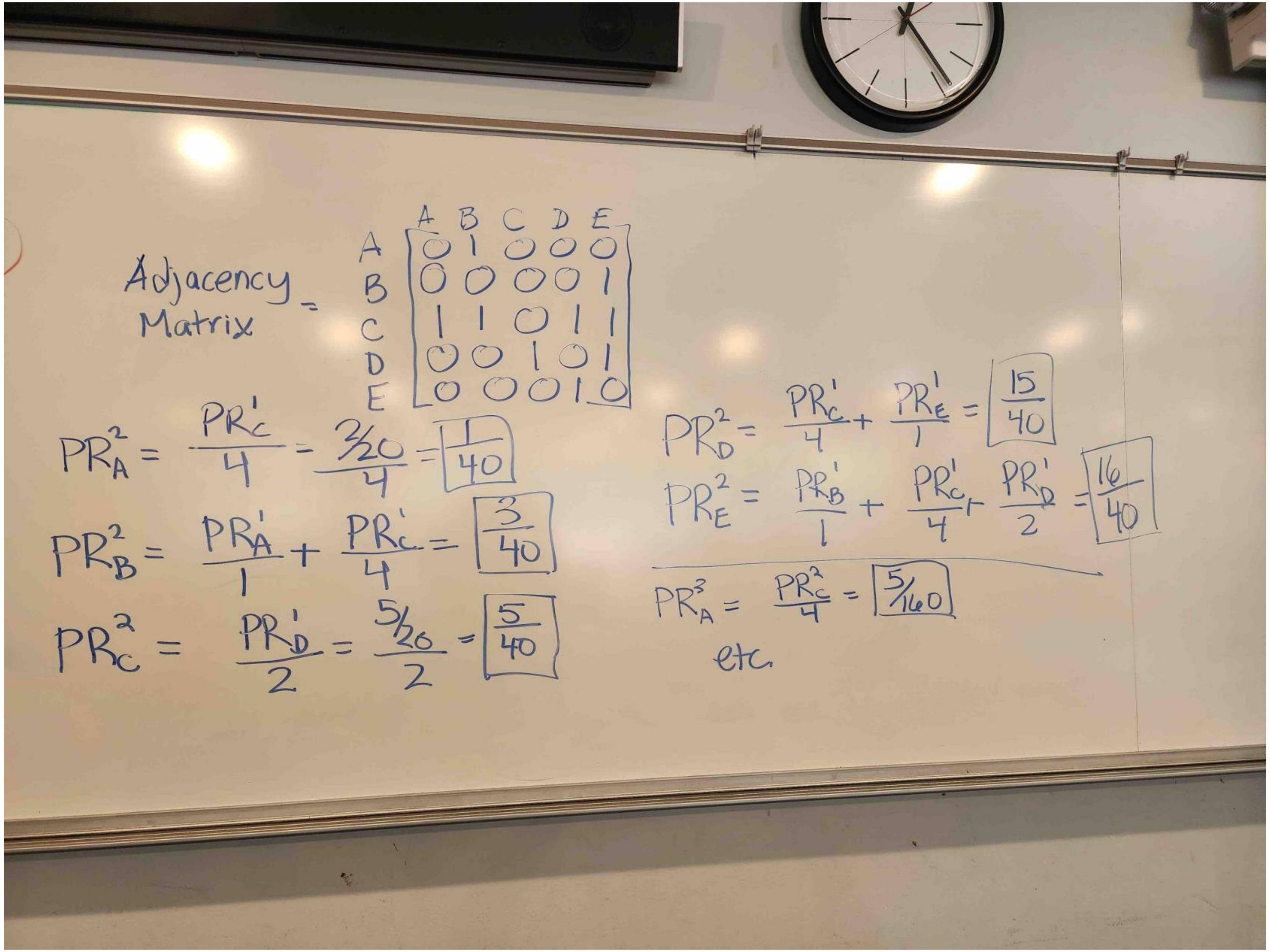
$$= \frac{1}{5}\left(0+0+0+1\cdot\frac{1}{2}+0\right) = \boxed{\frac{1}{10} = \frac{2}{20}}$$

$$PR_D' = \frac{1}{5}\left(a_{AD}\frac{1}{d_A^{out}} + a_{BD}\frac{1}{d_B^{out}} + a_{CD}\frac{1}{d_C^{out}} + a_{DD}\frac{1}{d_D^{out}} + a_{ED}\frac{1}{d_E^{out}}\right)$$

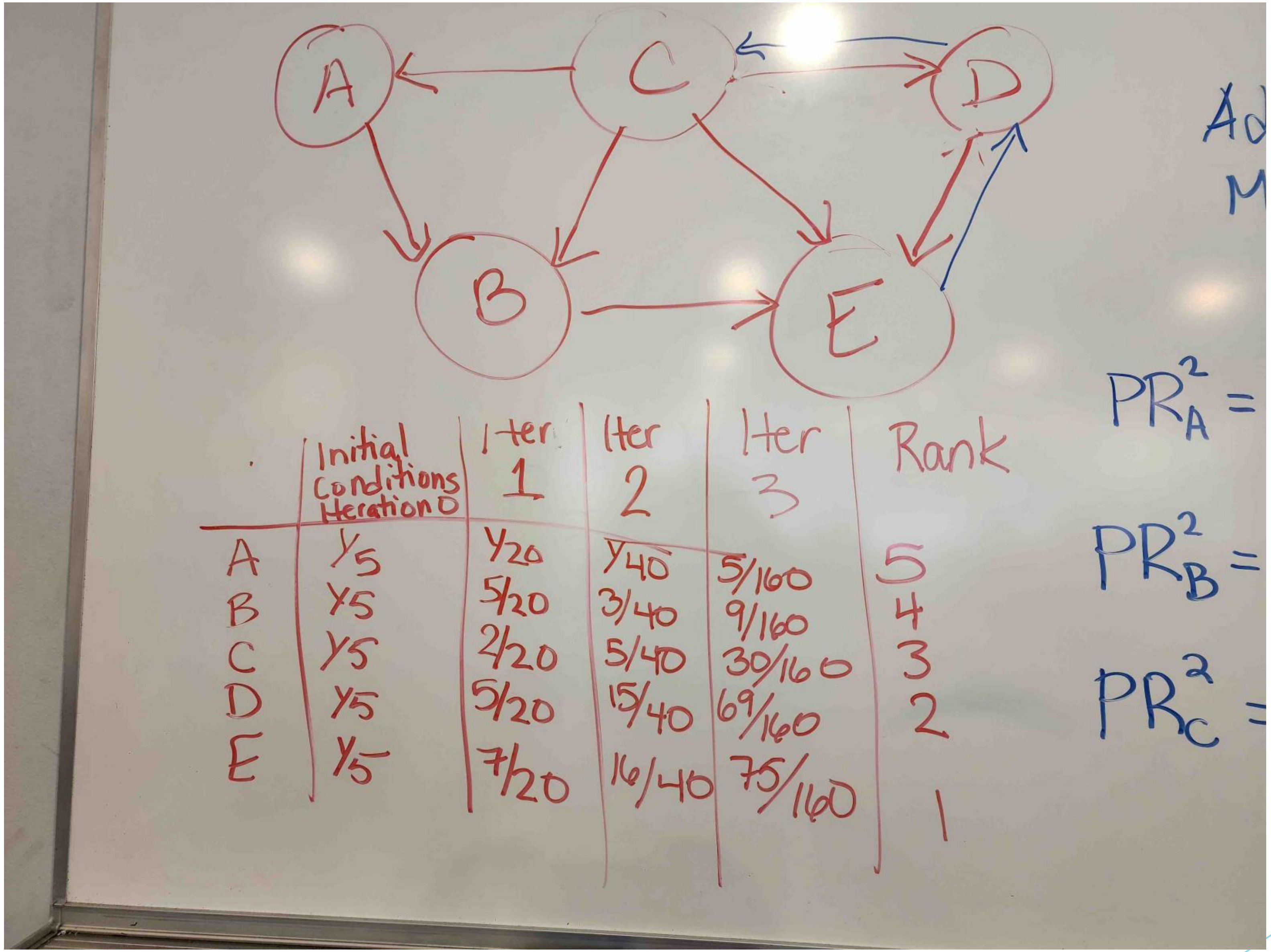
$$= \frac{1}{5}\left(0+0+1\cdot\frac{1}{4}+0+1\frac{1}{1}\right) = \frac{1}{5}\left(\frac{5}{4}\right) = \boxed{\frac{5}{20}}$$

$$PR_E' = \frac{1}{5}\left(a_{AE}\frac{1}{d_A^{out}} + a_{BE}\frac{1}{d_B^{out}} + a_{CE}\frac{1}{d_C^{out}} + a_{DE}\frac{1}{d_D^{out}} + a_{EE}\frac{1}{d_E^{out}}\right)$$

$$= \frac{1}{5}\left(0+1\cdot\frac{1}{1}+1\cdot\frac{1}{4}+1\cdot\frac{1}{2}+0\right) = \boxed{\frac{7}{20}}$$



Adjacency Matrix =
A B C D E
A 0 1 0 0 0
B 0 0 0 0 1
C 1 1 0 1 1
D 0 0 1 0 1
E 0 0 0 1 0
PR_A^2 = PR_C^1 / 4 = (2/20) / 4 = 1/40
PR_B^2 = PR_A^1 / 1 + PR_C^1 / 4 = 3/40
PR_C^2 = PR_D^1 / 2 = (5/20) / 2 = 5/40
PR_D^2 = PR_C^1 / 4 + PR_E^1 / 1 = 15/40
PR_E^2 = PR_B^1 / 1 + PR_C^1 / 4 + PR_D^1 / 2 = 16/40
PR_A^3 = PR_C^2 / 4 = 5/160
etc.


A
B
C
D
E
Initial Conditions Iteration 0
Iter 1
Iter 2
Iter 3
Rank
A 1/5 1/20 1/40 5/160 5
B 1/5 5/20 3/40 9/160 4
C 1/5 2/20 5/40 30/160 3
D 1/5 5/20 15/40 69/160 2
E 1/5 7/20 16/40 75/160 1
Ad
M
$PR_A^2 =$
$PR_B^2 =$
$PR_C^2 =$

Node Strength (for weighted graphs)

For an undirected network, the W matrix defines the strength s_i of a node as the sum of the weights of the edges incident on a node as follows:

$$s_i = \sum_{j \in \mathbb{N}} w_{ij} \tag{3.12}$$

If the network is directed, there are two factors that determine the node strength: the quantity of incoming (s_i^{in}) and outgoing (s_i^{out}) weighted edges:

$$\begin{aligned} s_i^{in} &= \sum_j w_{ji} \\ s_i^{out} &= \sum_j w_{ij} \\ s_i &= \sum_{j \in \mathbb{N}} w_{ij} \end{aligned} \tag{3.13}$$

Total strength is finally defined as:

$$s_i = s_i^{in} + s_i^{out} \tag{3.14}$$

Node Strength

- Similar to degree, can obtain the strength distribution P(s)
- In case of directed networks, two distributions
  - $P(s^{in})$
  - $P(s^{out})$
- Nodes with significantly greater strength than the rest of the network's nodes are referred to as hubs

Where might we use weighted graphs?

Recall Node's Neighborhood

- A nodes neighborhood includes the group of nodes that it is connected to

$$N_i(G) = \{j : g_{ij} = 1\}$$

- Studying the neighborhood aids in identifying nodes or communities of nodes that share specific characteristics
- When individuals are motivated to emulate the behavior of their neighbors, cascading effects can occur
  - Can be observed when a new behavior initiates with a small group of early adopters and then spreads radially outward through the network

Neighborhood Degree Sequence

The neighborhood degree sequence for the node i, s_i, is derived from:

$$s_i = \left\{k_1^i, k_2^i k_3^i \ldots, k_n^i\right\} \tag{3.19}$$

with k_i equal to the degree of the nodes to which i is connected. Being a sequence of degrees:

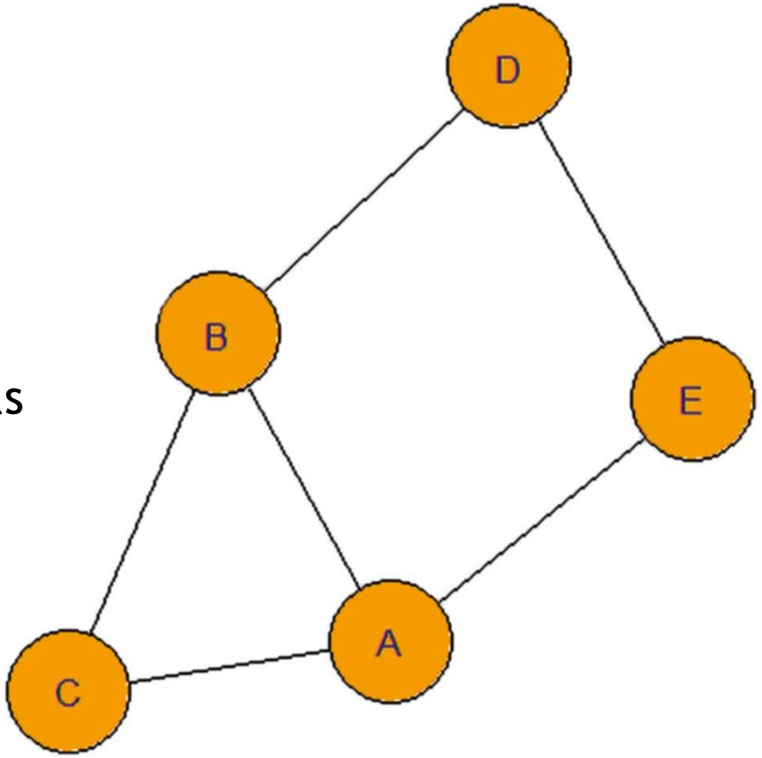
$$k_1^i \le k_2^i \le k_3^i \ldots \le k_n^i \tag{3.20}$$

Neighborhood Degree Sequence

- Analysis
  - May help in examining variance of neighborhood degree sequences
    - Context of hierarchical complexity
  - May analyze neighbors of same degree to determine how similar they are

Example Neighborhood Degree Sequence

Fig. 3.4 Graph example



Extra Credit #4: List degree sequences for each node

Section 1: in class only
Sections 2 and 3: submit to canvas by 9/25