Exam Review

CS 579 Online Social Network Analysis

Dr. Cindy Hood
10/7/25

Homework Assignments

- HW #4 – Chicago Community Areas + Census Data
  - You may work in groups up to 4 students (no exceptions) on this hw
  - Assigned this week
- Please contact TAs with questions

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

Exam 1

- 4 or 5 questions
  - Format tbd
- Any topic covered through 9/30 class is fair game
  - No GIS
  - No Census Data
- Sheet of Formulas provided
- Closed books
- Closed notes
- No electronics
  - All electronics will be put in front of class, no exceptions
  - Your exam will be collected if you use electronics during the exam
- No talking once exam is handed out
  - Your exam will be collected immediately if you talk

HW 4

HW #4 Using census data to explore Chicago Community Areas (CAs)

Each student has been assigned a Community Area (1-77) to explore. You will use data from census.gov to do an analysis of your CA using census data from the ACS (block group level) and decennial census (block level). Depending on the number of block groups in your CA, you may consider the broader geographic neighborhood of your CA as well (see instructions below). The deliverables are:

1. A data-based characterization of your CA including the recent history (back to 2010) of your CA
2. A description of the data (i.e. variables) that you have chosen to use as the basis of your study along with a discussion of why you chose these variables and any preliminary analysis you did to narrow the data. You will choose at least 6 variables that are available at the block group level (i.e. from the ACS).

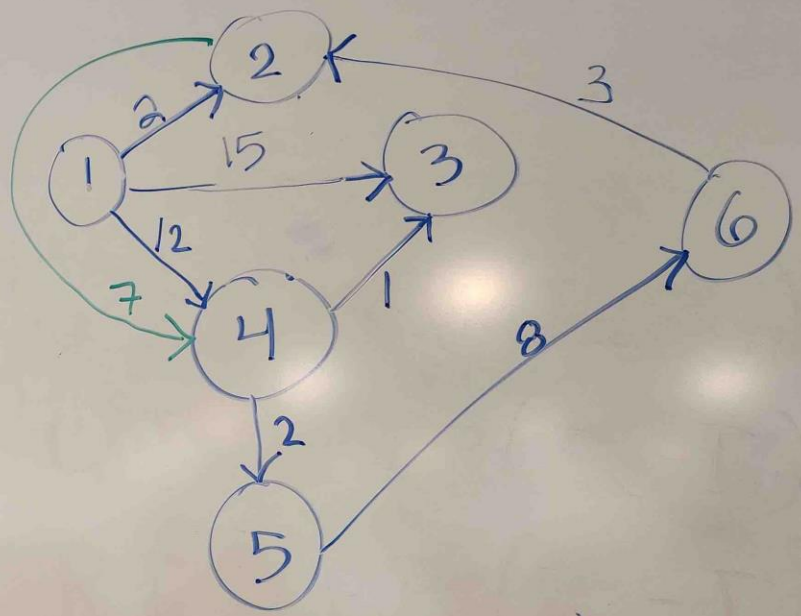
HW 4

3. An analysis of similarity of the block groups that comprise your CA. If your CA has less than 60 block groups, you will add the block groups of geographic neighbors to get at least 60 block groups. This analysis should include modeling the block groups as a network.
    a. You will describe the modeling and analysis you did, stating assumptions and justifying decisions.
    b. You will provide data-based arguments including visualizations to support why your CA is a community.
    c. You will provide data-based arguments including visualizations to support why your CA is not a community.
    d. You will provide a proposal for alternative community(ies). If your CA has 60 or more block groups, this will be a proposal for organization of communities within your CA. If your CA has <60 block groups, you will propose an organization of block groups from your CA and neighbors (resulting in analysis of 60 or more). This proposal should include visualizations.
4. If you are working in a team, there is an additional step where you will put all your data together and propose an organization of communities. Each team member must have a unique CA so if your team has duplicates, please email me asap.
5. The above should be compiled into a report that includes citations and transcripts of any AI assistance. You will submit a pdf and code.

Student Questions for Exam Prep

- Examples of Betweenness Centrality and Closeness Centrality on weighted graph
- Convergence of Page Rank calculation

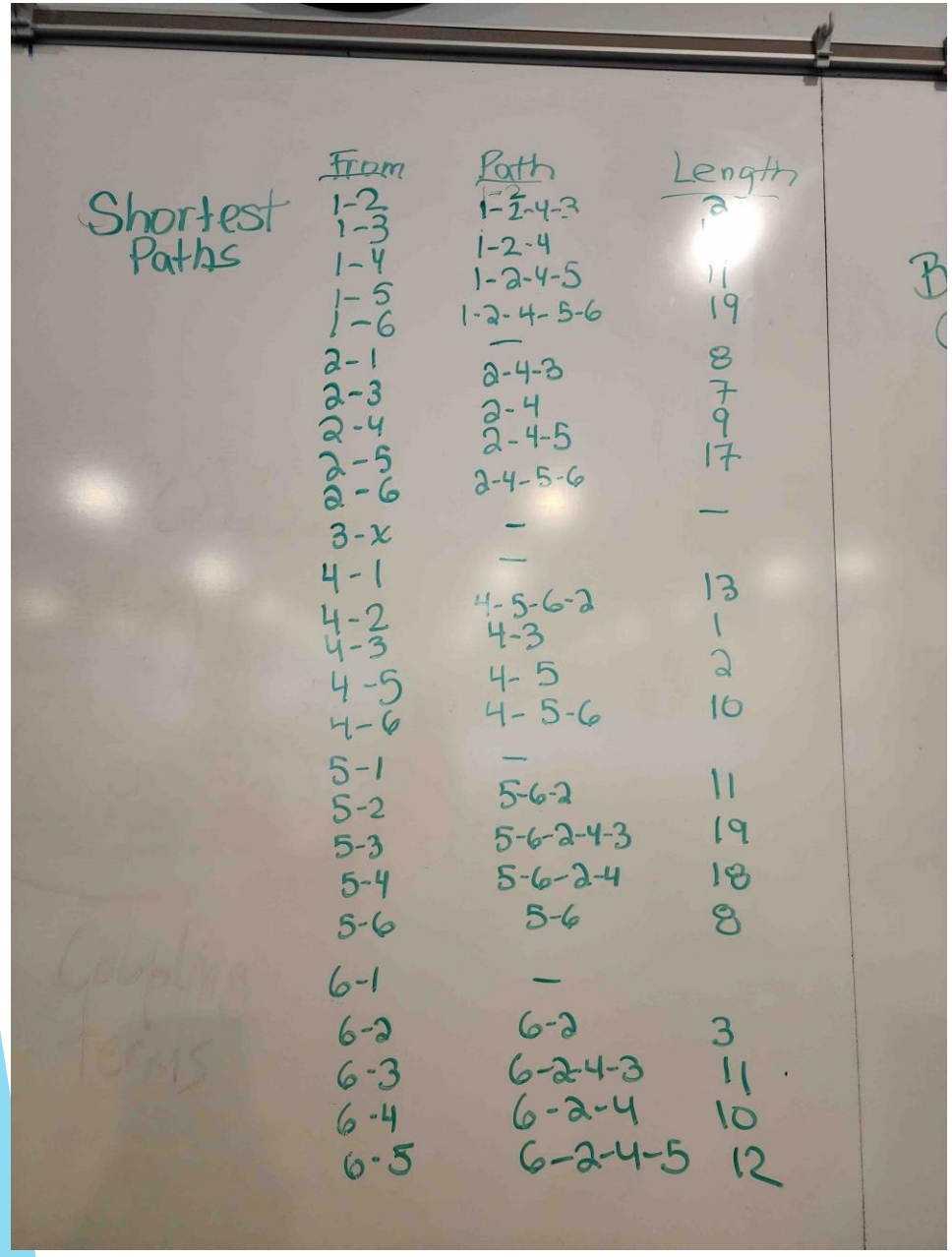
$$G = \begin{bmatrix} 0 & 2 & 15 & 12 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 \\ 0 & 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Graph is connected

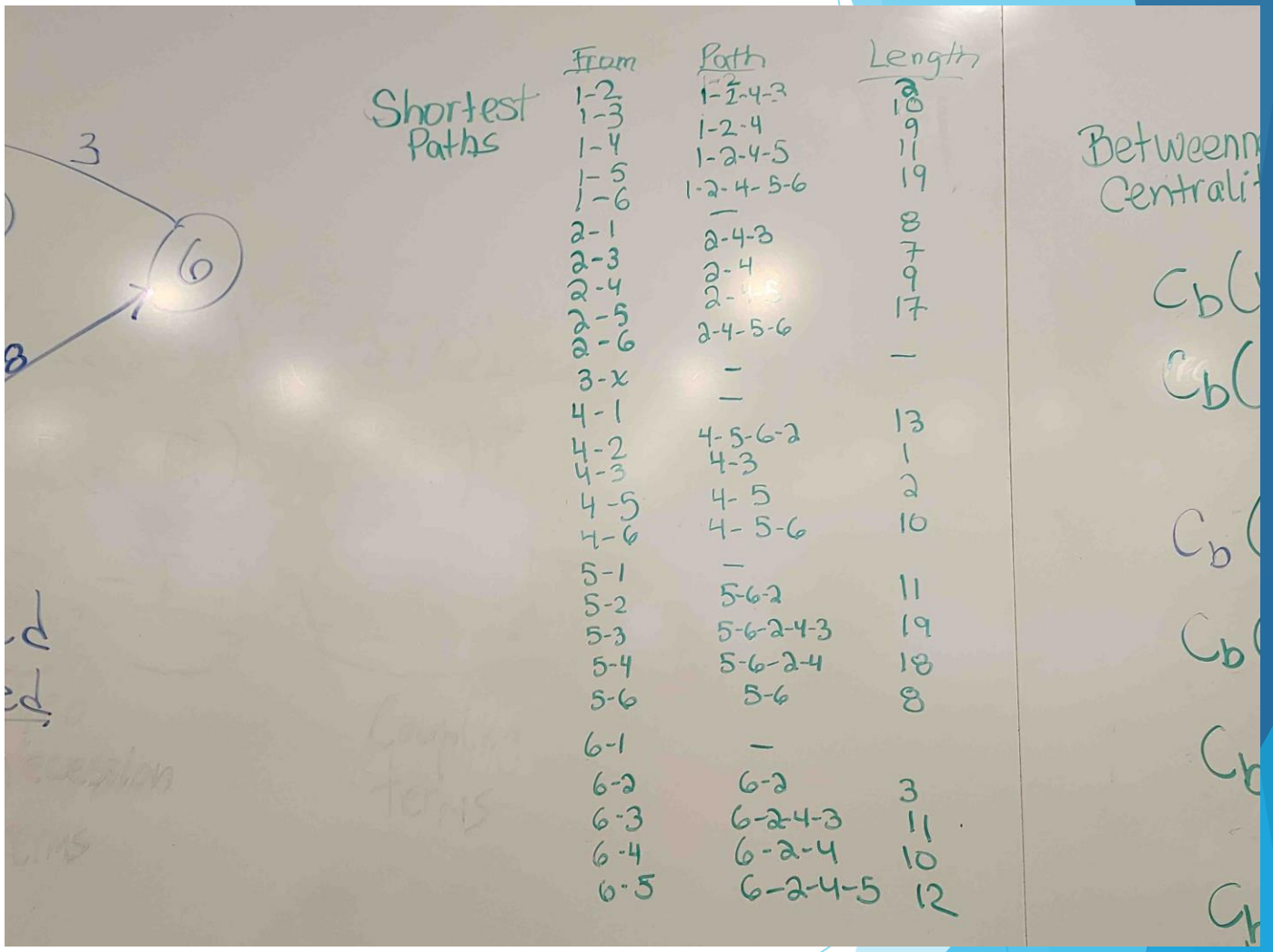
Weakly connected

Diameter = 19



Shortest Paths

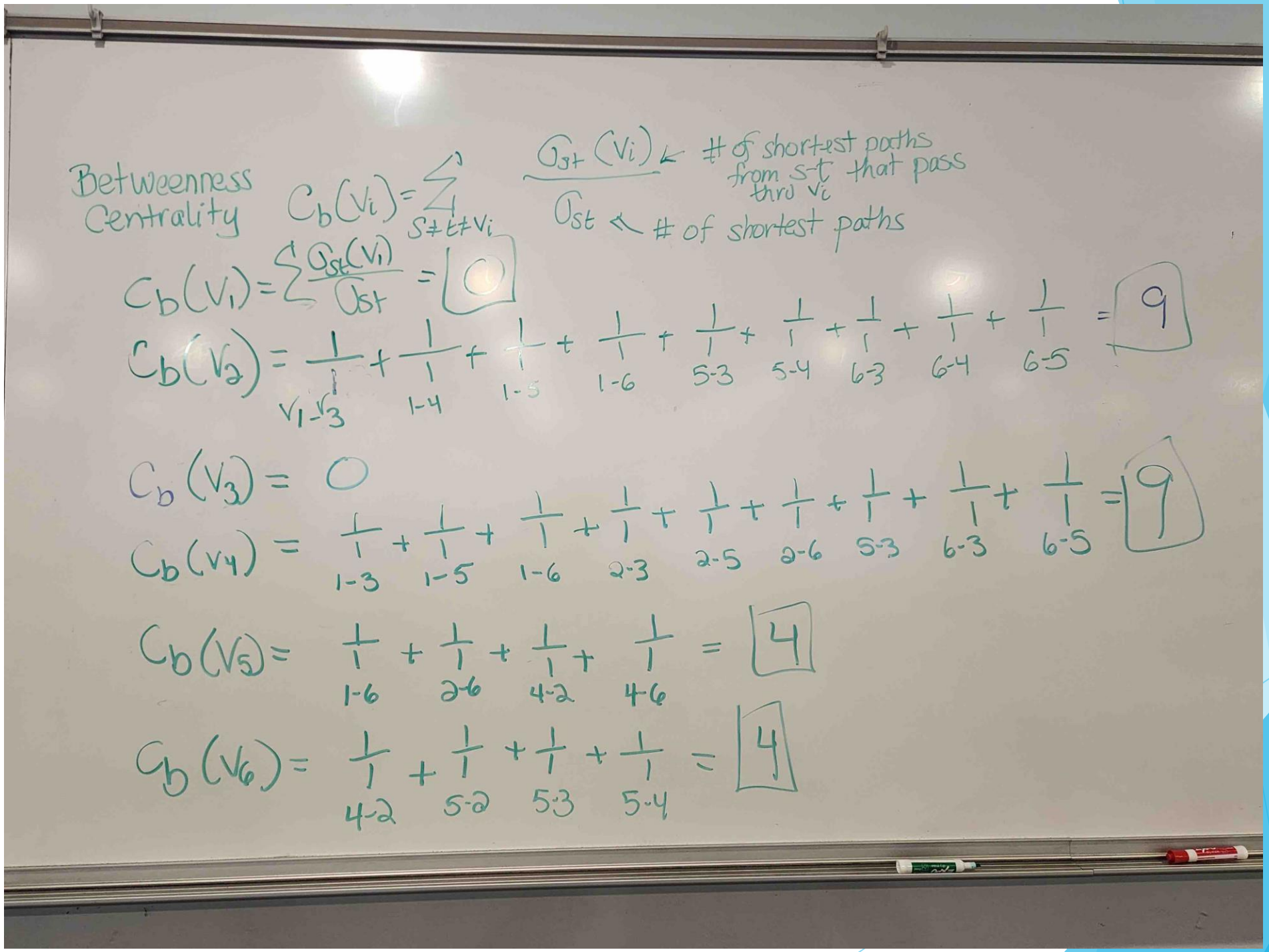
From	Path	Length
1-2	1-2	2
1-3	1-2-4-3	
1-4	1-2-4	
1-5	1-2-4-5	11
1-6	1-2-4-5-6	19
2-1	—	
2-3	2-4-3	8
2-4	2-4	7
2-5	2-4-5	9
2-6	2-4-5-6	17
3-x	—	—
4-1	—	
4-2	4-5-6-2	13
4-3	4-3	1
4-5	4-5	2
4-6	4-5-6	10
5-1	—	
5-2	5-6-2	11
5-3	5-6-2-4-3	19
5-4	5-6-2-4	18
5-6	5-6	8
6-1	—	
6-2	6-2	3
6-3	6-2-4-3	11
6-4	6-2-4	10
6-5	6-2-4-5	12

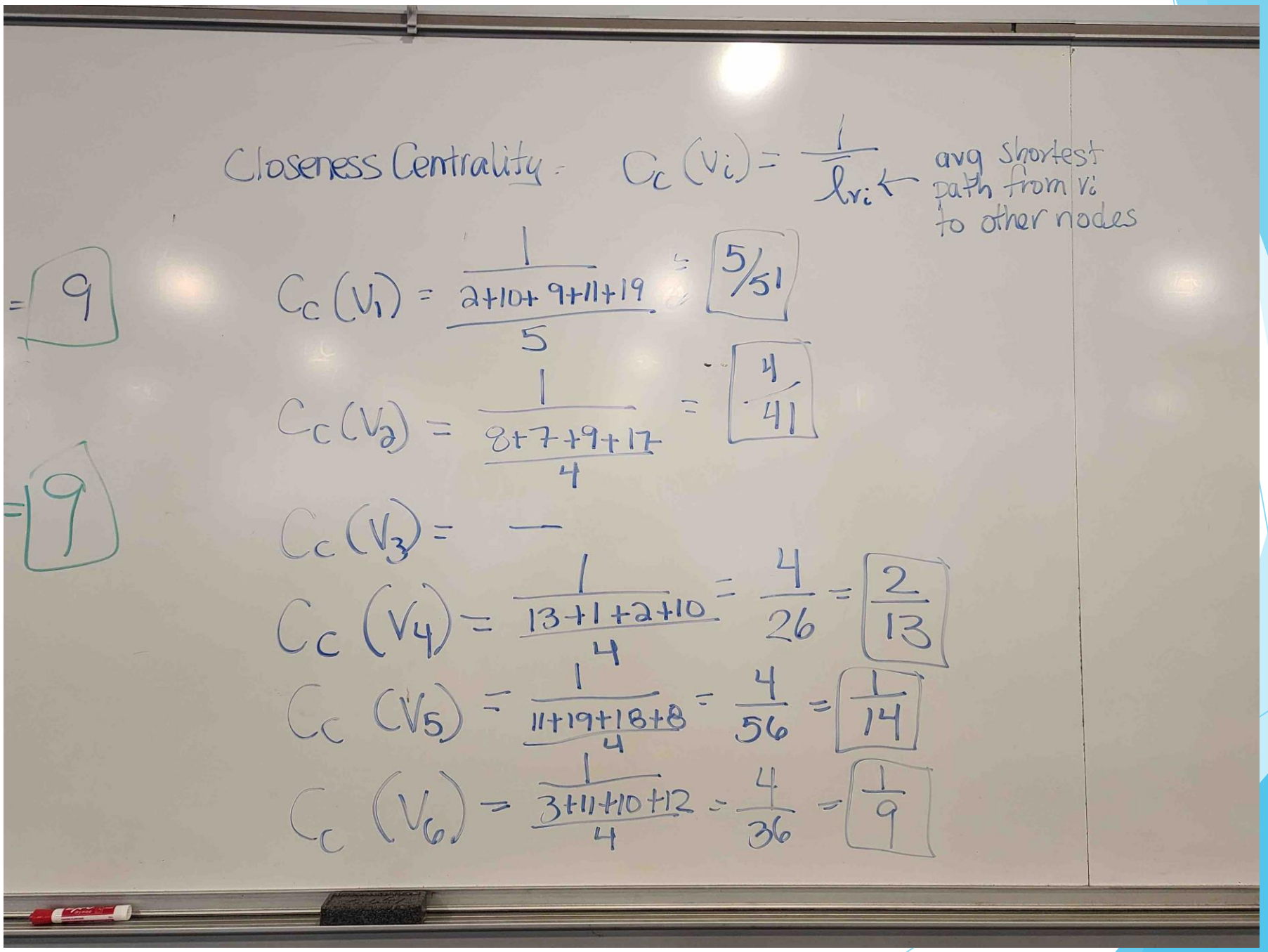


Shortest Paths

From	Path	Length
1-2	1-2	2
1-3	1-2-4-3	10
1-4	1-2-4	9
1-5	1-2-4-5	11
1-6	1-2-4-5-6	19
2-1	—	
2-3	2-4-3	8
2-4	2-4	7
2-5	2-4-5	9
2-6	2-4-5-6	17
3-x	—	—
4-1	—	
4-2	4-5-6-2	13
4-3	4-3	1
4-5	4-5	2
4-6	4-5-6	10
5-1	—	
5-2	5-6-2	11
5-3	5-6-2-4-3	19
5-4	5-6-2-4	18
5-6	5-6	8
6-1	—	
6-2	6-2	3
6-3	6-2-4-3	11
6-4	6-2-4	10
6-5	6-2-4-5	12

Two photos of same thing due to glare obscuring parts of each


Betweenness Centrality
$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$
$\sigma_{st}(v_i)$ ← # of shortest paths from s-t that pass thru v_i
σ_{st} ← # of shortest paths
$C_b(v_1) = \sum \frac{\sigma_{st}(v_1)}{\sigma_{st}} = 0$
$C_b(v_2) = \frac{1}{1}_{v_1\text{-}v_3} + \frac{1}{1}_{1\text{-}4} + \frac{1}{1}_{1\text{-}5} + \frac{1}{1}_{1\text{-}6} + \frac{1}{1}_{5\text{-}3} + \frac{1}{1}_{5\text{-}4} + \frac{1}{1}_{6\text{-}3} + \frac{1}{1}_{6\text{-}4} + \frac{1}{1}_{6\text{-}5} = 9$
$C_b(v_3) = 0$
$C_b(v_4) = \frac{1}{1}_{1\text{-}3} + \frac{1}{1}_{1\text{-}5} + \frac{1}{1}_{1\text{-}6} + \frac{1}{1}_{2\text{-}3} + \frac{1}{1}_{2\text{-}5} + \frac{1}{1}_{2\text{-}6} + \frac{1}{1}_{5\text{-}3} + \frac{1}{1}_{6\text{-}3} + \frac{1}{1}_{6\text{-}5} = 9$
$C_b(v_5) = \frac{1}{1}_{1\text{-}6} + \frac{1}{1}_{2\text{-}6} + \frac{1}{1}_{4\text{-}2} + \frac{1}{1}_{4\text{-}6} = 4$
$C_b(v_6) = \frac{1}{1}_{4\text{-}2} + \frac{1}{1}_{5\text{-}2} + \frac{1}{1}_{5\text{-}3} + \frac{1}{1}_{5\text{-}4} = 4$


Closeness Centrality - $C_c(v_i) = \frac{1}{l_{v_i}}$ ← avg shortest path from v_i to other nodes
$C_c(V_1) = \frac{1}{\frac{2+10+9+11+19}{5}} = \frac{5}{51}$
$C_c(V_2) = \frac{1}{\frac{8+7+9+17}{4}} = \frac{4}{41}$
$C_c(V_3) =$ —
$C_c(V_4) = \frac{1}{\frac{13+1+2+10}{4}} = \frac{4}{26} = \frac{2}{13}$
$C_c(V_5) = \frac{1}{\frac{11+19+18+8}{4}} = \frac{4}{56} = \frac{1}{14}$
$C_c(V_6) = \frac{1}{\frac{3+11+10+12}{4}} = \frac{4}{36} = \frac{1}{9}$
= 9
= 9