Exam Review

CS 579 Online Social Network Analysis

Dr. Cindy Hood

11/25/25

Remaining Exam and Deliverables

- HW #5 posted
  - Problems that will help you prepare for Exam 2
  - Due 11/24 (~~No late days~~)
    - Solutions posted soon
    - HW accepted until solutions posted
- Exam 2
  - Cumulative but will focus on material covered after Exam 1
  - 12/2
- Final project poster presentation/video (online students)
  - 12/4
  - Section 2 students please fill out survey if you haven't done so
- Final project report
  - Due by midnight on 12/10

Final Project Presentation

- Poster session
  - Outside of SB 104 during class time 11:25am – 12:40pm 12/4
    - Will start before class and continue after
    - I will send out an email for you to schedule a timeslot
    - Plan to be present for the entire class period
      - For you to complete your reviews
      - For others to complete their review of your poster
  - I have mounting putty/tape to stick your posters to the wall there
  - Pizza will be served
- Online video
  - Must be posted to google folder (I will send link) by 11:00am 12/4

Poster/Presentation Requirements

- Must use printed poster (on campus) or slides (online)
- Each team member must present and be visible for online presentation
- No more than 5 minutes
- Content
  - Title and names of all team members
  - Intro – What did you do
  - Design of project
  - Execution of project
    - Data utilized
  - Results
  - Discussion of results and conclusion
    - What worked/what didn't work?
    - How can you evaluate your project?
      - What did you do?
      - What could you do?
    - What surprised you?
    - What would you do if you had more time?

Focus on the most interesting parts of the project

Poster Creation

- Your poster should be 24” x 36”
- Powerpoint
  - https://designshack.net/articles/business-articles/how-to-make-a-poster-in-powerpoint/

Poster Printing

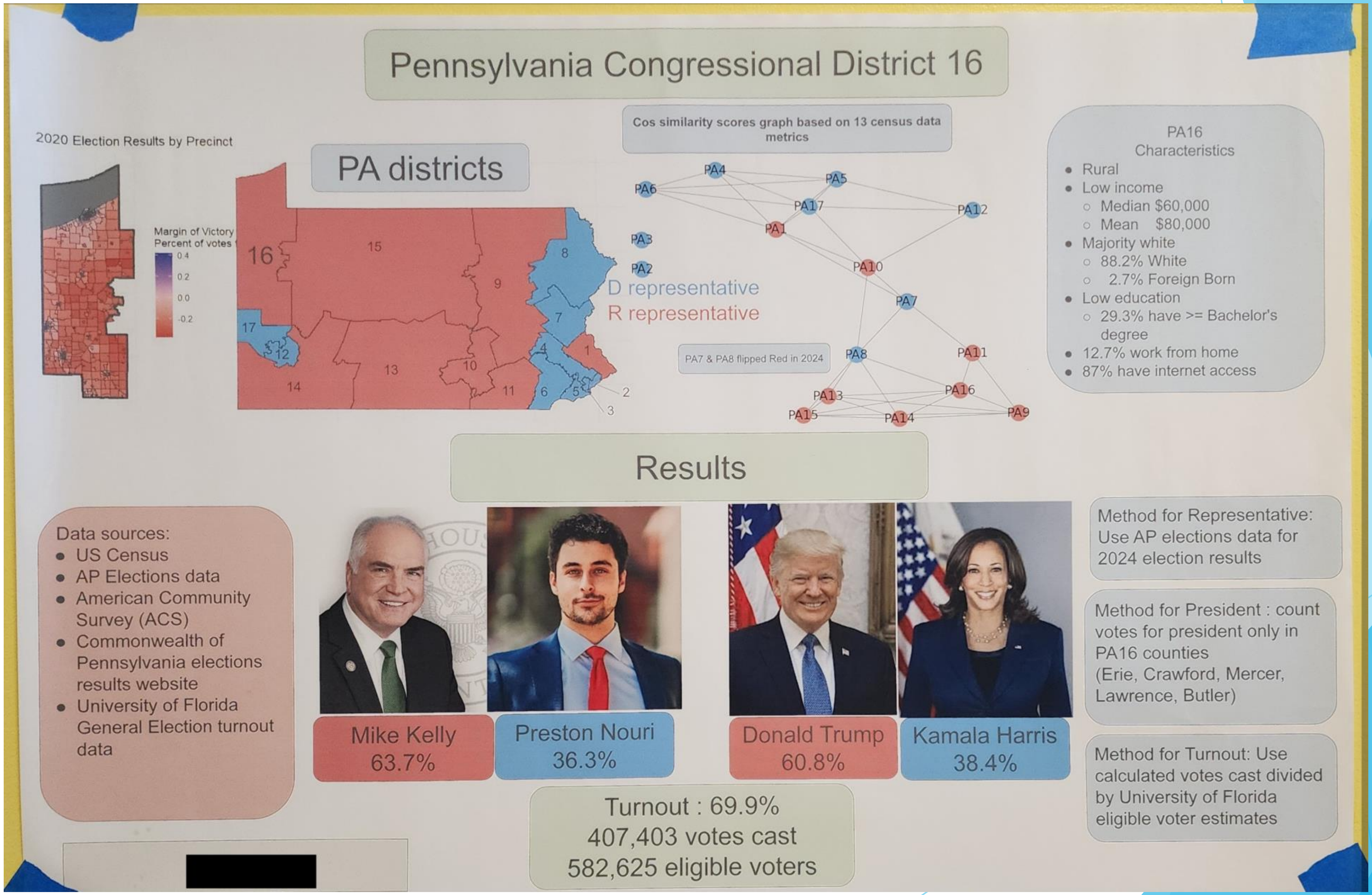
How to Print to a Plotter

Using one of the HP plotters on the Mies Campus (located in Crown and Tech North) takes some additional preparation to get a quality plot and avoid waisting your printing funds. The OTS staff have produced a series of four videos to help you learn how to plot properly. Please review each of these videos before attempting to use a plotter.

1. Plotter Introduction – Preparing Your File
2. Z6200 Plotter Instructional Video

https://www.iit.edu/ots/services/student-printing

Example from last fall



Project presentation review assignments

- All reviews must be completed on 12/4
  - On campus during class time
  - Online

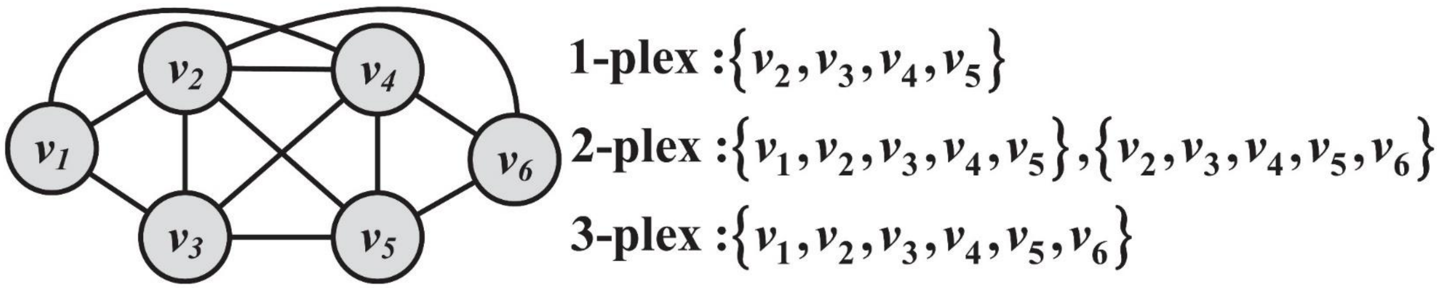
II. Relaxing Cliques

- ***k*-plex**: a set of vertices V in which we have

$$d_v \geq |V| - k, \forall v \in V$$

- d_v is the degree of v in the induced subgraph
  - Number of nodes from V that are connected to v

- Clique of size k is a 1-plex
- Finding the maximum k-plex: **NP-hard**
  - In practice, relatively easier due to smaller search space.



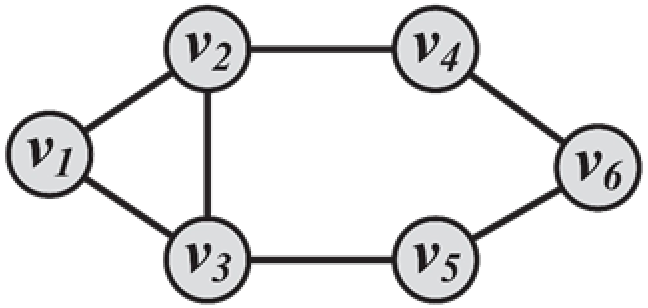
Maximal k-plexes

Special Subgraphs

1. ***k*-Clique**: a **maximal** subgraph in which the largest shortest path distance between any nodes is less than or equal to k

2. ***k*-Club**: follows the same definition as a k-clique
   - **Additional Constraint:** nodes on the shortest paths should be part of the subgraph (i.e., diameter)

3. ***k*-Clan:** a k-clique where for all shortest paths within the subgraph the distance is equal or less than k.
   - All k-clans are k-cliques, but not vice versa.



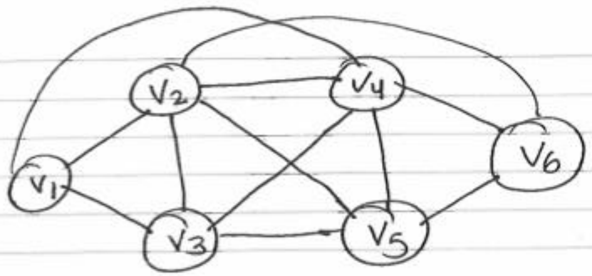
2-cliques : $\{v_1, v_2, v_3, v_4, v_5\}, \{v_2, v_3, v_4, v_5, v_6\}$

2-clubs : $\{v_2, v_3, v_4, v_5, v_6\}, \{v_1, v_2, v_3, v_4\}, \{v_1, v_2, v_3, v_5\}$

2-clans : $\{v_2, v_3, v_4, v_5, v_6\}$

K-Plex

(1)



A node is a member of a k-plex of size n if it has direct ties to n-k members

We know $\{v_2, v_3, v_4, v_5\}$ are a clique since they are fully connected.

We also know that $\{v_2, v_3, v_4, v_5\}$ are members of a 1-plex of size 4. Each node has ties to $\underbrace{4}_{n} - \underbrace{1}_{k} = 3$ nodes.

Lets add a node v_1. Then $n=5$ $\{v_1, v_2, v_3, v_4, v_5\}$ Is this a 1-plex? If it was then each node would need to have direct ties to $n-k = 5-1 = 4$ members. The degree of $v_1 = 3$ so These 5 nodes are not a 1-plex.

How about a 2-plex? Does each node have direct ties to $n-k = 5-2 = 3$ members? Yes so $\{v_1, v_2, v_3, v_4, v_5\}$ is a 2-plex

We can make the same case to say that $\{v_2, v_3, v_4, v_5, v_6\}$ is a 2-plex

(2)

What if we add both v_1 and v_6 to the original clique, do we have a 2-plex with $\{v_1, v_2, v_3, v_4, v_5, v_6\}$

$n=6 \quad k=2$

$n-k=4$

Since neither v_1 or v_6 has direct ties with 4 other nodes, this set is not a 2-plex.

How about if we try a 3-plex for $\{v_1, v_2, v_3, v_4, v_5, v_6\}$

$n=6 \quad k=3$

$n-k=3$

Yes, each node has direct ties to at least 3 nodes in the 3-plex.

N-cliques

- The strict clique definition (maximal fully-connected sub-graph) may be too strong for many purposes.
  - It insists that every member or a sub-group have a direct tie with each and every other member.
- You can probably think of cases of "cliques" where at least some members are not so tightly or closely connected.
- There are two major ways that the "clique" definition has been "relaxed" to try to make it more helpful and general.
- One alternative is to define an actor as a member of a clique if they are connected to every other member of the group at a distance greater than one.
- Usually, the path distance two is used. This corresponds to being "a friend of a friend."
- This approach to defining sub-structures is called N-clique, where N stands for the length of the path allowed to make a connection to all other members.

https://faculty.ucr.edu/~hanneman/nettext/C11_Cliques.html

N-Clans

- The N-clique approach tends to find long and stringy groupings rather than the tight and discrete ones of the maximal approach.
- In some cases, N-cliques can be found that have a property that is probably undesirable for many purposes: it is possible for members of N-cliques to be connected by actors who are not, themselves, members of the clique.
  - For most sociological applications, this is quite troublesome.
- To overcome this problem, some analysts have suggested restricting N-cliques by insisting that the total span or path distance between any two members of an N-clique also satisfy a condition.
- The additional restriction has the effect of forcing all ties among members of an n-clique to occur <u>by way of other members of the n-clique</u>. This is the n-clan approach.

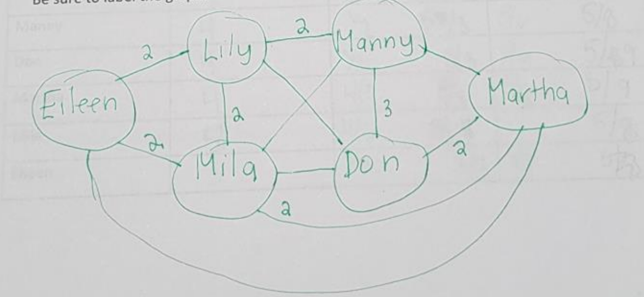
https://faculty.ucr.edu/~hanneman/nettext/C11_Cliques.html

Problem from previous exam

You are given the following dataset:

Name	Age	Political Affiliation	Causes interested in	Voted in the last election	Amount donated in the last year
Lily	67	Democrat	Anti-war, civil rights, animal rights, religious liberty	Yes	$12,000
Manny	53	Republican	Anti-war, environment, religious liberty, freedom	Yes	$50
Don	28	Libertarian	Environment, anti-tax, religious liberty, freedom	No	$400
Martha	34	Independent	Environment, anti-tax, food security	No	$750
Mila	60	Democrat	Anti-war, anti-tax, animal rights, food security	Yes	$25
Eileen	25	Independent	Civil rights, animal rights, food security	No	0

1. (20 pts) Create a unimodal, connected network graph from this dataset and draw the graph. Be sure to label the graph.



4. (20 pts) K-clique and k-plex structures provide different ways of relaxing the clique structure.
   a. (10 pts) Provide the maximal 2-cliques of the network graph in problem 1.

{Lily, Manny, Don, Martha, Mila, Eileen}

   b. (10 pts) Provide the maximal 2-plexes of the network graph in probem 1.

{Lily, Manny, Don, Martha, Mila}
{Lily, Don, Martha, Mila, Eileen}

5. (10 pts) You are going pay one of the people in the network to be an influencer. For maximum effectiveness, which person would you choose and why?

Mila

Member-Based Community Detection

- Look at node characteristics; and
- Identify nodes with similar characteristics and consider them a community

Node Characteristics

A. Degree

- Nodes with same (or similar) degrees are in one community
- Example: cliques

B. Reachability

- Nodes that are close (small shortest paths) are in one community
- Example: k-cliques, k-clubs, and k-clans

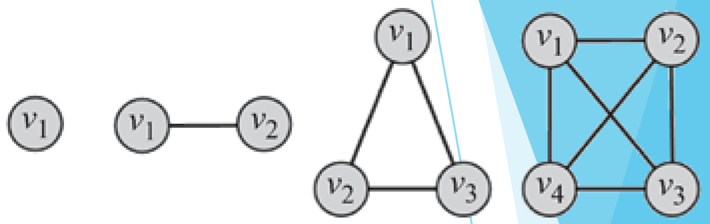
C. Similarity

- Similar nodes are in the same community

A. Node Degree

Most common subgraph searched for:

- **Clique**: a maximum complete subgraph in which all nodes inside the subgraph adjacent to each other



Find communities by searching for

1. **The maximum clique**: **the one** with the largest number of vertices, or
2. **All maximal cliques**: cliques that are not subgraphs of a larger clique; i.e., cannot be further expanded

Both problems are NP-hard

To overcome this, we can

I. Brute Force
II. Relax cliques
III. Use cliques as the core for larger communities

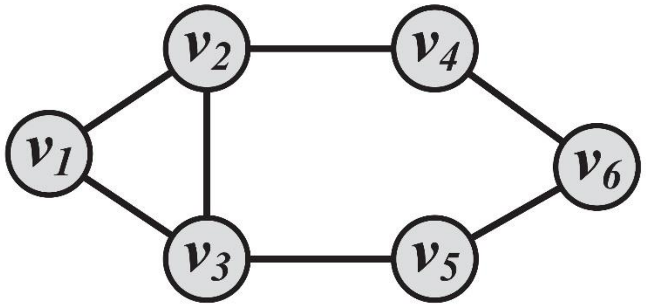
Node Similarity (Structural Equivalence)

Jaccard Similarity

$$\sigma_{\text{Jaccard}}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|}$$

Cosine similarity

$$\sigma_{\text{Cosine}}(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{\sqrt{|N(v_i)||N(v_j)|}}$$



$$\sigma_{\text{Jaccard}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{|\{v_1, v_3, v_4, v_6\}|} = 0.25$$

$$\sigma_{\text{Cosine}}(v_2, v_5) = \frac{|\{v_1, v_3, v_4\} \cap \{v_3, v_6\}|}{\sqrt{|\{v_1, v_3, v_4\}||\{v_3, v_6\}|}} = 0.40$$

Ratio Cut and Normalized Cut

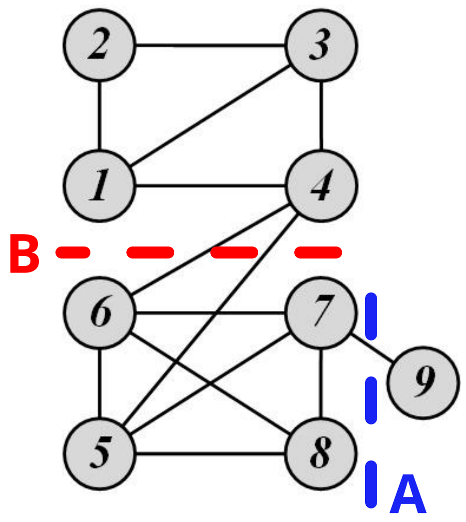
- To mitigate the min-cut problem we can change the objective function to consider community size

$$\text{Ratio Cut}(P) = \frac{1}{k}\sum_{i=1}^{k}\frac{\text{cut}(P_i,\bar{P}_i)}{|P_i|}$$

$$\text{Normalized Cut}(P) = \frac{1}{k}\sum_{i=1}^{k}\frac{\text{cut}(P_i,\bar{P}_i)}{\text{vol}(P_i)}$$

- $\bar{P}_i = V - P_i$ is the complement cut set
- $cut(P_i, \bar{P}_i)$ is the size of the cut
- $vol(P_i) = \sum_{v \in P_i} d_v$

Ratio Cut & Normalized Cut: Example



For Cut A

$$\text{Ratio Cut}(\{1,2,3,4,5,6,7,8\},\{9\}) = \frac{1}{2}\left(\frac{1}{1}+\frac{1}{8}\right) = 9/16 = 0.56$$

$$\text{Normalized Cut}(\{1,2,3,4,5,6,7,8\},\{9\}) = \frac{1}{2}\left(\frac{1}{1}+\frac{1}{27}\right) = 14/27 = 0.52$$

For Cut B

$$\text{Ratio Cut}(\{1,2,3,4\},\{5,6,7,8,9\}) = \frac{1}{2}\left(\frac{2}{4}+\frac{2}{5}\right) = 9/20 = 0.45 < 0.56$$

$$\text{Normalized Cut}(\{1,2,3,4\},\{5,6,7,8,9\}) = \frac{1}{2}\left(\frac{2}{12}+\frac{2}{16}\right) = 7/48 = 0.15 < 0.52$$

Both ratio cut and normalized cut prefer a balanced partition