Community Analysis

CS 579 Online Social Network Analysis

Dr. Cindy Hood

10/30/25

Exam 1

- Most graded and posted
  - Still finalizing section 3

- Preliminary statistics – will post final statistics when calculated
  - Average in the low 70s
- Will post solutions
  - Hold exam questions until after you view the solutions
    - Unless I made an error in calculating your score

- No curving of exam, any curving done for final grade
- But, if I had to assign grades only on exam
  - A >= 85
  - 65 >= B < 85
  - 50 >= C < 65
  - E < 50

Homework Assignments

- HW #4 – Chicago Community Areas + Census Data
  - You may work in groups up to 4 students (no exceptions) on this hw
  - Due date now 11/3

- Please contact TAs with questions

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

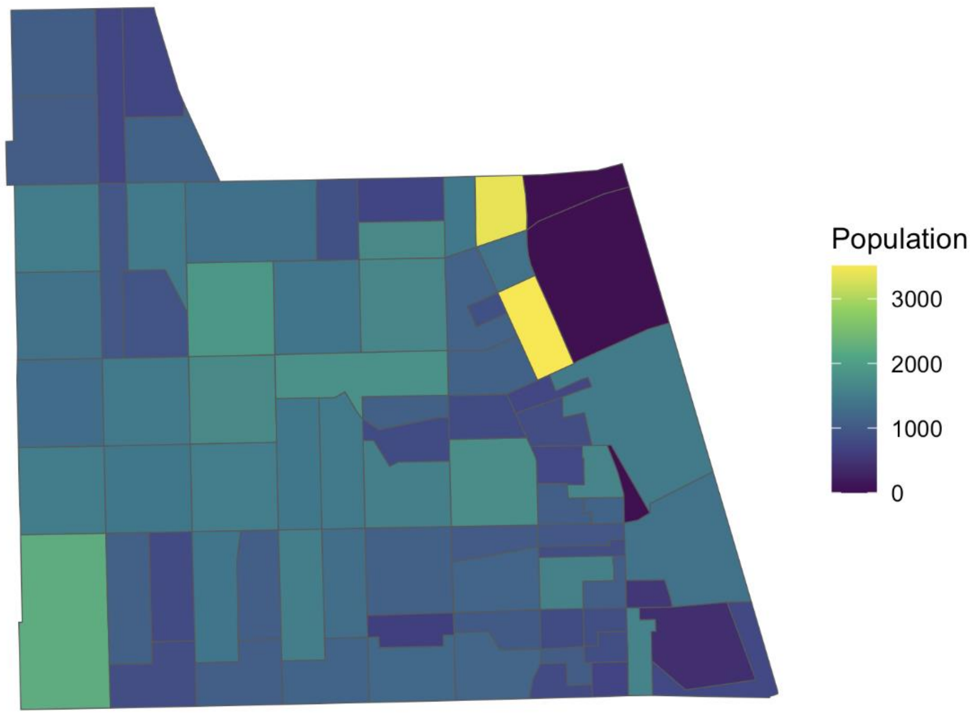
I suggest you use compare a few variables from the decennial census in 2010 and the decennial census in 2020

- Population
- Demographics

This must be done at the block group level for all of the block groups you are including in your study

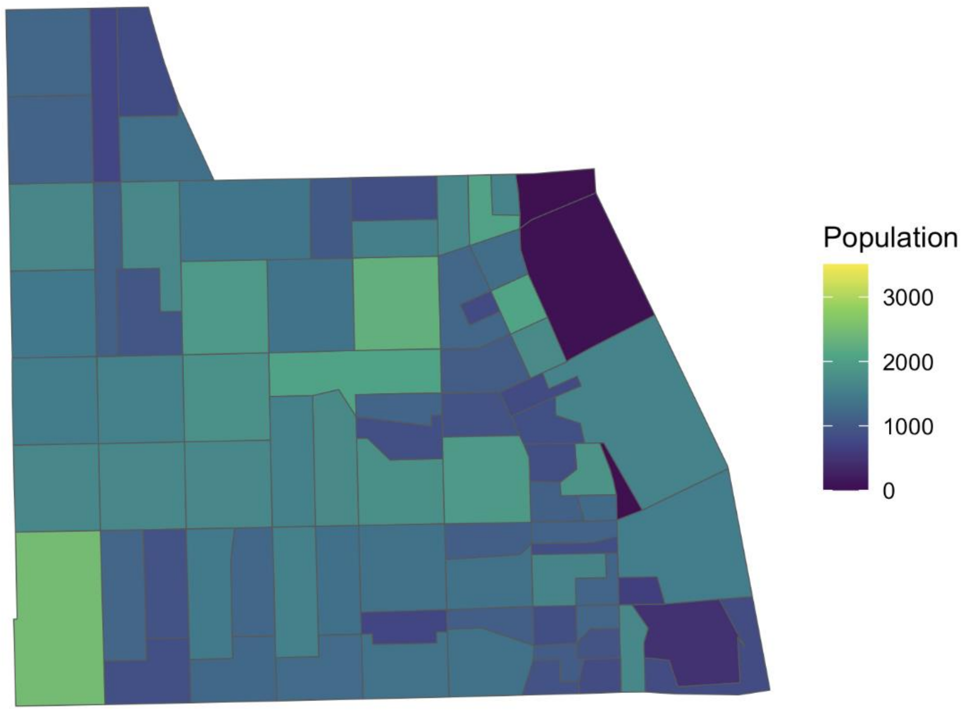
Example of what I'm looking for

Population in Lakeview CCA by Block Group, 2010



Total population = 94368. Data source: 2010 Decennial Census, US Census Bureau

Population in Lakeview CCA by Block Group, 2020



Total population = 103050. Data source: 2020 Decennial Census, US Census Bureau

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).
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.

Teams will submit one report. Be sure to describe which Cas each team member analyzed

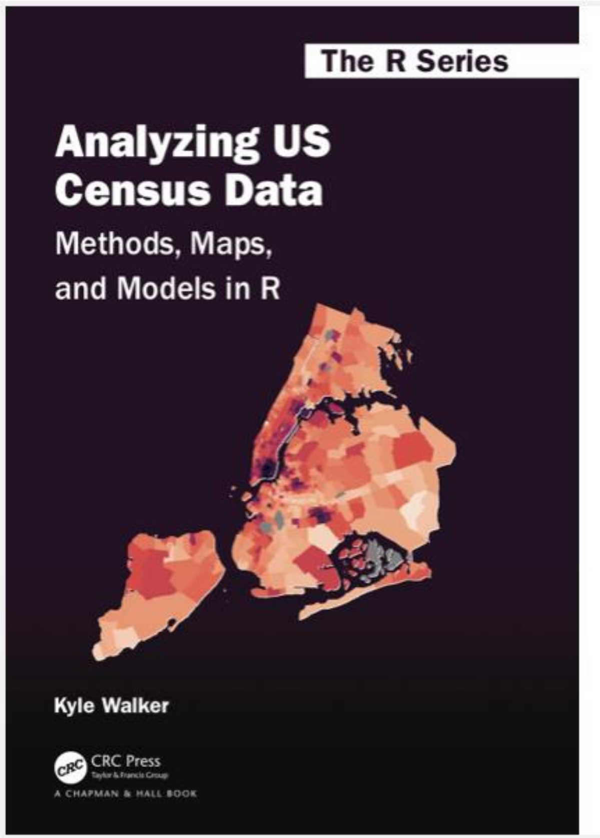
Exams and Final Project Poster Presentation

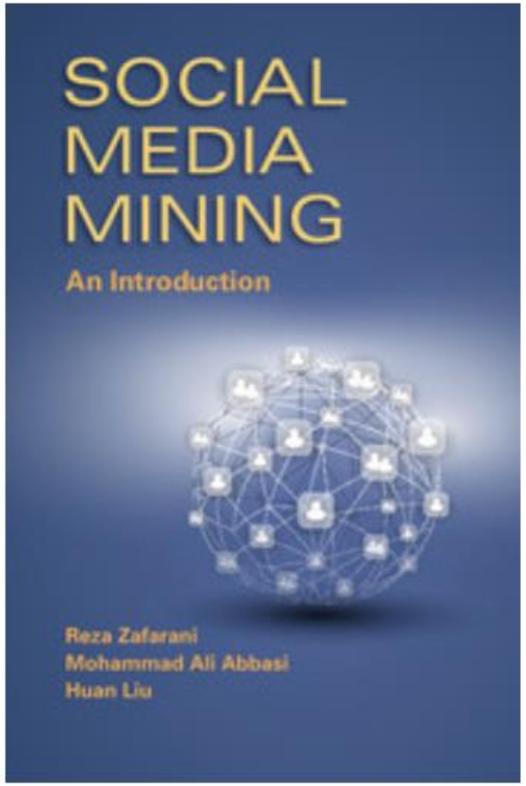
- Exam 2 – Dec 2 in class
- Final Project Poster Session – Dec 4 in class
- Online students (sections 2 and 3) will have remote options

Final Project

- Extension of HW 4
- Teams of up to 4 permitted
- Proposal due 11/8 <- Change
  - Detailed requirements posted soon

References





https://walker-data.com/census-r/

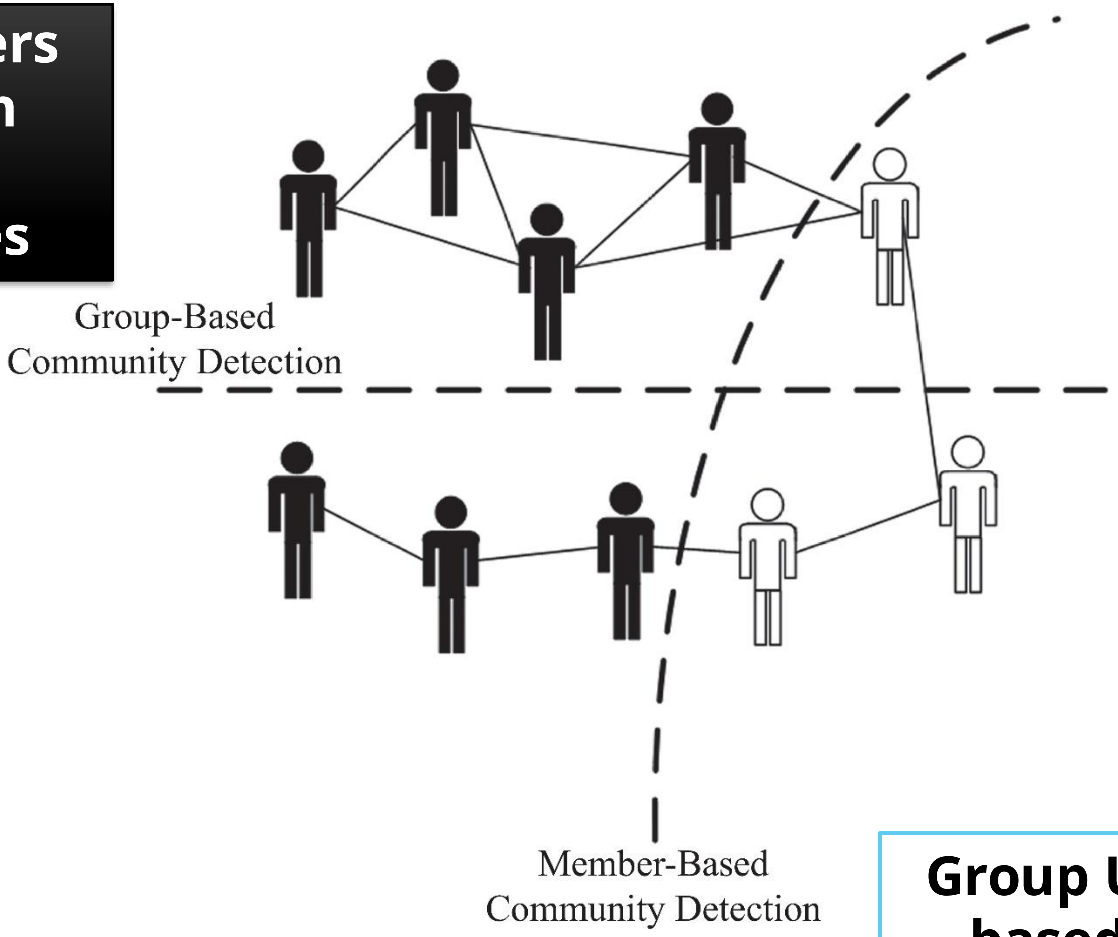
http://www.socialmediamining.info

Some additional resources

- Myatt and Johnson (2014), *Making Sense of Data I*, 2nd Edition, Wiley, ISBN: 978-1-118-40741-7
  - https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781118422007
- Networks, Crowds, and Markets: Reasoning About a Highly Connected World by David Easley and Jon Kleinberg.
  - http://www.cs.cornell.edu/home/kleinber/networks-book/
- Hanneman, Robert A. and Mark Riddle. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside
  - https://faculty.ucr.edu/~hanneman/nettext/

Community Detection Algorithms

Group Users based on Group attributes



Group Users based on Member attributes

Group-Based Community Detection

Group-Based Community Detection

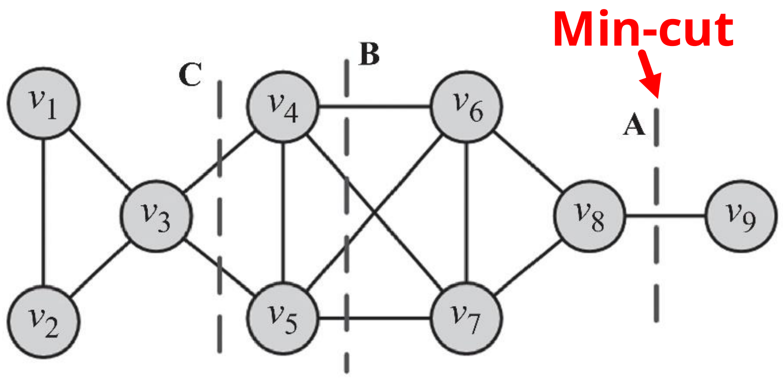
Group-based community detection: finding communities that have certain **group properties**

Group Properties:

I. **Balanced:** Spectral clustering
II. **Robust:** k-connected graphs
III. **Modular:** Modularity Maximization
IV. **Dense:** Quasi-cliques
V. **Hierarchical:** Hierarchical clustering

I. Balanced Communities

- Community detection can be thought of *graph clustering*
- **Graph clustering:** we cut the graph into several partitions and assume these partitions represent communities
- **Cut**: partitioning (*cut*) of the graph into two (or more) sets (*cutsets*)
  - **The size of the cut** is the number of edges that are being cut
- **Minimum cut (min-cut) problem**: find a graph partition such that the number of edges between the two sets is minimized



Min-cuts can be computed efficiently using the max-flow min-cut theorem

Min-cut often returns an imbalanced partition, with one set being a singleton

Ratio Cut and Normalized Cut

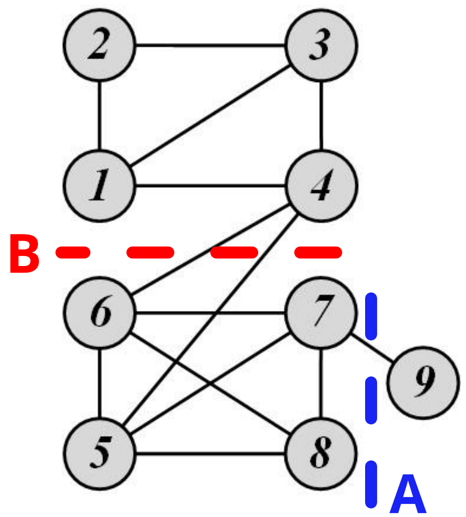
- To mitigate the min-cut problem we can change the objective function to consider community size

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

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

- $\bar{P}_i = V - P_i$ 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

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

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

For Cut B

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

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

Both ratio cut and normalized cut prefer a balanced partition

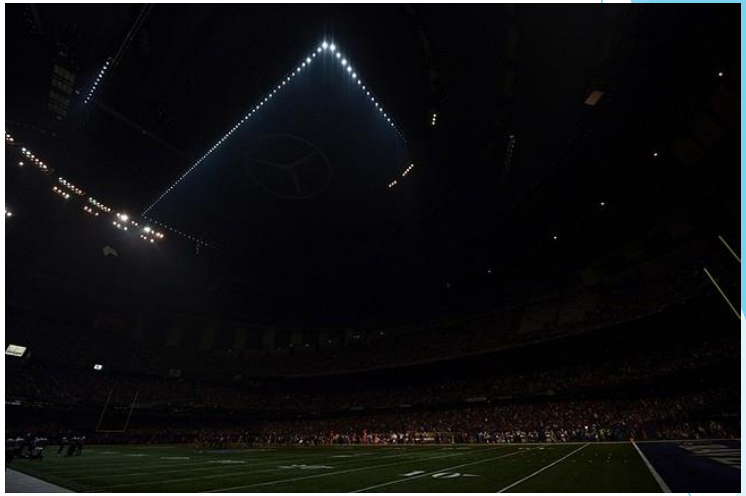
Information Diffusion

Chapter 7

Definition

In February 2013, during the third quarter of Super Bowl XLVII, a power outage stopped the game for 34 minutes.

- Oreo, a sandwich cookie company, tweeted during the outage:
  - “Power out? No Problem. You can still dunk it in the dark”.
- The tweet caught on almost immediately, reaching
  - 15,000 retweets and 20,000 likes on Facebook in less than 2 days.
- Cheap advertisement reaching a large population of individuals.
  - companies spent as much as 4 million dollars to run a 30 second ad during the super bowl.



Example of **Information Diffusion**

Information Diffusion

- Information diffusion: process by which a piece of information (knowledge) is spread and reaches individuals through interactions.
- Studied in a plethora of sciences.
- We discuss methods from
  - Sociology, epidemiology, and ethnography
  - All are useful for social media mining.
- We focus on techniques that can model information diffusion.

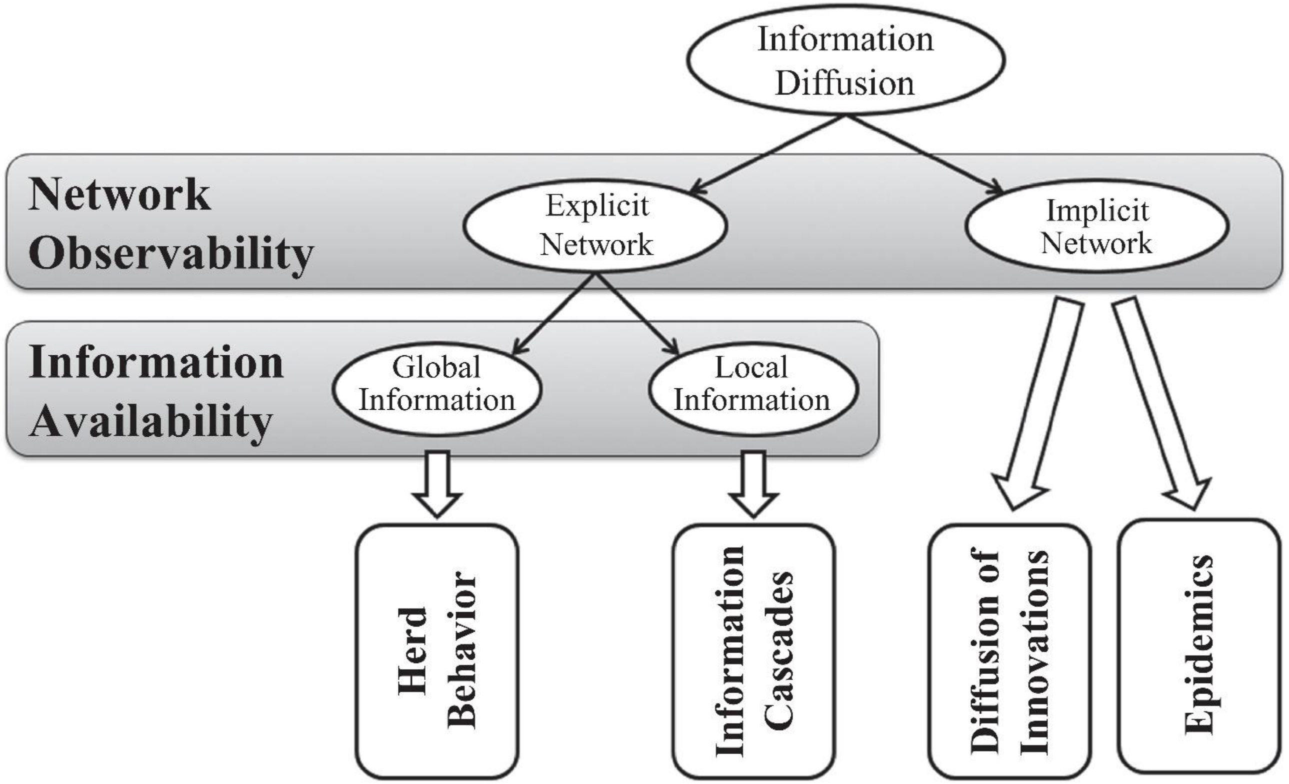
Information Diffusion

- **Sender(s).** A sender or a small set of senders that initiate the information diffusion process;

- **Receiver(s).** A receiver or a set of receivers that receive diffused information. Commonly, the set of receivers is much larger than the set of senders and can overlap with the set of senders; and

- **Medium.** This is the medium through which the diffusion takes place. For example, when a rumor is spreading, the medium can be the personal communication between individuals

Information Diffusion Types



We define the process of interfering with information diffusion by expediting, delaying, or even stopping diffusion as Intervention

Herd Behavior

- Network is observable
- Only public information is available

Herd Behavior Example

- Consider people participating in an online auction.
- In this case, individuals can observe the behavior of others by monitoring the bids that are being placed on different items.
- Individuals are connected via the auction's site where they can not only observe the bidding behaviors of others, but can also often view profiles of others to get a feel for their reputation and expertise.
- In these online auctions, it is common to observe individuals participating actively in auctions, where the item being sold might otherwise be considered unpopular.
- This is due to individuals trusting others and assuming that the high number of bids that the item has received is a strong signal of its value. In this case, Herd Behavior has taken place.

Herd Behavior: Popular Restaurant Example

- Assume you are on a trip in a metropolitan area that you are less familiar with.
- Planning for dinner, you find restaurant **A** with excellent reviews online and decide to go there.
- When arriving at **A**, you see **A** is almost empty and restaurant **B**, which is next door and serves the same cuisine, almost full.
- Deciding to go to **B**, based on the belief that other diners have also had the chance of going to **A**, is an example of herd behavior

Herd Behavior: Milgram's Experiment

Stanley Milgram asked one person to stand still on a busy street corner in New York City and stare straight up at the sky

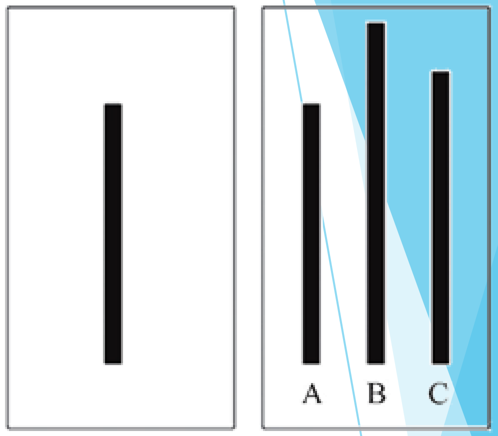
- About 4% of all passersby stopped to look up.

When 5 people stand on the sidewalk and look straight up at the sky, 20% of all passersby stopped to look up.

Finally, when a group of 18 people look up simultaneously, almost 50% of all passersby stopped to look up.

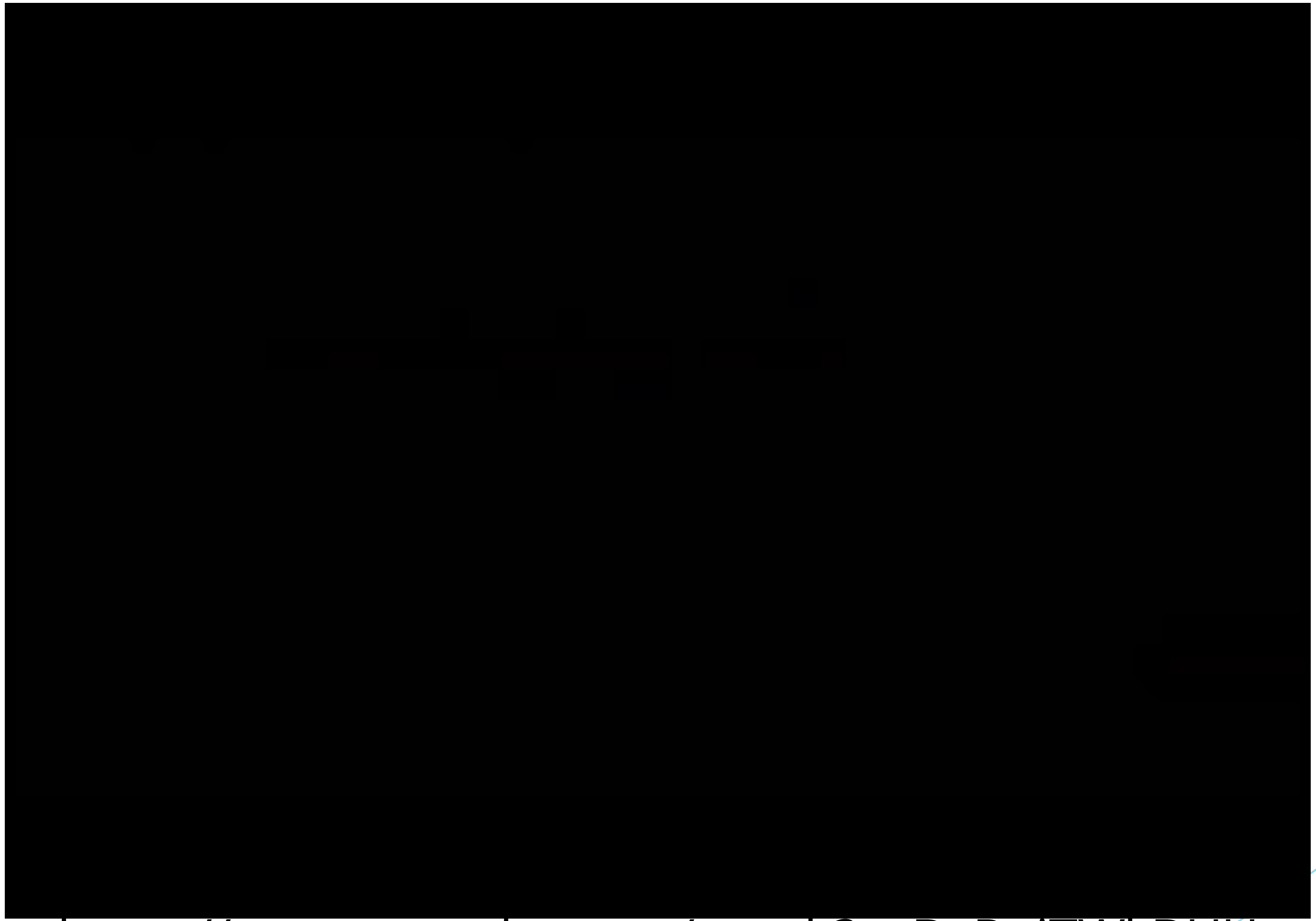
Herd Behavior: Solomon Asch's Experiment

- Groups of students participated in a vision test
- They were shown two cards, one with a single line segment and one with 3 lines
- The participants were required to match line segments with the same length
- Each participant was put into a group where all other group members were collaborators with Asch.



- These collaborators were introduced as participants to the subject.
- Asch found that in control groups with no pressure to conform, only 3% of the subjects provided an incorrect answer.
- However, when participants were surrounded by individuals providing an incorrect answer, up to 32% of the responses were incorrect.

Herding: Asch Elevator Experiment



https://www.youtube.com/watch?v=BgRoiTWkBHU

Herd Behavior

Herd behavior describes when a group of individuals performs actions that are highly correlated without any plans

Main Components of Herd Behavior

- Connections between individuals
- A method to transfer behavior among individuals or to observe their behavior

Examples of Herd Behavior

- Flocks, herds of animals, and humans during sporting events, demonstrations, and religious gatherings

Network Observability in Herb Behavior

In herd behavior, individuals make decisions by observing all other individuals' decisions

- In general, herd behavior's network is close to a complete graph where nodes can observe at least most other nodes and they can observe public information
  - For example, they can see the crowd

Designing a Herd Behavior Experiment

1. There needs to be a decision made.
   - In our example, it is going to a restaurant

2. Decisions need to be in sequential order

3. Decisions are not mindless and people have private information that helps them decide

4. No message passing is possible. Individuals don’t know the private information of others, but can infer what others know from what is observed from their behavior.

Herding: Urn Experiment

- There is an urn in a large class with three marbles in it

B B R	R R B
Majority Blue	**Majority Red**
P[B,B,R]=50%	**P[R,R,B]=50%**

- During the experiment, each student comes to the urn, picks one marble, and checks its color in private.

- The student predicts **majority blue** or **majority red**, writes her prediction on the blackboard, and puts the marble back in the urn.

- Students cannot see the color of the marble taken out and can only see the predictions made by different students regarding the majority color on the board.

Urn Experiment: First and Second Student

- First Student:
  - *Board: -*
    - Observed: **B** → Guess: **B**

    -or-
    - Observed: **R** → Guess: **R**
- Second Student:
  - *Board: **B***
    - Observed: **B** → Guess: **B**

    -or-
    - Observed: **R** → Guess: **R**/**B** (flip a coin)

Urn Experiment: Third Student

- *If board: **B**, **R***
  - Observed: **B** → Guess: **B**, or
  - Observed: **R** → Guess: **R**

- *If board: **B**, **B***
  - Observed: **B** → Guess: **B**, or
  - **Observed: R → Guess: B (Herding Behavior)**

The forth student and onward

- Board: **B**,**B**,**B**
- **Observed: B/R → Guess: B**

Bayes's Rule in the Herding Experiment

Each student tries to estimate the conditional probability that the urn is **majority-blue** or **majority-red**, given what she has seen or heard

- She would guess **majority-blue** if:

P[majority-blue|what she has seen or heard] > 1/2

- From the setup of the experiment we know:

P[majority-blue] = P[majority-red]=1/2

P[blue|majority-blue] = P[red|majority-red]=2/3

Bayes's Rule in the Herding Experiment

P[majority-blue | blue] = P[blue | majority-blue] x P[majority-blue] / Pr[blue]

P[blue] = P[blue | majority-blue] $\times$ P[majority-blue]
+ P[blue | majority-red] $\times$ P[majority-red]
= 2/3 $\times$ 1/2 + 1/3 $\times$ 1/2 = 1/2

P[majority-blue | blue] = (2/3 $\times$ ½)/(1/2)

- So the first student should guess **blue** when she sees **blue**
- The same calculation holds for the second student

Third Student

P[majority-blue | blue, blue, red] =
P[blue, blue, red | majority-blue] × P[majority-blue] / P[blue, blue, red]

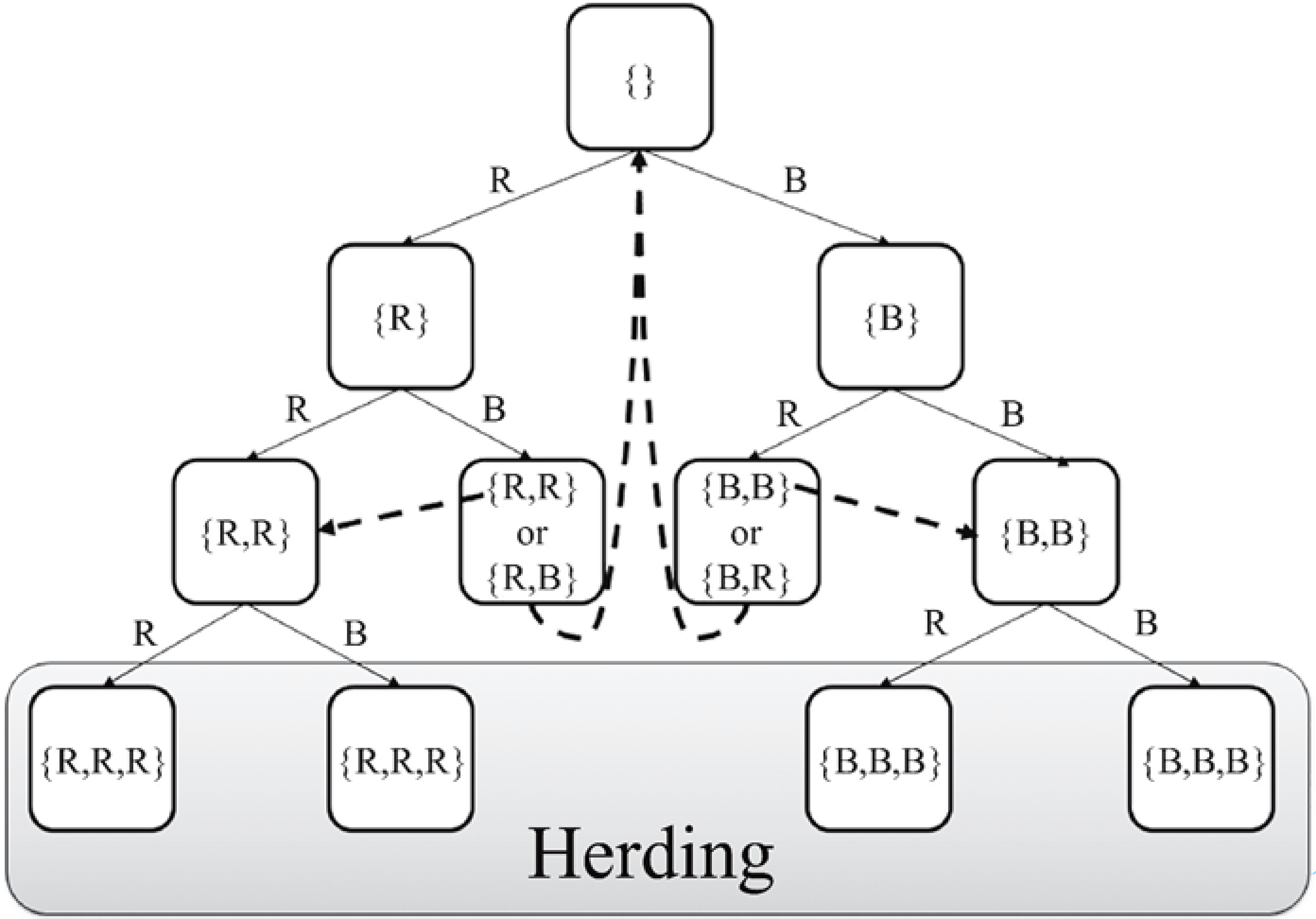
P[blue, blue, red | majority-blue] = 2/3 × 2/3 × 1/3 = 4/27

P[blue, blue, red] = P[blue, blue, red | majority-blue] × P[majority-blue]
+ P[blue, blue, red | majority-red] × P[majority-red]
= (2/3 × 2/3 × 1/3) × 1/2 + (1/3 × 1/3 × 2/3) × 1/2 = 1/9

P[majority-blue | blue, blue, red] = (4/27 × 1/2) / (1/9) = 2/3

- So the third student should guess **blue** even when she sees **red**
- All future students will have the same information as the third student

Urn Experiment

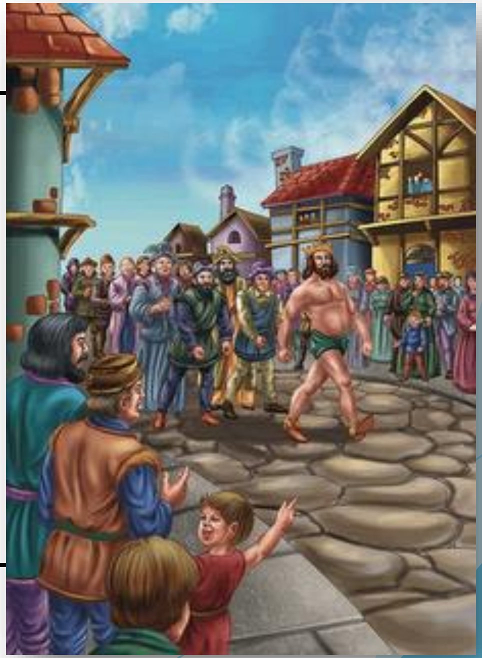


Herding Intervention

Herding: we only have access to public information

- Herding may be intervened by releasing private information which was not accessible before

The little boy in
"**The Emperor's New Clothes**"
story *intervenes* the herd by
shouting
"**There is no clothe**"



Herding Intervention

To intervene the herding effect, we need one person to tell the herd that there is nothing in the sky and the first person is doing this to stop his nose bleeding

How Does Intervention Work?

- When a new piece of private information releases,
  - The herd reevaluate their guesses and this may create completely new results
- The Emperor's New Clothes
  - When the boy gives his private observation, other people compare it with their observation and confirm it
  - This piece of information may change others guess and ends the herding effect
- In urn experiment, intervention is possible by
  1. A private message to individuals informing them that the urn is majority blue or
  2. Writing the observations next to predictions on the board.