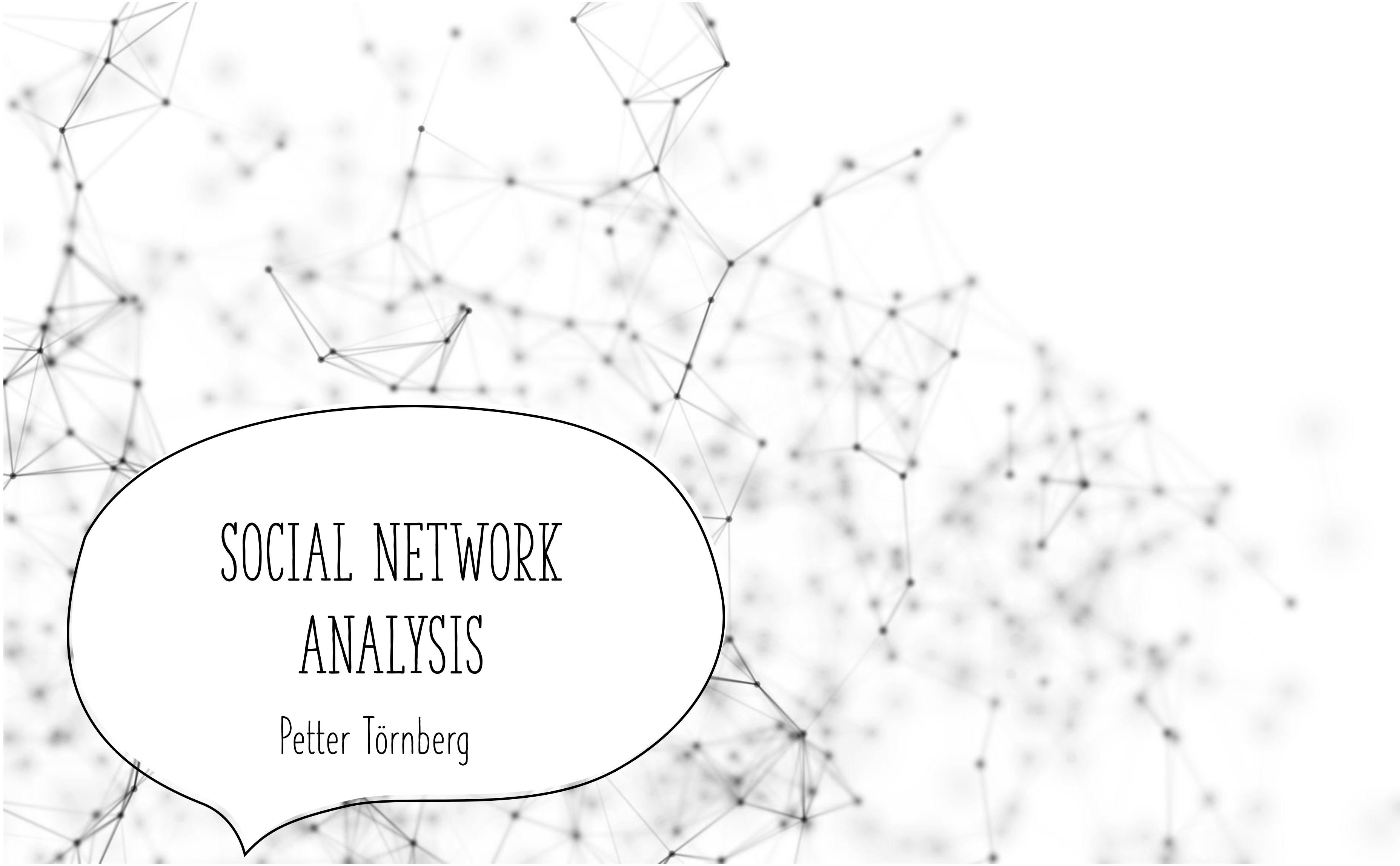
SOCIAL NETWORK ANALYSIS

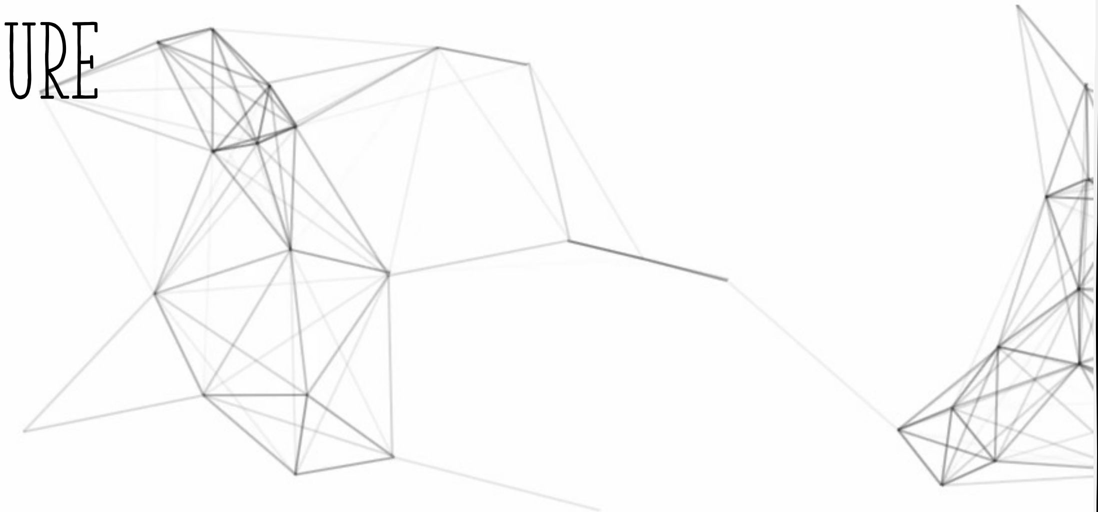
Petter Törnberg



LEARNING OUTCOMES

- Thinking relationally
- Foundational terms: nodes, edges, bridges, …
- How to represent a network: adjacency matric, adjacency list, …
- How to answer research questions using SNA

STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Visual Network Analysis
6. Community detection
7. 6 degrees of separation and the strength of weak ties
8. Case study: Zwarte Piet debate on Twitter

THE THREE TYPES OF DATA IN THE SOCIAL SCIENCES

1. Attribute data:

E.g., survey, demographics

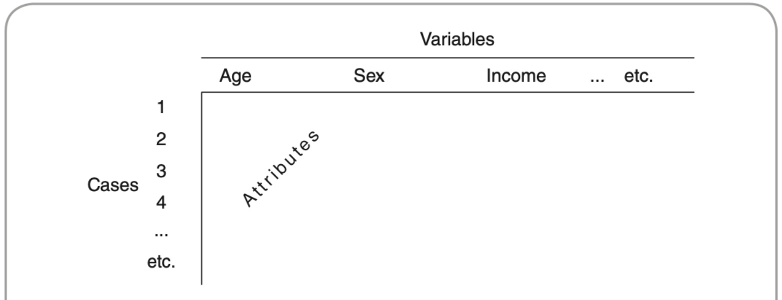
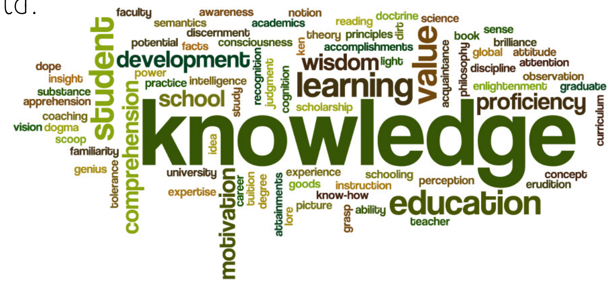


Figure 4.1 A data matrix for variable analysis

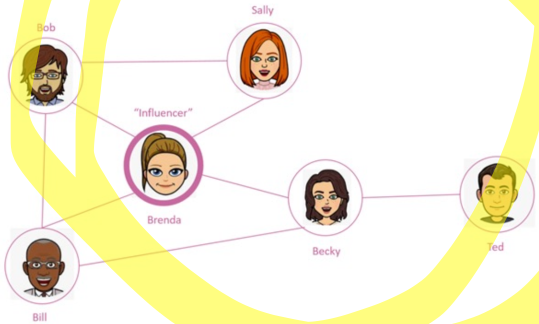
2. Ideational data:

E.g., text, speech

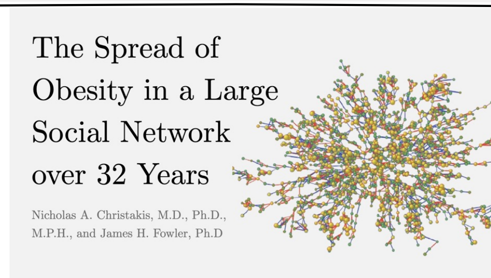


3. Relational data:

e.g. friendships, connections, etc.



IS OBESITY CONTAGIOUS? (CHRISTAKIS & FOWLER, 2007)



Data: Framingham Heart Study, a long-term cardiovascular study initiated in 1948, which tracks the health and lifestyle of participants and their families over multiple generations. Longitudinal over 32 years.

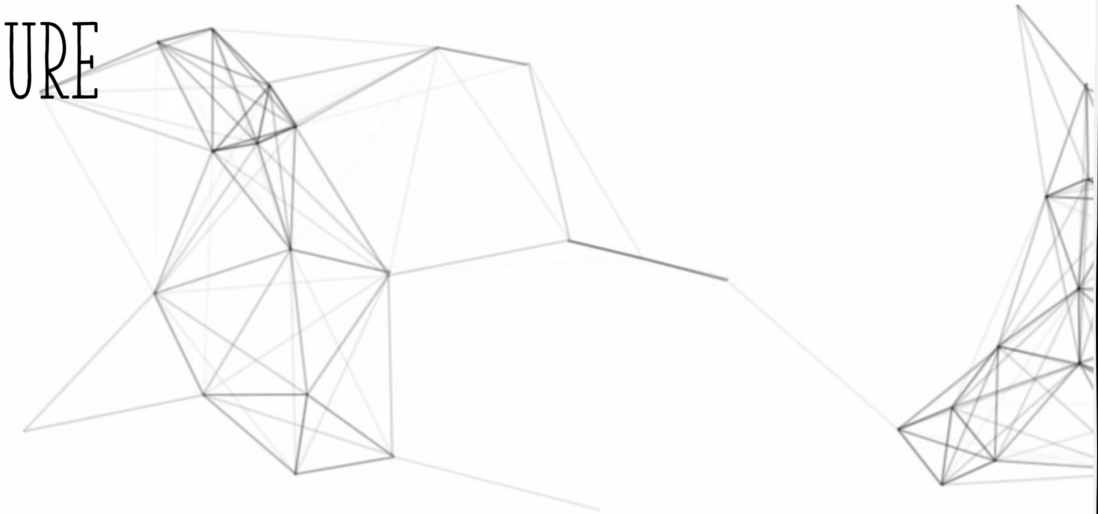
Network: Social networks based on familial, friendship, and spousal relationships within the Framingham cohort.

Finding: If a person became obese, their friends, siblings, and even distant social contacts were more likely to become obese as well.

Obesity tends to "cluster" in social groups. If an individual's close friend became obese, the person was 57% more likely to become obese themselves. This likelihood increased if multiple people in the network were obese.

Obesity is not merely correlated with social relationships, but it spreads contagiously through social networks.

STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Visual Network Analysis
6. Community detection
7. 6 degrees of separation and the strength of weak ties
8. Case study: Zwarte Piet debate on Twitter

WHAT ARE SOME RESEARCH QUESTIONS YOU CAN ANSWER WITH SNA?

How do friendship networks evolve over time in schools?

Who are the information brokers in the spread of innovation within a sector?

What are the patterns of professional networking that lead to successful job placement?

Which influencer should you buy if you want to maximize your social media viral ads in a particular social group?

Who is the central person within your friendship community?

How do friendship networks evolve over time in schools?

Centrality Analysis: Identifies the most important or influential individuals within a network.

Community Detection: Unearths the structure of the network by identifying clusters or communities within it.

Network Dynamics and Change Over Time: Examines how social networks evolve, and who forms ties with whom.

Structural Holes and Network Brokers: Analyzes the gaps in a social network (structural holes) and identifies brokers who bridge these gaps.

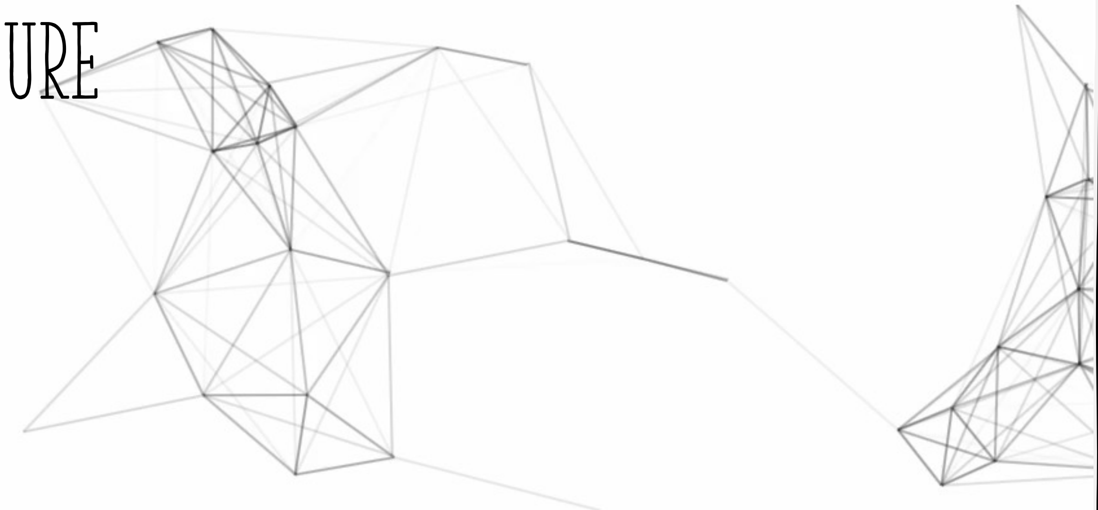
Network Cohesion and Subgroup Analysis: Studies the strength of relationships within a network to identify tight-knit groups or subgroups. This can help in understanding the social cohesion or fault lines within a community.

Role Analysis: Looks at the roles individuals play within a network based on their relationships and interactions. It can identify roles like leaders, gatekeepers, or peripheral members.

Diffusion of Information and Contagion: Examines how information, behaviors, or innovations spread through a network. This is particularly relevant for understanding phenomena like viral marketing, spread of rumors, or adoption of new technologies.

Social Capital Analysis: Explores the resources available to individuals through their social networks. It can shed light on how social ties contribute to personal or organizational success.

STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Visual Network Analysis
6. Community detection
7. 6 degrees of separation and the strength of weak ties
8. Case study: Zwarte Piet debate on Twitter

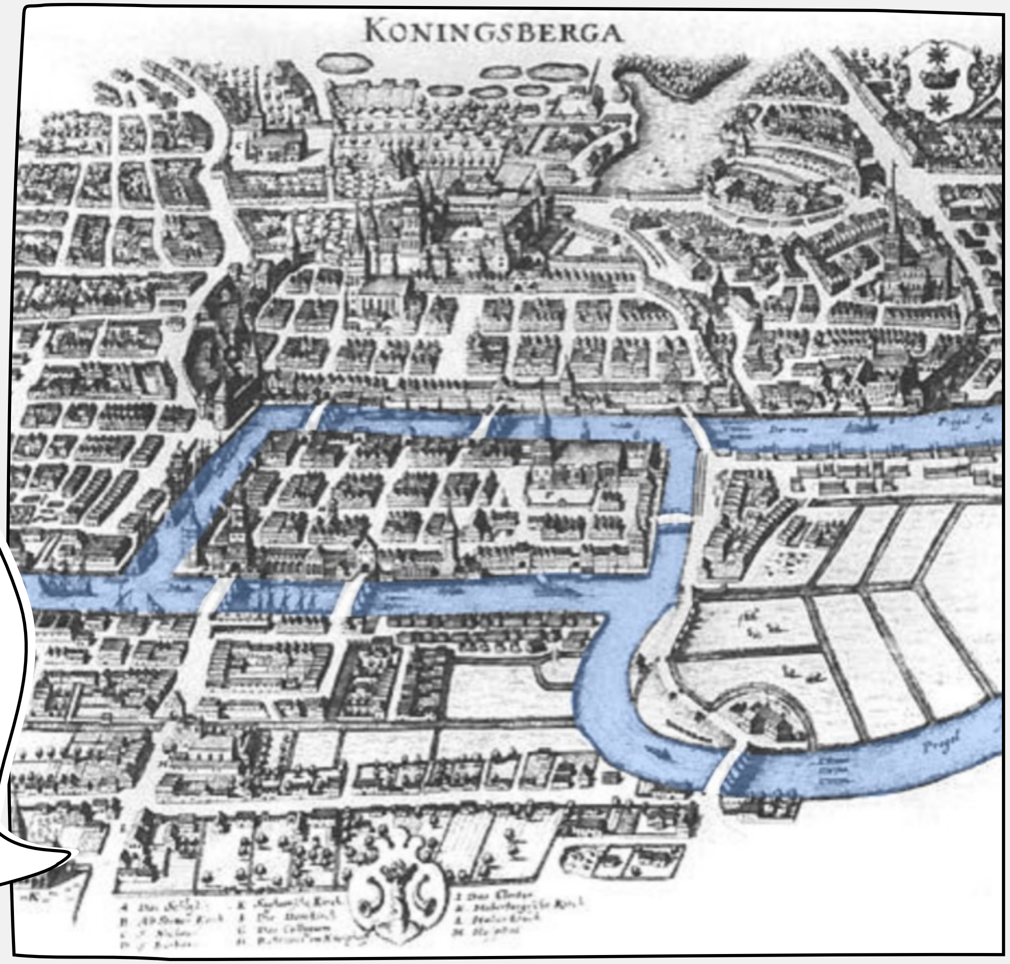
MATHEMATICAL ORIGINS OF SOCIAL NETWORKS

Graph == network. They're just different words for the same thing, but "graph" is used in math, and "network" in more applied areas.

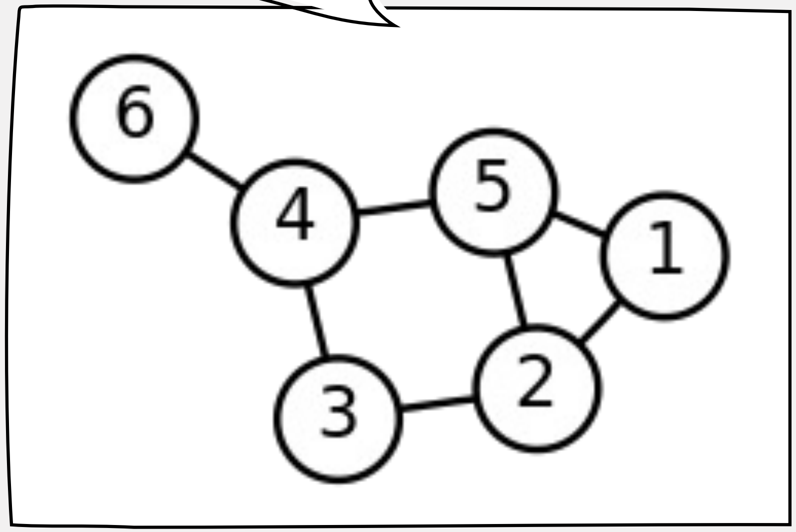
Graph theory: Leonhard Euler in 1736.

The Königsberg Bridge problem: Can you walk through the town in such a way that each bridge would be crossed exactly once?

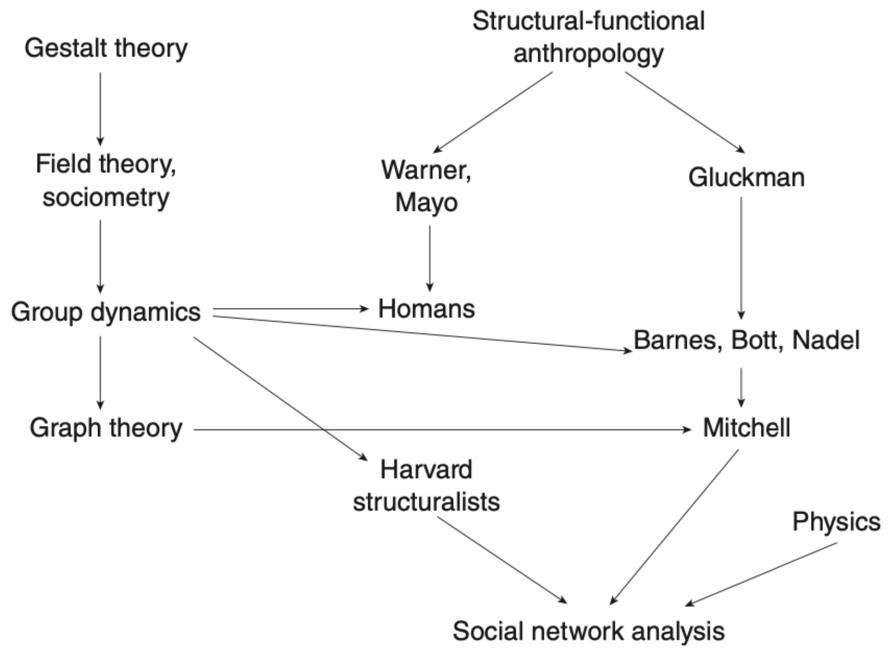
Euler: Nope.



GRAPHS ARE CENTRAL MATHEMATICAL CONCEPT



- Graphs are used in computer science, linguistics, physics, chemistry, biology
- Can be used to capture a range of phenomena
- Social networks are social science graphs



ORIGINS AND EVOLUTION OF SNA IN SOCIAL SCIENCES

Early Developments: Sociologists, anthropologists, psychologists explored social relationships (Early 20th Century). Georg Simmel, Émile Durkheim emphasized studying relationship patterns.

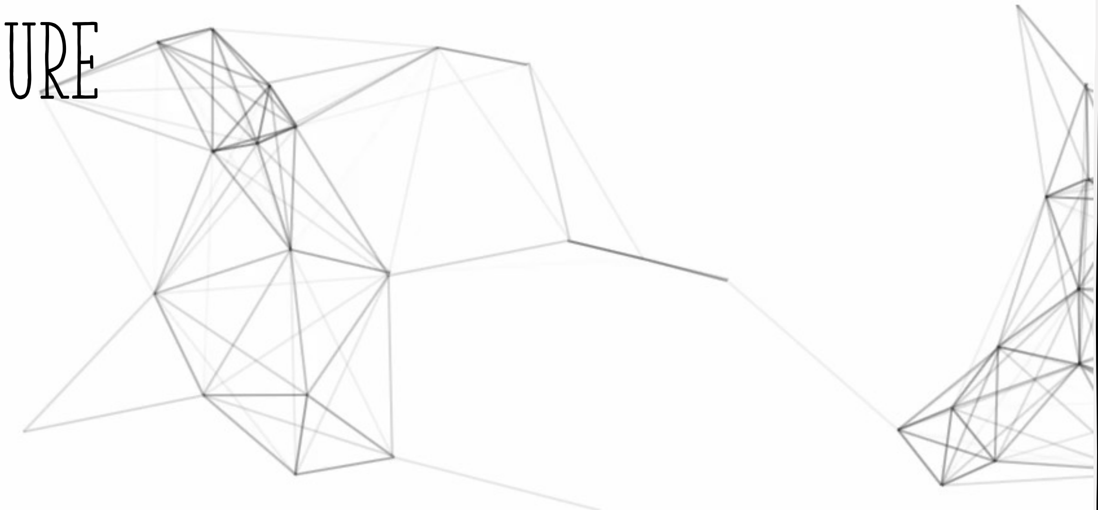
Formalization: Introduction of sociograms by Jacob Moreno (1930s).

Growth Period: Theoretical advancements by Harrison White and others (1940s-1970s).

Systematic Analysis Expansion: Contributions by Ronald Burt, Barry Wellman, and others (1970s onwards).

Recent digital explosion: Expansion due to computers and the internet; application in economics, epidemiology, political science. Advancements by Duncan J. Watts, Albert-László Barabási

STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Visual Network Analysis
6. Community detection
7. 6 degrees of separation and the strength of weak ties
8. Case study: Zwarte Piet debate on Twitter

IS THERE A THEORY OF SOCIAL NETWORKS?

Several theories and movements *align* with SNA lens:

- Theory of social capital: Robert Putnam 2000 "Bowling alone". But reduces networks to individual value
- Actor-Network Theory. Bruno Latour 2005. "actors" are not to be equated with human individuals or even groups but are to be seen as constituted by the relations that connect individuals to material objects
- Manifesto for *relational sociology*. Emirbayer 1997: we should focus less on sociology as people with attributes, and more as a process of relationships.

IS THERE A THEORY OF SOCIAL NETWORKS? NOT REALLY.

- Social networks cannot be reduced to a single theory
- They are a general set of techniques with many theoretical interpretations.
- A form of data or "a way of seeing".

- Since they can be applied to very different things, we must always add an interpretation for what the measures and representation means!

WHY SNA?

Relationships and how we connect with one another matter!

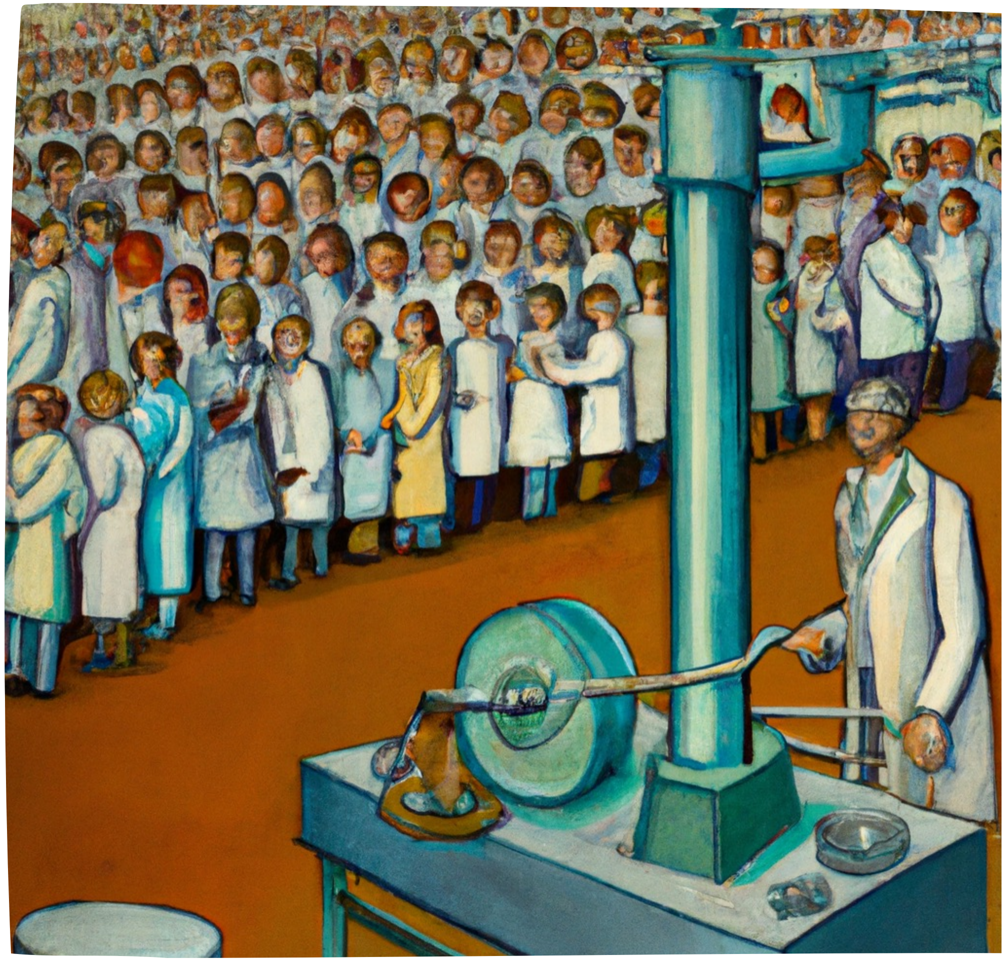
The way networks are patterned and structured also matters

Surveys assume independence - but we are not independent! We are social beings!

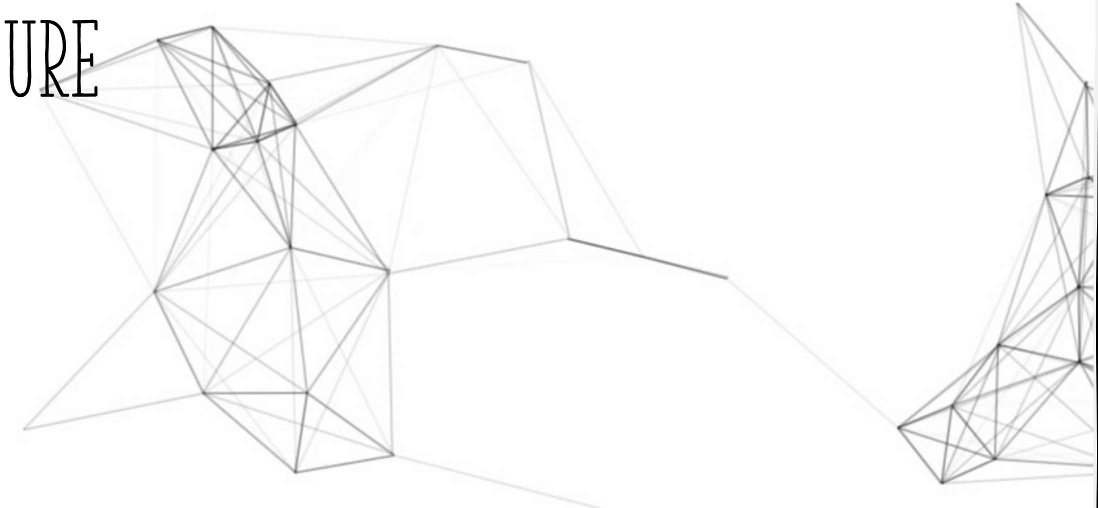
"The whole is more than the sum of its parts". Nonlinearity. Complexity

"For the last thirty years, empirical social research has been dominated by the sample survey. But as usually practiced, … the survey is a sociological meat grinder, tearing the individual from his social context and guaranteeing that nobody in the study interacts with anyone else in it."

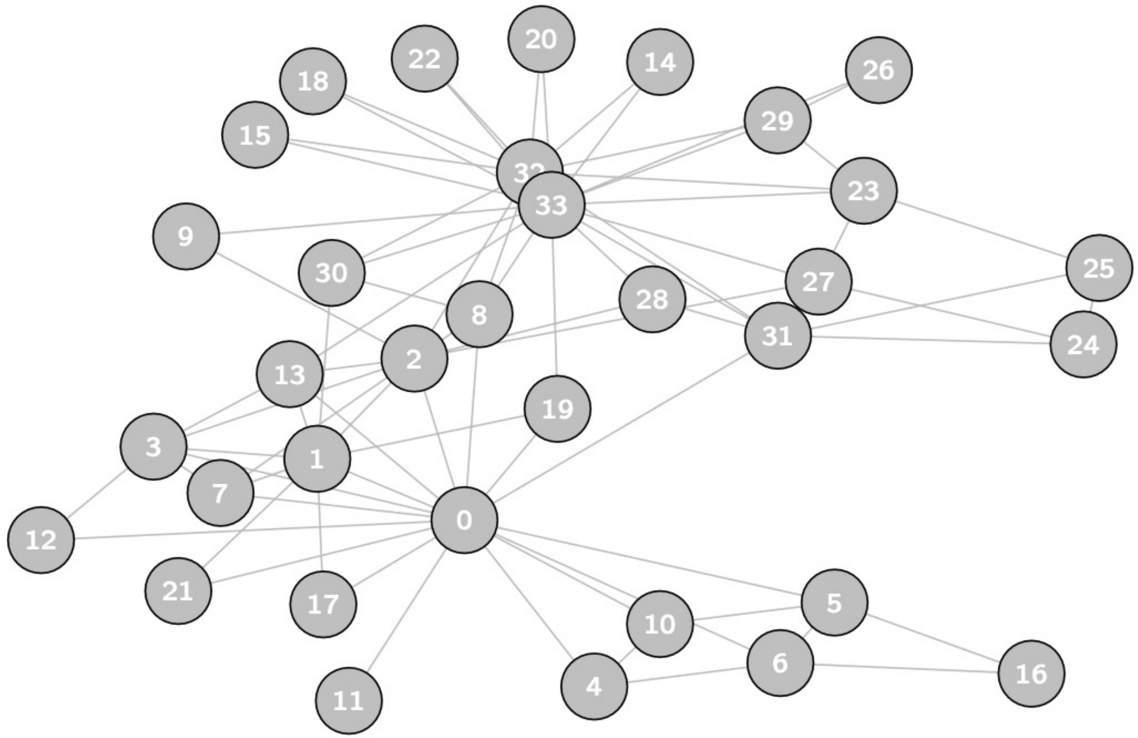
Allen Barton, 1968



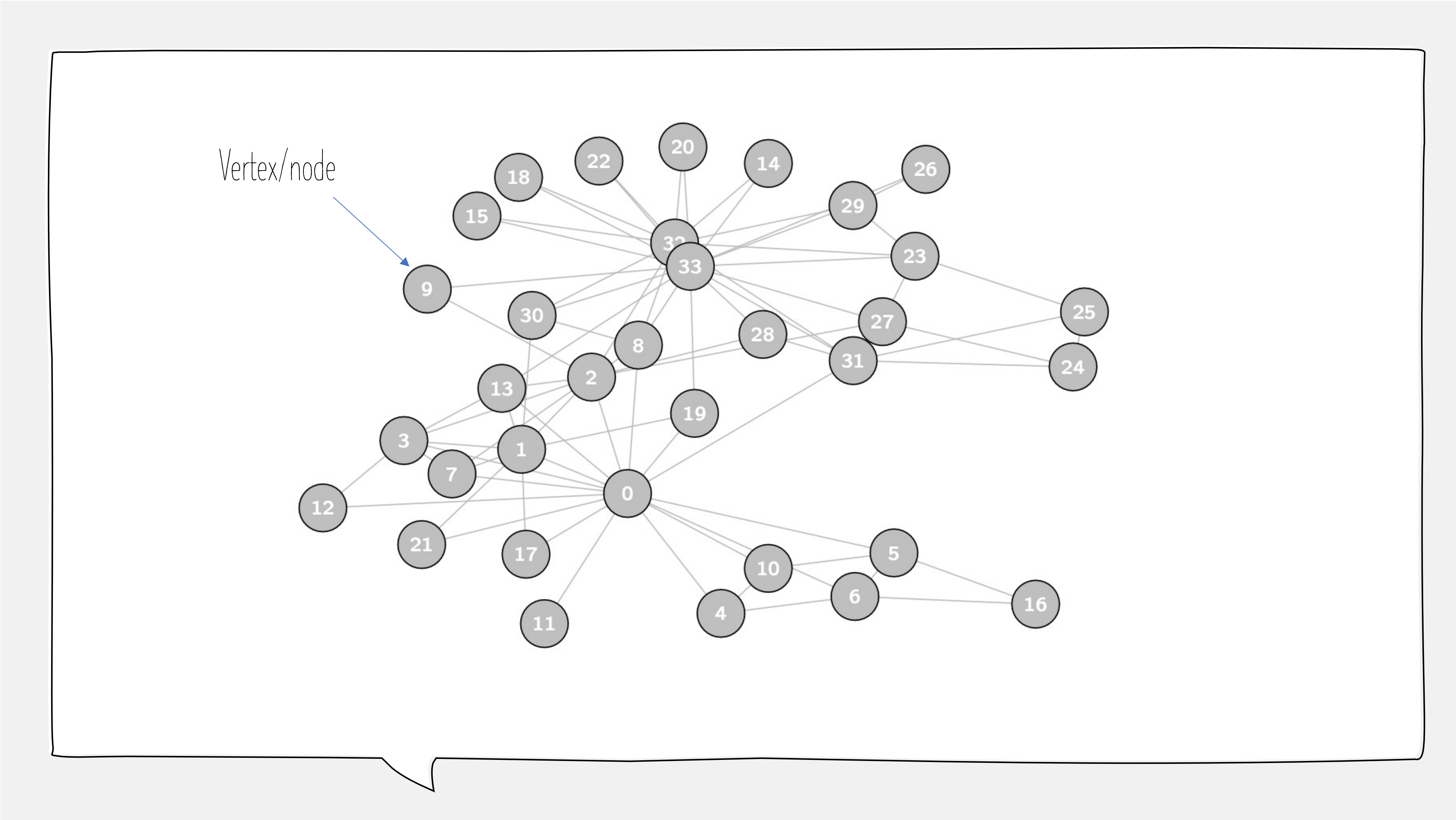
STRUCTURE OF THIS LECTURE

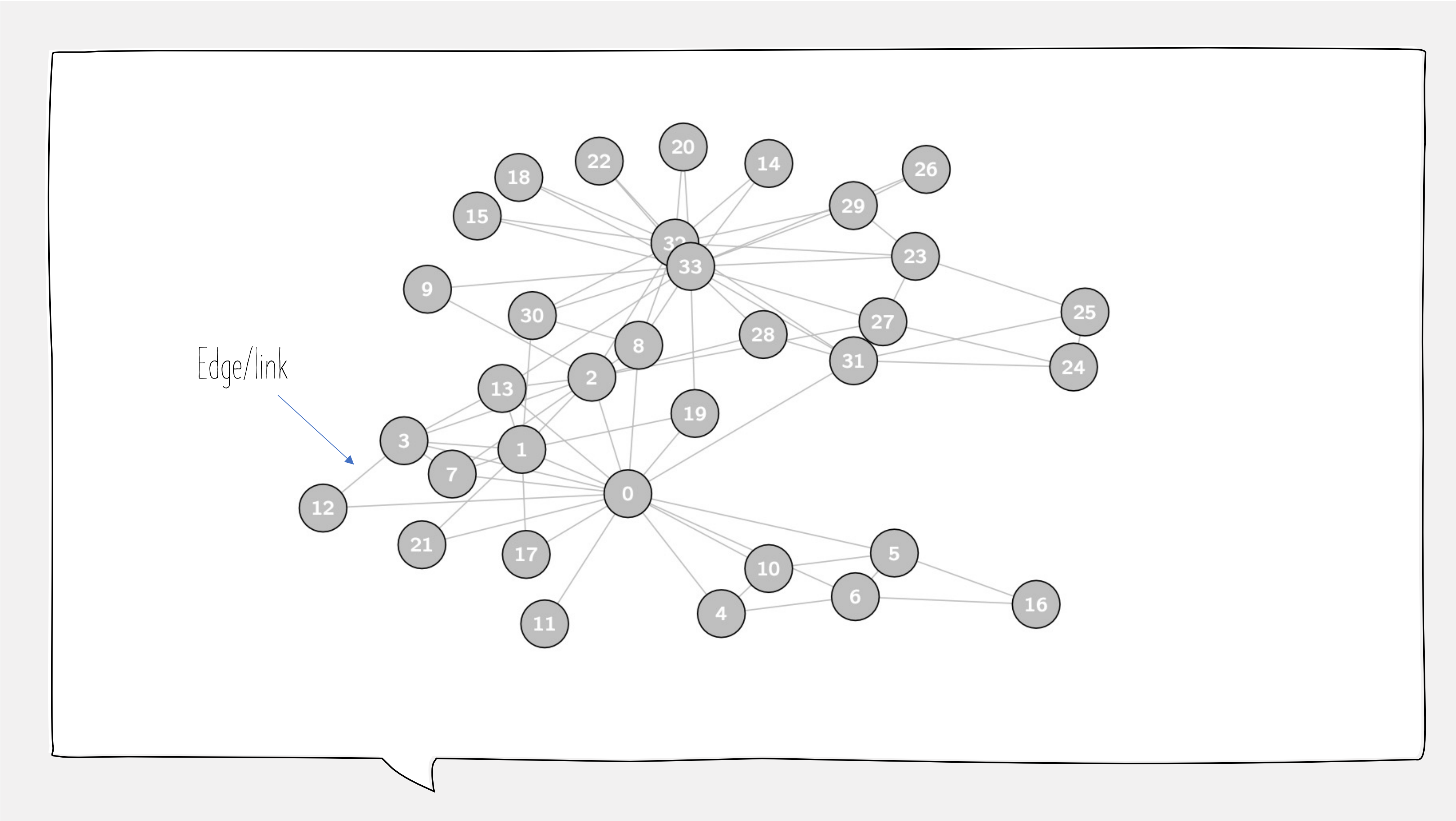


1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Visual Network Analysis
6. Community detection
7. 6 degrees of separation and the strength of weak ties
8. Case study: Zwarte Piet debate on Twitter



0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33


Vertex/node
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33


Edge/link
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

What do the nodes and links represent?

Nodes

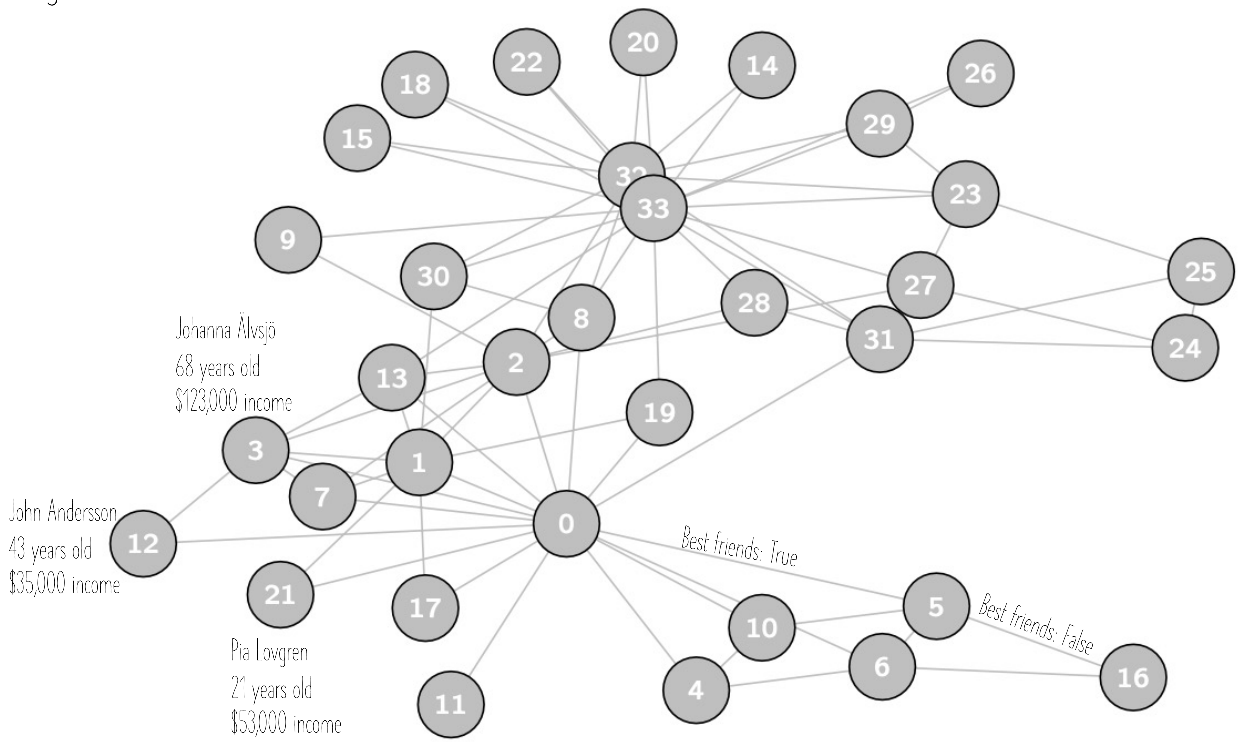
- People
- Organisations
- Blogs
- Countries

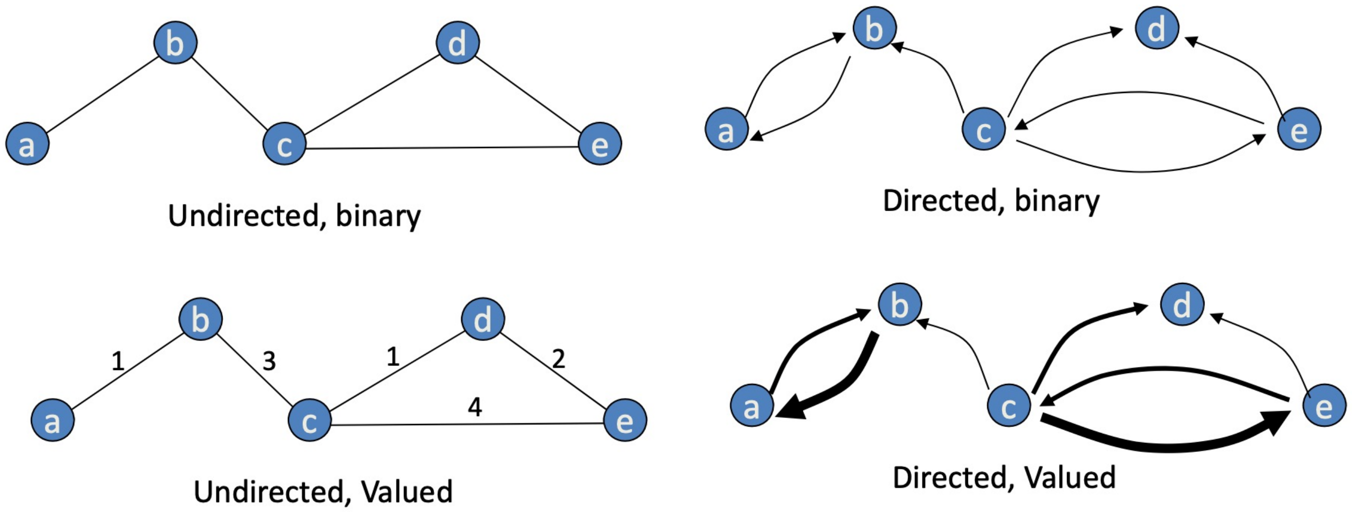
Links

- Communication
- Collaboration
- Hyperlinks
- Alliances

Or basically anything else!

Vertices and edges can have information associated to them

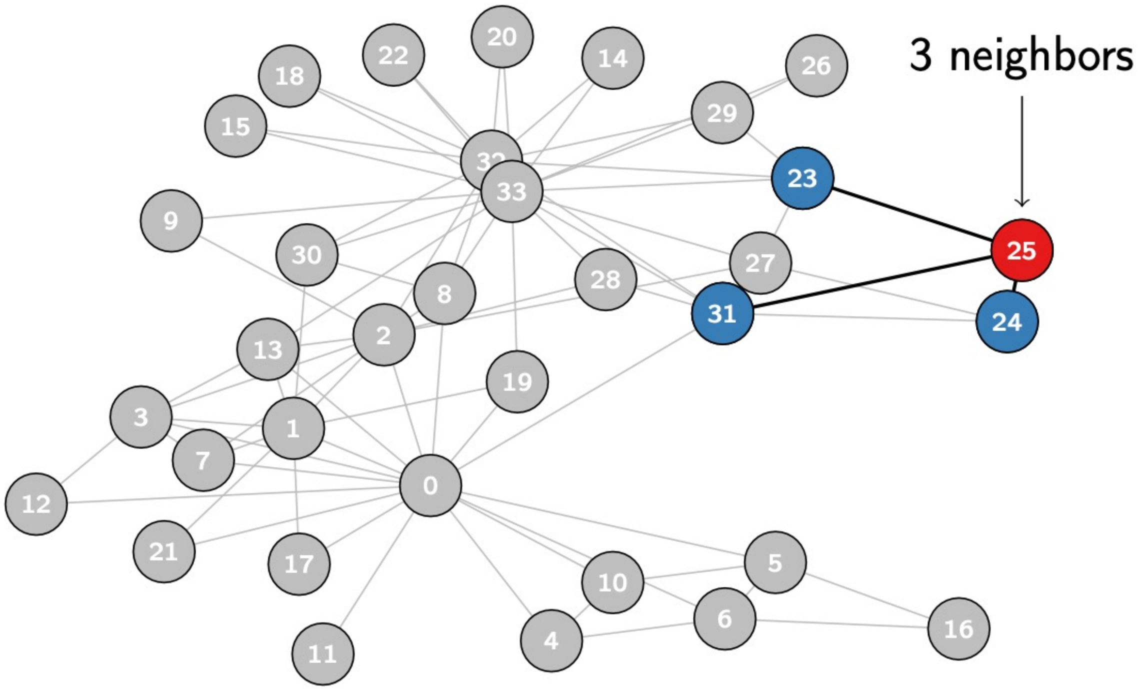



b
d
a
c
e
Undirected, binary
b
d
a
c
e
Directed, binary
b
1
3
1
d
2
4
a
c
e
Undirected, Valued
b
d
a
c
e
Directed, Valued

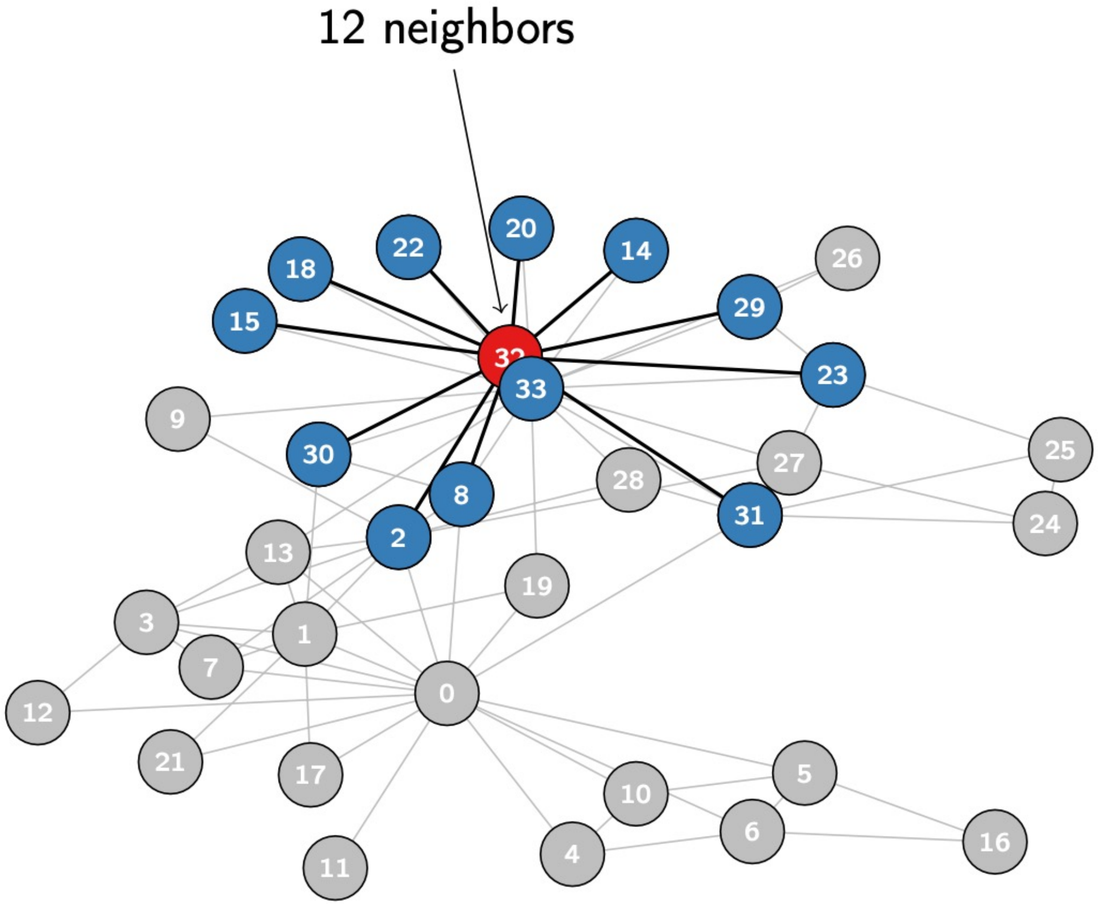
Do you have some network data in your project?
Or can you think of some way of representing it as a network?

- Is it directed?
- Is it weighted?
- What attributes are associated to nodes/edges?

DEGREE

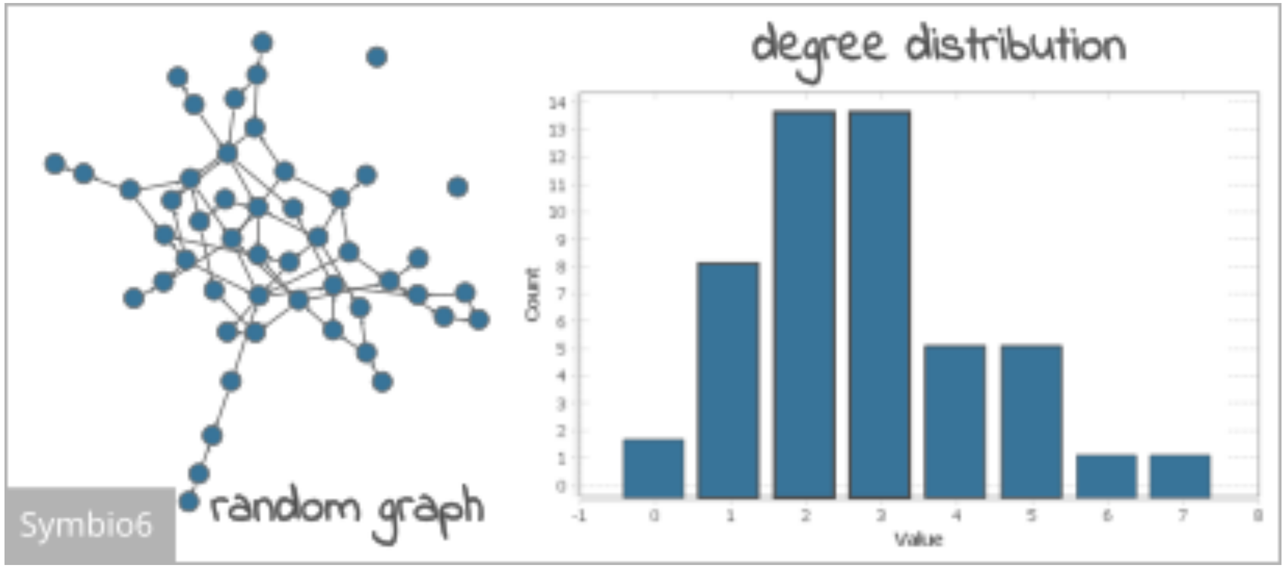

3 neighbors
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
33

DEGREE

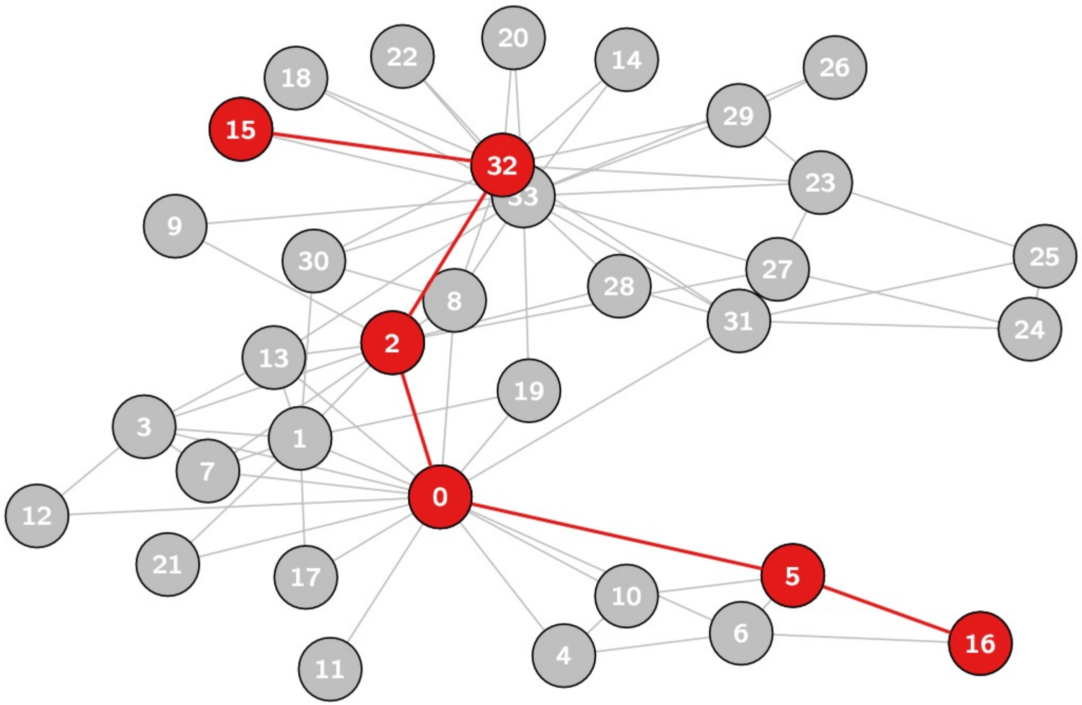

12 neighbors
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

DEGREE DISTRIBUTION

The number of nodes in a network with each possible degree.



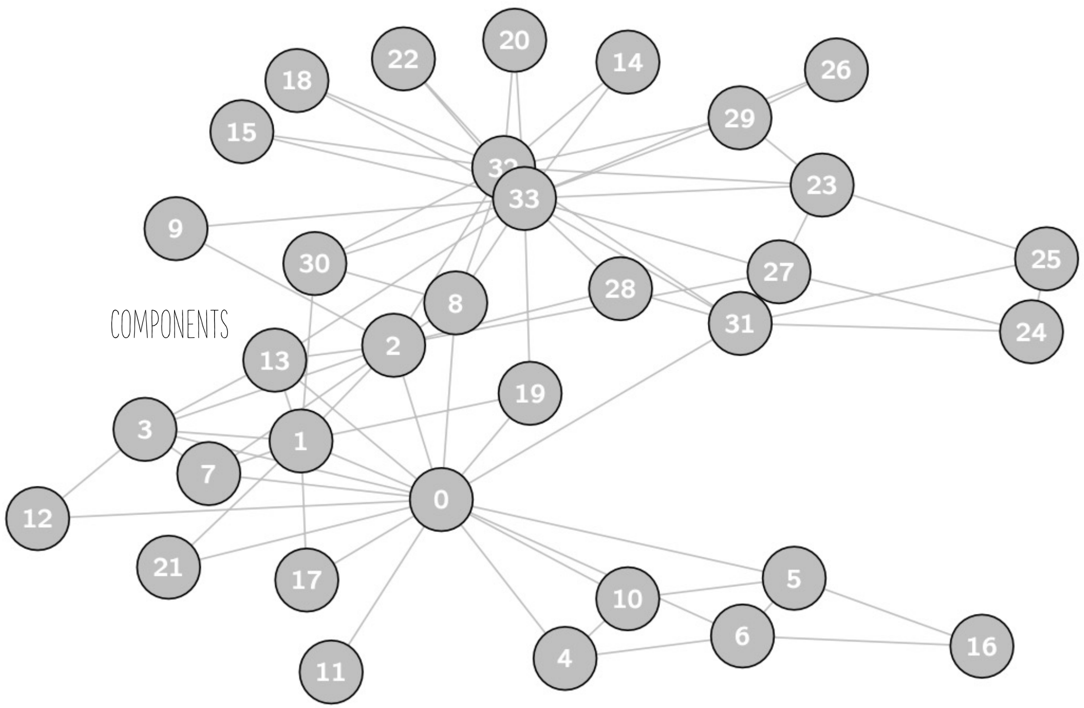
PATHS



0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

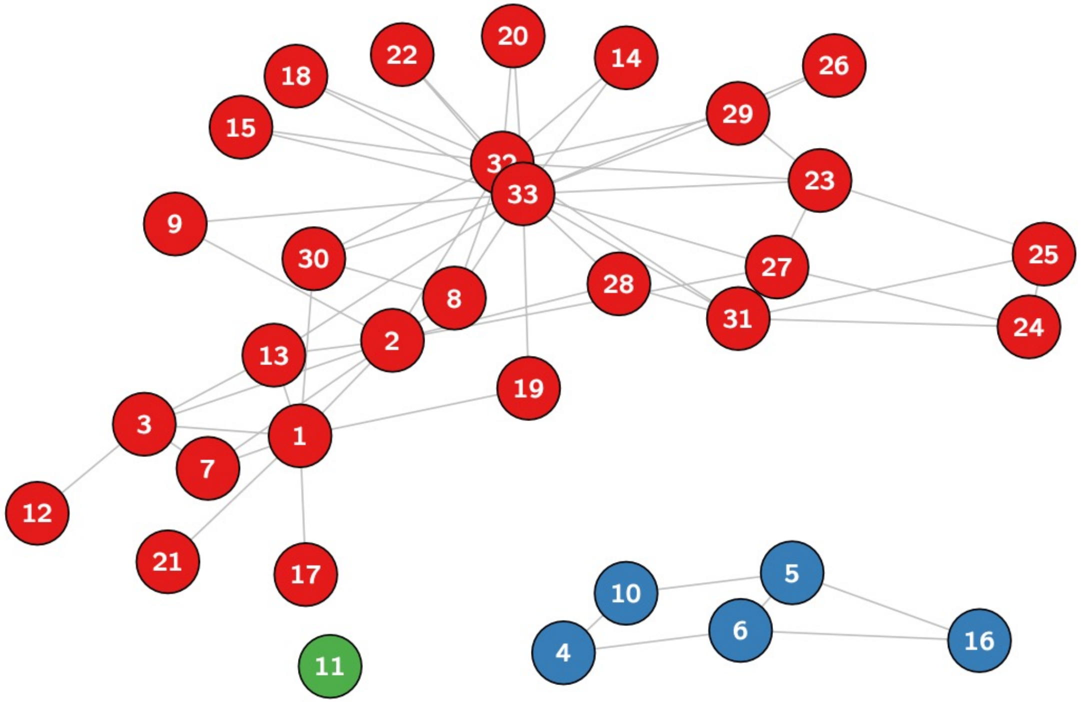
COMPONENTS

Delete node 0



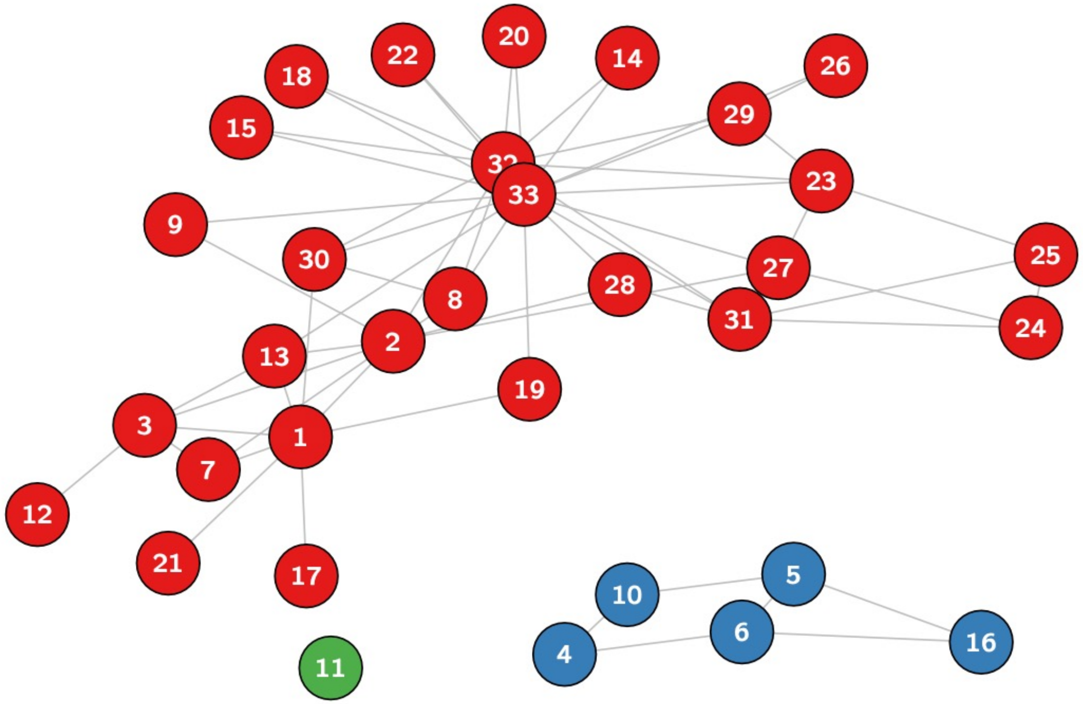
COMPONENTS

Node 0 functioned as a bridge



COMPONENTS

Usually one *giant component*.

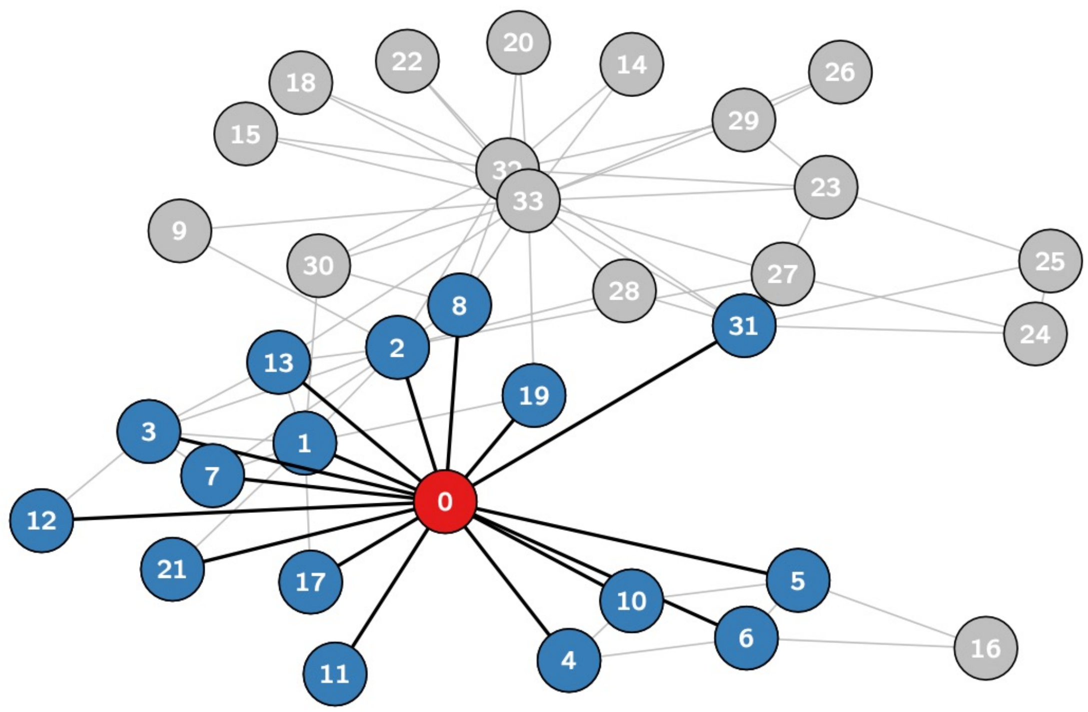


CLUSTERING

This measures the degree to which the neighbors of a specific node are interconnected.

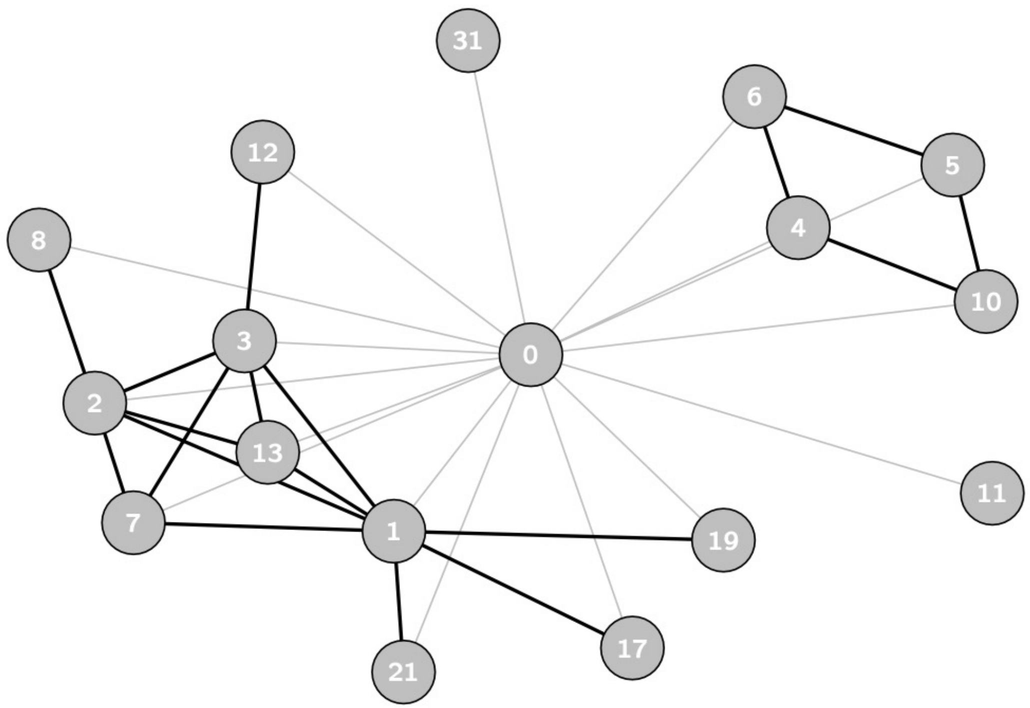
Clustering refers to the tendency of individuals to form tightly-knit groups characterized by relatively high interconnectivity.

Zoom in on node 0



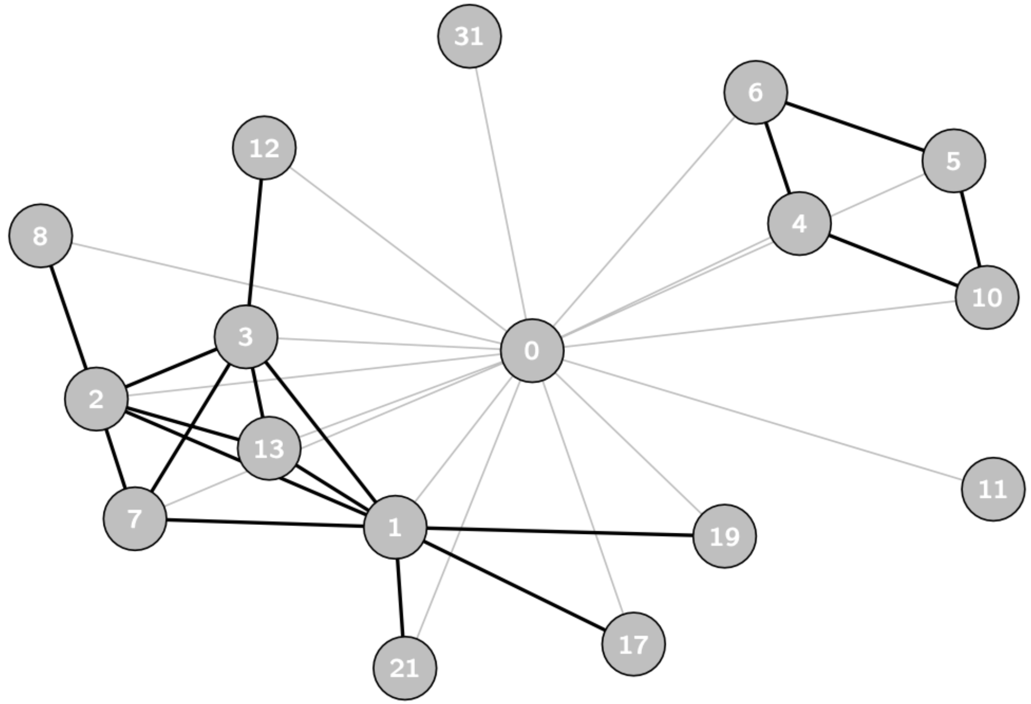
CLUSTERING

Ego network of node 0



CLUSTERING

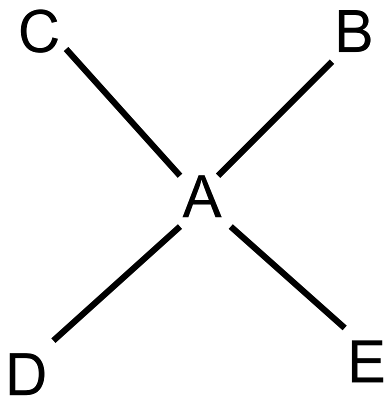
$$\text{Clustering} = \frac{18}{120} = 0.15$$



DENSITY

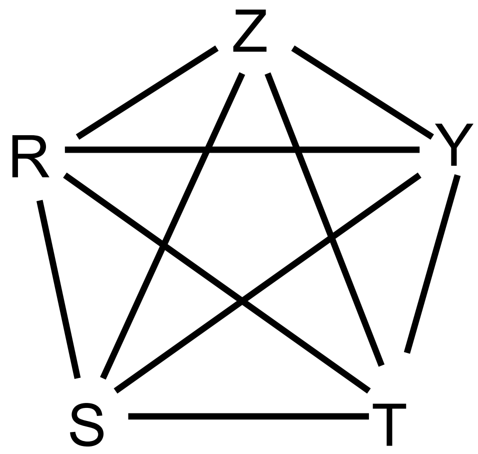
The density of a network is a measure of how many edges are in the network compared to the maximum possible number of edges.

1a



5 nodes, 4 edges

1b

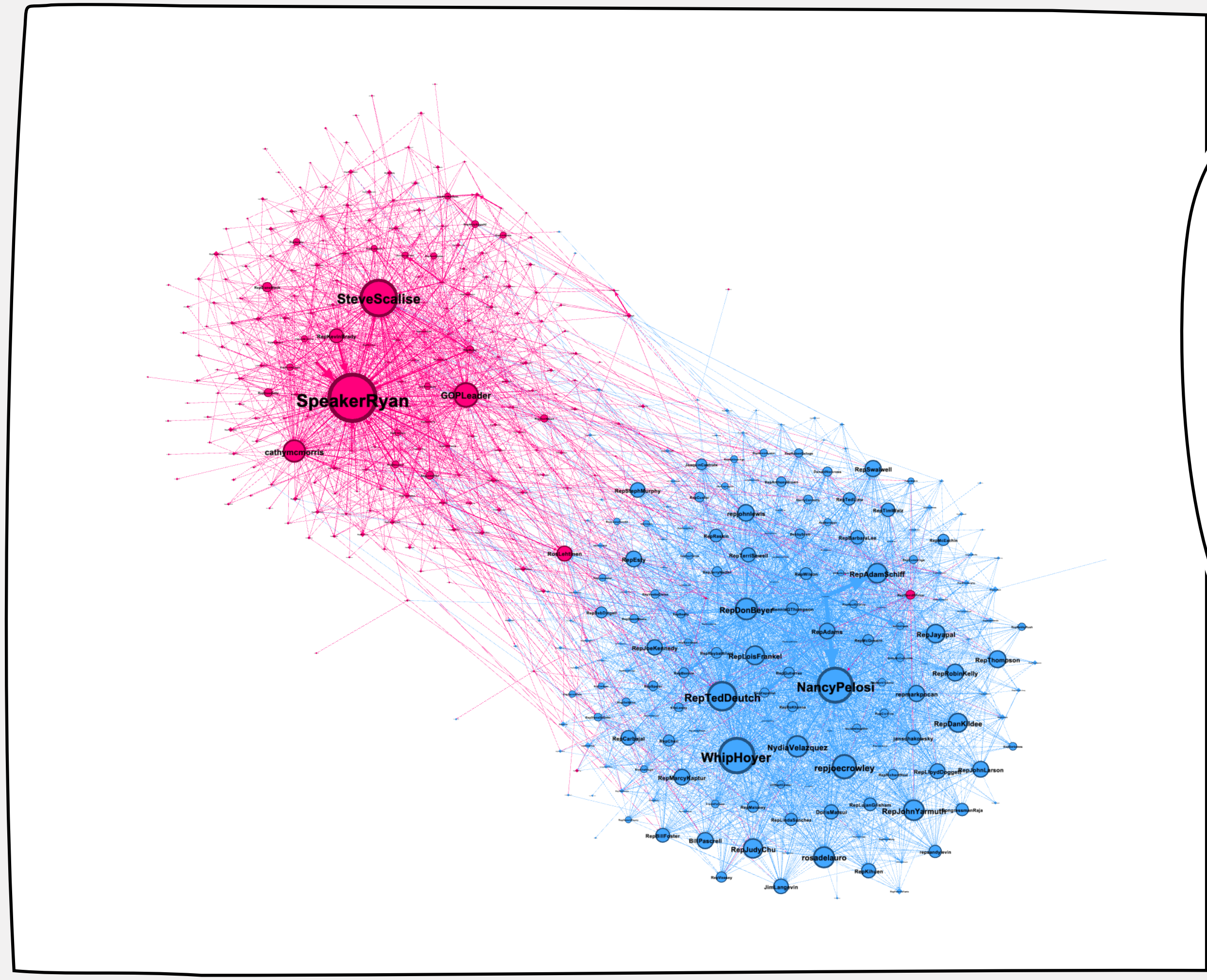


What's the density of this network?

(2*4)/(N(N−1)) = (2*4)/(5(5−1)) = 0.4

5 nodes, 10 edges

Maximal possible edges: (N * (N−1)) / 2



HOMOPHILY:

BIRDS OF A FEATHER FLY TOGETHER

Republicans retweet Republicans

Democrats retweet Democrats

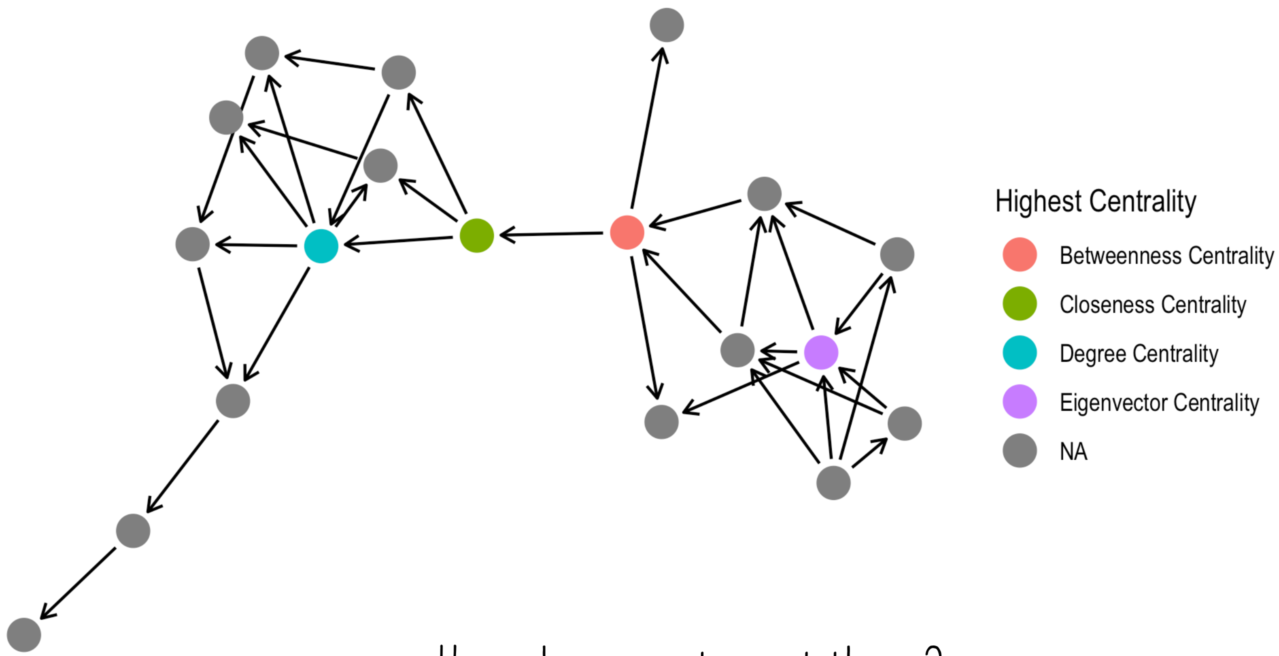
HOW TO MEASURE POWER / INFLUENCE? CENTRALITY!

Betweenness centrality: the number of times a node acts as a bridge along the shortest path between two other nodes.

Closeness centrality: how quickly information can flow from a given node to all other nodes in the network, calculated as the inverse of the sum of the shortest distances from that node to all other nodes

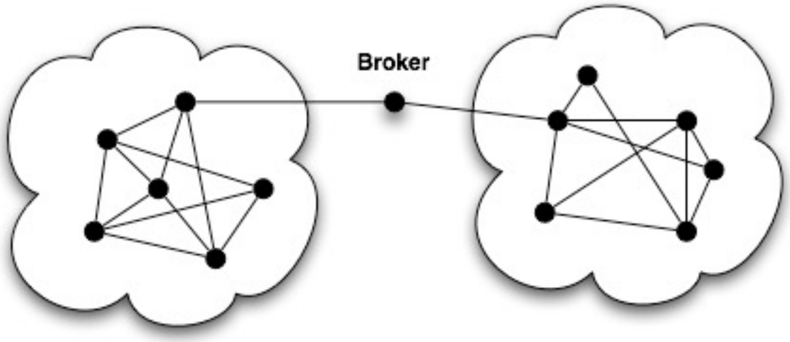
Degree centrality: the number of direct connections (or edges) a node has with other nodes in the network.

Eigenvector centrality / PageRank: how well-connected your friends are, and how well-connected they are. The probability of encounters in a random walk.



How do you interpret these?

BROKERAGE



Brokers can be powerful,
as they control the flow of information

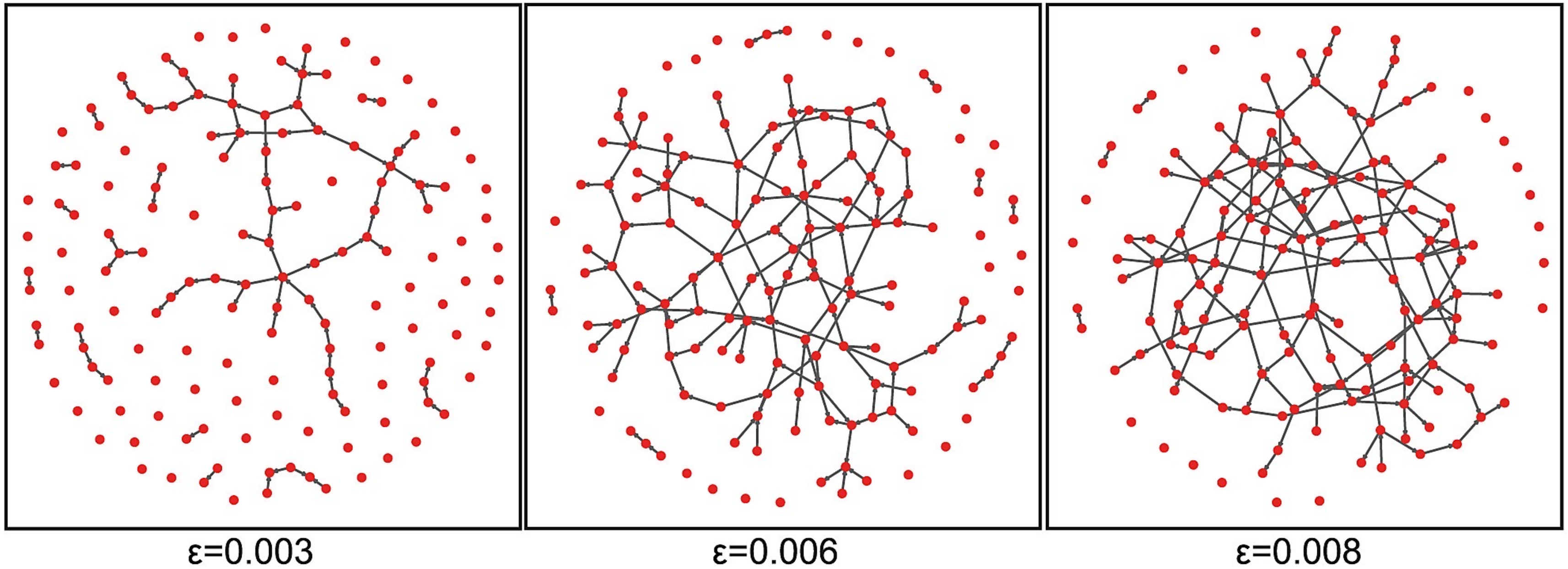
GENERATING NETWORKS

We can use simple algorithms that produce networks with certain properties.

There are several famous such algorithms.

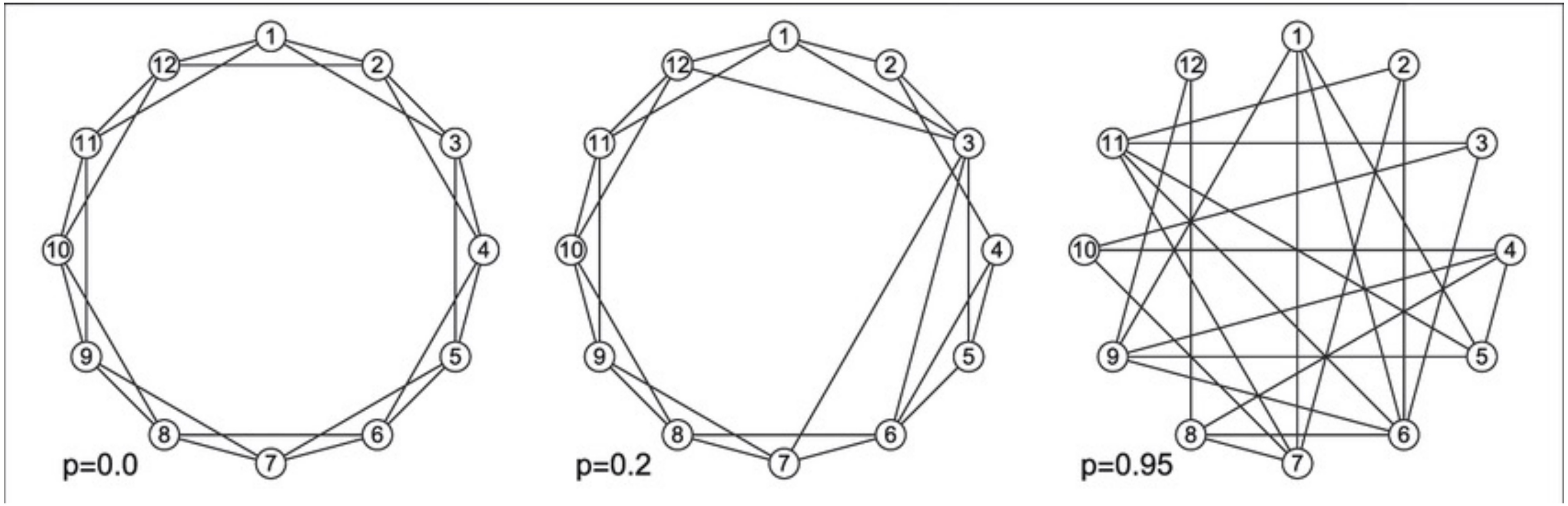
ERDÓS-RÉNYI GRAPH (RANDOM GRAPH):

Nodes are connected randomly with a fixed probability.



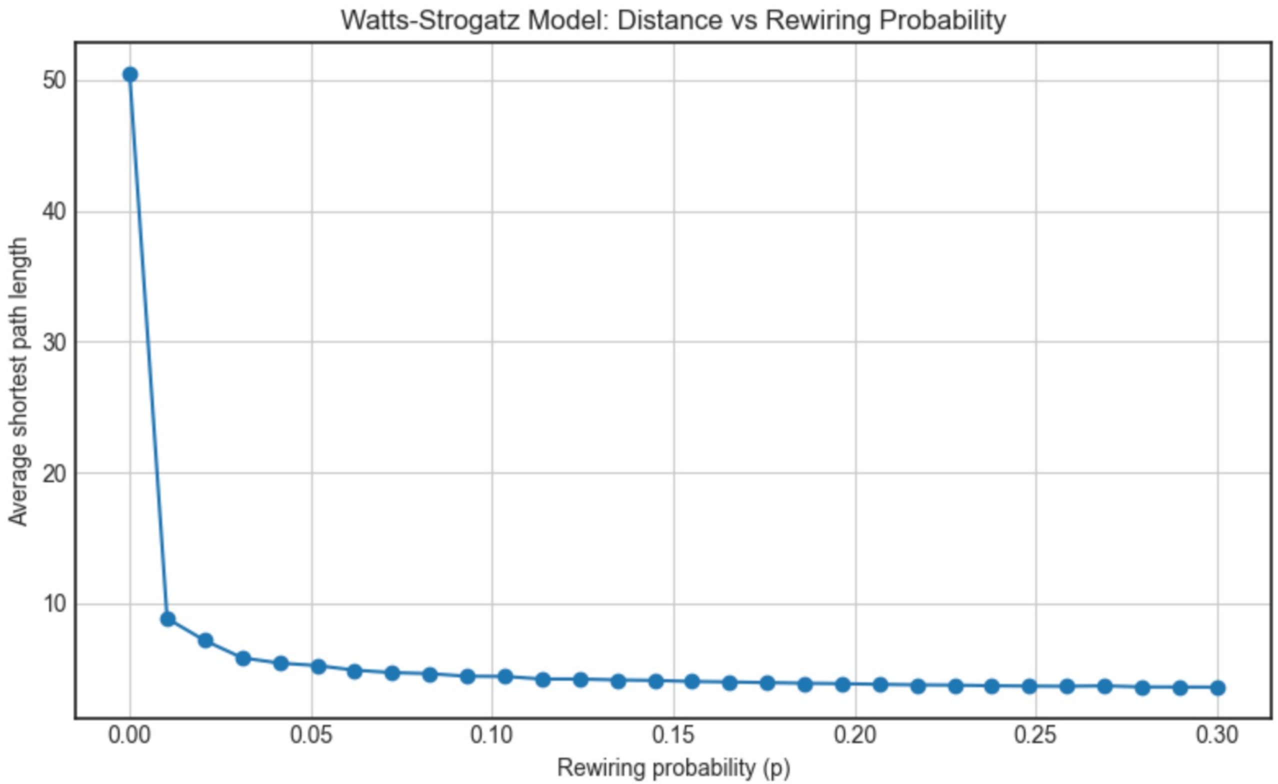
WATTS-STROGATZ SMALL-WORLD NETWORK

Make a circle. Then reconnect ties randomly.



HOW DOES THE AVERAGE SHORTEST PATH CHANGE WITH REWIRING?

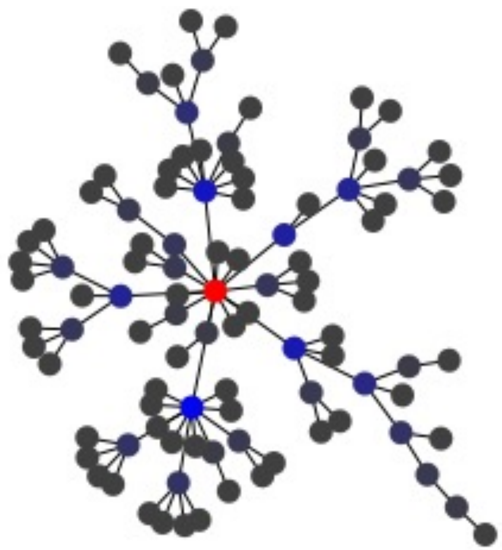
What does this mean?

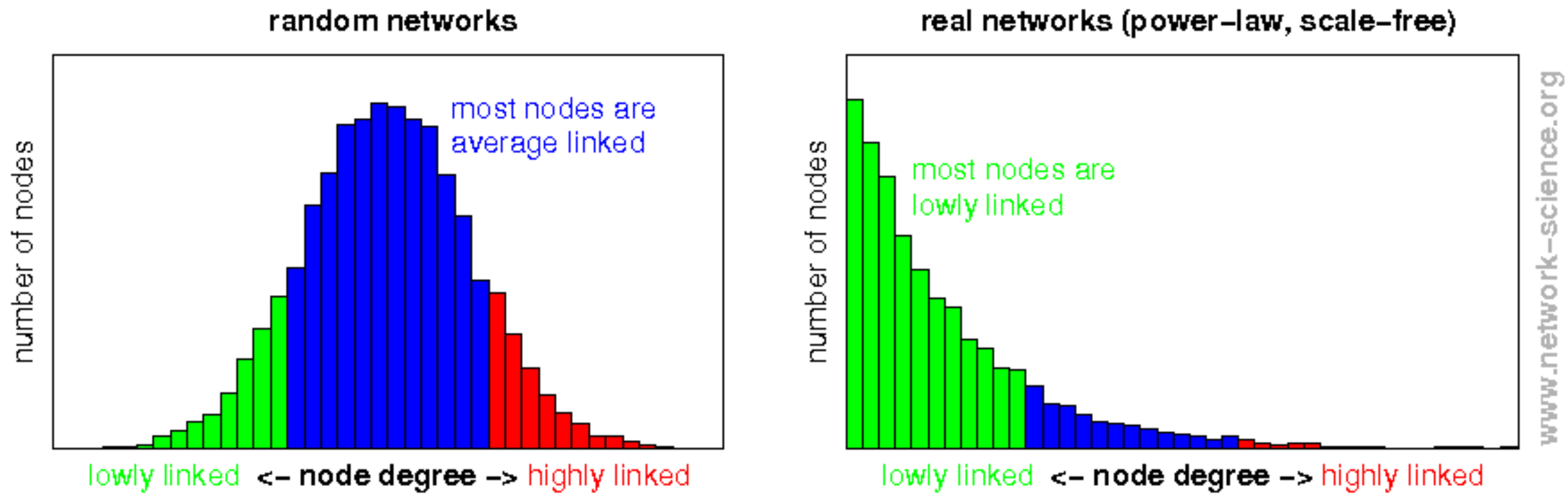


BARABÁSI-ALBERT - PREFERENTIAL ATTACHMENT

Each new node connects to existing nodes with a probability proportional to how many links they already have–"the rich get richer."

Captures realistic feature of online networks





The central feature of the internet and computational social science is that we no longer can expect normal distributions.

We have power-law distributions! Few actors control nearly all resources and

OTHER IMPORTANT GRAPHS

Caveman Graph: Several tightly connected communities with few interconnecting edges.

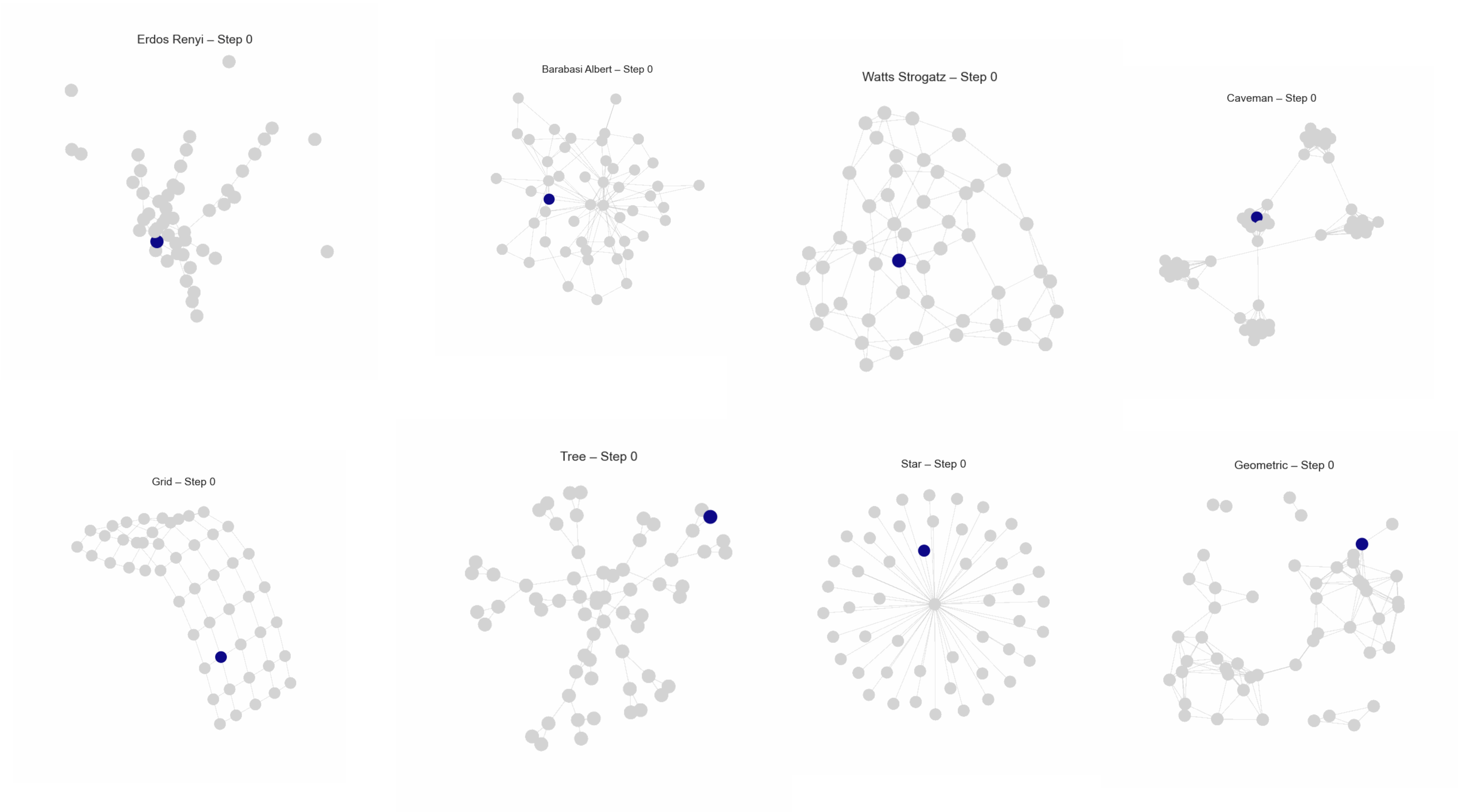
Geometric Graph: Nodes are connected if they are within a certain physical distance in space.

Grid Graph: Nodes are arranged in a grid, each connected to its direct neighbors.

Star Graph: One central node connected to all other nodes, with no other connections between peripheral nodes.

DIFFUSION IN NETWORKS

If disease and obesity spread in networks, do the network structure affect the outcome?


Erdos Renyi – Step 0
Barabasi Albert – Step 0
Watts Strogatz – Step 0
Caveman – Step 0
Grid – Step 0
Tree – Step 0
Star – Step 0
Geometric – Step 0

RESEARCH ARTICLE

Echo chambers and viral misinformation: Modeling fake news as complex contagion

Petter Törnberg

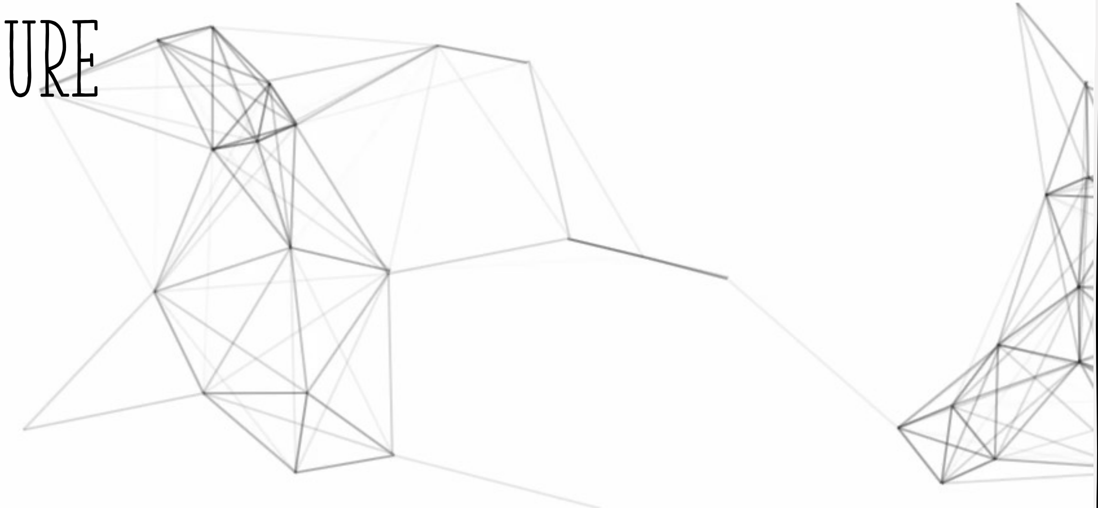
Published: September 20, 2018 • https://doi.org/10.1371/journal.pone.0203958

Misinformation spreads like a virus, but not just any virus–it behaves more like a *complex* contagion. They often need to see it multiple times from trusted people in their own group before they believe and spread it.

Method: Network simulation with diffusion.

Finding: Echo chambers make fake news spread faster and farther. Separate clusters counterintuitively intensify spread!

STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Visual Network Analysis
6. Community detection
7. 6 degrees of separation and the strength of weak ties
8. Case study: Zwarte Piet debate on Twitter

VISUAL NETWORK ANALYSIS

Visual network analysis involves the use of graphical representations to understand, interpret, and analyze the structure and characteristics of networks. This approach is particularly useful in network science, as it provides a way to intuitively comprehend complex relationships and patterns that might be difficult to discern through numerical data alone.

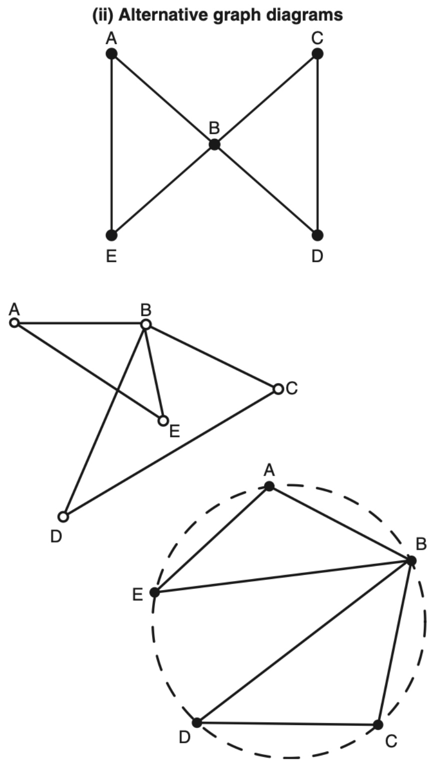


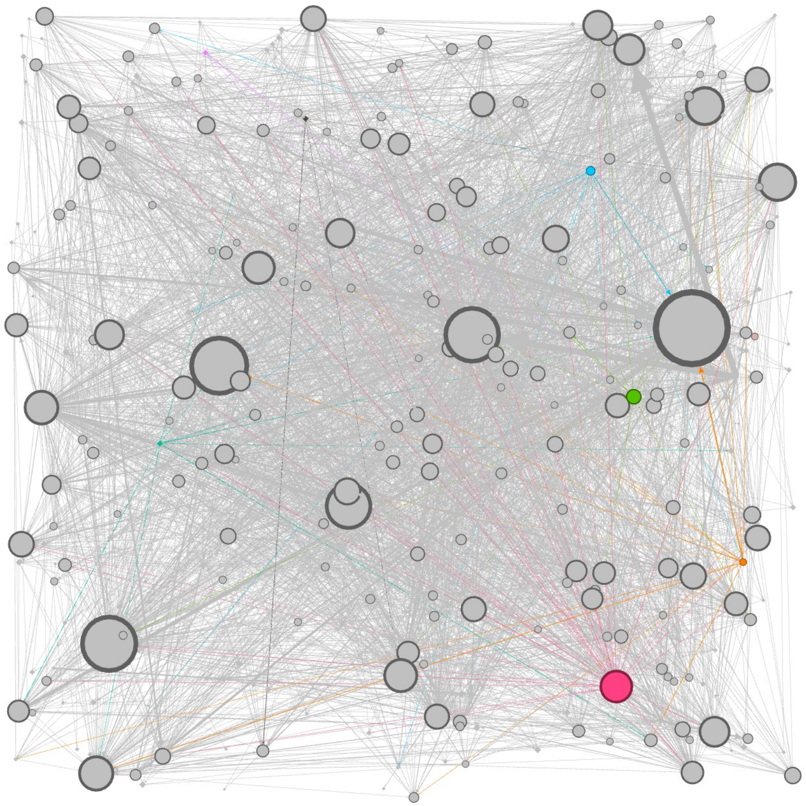
Figure 5.1 Alternative drawings of a graph

LAYOUT ALGORITHMS

Layout algorithms in network analysis are used to determine the position of nodes and the arrangement of edges in a network visualization. Force-Directed Layouts: These simulate physical forces, treating edges as springs and nodes as objects repelling each other.

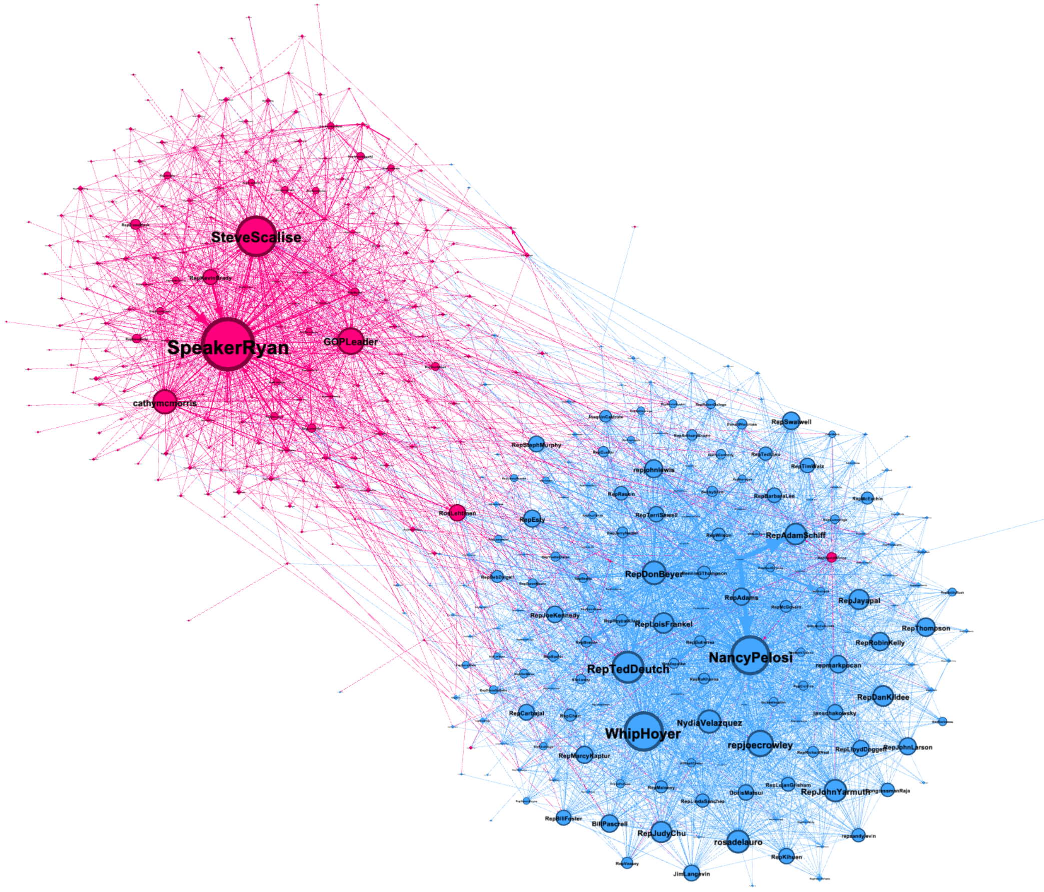
- Circular Layouts: These place nodes in a circle, which can be useful for highlighting cyclic structures and ensuring that all nodes are equally visible.
- ForceAtlas 2 powerful and versatile force-directed layout.
- Fruchterman-Reingold: A type of force-directed algorithm that aims to produce aesthetically pleasing layouts by considering attractive forces (like springs) and repulsive forces (like charged particles).
- Kamada-Kawai: Another force-directed approach, focusing on accurately representing the distance between nodes in the layout, ideal for emphasizing the global structure of the network.
- Hierarchical Layouts: These are useful for directed networks (like organizational charts or decision trees), where nodes are arranged in layers, emphasizing the directionality of connections.
- Grid Layouts: These arrange nodes in a regular grid, which can be useful for networks with an inherent grid-like structure or when uniformity and alignment are important.

PARLIAMENTARY RETWEET NETWORK

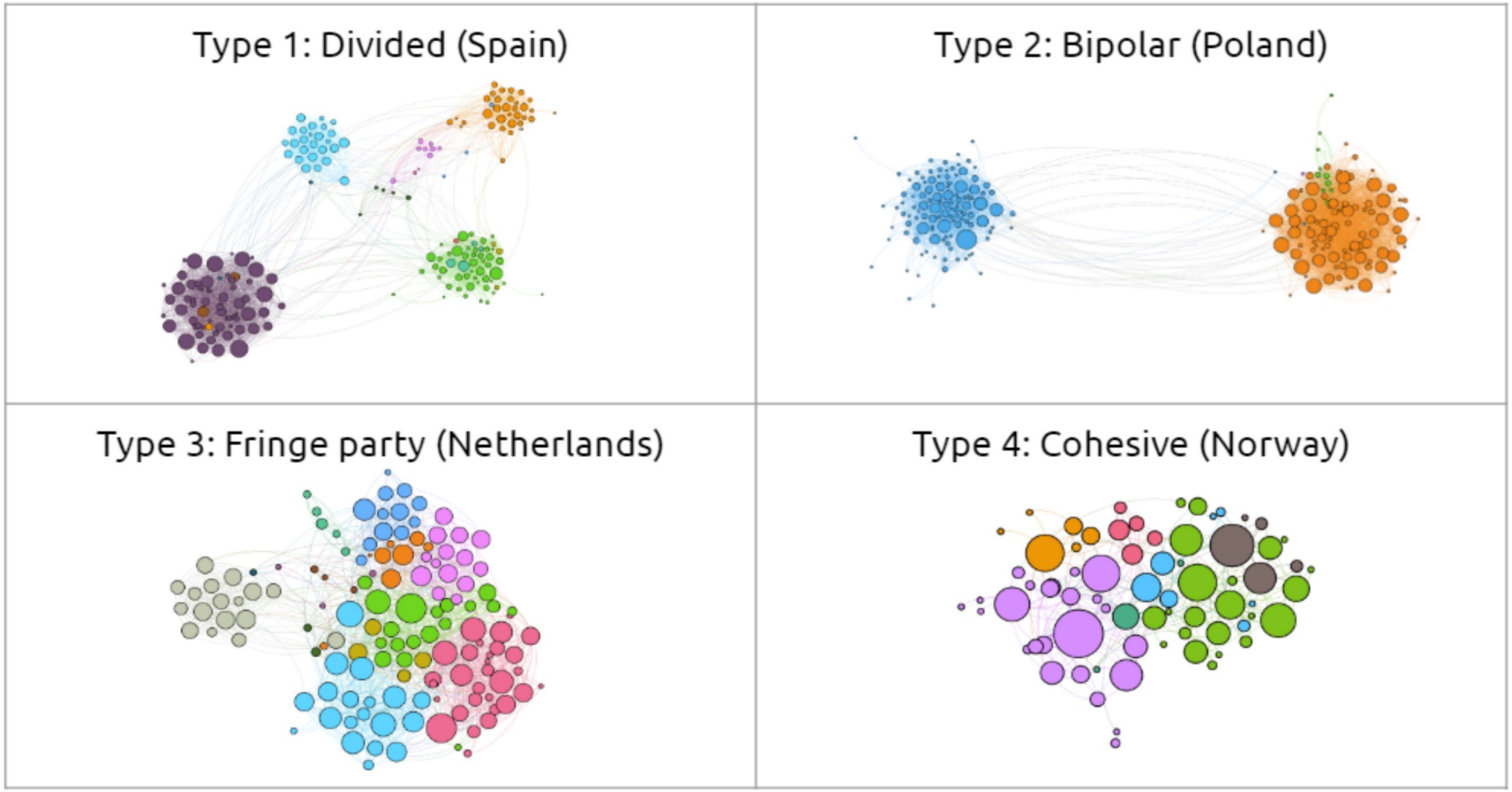


PARLIAMENTARY RETWEET NETWORK

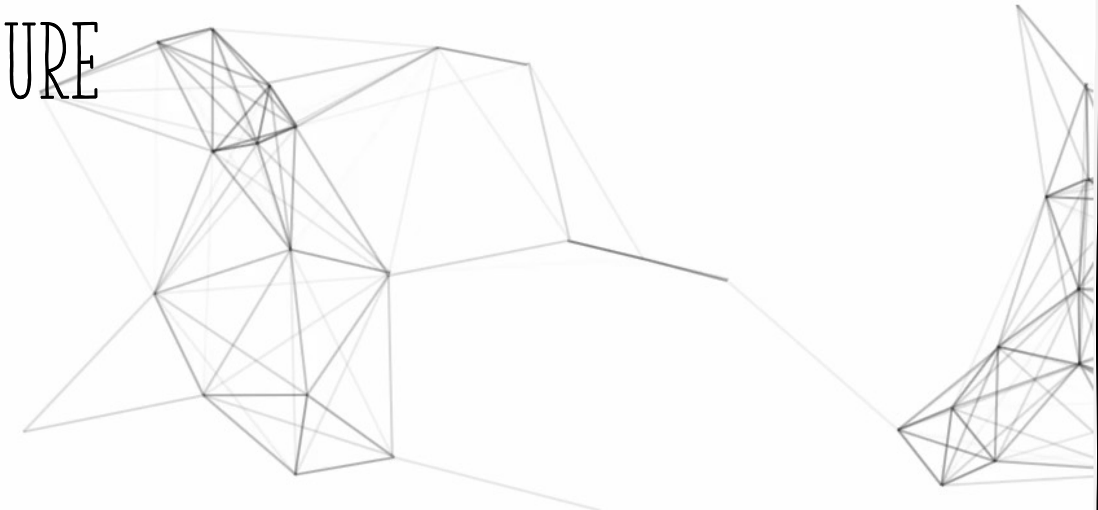
What can we say based on what we see here?



How does polarization look in multi-party systems?



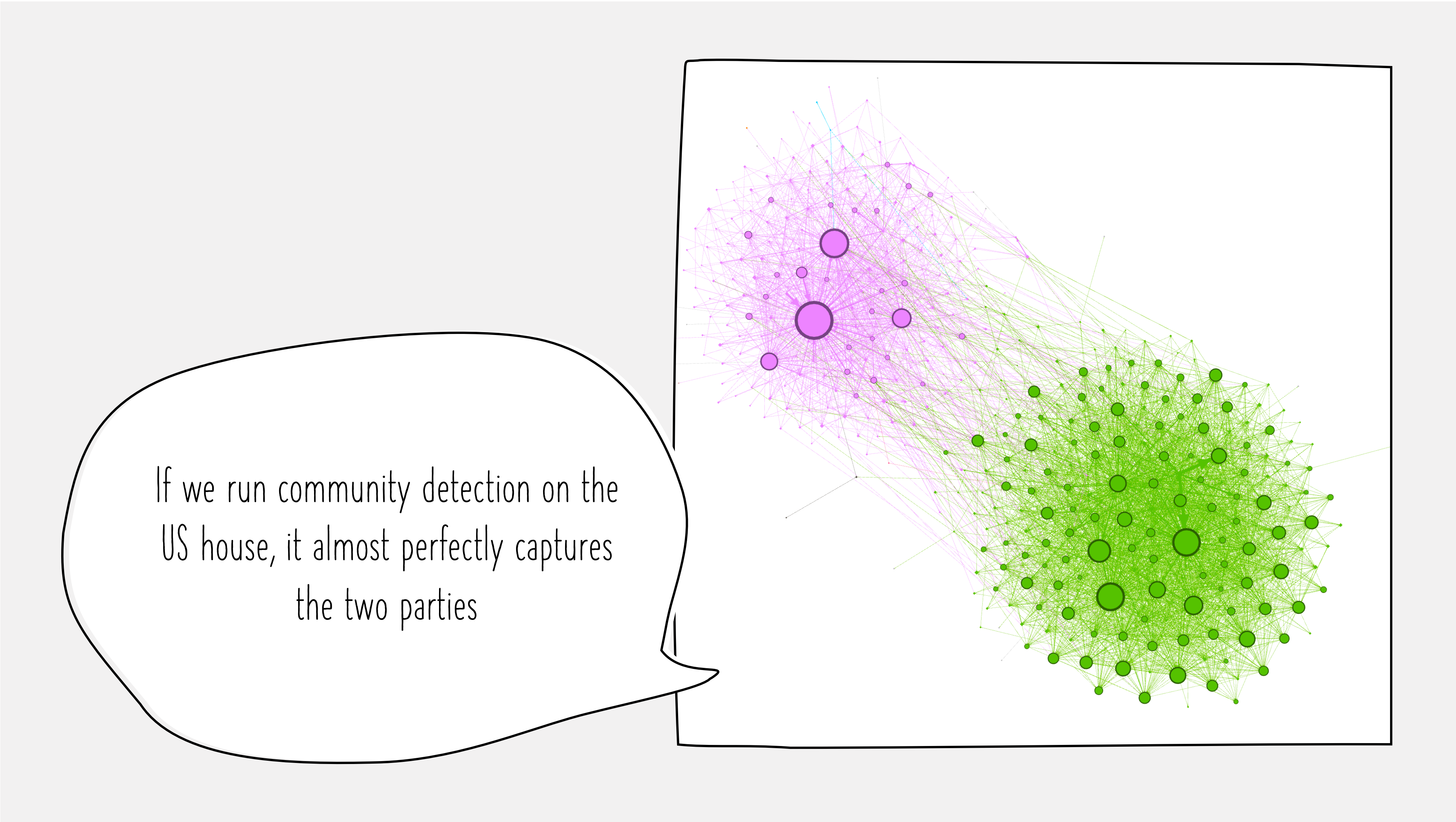
STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Representing a network: data and formats
6. Visual Network Analysis
7. Community detection
8. 6 degrees of separation and the strength of weak ties
9. Case study: Zwarte Piet debate on Twitter

COMMUNITY DETECTION

- Identify communities: groups of nodes that are more densely connected to each other than to the rest of the network.
- These communities often represent real-world structures like social circles, friend groups, or groups with common interests.
- Detecting these communities can be crucial for understanding the structure and function of social networks.

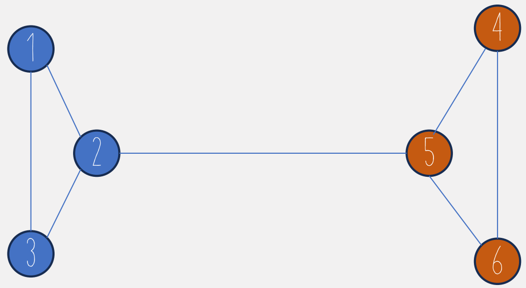

If we run community detection on the US house, it almost perfectly captures the two parties

CALCULATING MODULARITY

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

- A_{ij} is the element of the adjacency matrix (1 if nodes i and j are connected, 0 otherwise).
- k_i and k_j are the degrees of nodes i and j.
- m is the total number of edges in the network.
- c_i and c_j are the communities of nodes i and j.
- $\delta(c_i, c_j)$ is the Kronecker delta function (1 if $c_i = c_j$, 0 otherwise).

NOTE: Modularity unfortunately doesn't work well to compare large and small networks.

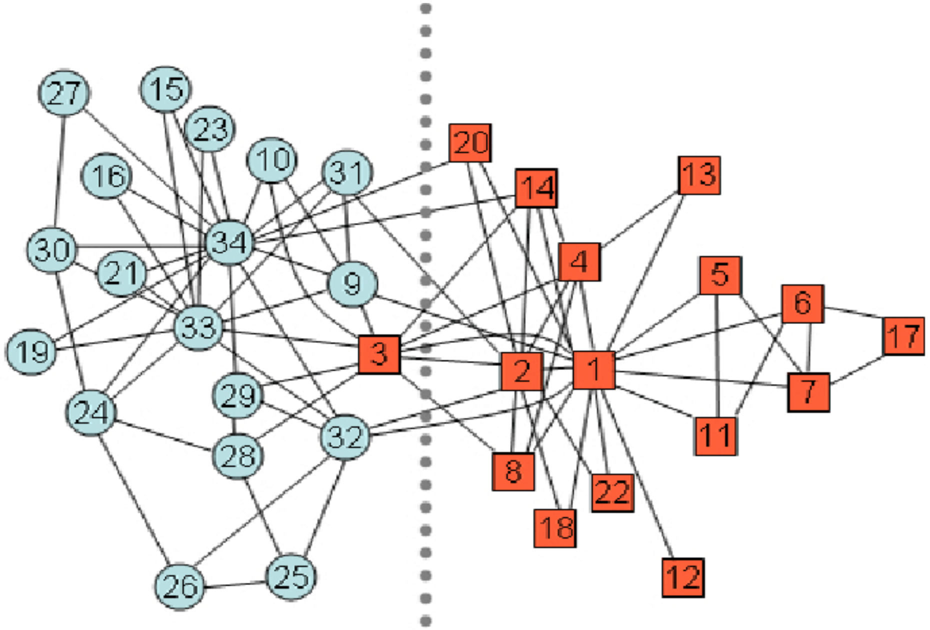


Basically, you calculate the number of edges between the communities and the edges within, and look at the fraction - while taking into consideration that some nodes are highly connected.

COMMUNITY DETECTION METHODS:

- Modularity-based methods try to optimize modularity.
- Clique Percolation Method (CPM): find overlapping subgroups (cliques) within the network and then building communities from these cliques.

The Leiden Method (builds on Louvain) is a popular approach: it's a hierarchical clustering algorithm that iteratively groups nodes to optimize the modularity score.



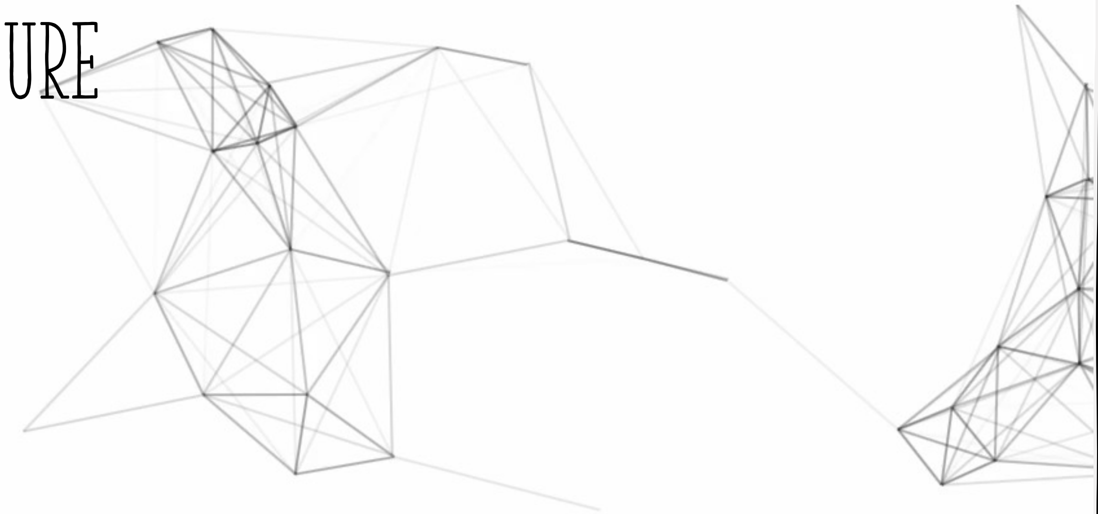
ZACHARY'S KARATE CLUB

A social network of a karate club was studied by Wayne W. Zachary for a period of three years from 1970 to 1972. The network captures 34 members of a karate club, documenting links between pairs of members who interacted outside the club.

By analyzing the network Zachary managed to predict exactly where the split would take place.

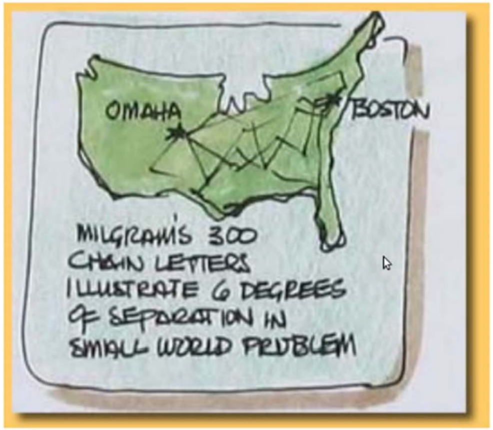
"An Information Flow Model for Conflict and Fission in Small Groups" by Wayne W. Zachary

STRUCTURE OF THIS LECTURE

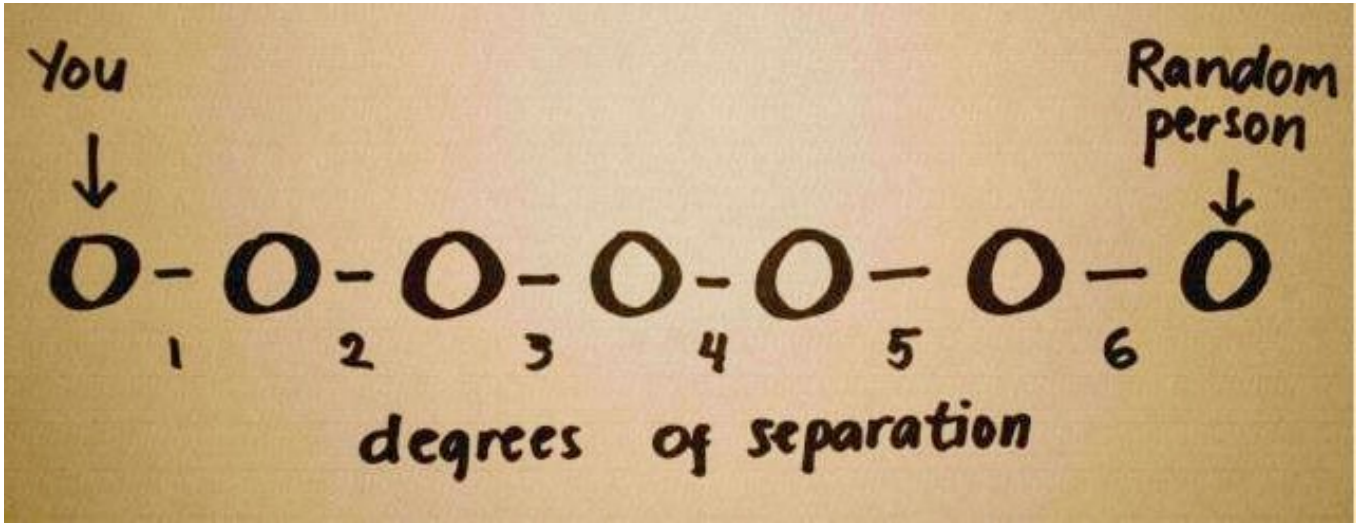


1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Representing a network: data and formats
6. Visual Network Analysis
7. Community detection
8. 6 degrees of separation and the strength of weak ties
9. Case study: Zwarte Piet debate on Twitter

SMALL WORLD EXPERIMENT



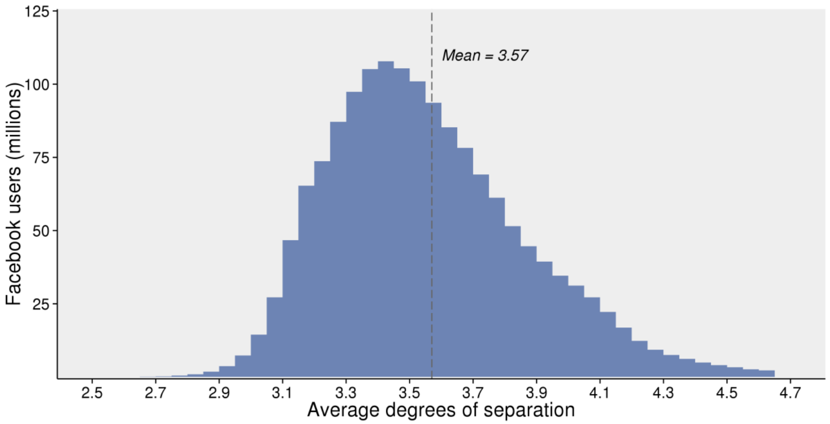
- Stanley Milgram
- Picked 300 random people in Omaha
- Ask them to get a letter to a stock-broker in Boston by passing it through friends
- How many steps did it take?
- Letters arrived after on average 6.2 steps
- Total of 64 chains completed

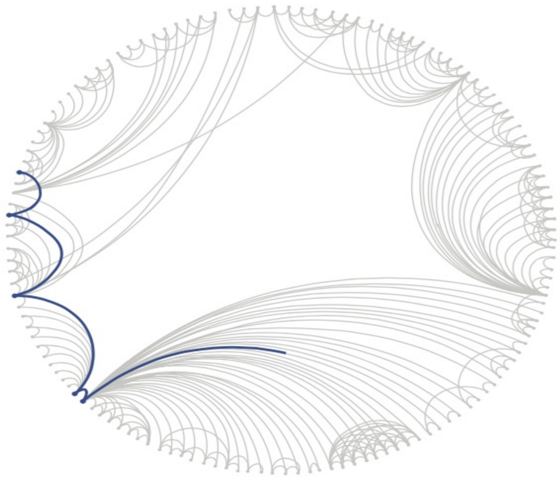

You
Random person
1
2
3
4
5
6
degrees of separation

1. You have 100 friends	100
2. Your friends have 100 friends	100*100 = 10,000
3. Your friends' friends have 100 friends	100*100 = 10,000
4. Your friends' friends' friends have 100 friends	100*100*100 = 1,000,000
5. Your friends' friends' friends' friends have 100 friends	100*100*100*100 = 100,000,000
6. Your friends' friends' friends' friends' friends have 100 friends	100*100*100*100*100 = 10,000,000,000
....	
n. Your friends' ... friends have 100 friends	100^n


facebook
December 2010

- Each person in the world (at least among the 1.59 billion people active on Facebook) is connected to every other person by an average of three and a half other people.
- The average distance we observe is 4.57, corresponding to 3.57 intermediaries or "degrees of separation."
- Within the US, people are connected to each other by an average of 3.46 degrees.

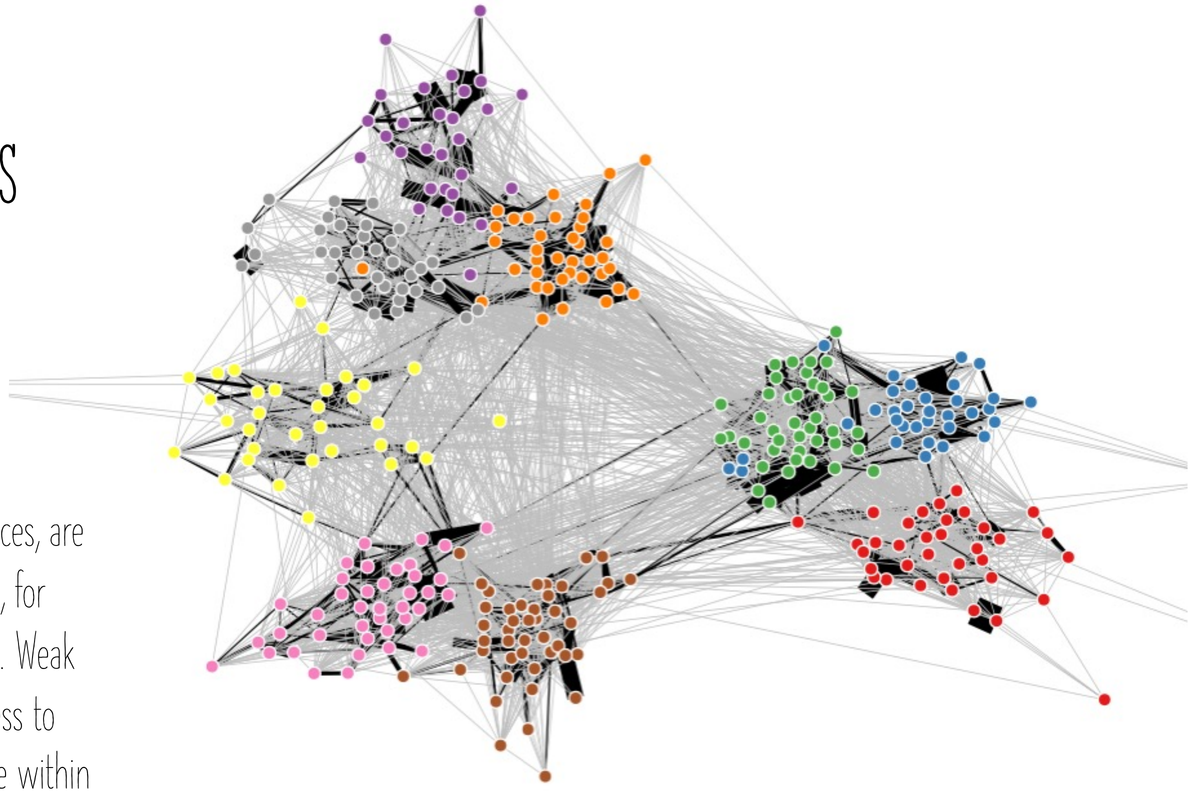




https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

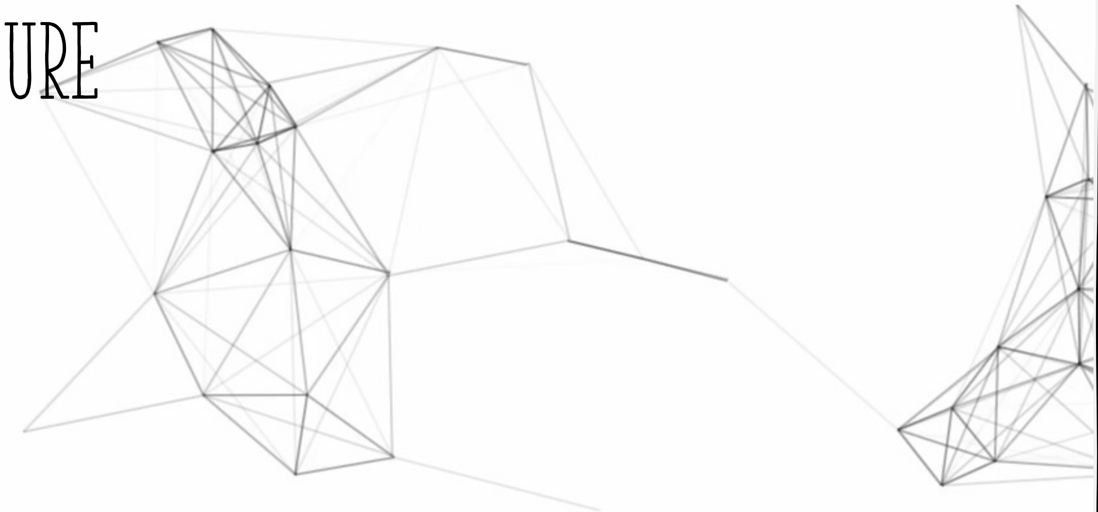
THE STRENGTH OF WEAK TIES

- Weak ties are links that have a low weight.
- Strong ties are our close friends. These tend to be clustered.



<u>Mark Granovetter</u>: weak social ties, such as acquaintances, are often more valuable than strong ties, like close friends, for accessing new information, opportunities, or resources. Weak ties form bridges to different networks, providing access to diverse, non-redundant information that isn't available within one's close-knit social circle.

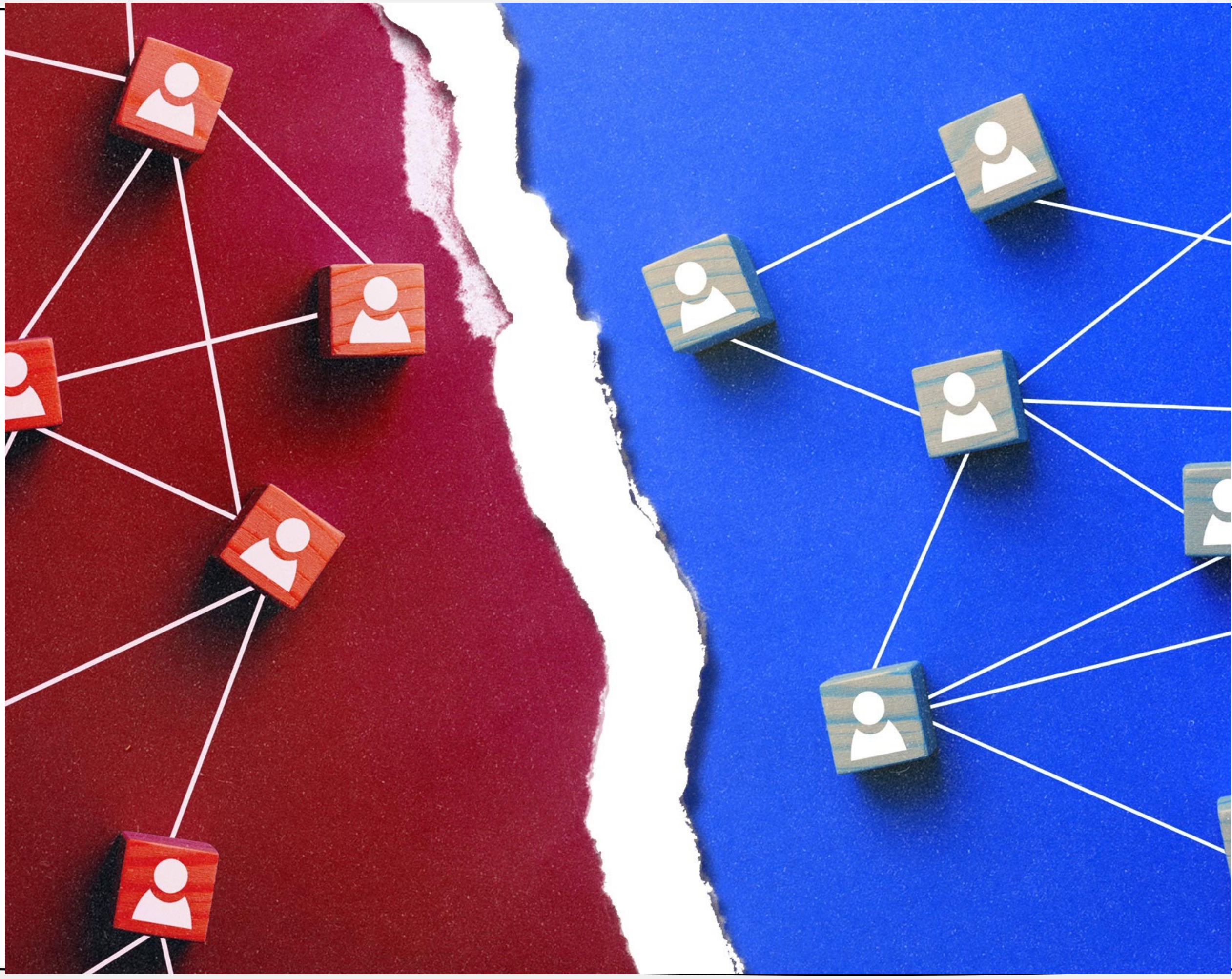
STRUCTURE OF THIS LECTURE



1. Questions you can answer with SNA
2. A brief history of social networks
3. Is there a theory of social networks?
4. Network basics: terms and foundations
5. Representing a network: data and formats
6. Visual Network Analysis
7. Community detection
8. 6 degrees of separation and the strength of weak ties
9. Case study: Zwarte Piet debate on Twitter

CASE STUDY:

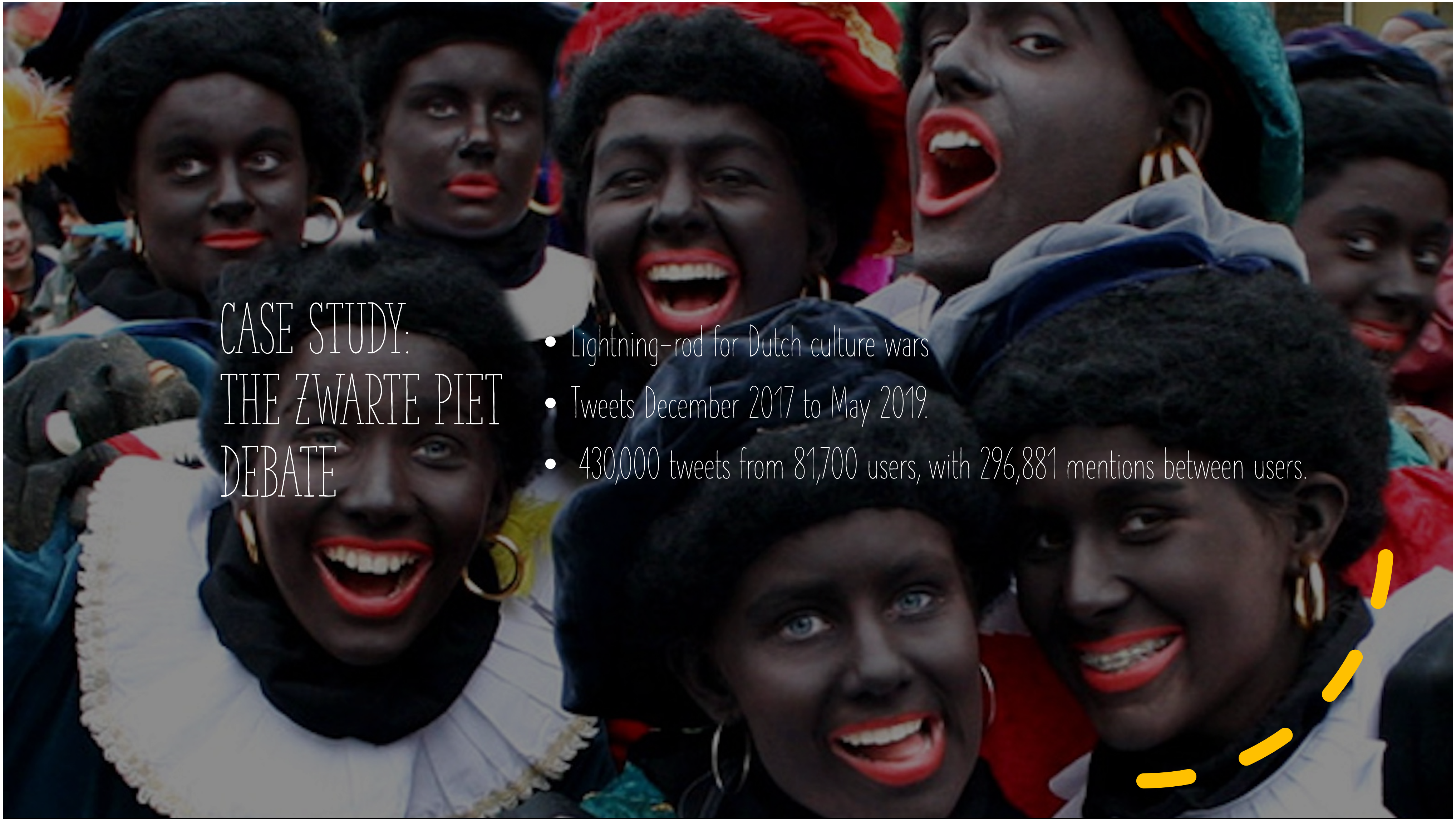
Are there echo chambers on Twitter?



"ECHO CHAMBER" OR "FILTER BUBBLE"

- Cass Sunstein: social media allow us to segregate into homogeneous groups with likeminded others (Sunstein 1999) Deliberative enclaves or "echo chambers": "breeding ground for group polarization and extremism" (2002)

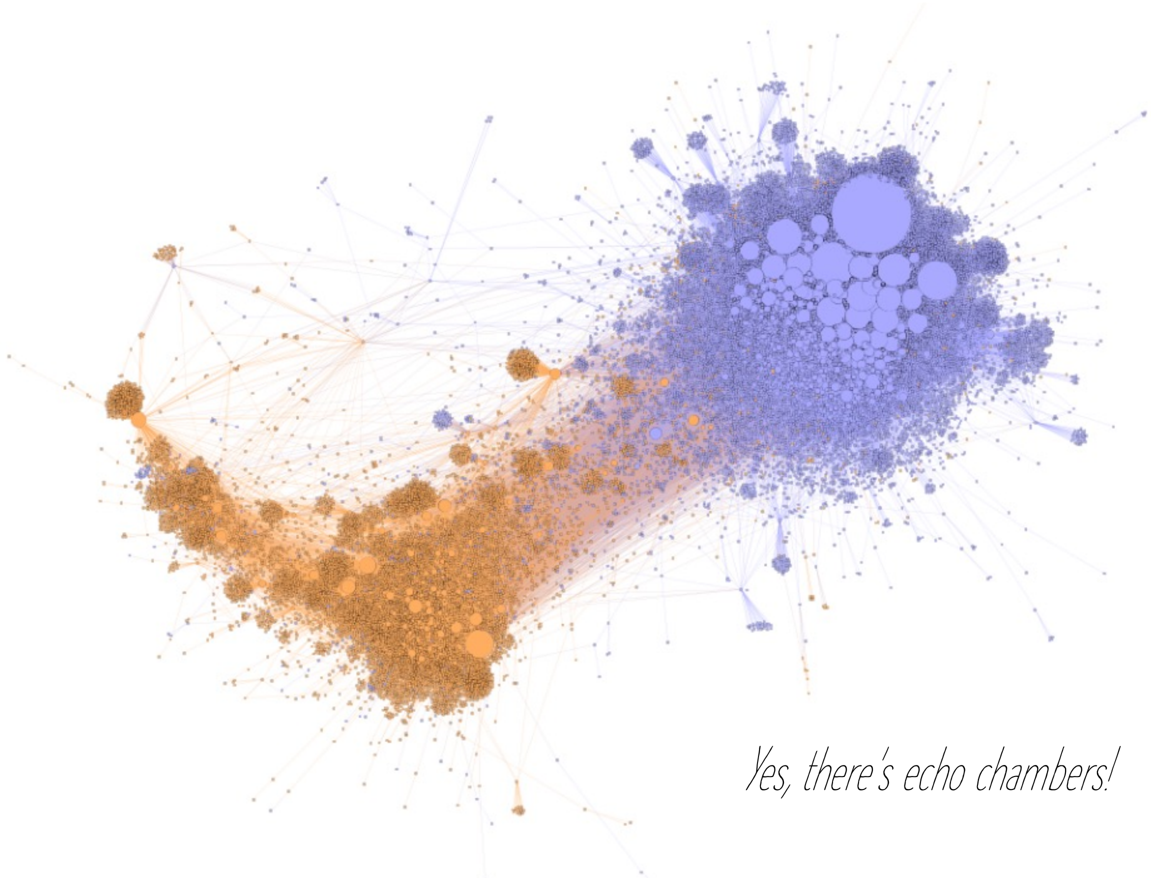
- One-sided arguments: Feedback loop between isolation and radicalization
- More extreme opinions



CASE STUDY: THE ZWARTE PIET DEBATE

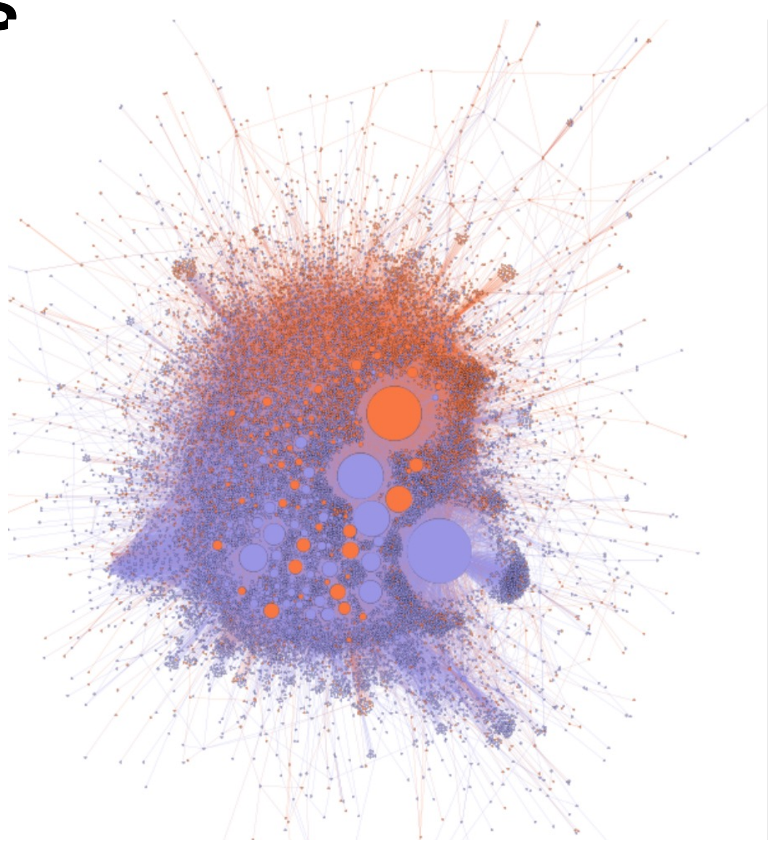
- Lightning-rod for Dutch culture wars
- Tweets December 2017 to May 2019.
- 430,000 tweets from 81,700 users, with 296,881 mentions between users.

Retweets



Yes, there's echo chambers!

Mentions

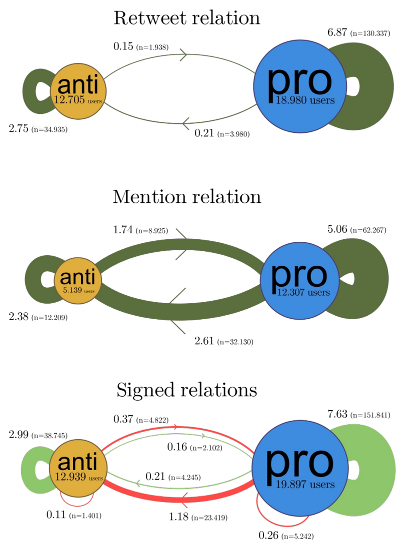


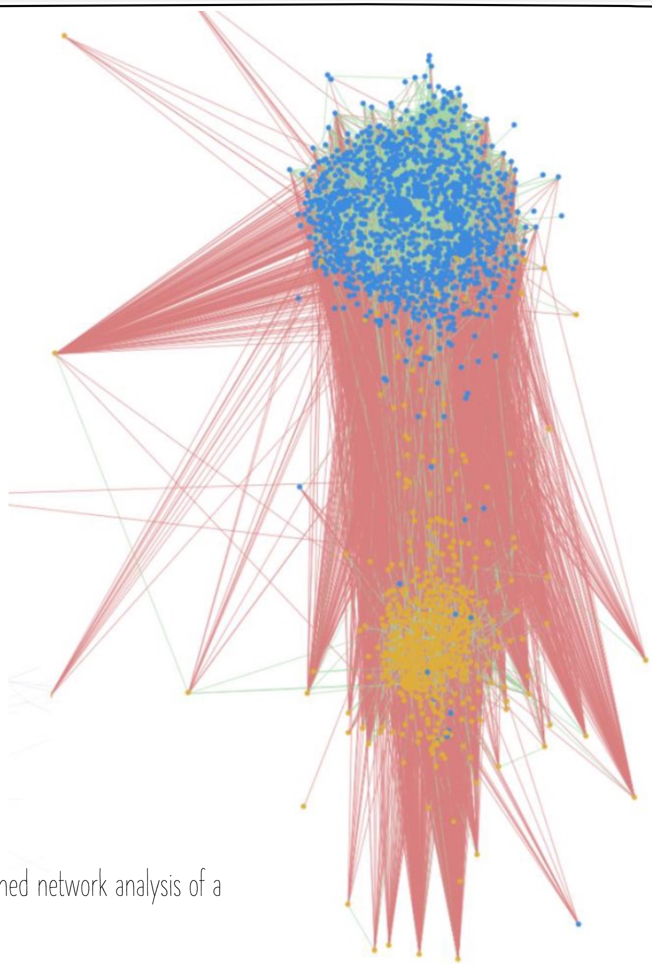
No, there's no echo chambers!

Communities of conflict

- Trained a machine learning algorithm to identify whether links were positive or negative.

- The groups are defined as much by external attacks as by internal support
- Conversation is not rational exchange, but predominantly conflictual and often uncivil – 'outgroup derogation'





Keuchenius, A., Törnberg, P., & Uitermark, J. (2021). Why it is important to consider negative ties when studying polarized debates: A signed network analysis of a Dutch cultural controversy on Twitter. PloS one, 16(8).

SOME TERMS WE COVERED

Node?
Edge?
Bridge?
Weighted?
Structural hole?
Component?
Homophily?

Degree?
Density?
Clustering?
Eigenvector centrality?
Reciprocity?
Assortativity?
Transitivity?

HOW DOES SOCIAL NETWORK ANALYSIS FIT YOUR PROJECT?

Centrality Analysis: Identifies the most important or influential individuals within a network.

Community Detection: Unearths the structure of the network by identifying clusters or communities within it.

Network Dynamics and Change Over Time: Examines how social networks evolve, and who forms ties with whom.

Structural Holes and Network Brokers: Analyzes the gaps in a social network (structural holes) and identifies brokers who bridge these gaps.

Network Cohesion and Subgroup Analysis: Studies the strength of relationships within a network to identify tight-knit groups or subgroups. This can help in understanding the social cohesion or fault lines within a community.

Role Analysis: Looks at the roles individuals play within a network based on their relationships and interactions. It can identify roles like leaders, gatekeepers, or peripheral members.

Diffusion of Information and Contagion: Examines how information, behaviors, or innovations spread through a network. This is particularly relevant for understanding phenomena like viral marketing, spread of rumors, or adoption of new technologies.

Social Capital Analysis: Explores the resources available to individuals through their social networks. It can shed light on how social ties contribute to personal or organizational success.