Network Analysis (INFOMNWA-2021)

Lecture 5: Social network and its measures

Jiamin Ou

Introducing myself

- **Interdisciplinary researcher** in energy, environment and human systems
- **Current project**: **EN**gaging **R**esidents in **G**reen energy **I**nvestments through **S**ocial networks, compl**E**xity, and **D**esign. (http://enrgised.nl/, funded by NWO)
- **Previously:**

  International Development (PhD), University of East Anglia, UK

  International Institute for Applied Systems Analysis (IIASA), Austria

  University of Texas at Austin, US

- **Awards:**

  Mikahlevich Award, by IIASA, 2018

  Green Talent, by the German Federal Ministry of Education and Research (BMBF), 2020

Also in this course

Kevin Wittenberg

PhD candidate Sociology

Project on sustainable cooperation: citizen collectives for health care (Dutch: zorgcoöperaties)

Research interests

- Social networks, cooperative relations, statistics, opinion dynamics, philosophy of science.

Thom Volker

Second-year MSc. Methods and Statistics

First-year MSc. Sociology and Social Research

Research interests (theses) are focused on

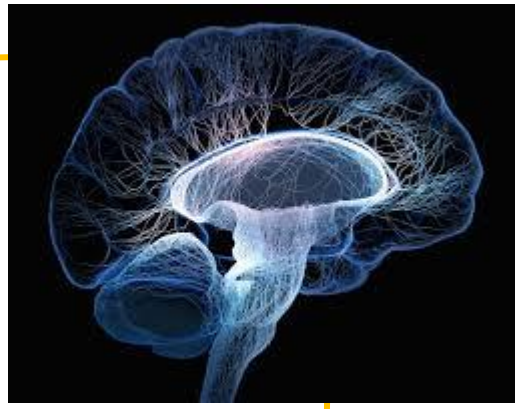
- Aggregating results from studies with divergent designs
- Cooperation in social dilemmas (e.g., Prisoner's Dilemma)
- Creation of synthetic data where privacy and confidentiality constraints prehibit sharing the observed research data

Lecture 1: Principles of deep learning in artificial networks

Lecture 2: Deep learning in biological neurons and networks

Lecture 3: Early and feedforward visual processing

Lecture 4: Higher and recurrent visual processing



Neural Networks
Unit: neurons

Lecture 5: Social network and its measures

Lecture 6: Network formation

Lecture 7: Simple and complex contagion

Lecture 8: Influence manipulation

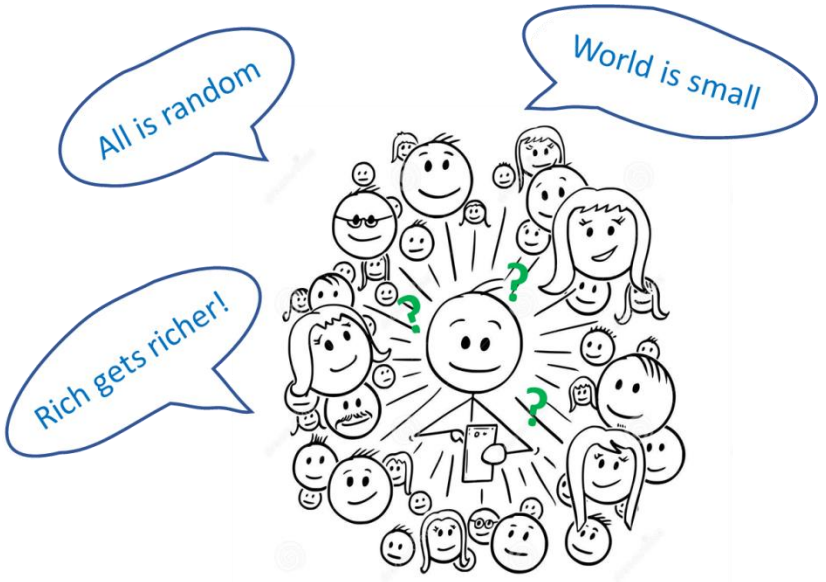
Social Networks
Unit: people, organization

Lecture 5: Social network and its measures

What is social network and how to measure it?

Lecture 6: Network formation

How are individuals or complex systems wired together?

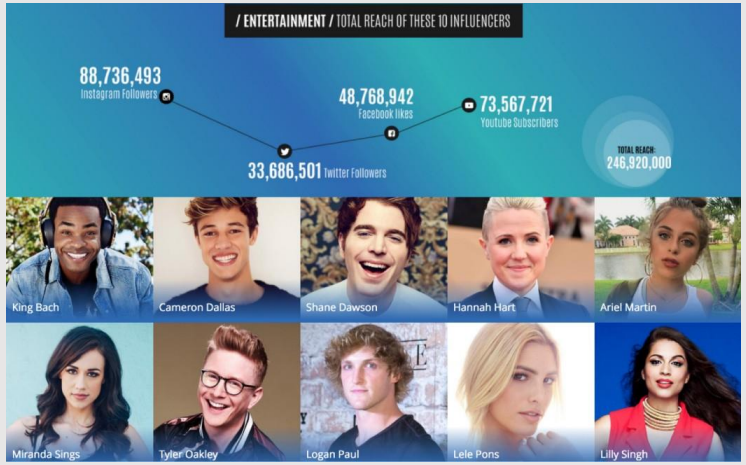


Lecture 7: Simple and complex contagion

How does influence transmit through the network?

Lecture 8: Influence manipulation

How can we model and control the spread of influence?



Course goals

- What to look for in a social network: the perspectives and metrics
- Social contagion theory and empirical models (Independent Cascade, Threshold model…)
- Apply the knowledge, theory and models in social networks as a descriptive and predictive tool for real-world problems

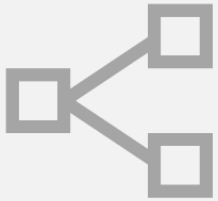
Program for today

Introduction of social network

-Why we need network thinking

-Definition of social network

-Development of social network analysis



Measuring social network

-Directed and undirected graph

-Node-level metrics

-Group-level metrics

-Example of Facebook user network

WE ALL CONNECT, LIKE A NET WE CANNOT SEE

—— *Mickenberg and Dugan, Taxi Driver Wisdom, 1995*

Imagine: To study companies' possibilities to support an NGO

Standard approaches:

- Define your potential targets
- By interview and survey to measure their variety (e.g., size, industry, profitability, their history to support NGO…)

Anything missing?

Imagine: To study companies' possibilities to support an NGO

Standard approaches:

- Define your potential targets
- By interview and survey to measure their variety (e.g., size, industry, profitability, , their history to support NGO …)

+ A network perspective

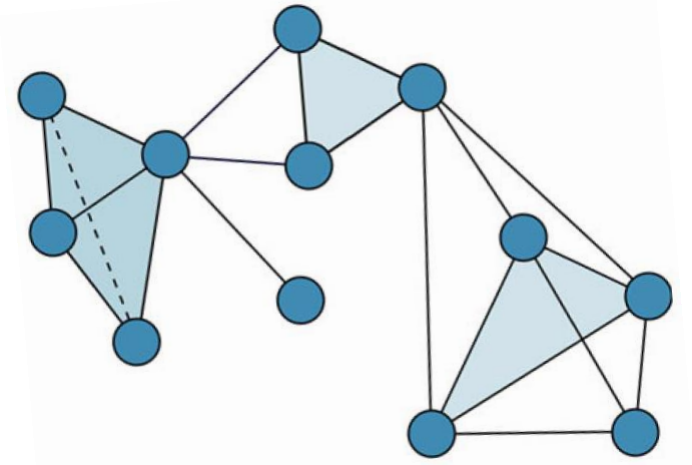
Within/between companies relations (e.g., memberships on each others' boards of directors, acquaintanceships of corporate officers, joint business dealings)

Manifestation of the social relations: A social network

Definition of a social network: “a graph made up of a set of social actors (such as individuals or organizations), and other social interactions between actors”

Significance of social network analysis: provides a mathematical statements of some social concepts, and make theory more testable

Individuals; organizations; other social players



Social interaction

Sociogram, Moreno, 1953: Precursor to SNA

L - 17

EMOTIONS MAPPED BY NEW GEOGRAPHY

Charts Seek to Portray the Psychological Currents of Human Relationships.

FIRST STUDIES EXHIBITED

Colored Lines Show Likes and Dislikes of Individuals and of Groups.

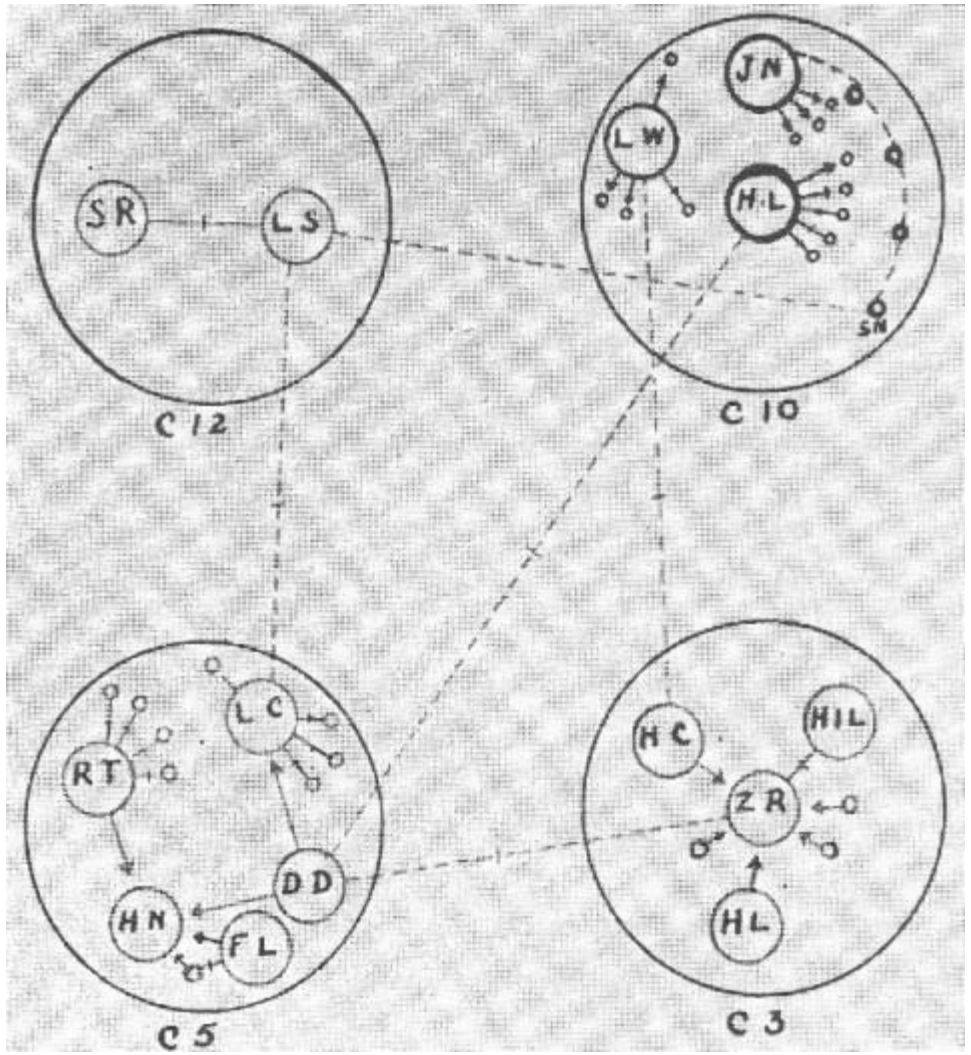
MANY MISFITS REVEALED

Dr. J. L. Moreno Calculates There Are 10 to 15 Million Isolated Individuals in Nation.

A new science, named psychological geography, which aims to chart the emotional currents, cross-currents and under-currents of human relationships in a community, was introduced here yesterday at the scientific exhibit of the Medical Society of the State of New York, which opens its 127th annual meeting here today at the Waldorf-Astoria.

The first series of maps of the new human geography were shown by Dr. Jacob L. Moreno of New York, consulting psychiatrist of the National Committee of Prisons and Prison Labor and director of re-

April 3, 1933
The New York Times



Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama.

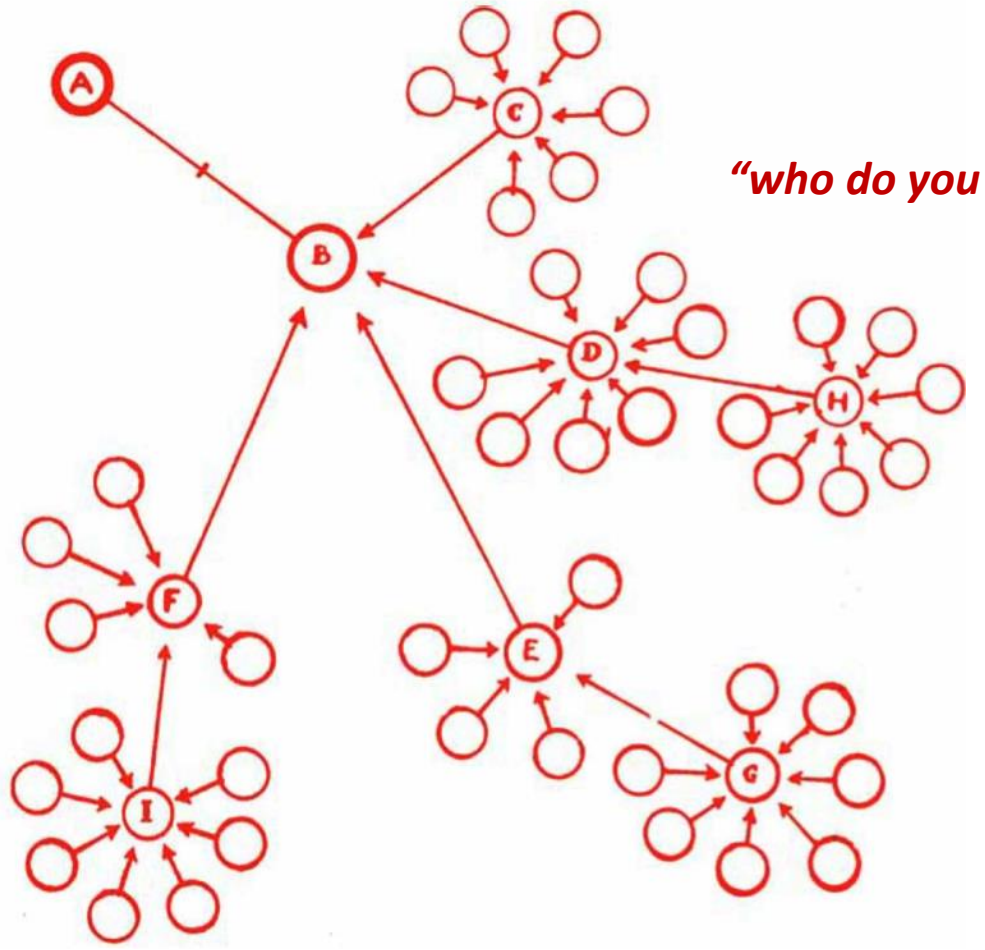
An epidemic of runaways among the girls at the Hudson school:14 girls runaway;

less to do with individual factors than with the positions of the runaways in an underlying social network

Emotional channels between girls run away

It was their location in the social network that determined whether and when they ran away

Sociogram, Moreno, 1953: Precursor to SNA



"who do you like"

Among a population of 435 persons, 204, or 46.5%, remained unchosen after the 1st choice; 139, or 30%, after the 2d choice; 87, or 20%, after the 3rd choice; 74, or 17%, after the 4th choice; and 66, or 15%, after the 5th choice. (Moreno 1934, p. 249)

15% of people being left behind.

Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama.

How SNA evolves over time…

1940's and 1950's, experimental structure to study group process, using points to depict actors and lines to depict channels of communications

1950's, used matrices to study social network

1950's 1960's, introduced the formal concepts of social network analysis (e.g., density, span, connectedness, clusterability…)

SNA: "developed out of a propitious meeting of social theory and application, with formal mathematical, statistical and computing methodology" *(Inherently interdisciplianry)*

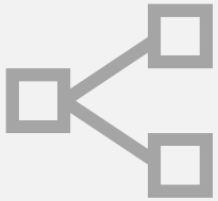
Program for today

Introduction of social network

-Why we need network thinking

-Definition of social network

-Development of social network analysis



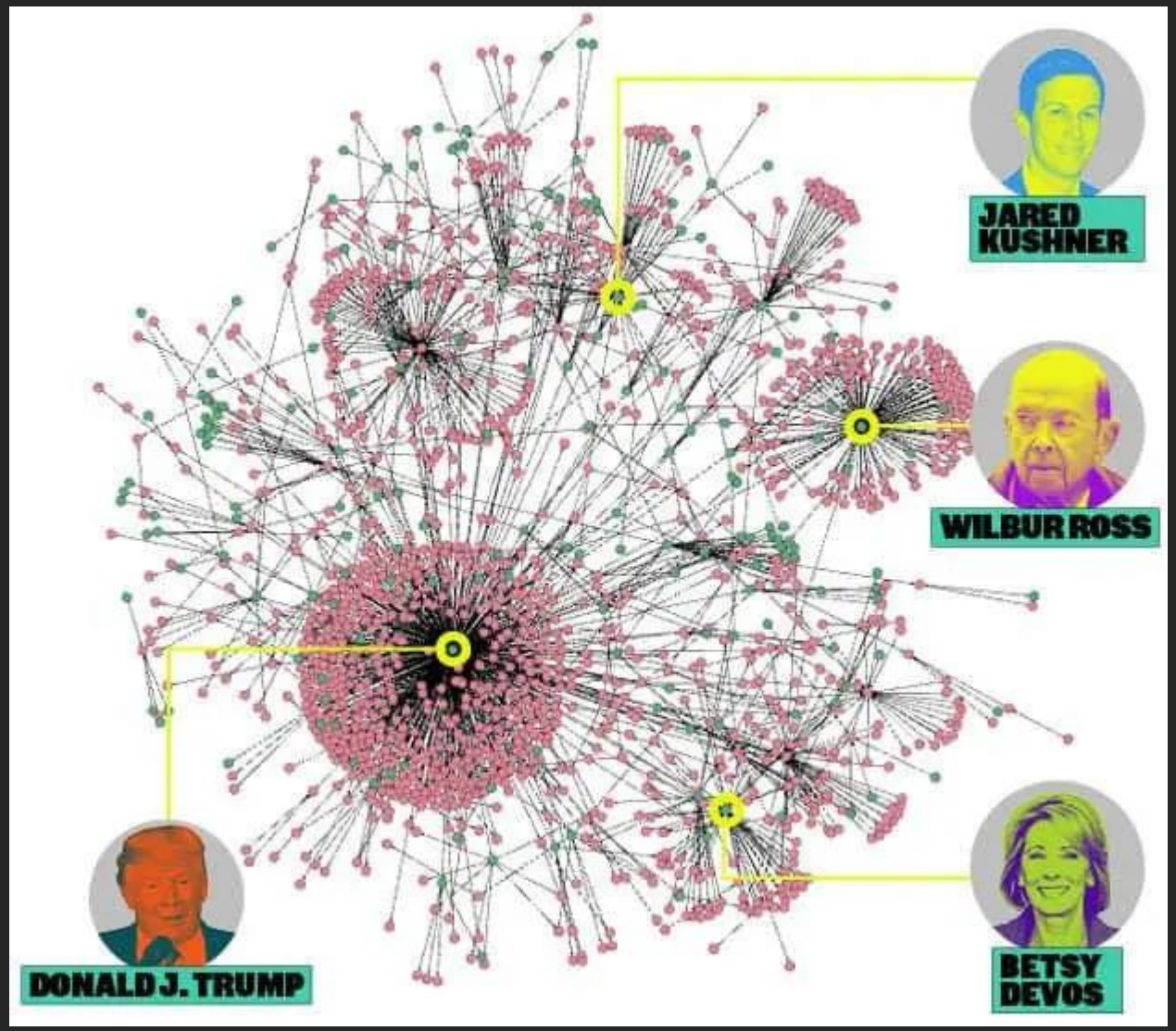
Measuring social network

-Directed and undirected graph

-Node-level metrics

-Group-level metrics

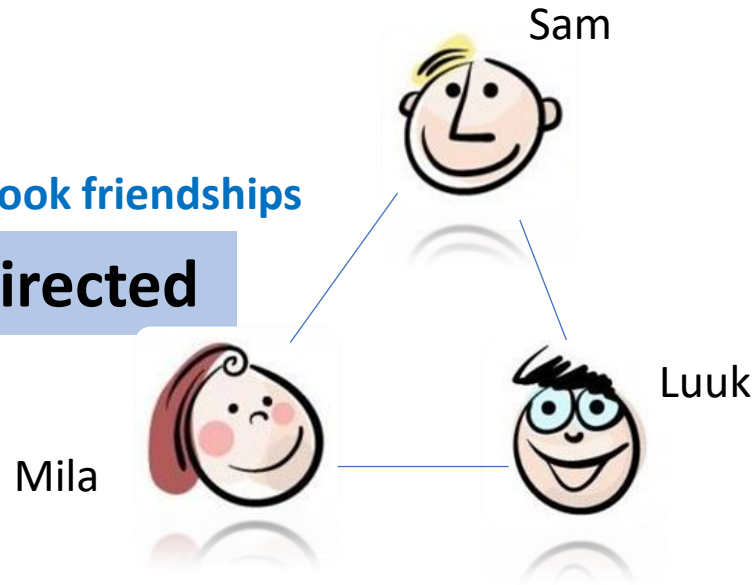
-Example of Facebook user network



When social network data is booming now…

- *What are the things you want to know in a network?*
- *How can you measure them?*

Directed and undirected graph



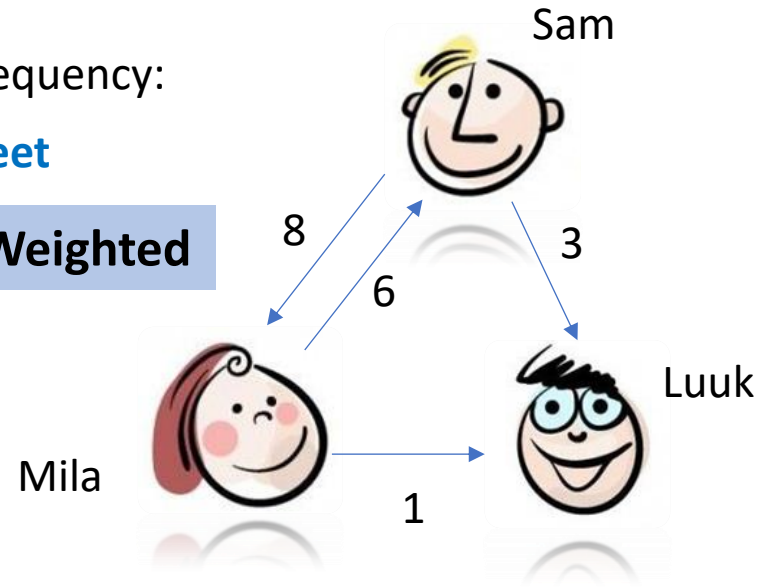
Facebook friendships

Undirected

Consider the frequency:

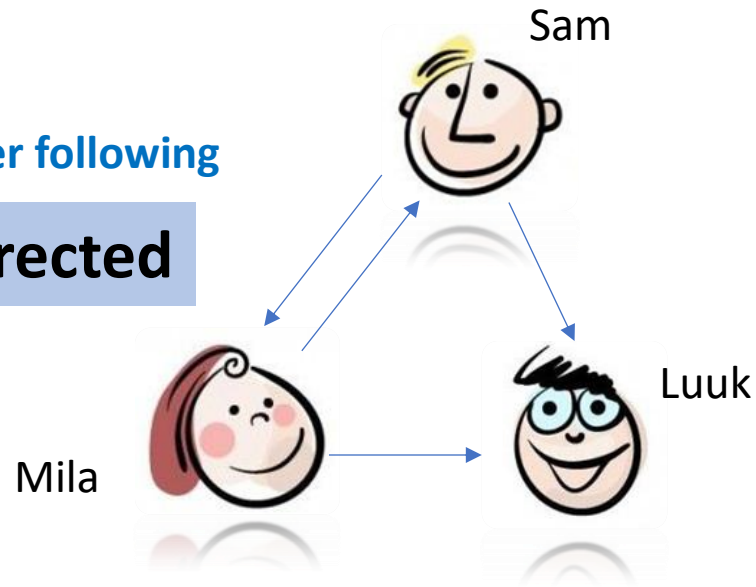
Retweet

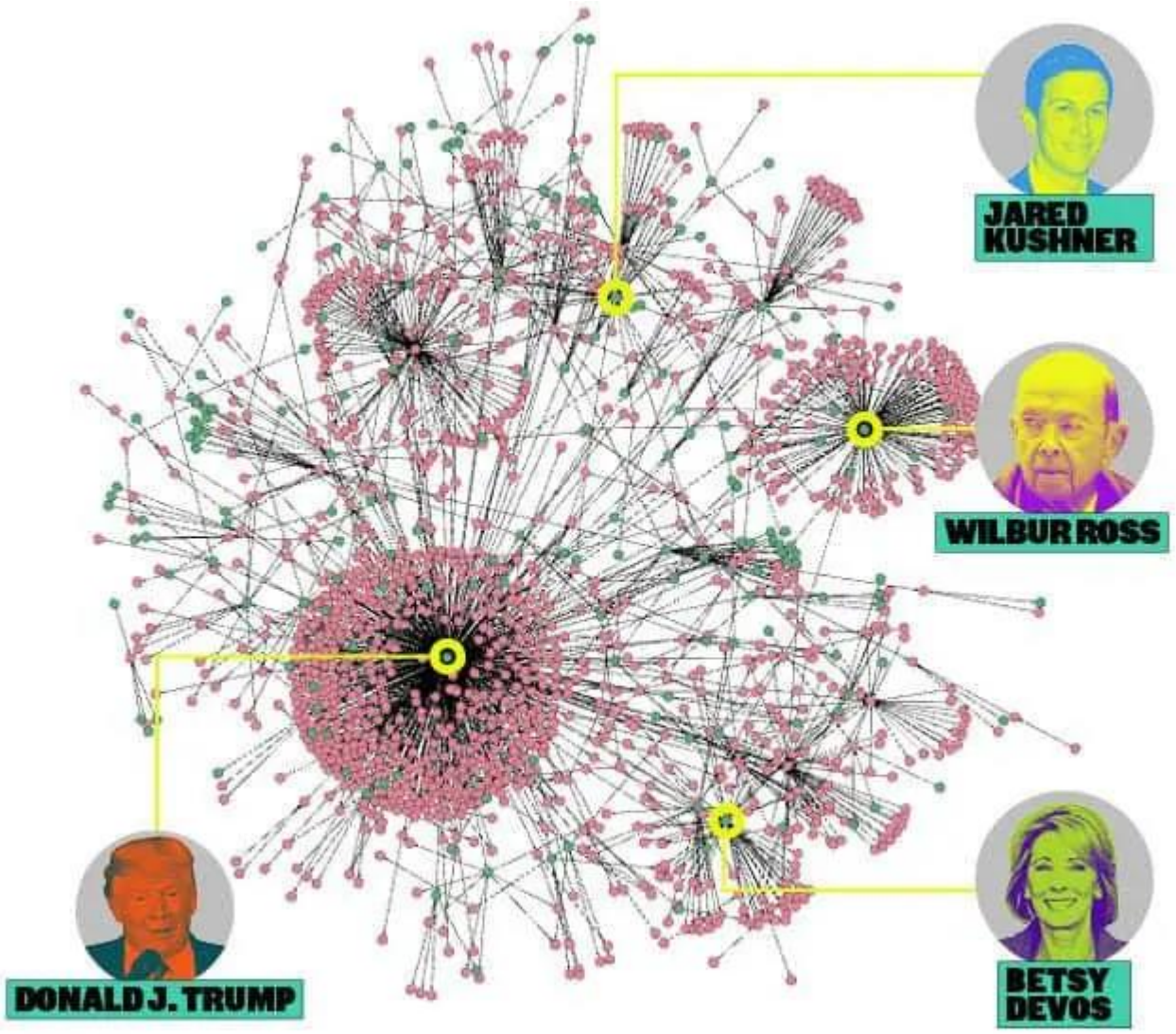
Directed and Weighted



Twitter following

Directed



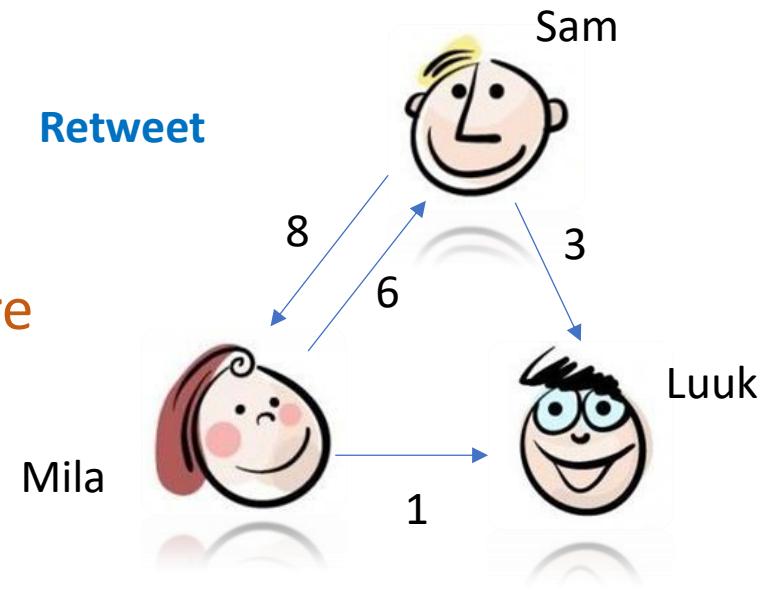


Who is the most central one in this network?

Node level network measures

- Degree (number of links)
  - Indegree
  - Outdegree
- Degree centrality
- Closeness centrality
- Betweenness centrality
- Eigenvector centrality
- Local clustering
- ….

How central you are

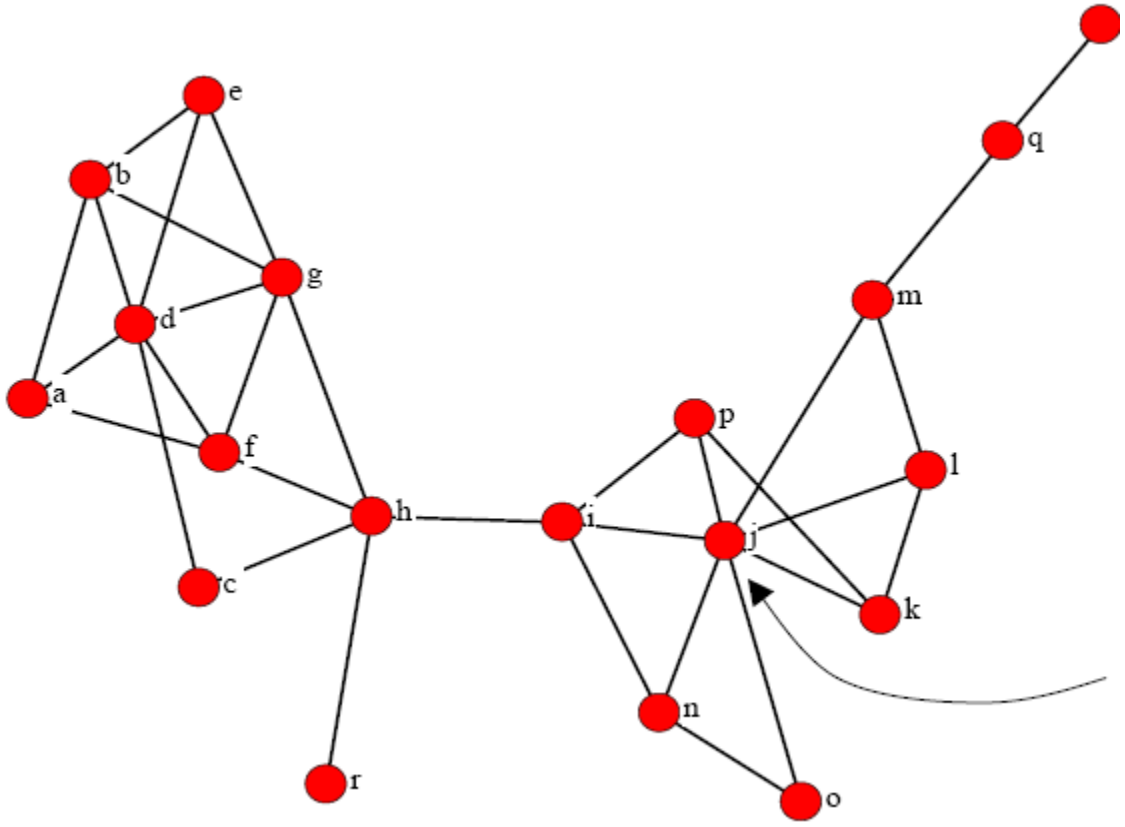


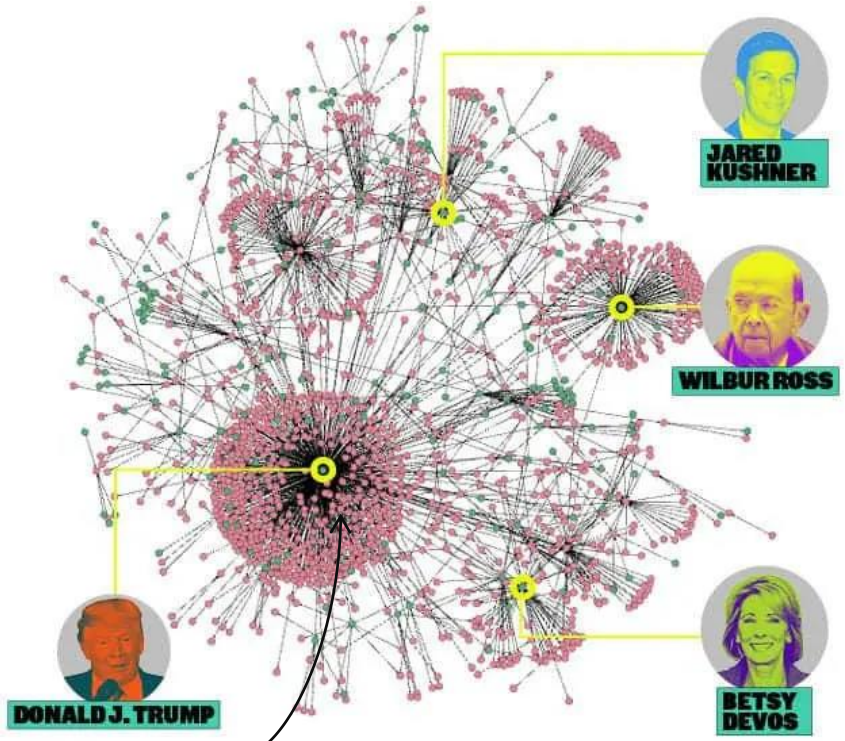
Mila's indegree: 8
Outdegree: 7

"It's all about location, location and location"

Degree Centrality

The number of nodes adjacent to a given node

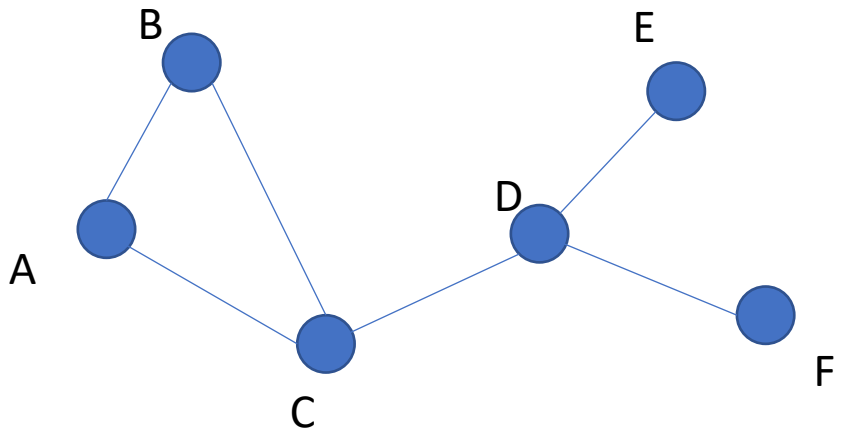




Highest degree centrality

Closeness centrality: the closest one to all other nodes

Sum of geodesic distances (shortest paths) to all other nodes



Shortest paths between A and others:

A to B: A-B 1
A to C: A-C 1
A to D: A-C-D 2
A to E: A-C-D-E 3
A to F: A-C-D-F 3

Sum of the shortest path from A to others: 10

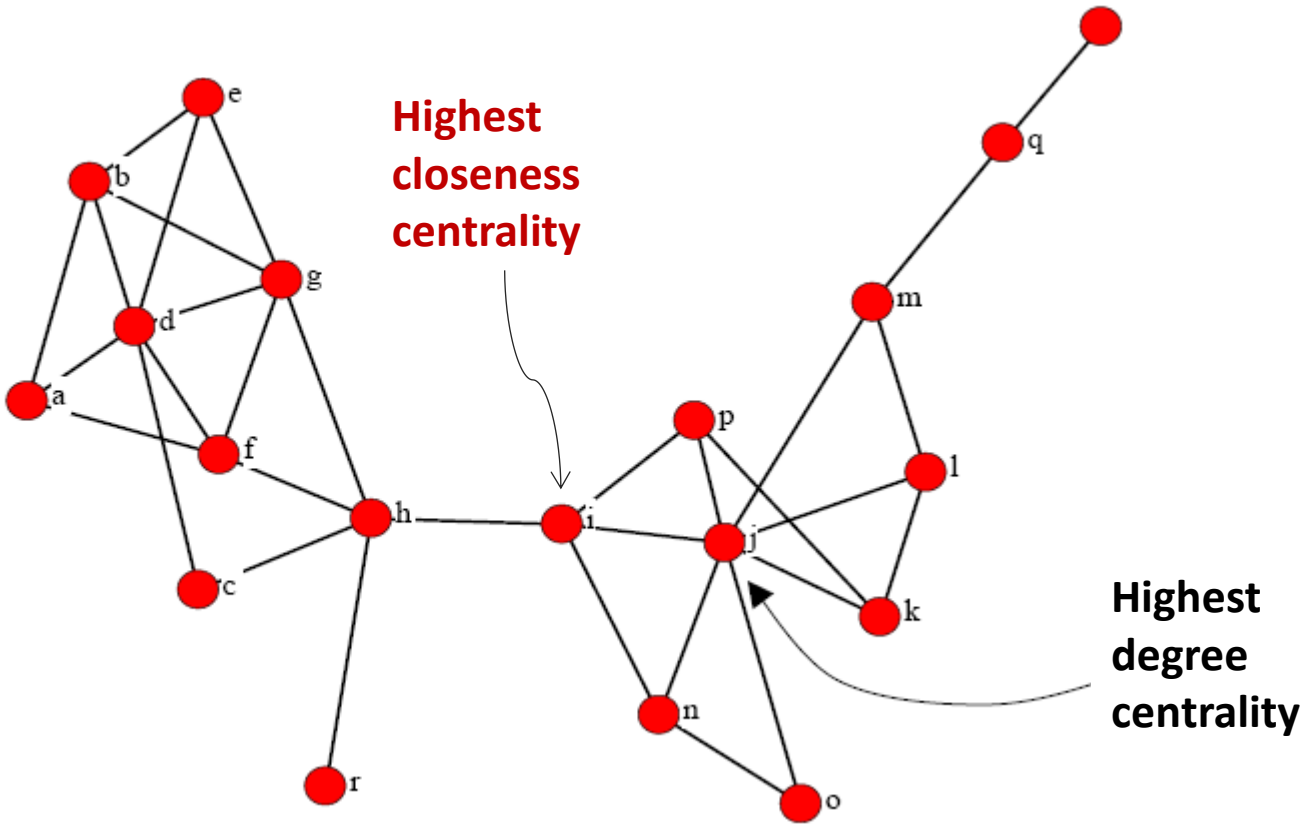
Shortest paths between C and others:

C to A: C-A 1
C to B: C-B 1
C to D: C-D 1
C to E: C-D-E 2
C to F: C-D-F 2

Sum of the shortest path from C to others: 7

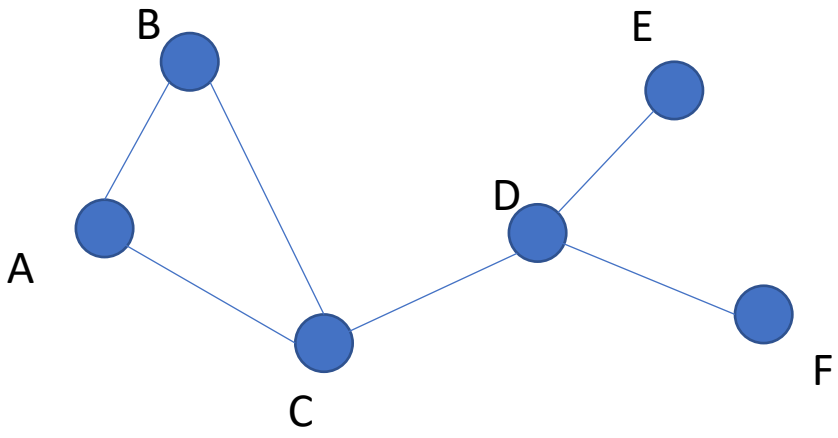
Node level measures

It takes the fewest steps for anyone to reach node *i*



Betweenness centrality: how often a node is a bridge between other nodes

Number of times that a node lies along the shortest path between two others



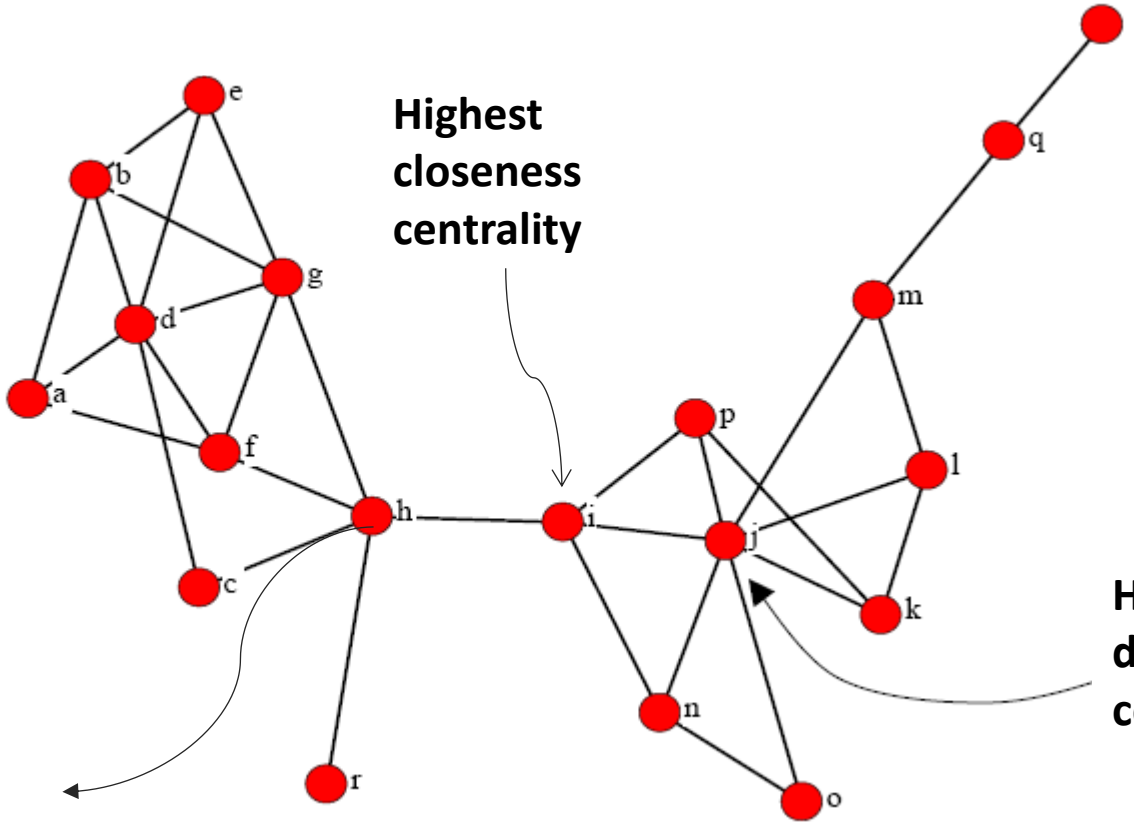
Shortest paths between every nodes:

A to B: A-B
A to C: A-C
A to D: A-C-D
A to E: A-C-D-E
A to F: A-C-D-F
B to C: B-C
B to D: B-C-D
B to E: B-C-D-E
B to F: B-C-D-F
C to D: C-D
C to E: C-D-E
C to F: C-D-F
D to E: D-E
D to F: D-F

C appears 6 times
D appears 2 times
C has the highest betweenness

Node level measures

Nodes with high betweenness centrality are often important controller of power or information (check point)



Highest closeness centrality

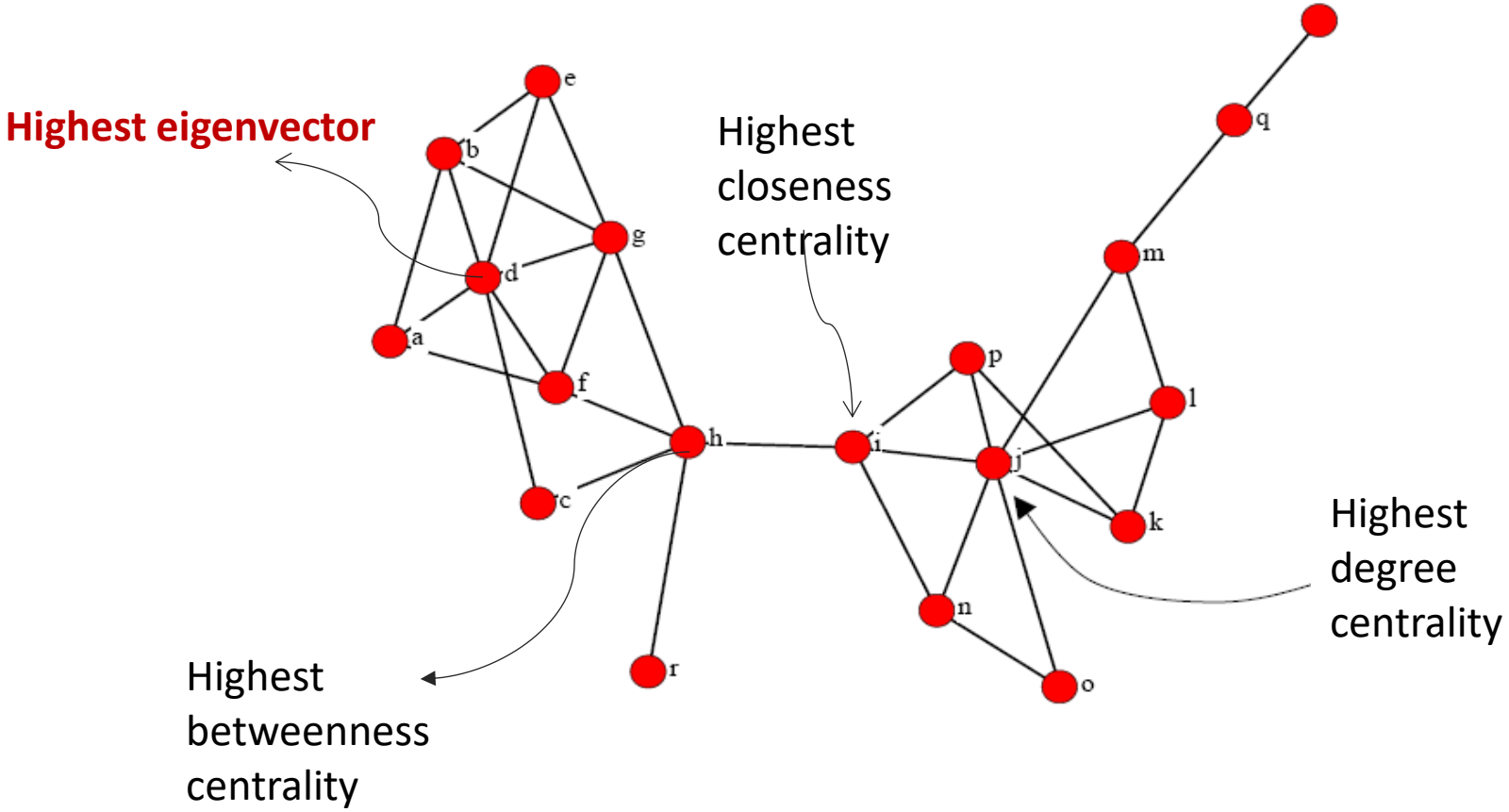
Highest degree centrality

Highest betweenness centrality

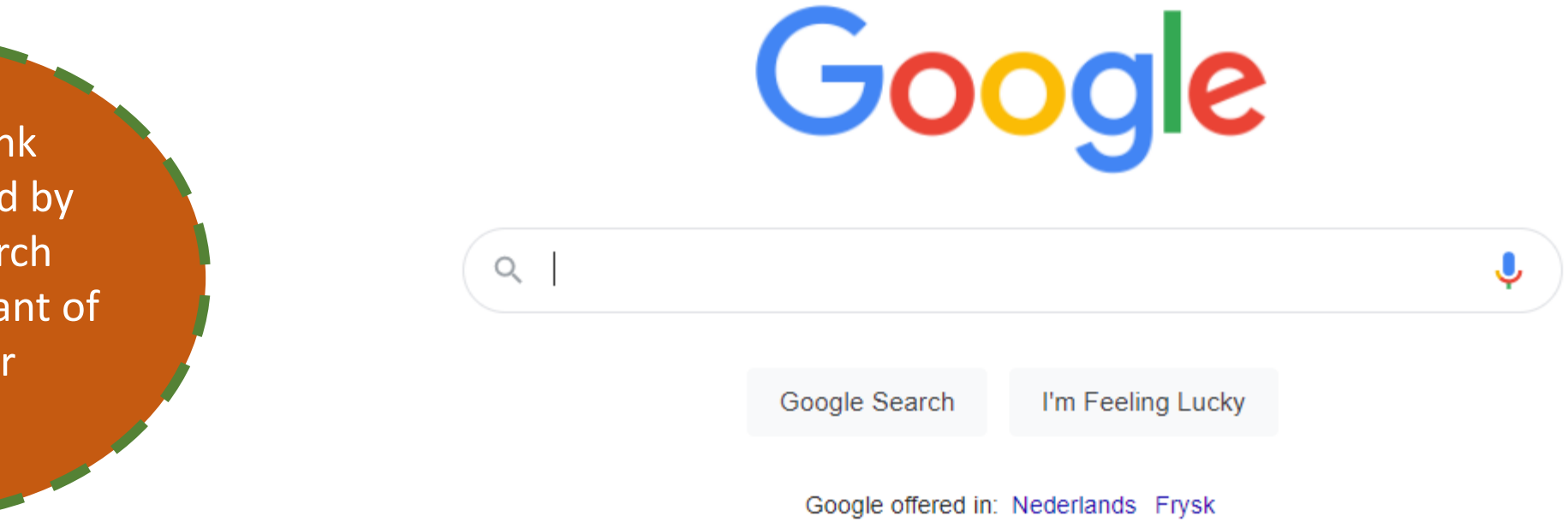
Eigenvector centrality

How central you are depends on how central your neighbors are (The sum of the centrality of your neighbours) (Triumph of Quality Over Quantity!)

A node with 30 unpopular (e.g., low degree) friends
< A node with 10 very popular friends (like Barack Obama)



The PageRank algorithm used by Google's search engine is a variant of Eigenvector Centrality.

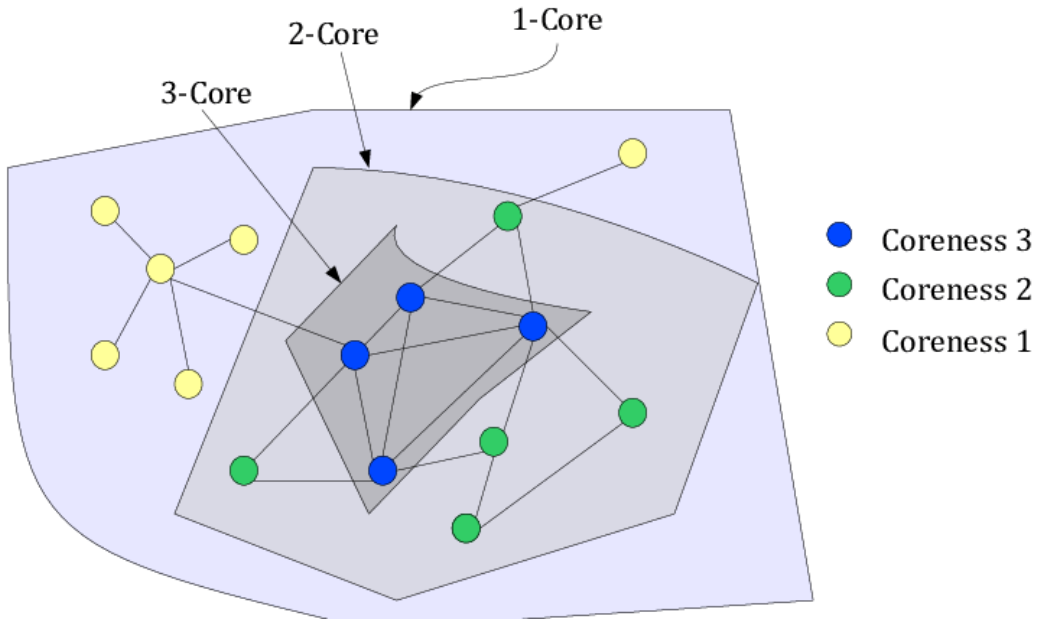


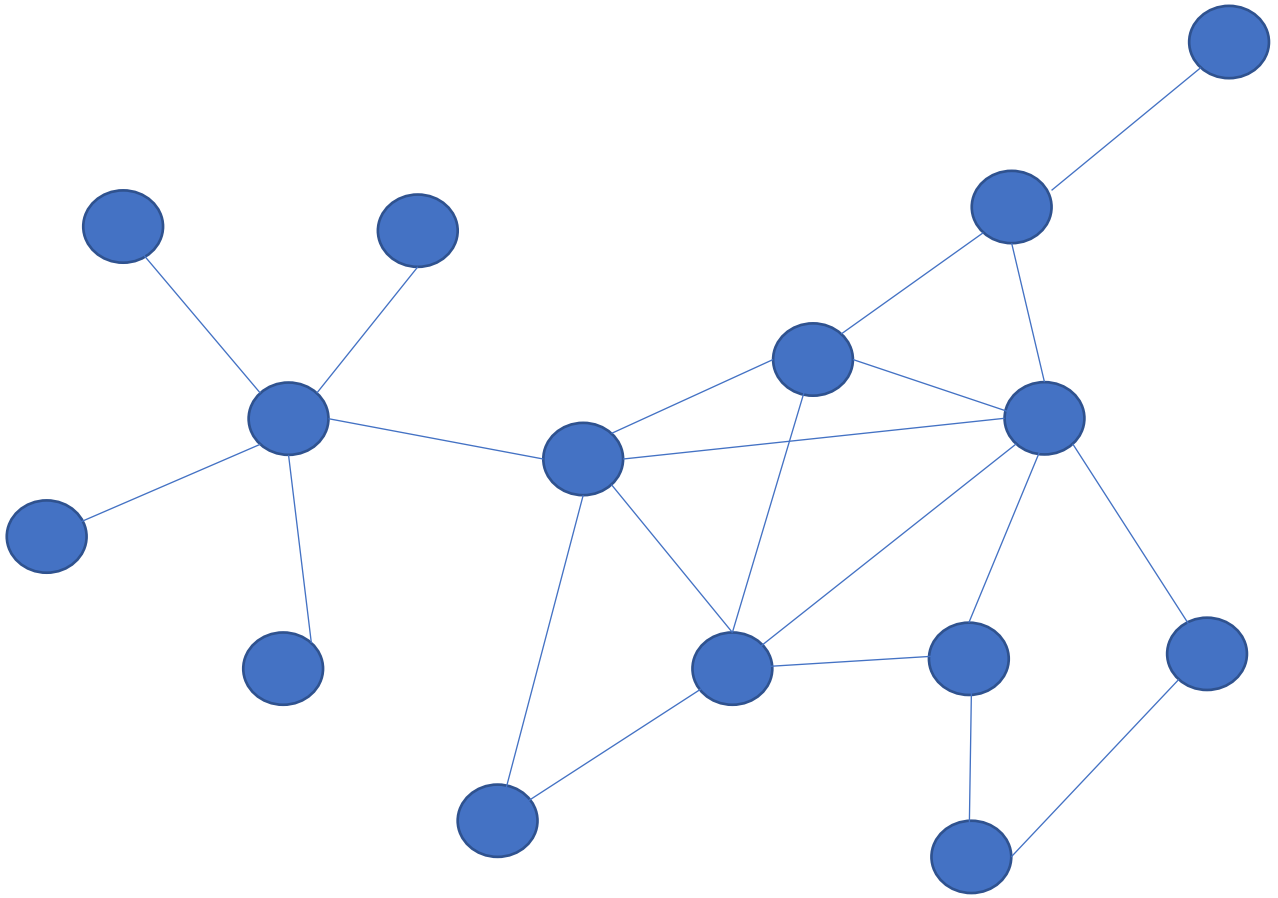
"PageRank works by ***counting the number and quality*** of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites"

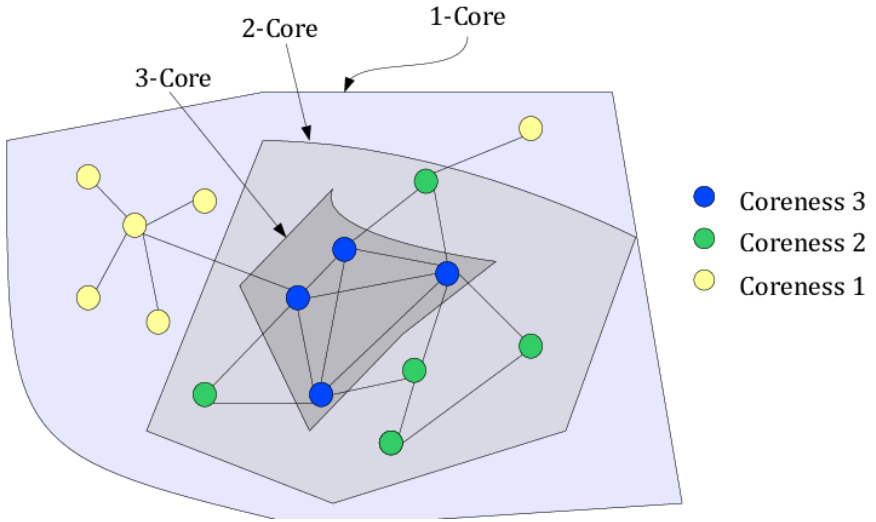
Facts about Google and Competition: https://web.archive.org/web/20111104131332/https://www.google.com/competition/howgooglesearchworks.html

k-cores

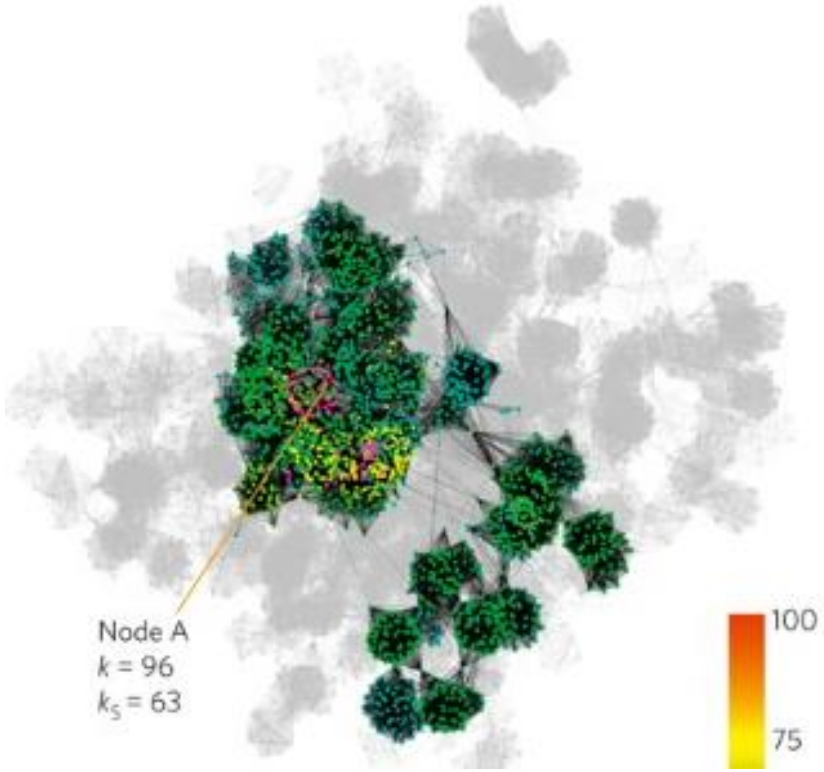
- Obtained by recursively removing all nodes of degree smaller or equal to *k*, until the degree of all remaining vertices is larger than *k*.



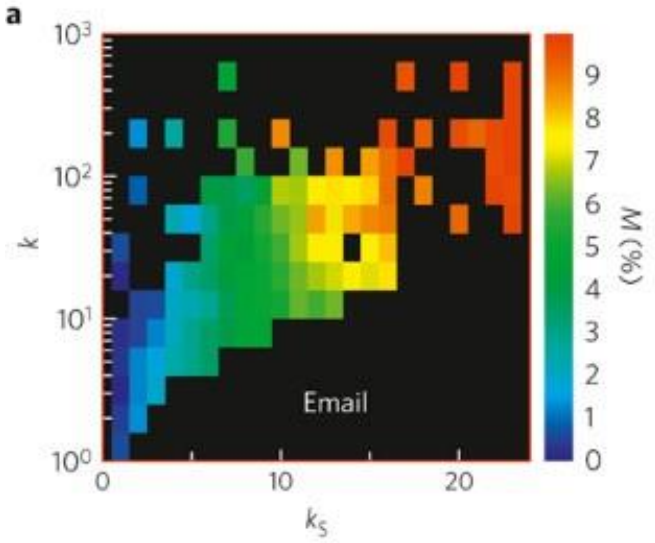




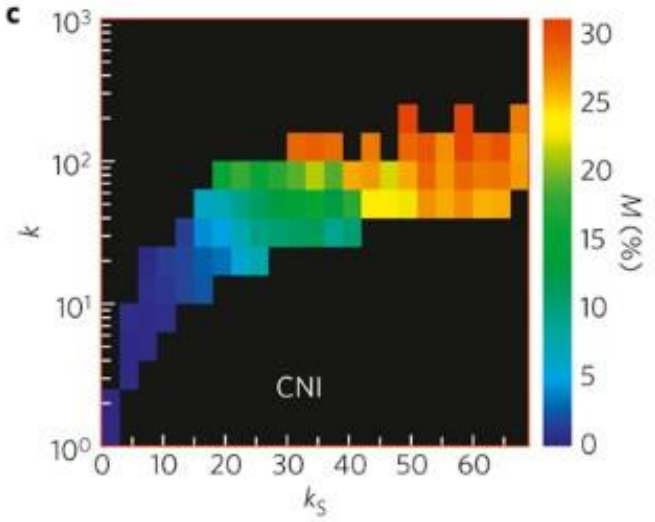
- High ks: at the centre of network
- Low ks: at the outer ring of network



The contact network of hospital patients (CNI)



The network of email contacts in the Computer Science Department of University College London



The contact network of hospital patients (CNI)

Degree is the most obvious metrics of centrality, but it shouldn't be the only one

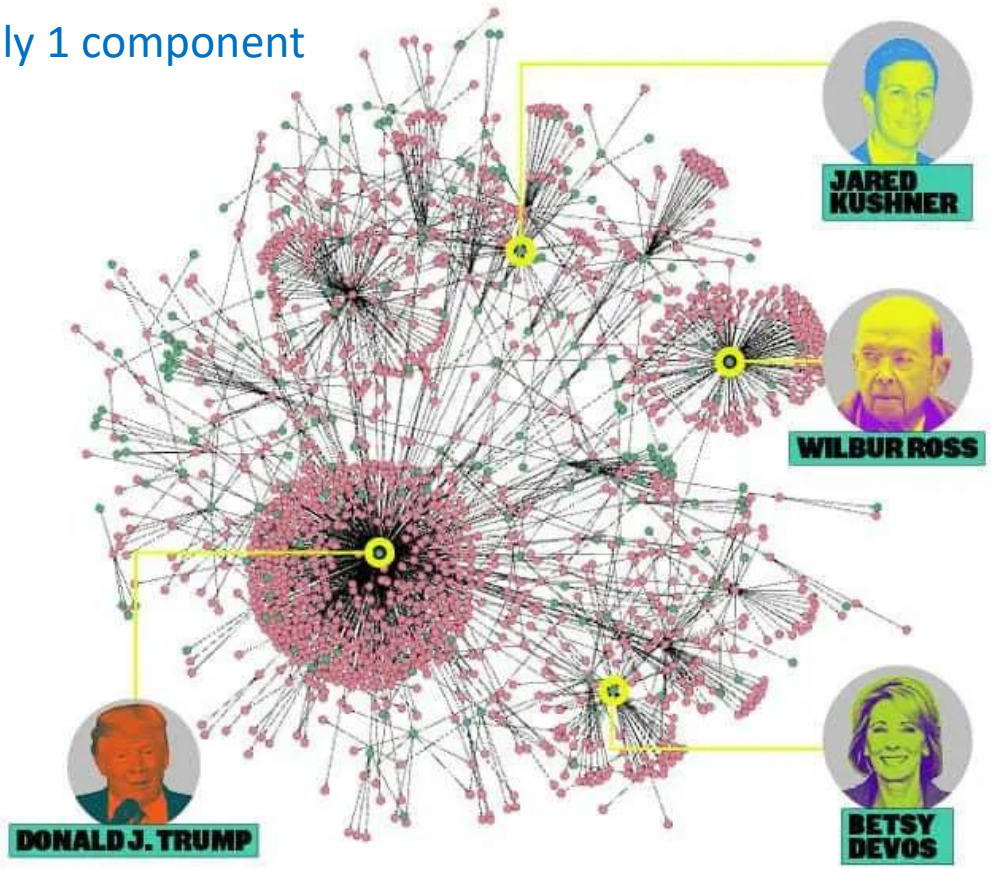
Network-level measures

- Component
- Community
- Degree distribution
- Density
- Global clustering
- Centralization

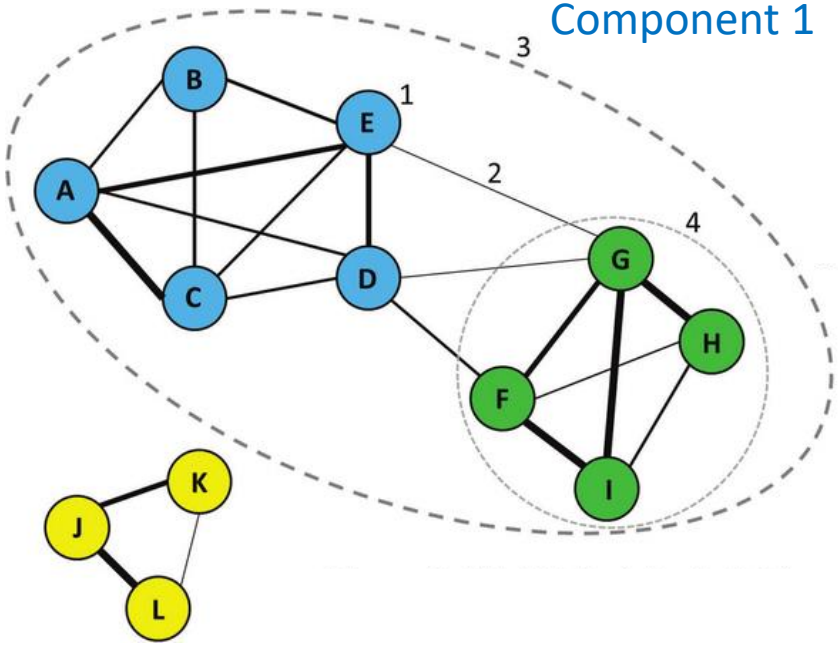
Component

No path between the nodes in different subsets

Only 1 component



Component 1



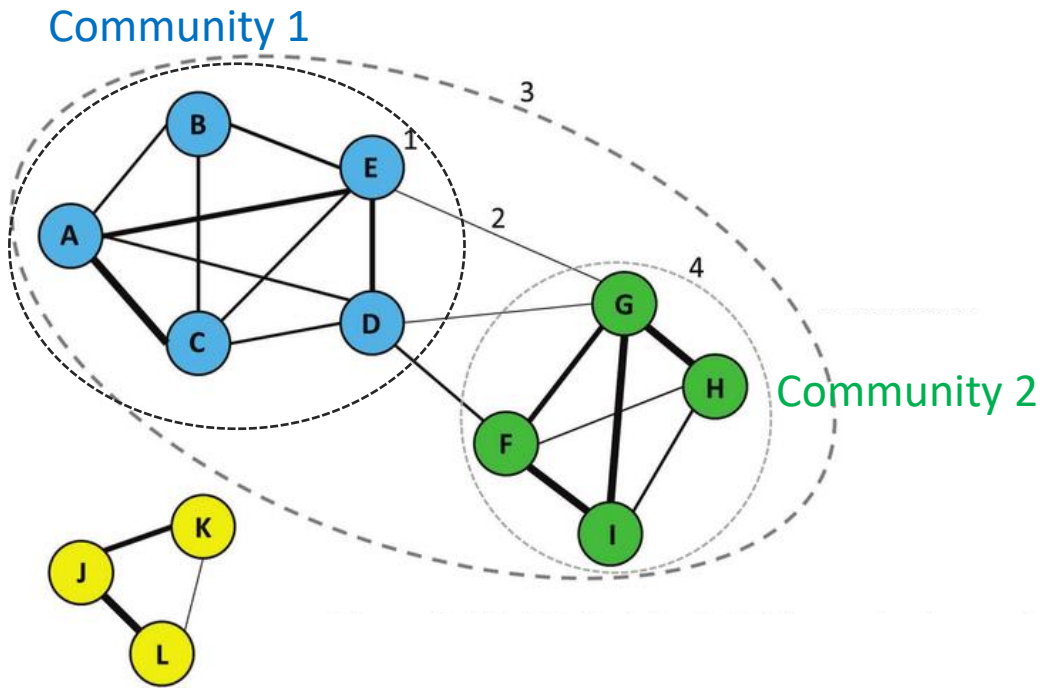
Component 2

Community

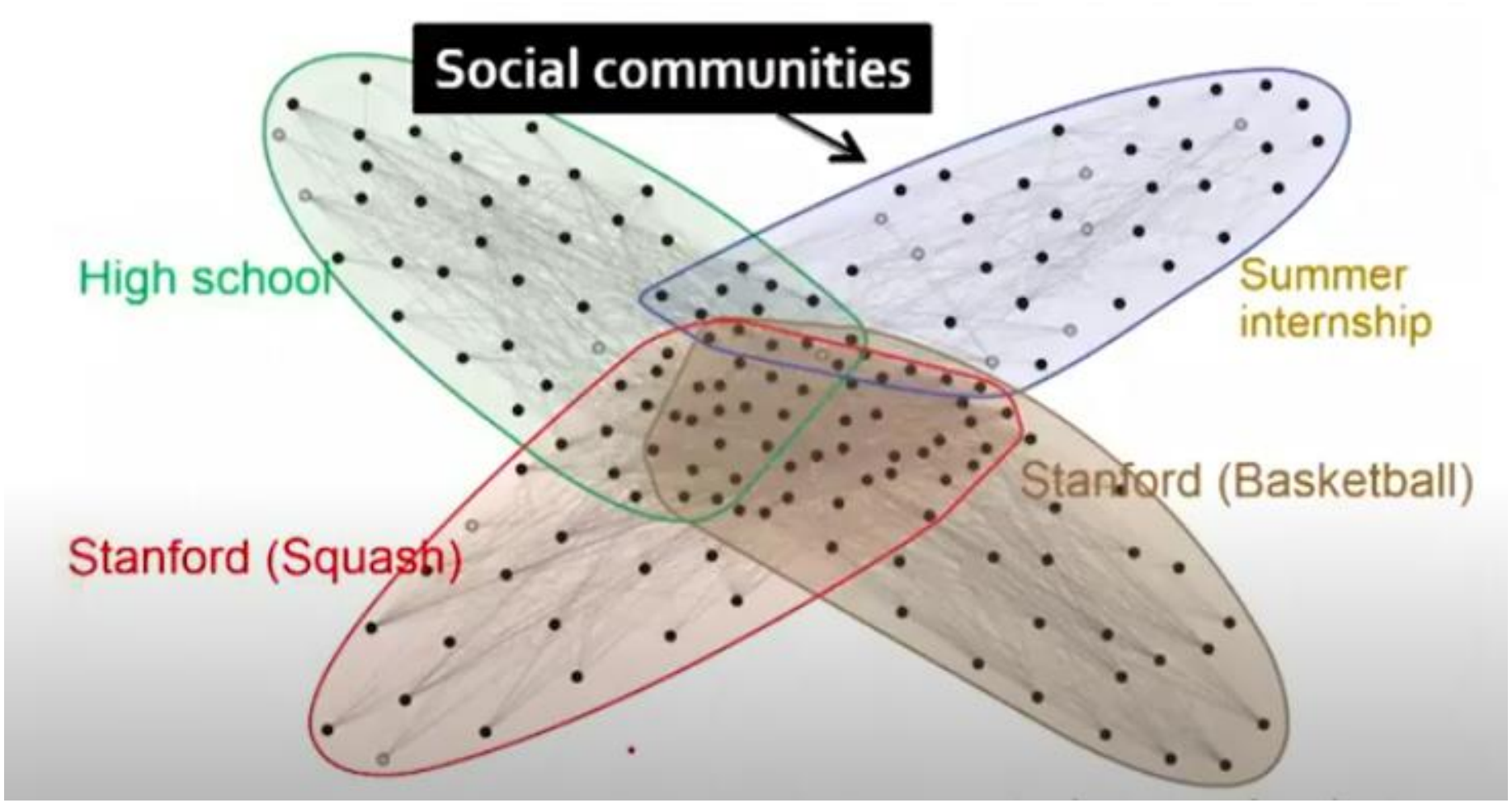
Most real networks contain parts in which the nodes are **more connected to each other than to the rest of the network**: clusters, communities, cohesive groups or modules (no widely accepted, unique definition)

Why important: a signature of the hierarchical nature of complex systems

Non-overlapping communities

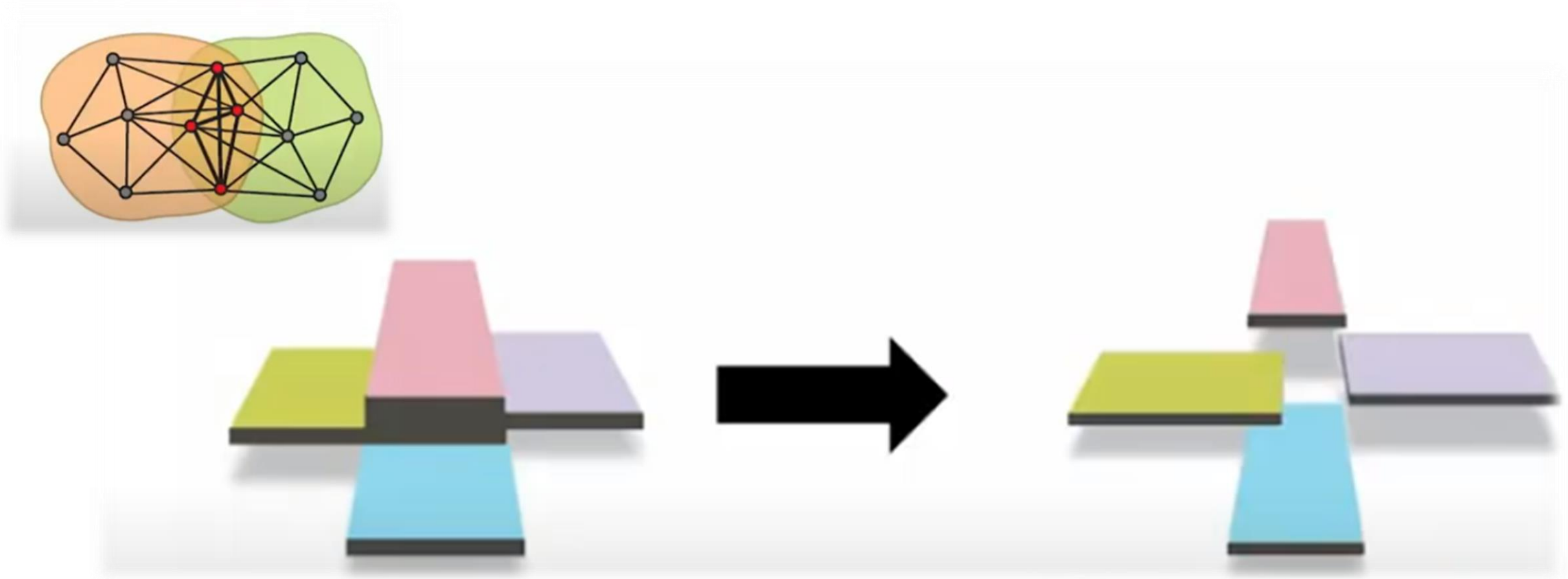


Overlapping communities



From Leskovec, Rajaraman and Ullman, Stanford University

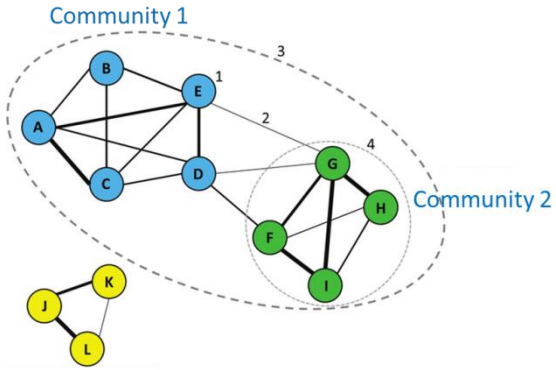
Community detection algorithms

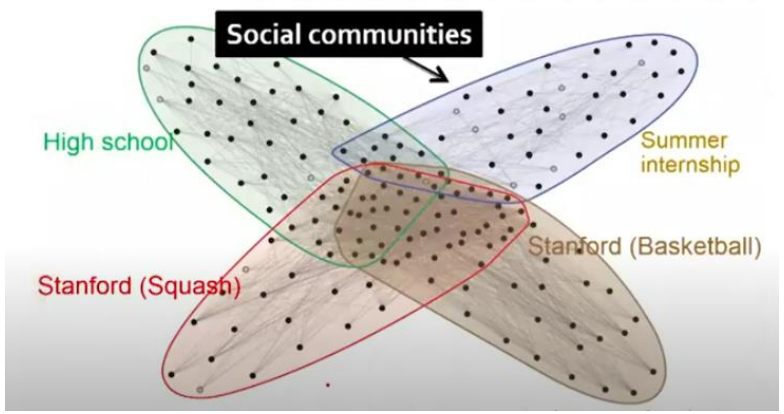


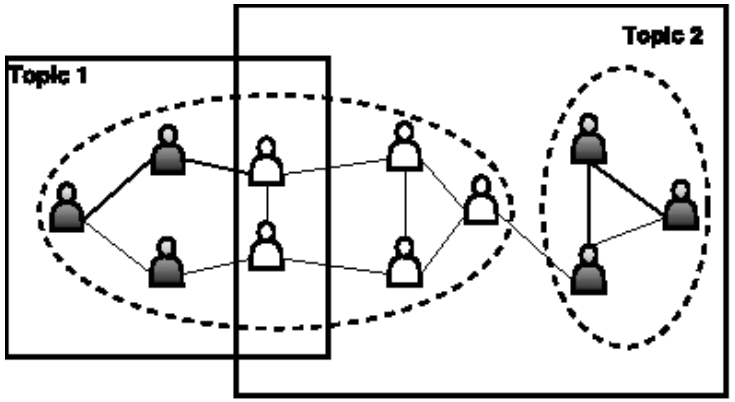
“community over community”
“tiles on top of tiles”

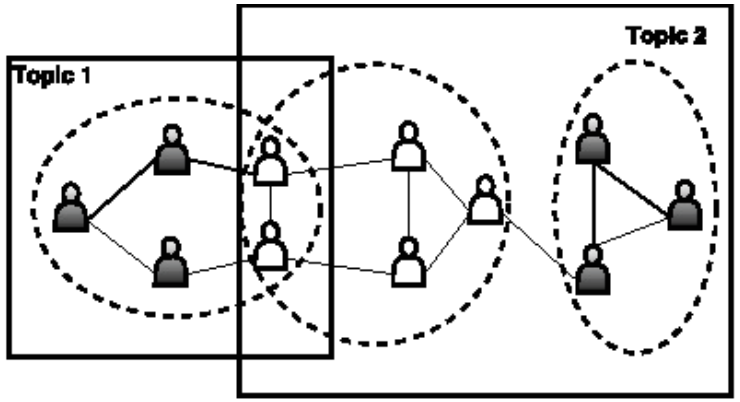
From Leskovec, Rajaraman and Ullman, Stanford University

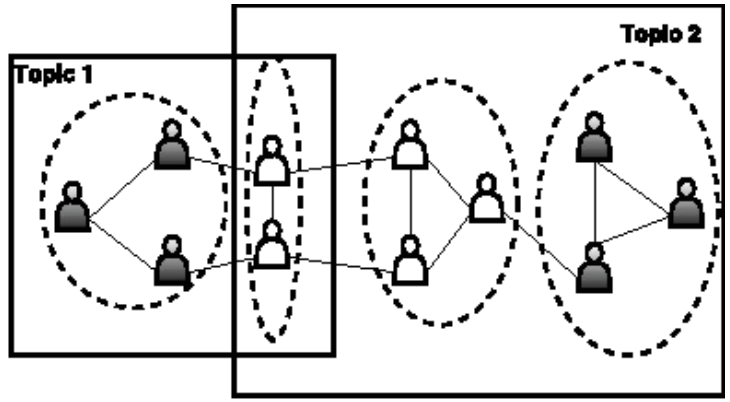
You can use many criteria to design community detection algorithms











Only based on network structure

Network structure+ Topic

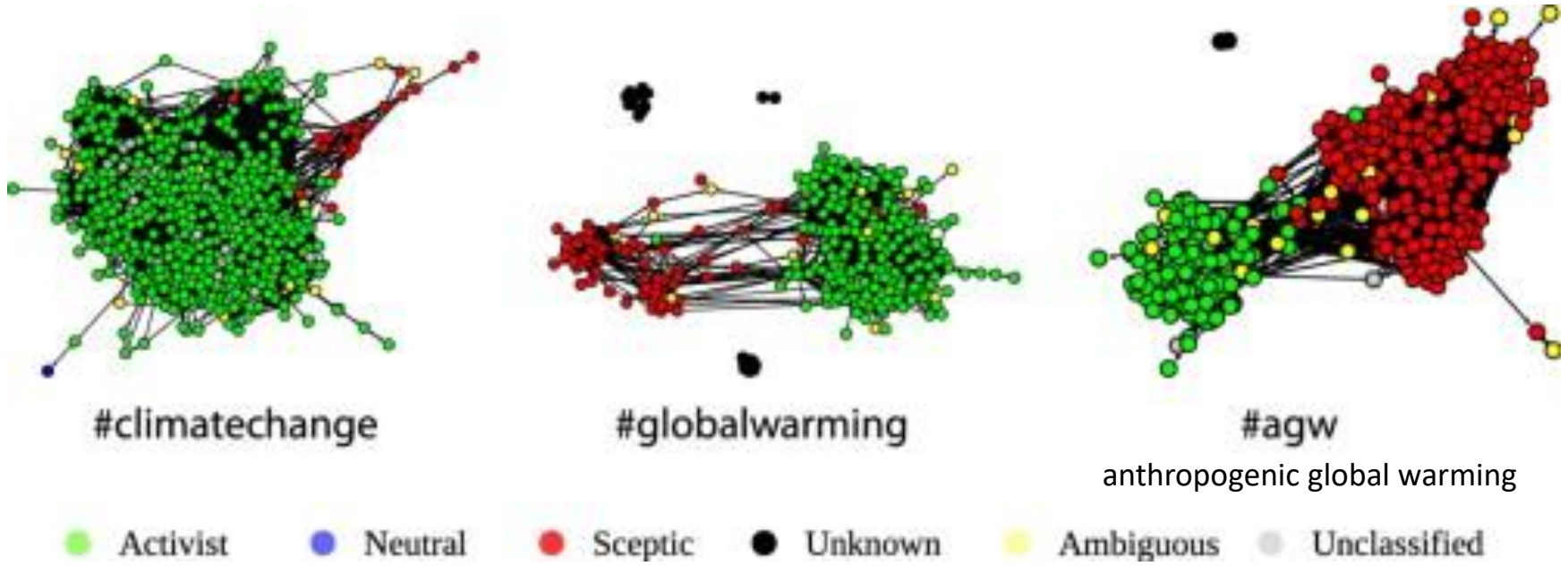
Network structure+ Topic+ Opinion

Negative opinion holder **Positive opinion holder** **Common topics** **Social communities**

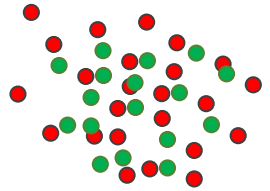
More: Behavior base community….

Opinion-based communities: Echo chambers

Echo chamber: "an environment where a person only encounters information or opinions that reflect and reinforce their own.“ (you hear the same voice again and again…)



When there is no echo chamber effect:



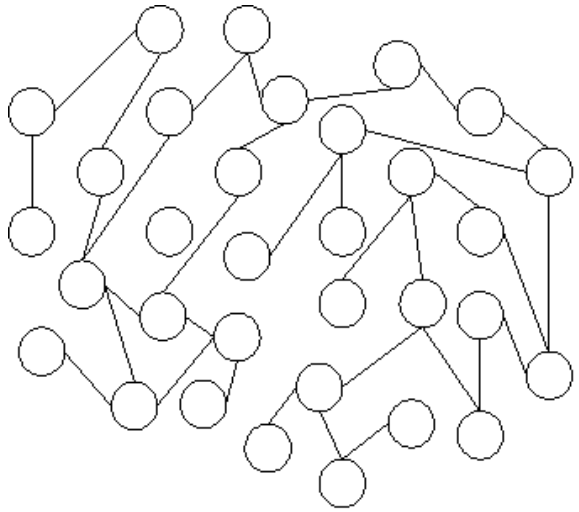
Others: polarization of opinion (pro-vaccination and anti-vaccination)

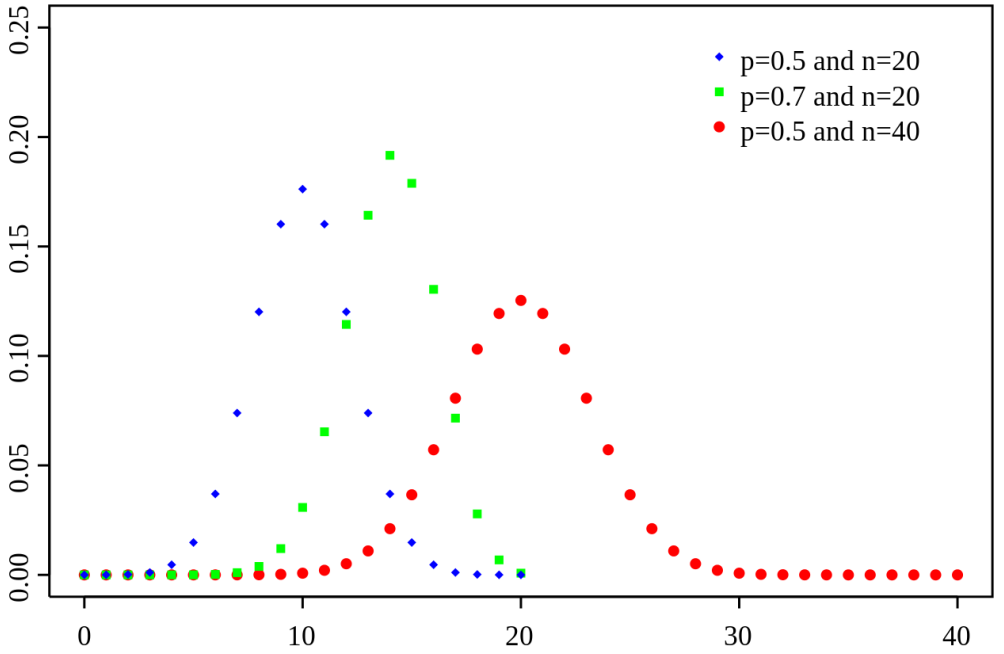
Hywel T.P.William et al., 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global Environmental Change, 32, :126-138.

Degree distribution

How to summarize the degree of all nodes of a network--Degree distribution

The degree by nodes (d_i) follow a distribution of $\{(p_{k(i)}, d_i), i \in N\}$





Each of *n* nodes is independently connected (or not) with probability *p* (or 1 − *p*) (the Erdős–Rényi random graph)

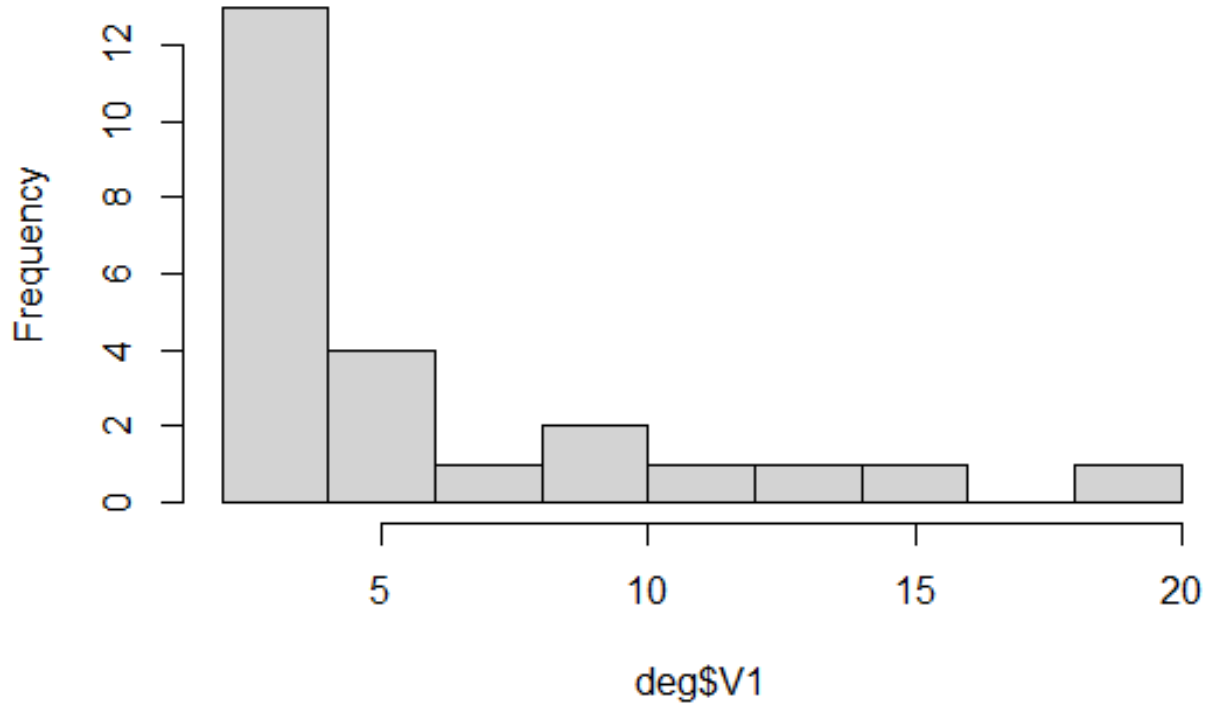
A binomial distribution

Does degree distribution of real-world social networks look like this?

Degree distribution of class network of INFOMNWA-2021

3. Name the people that you know in this class.

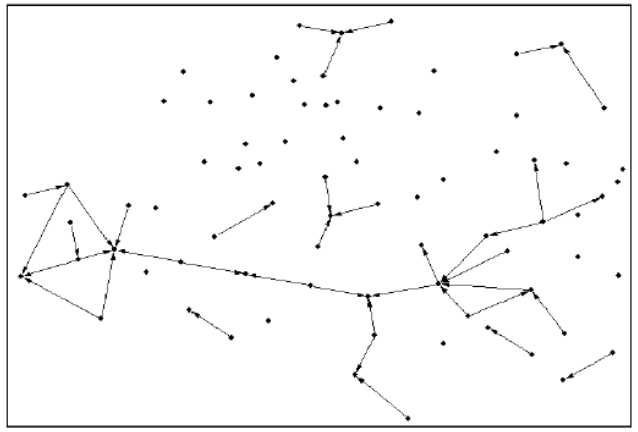
Enter your answer



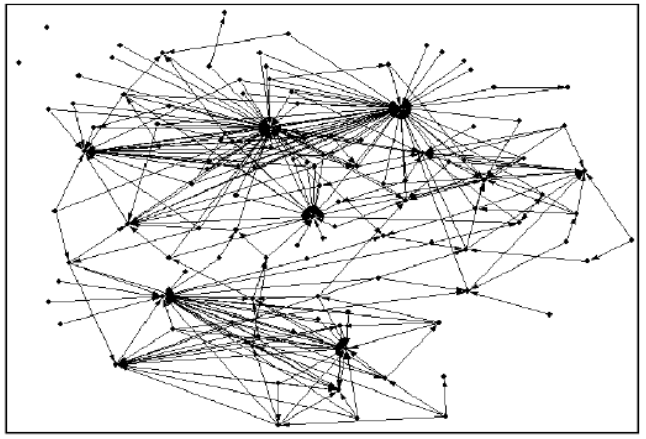
Density of network: how well connected

The number of connections a network has, divided by the total possible connections a network could have $\binom{n}{2}$ = *n(n-1)/2* possible unordered pairs of nodes

Help With the Rice Harvest



Village 1



Village 2

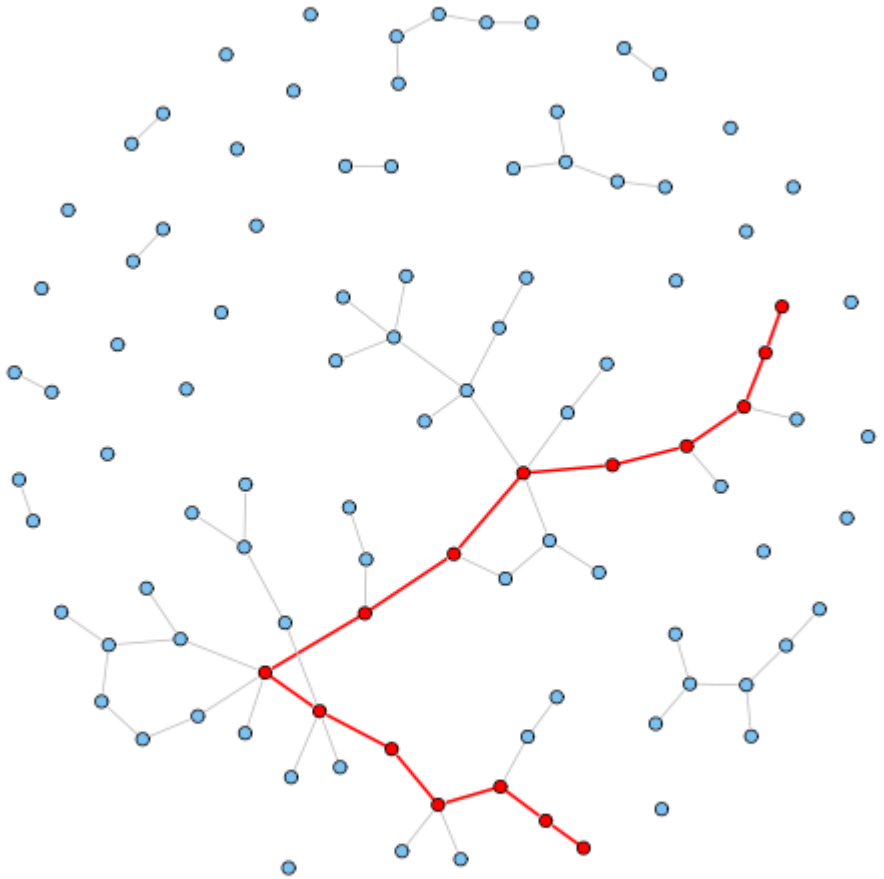
Which village has a higher density?

The redundancy of connection is important.

If there are many efficient paths connecting two actors, the odds are improved that a signal will get from one to the other.

Diameter of network: how big it is

The longest of the shortest path between two nodes



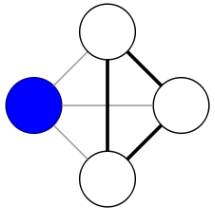
14 or what? (~60)

Clustering: The probability that your friends know each other

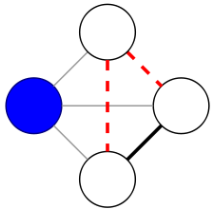
(Transitivity)

Local clustering coefficient of a node in a graph quantifies how close its neighbours are to being a clique (complete graph)

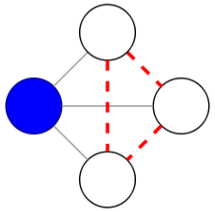
Global clustering coefficient counts the percentage of close triangles in the whole network.



c = 1



c = 1/3



c = 0

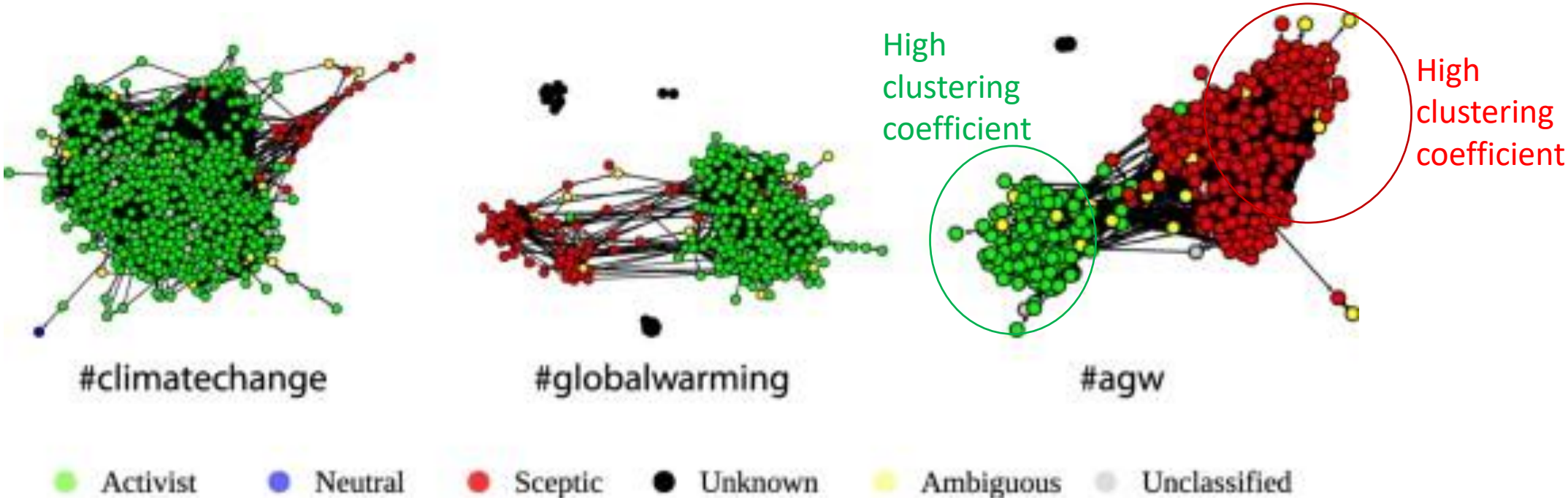
How connected are your friends?

- If your “friends” all know each other, you have a high clustering coefficient. If your “friends” don’t know each other, then you have a low clustering coefficient and you need to introduce them to each other!

Echo chambers & clustering coefficient

People tend to know each other very well in a liked-mined communities

But which one is the cause ??

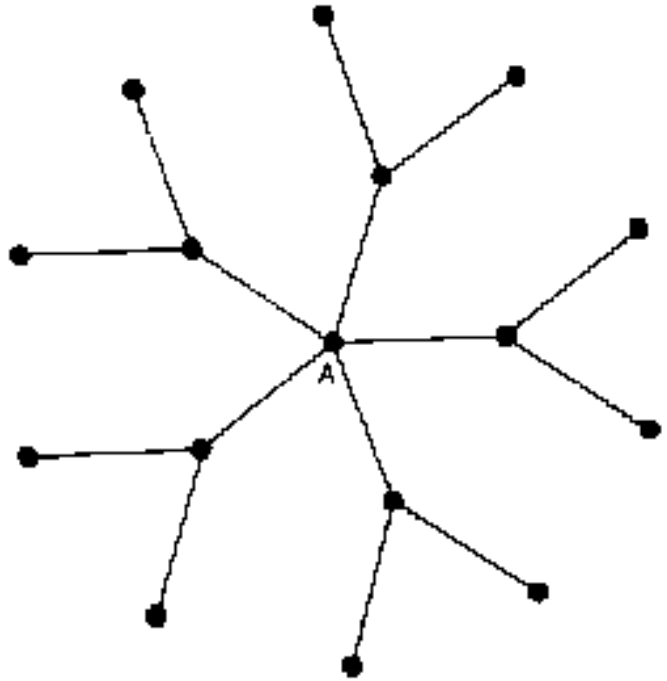


Hywel T.P.William et al., 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. Global Environmental Change, 32, :126-138.

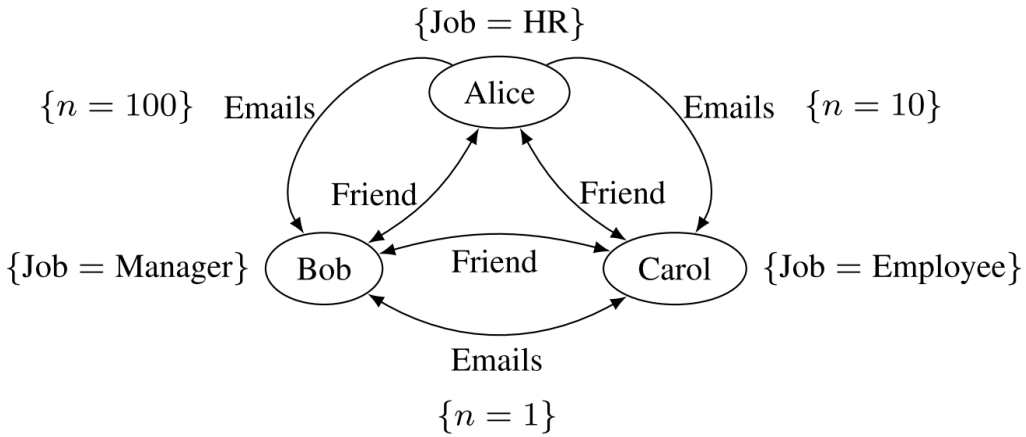
Centralization: to which degree a network revolves around a single node

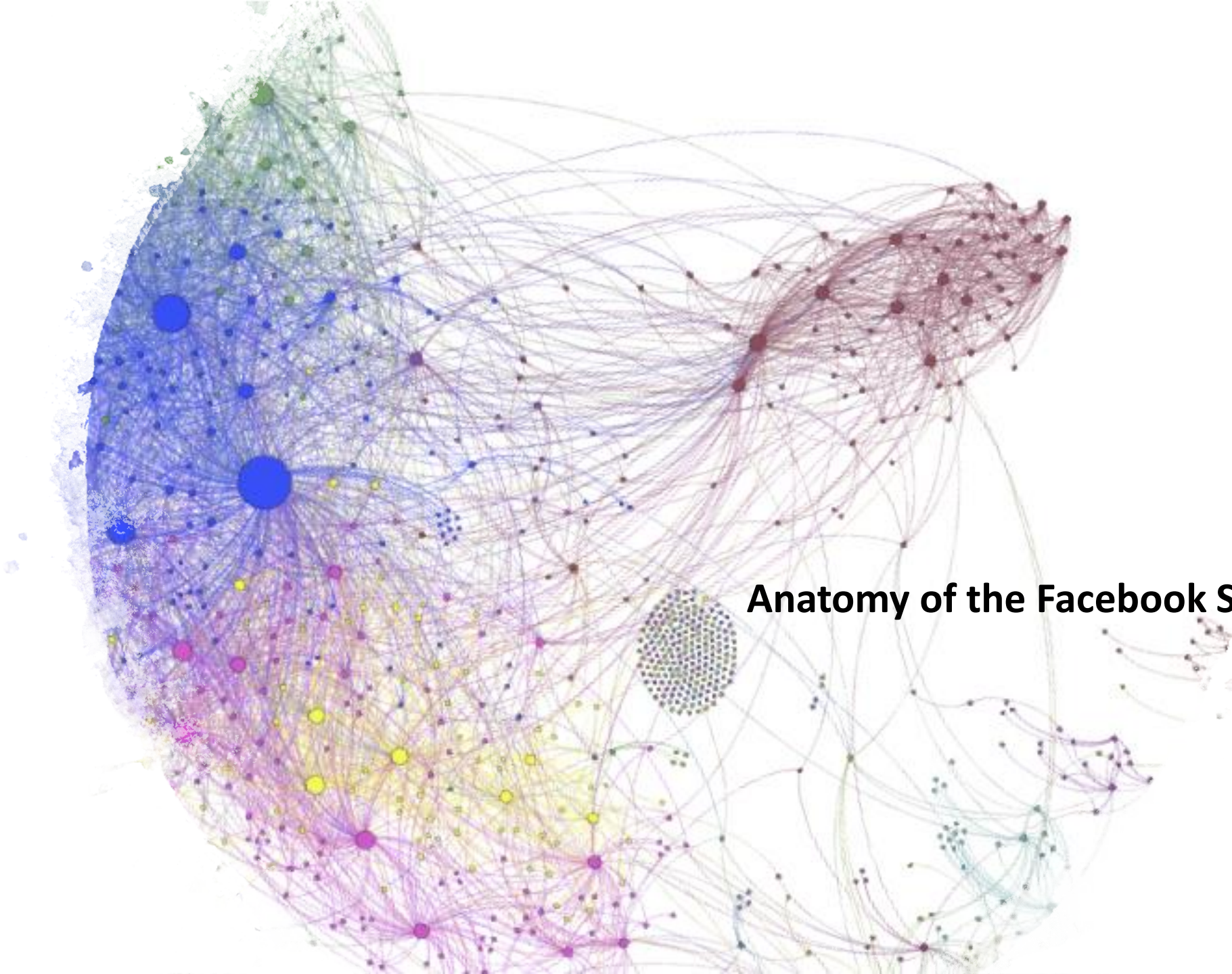
Opinion leader; information controller

Which network is more centralized?



Think about nowadays news media…



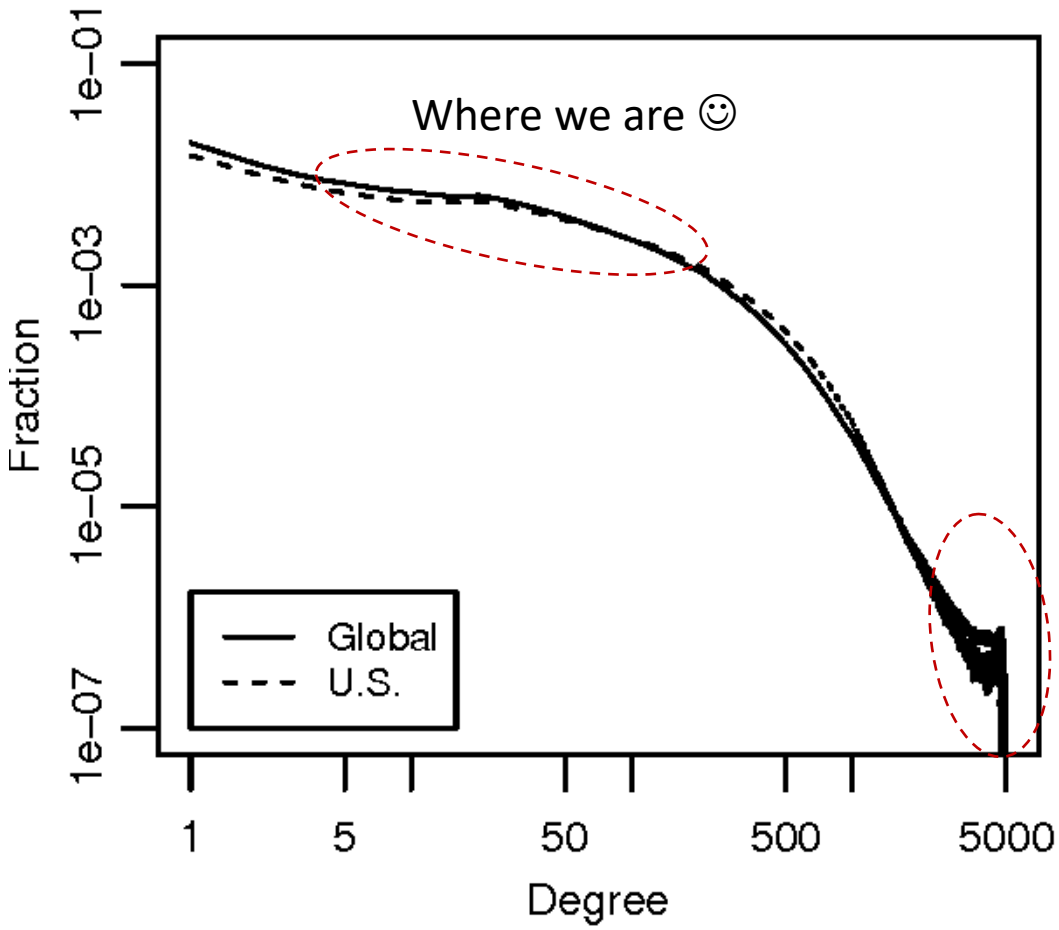


Anatomy of the Facebook Social Network

- n = 721 million active uers in May 2011
- 68.7 billion friendship edges

World's population in 2011: 6.9 billion

(1) degree distribution p_k



- Right-skewed with a high variance, but doesn't fit strict power-law distribution
- Nearly monotonically decreasing, except for a small anomaly near 20 friends. (Facebook encourage low friend count individuals to gain more friends until they reach 20 friends.)
- Most individuals have a moderate number of friends on Facebook, less than 200. (media=99)
- Small population of users with abnormally high degrees: ***hubs***

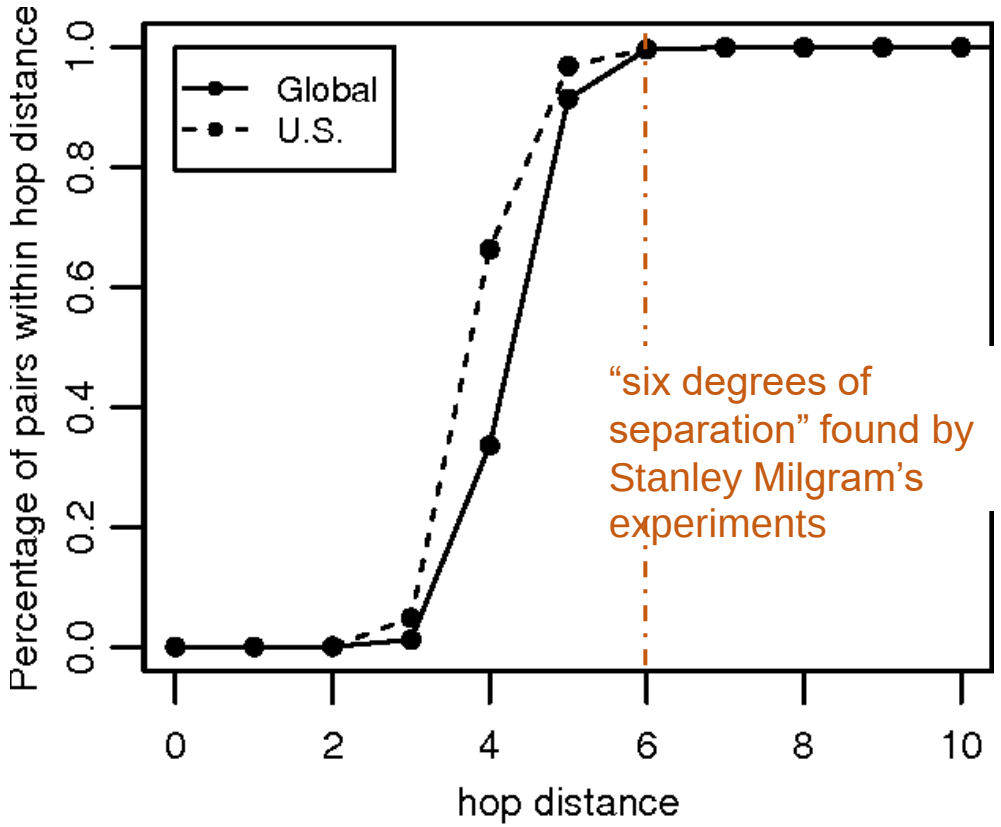
J. Ugander, B. Karrer, L. Backstrom, C. Marlow. The anatomy of the Facebook social graph.

(2) Path lengths

an extremal measure easily distorted by a single ill-connected path in some peripheral regions of the graph

Diameter: the maximum distance between any pair of vertices in the graph.

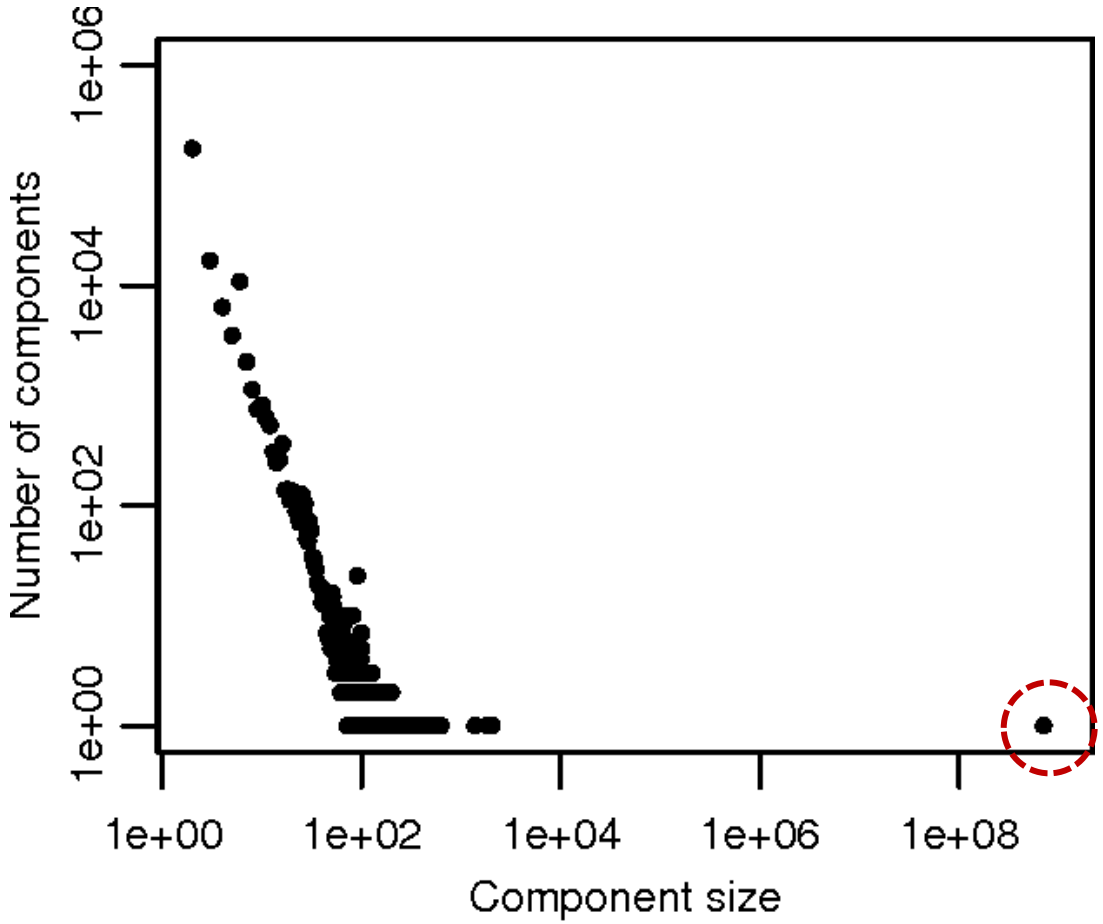
Hop distance: for a pair of vertices (u, v), u is reachable from v along a path in the network with *h* edges or less. (from the neighborhood function N(h))



- The average distance between pairs of users was 4.7 for global users and 4.3 for U.S. users.
- 92% of all pairs of Facebook users were within five degrees of separation, and 99.6% were within six degrees.
- For only U.S. users, 96% were within five degrees and 99.7% were within six degrees.

J. Ugander, B. Karrer, L. Backstrom, C. Marlow. The anatomy of the Facebook social graph.

(3) Component size



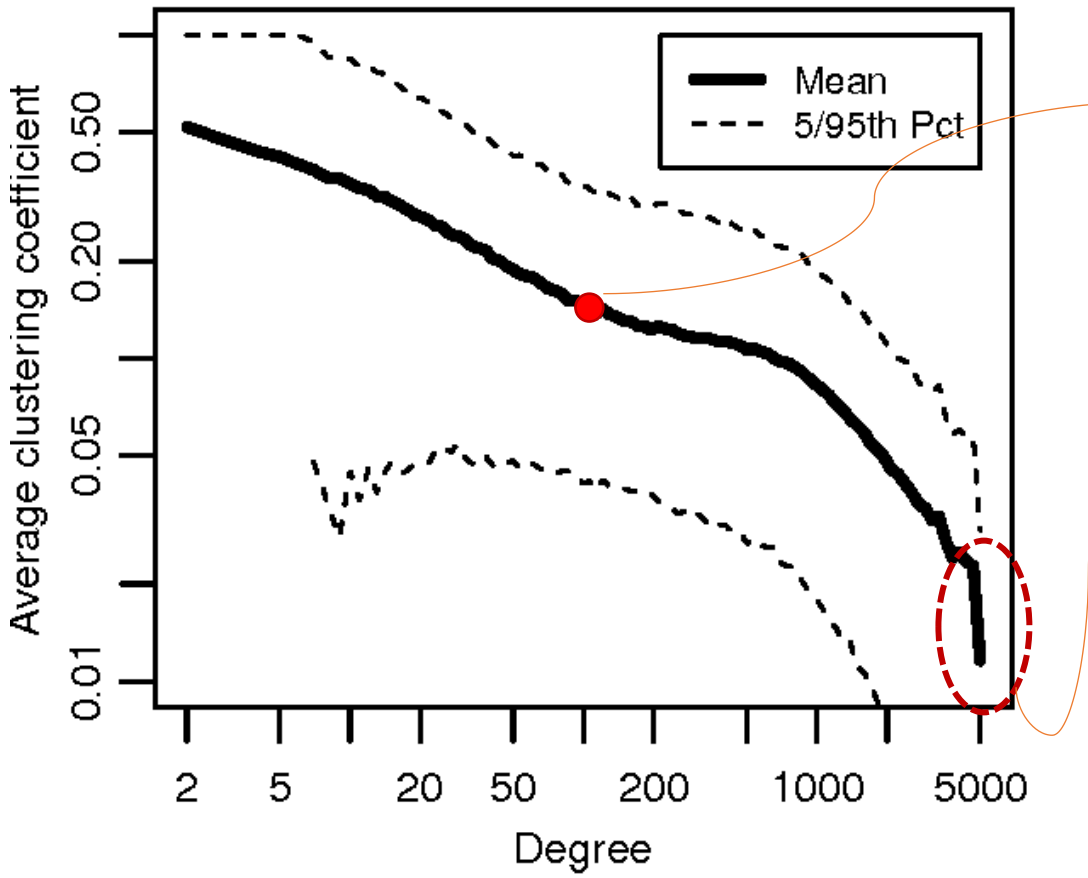
- There are many connected components, most of these components are extremely small.
- The second-largest connected component only has just over 2000 individuals
- The largest connected component, the outlier all the way on the right-hand side of the figure, consists of 99.91% percent of users.

99.91% of individuals belonging to a single large connected component

J. Ugander, B. Karrer, L. Backstrom, C. Marlow. The anatomy of the Facebook social graph.

(4) Local clustering coefficient

Average local clustering coefficient for users as a function of degree:



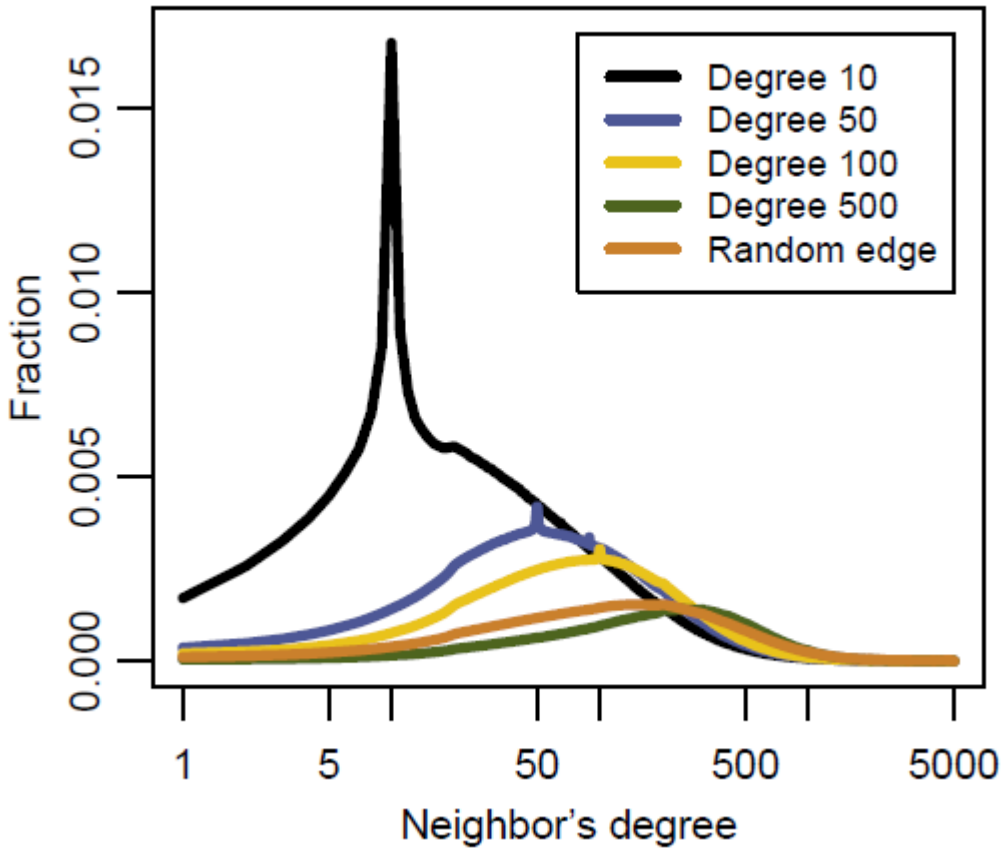
For users with 100 friends, the average local clustering coefficient is 0.14, indicating that for a median user, 14% of all their friend know each other.

The clustering coefficient drops rapidly for users with close to 5000 friends, indicating that these users are likely using Facebook for less coherently social purposes and friending users more indiscriminately.

>5 times the clustering coefficient for the graph of MSN messenger correspondences in the same neighborhood size

J. Ugander, B. Karrer, L. Backstrom, C. Marlow. The anatomy of the Facebook social graph.

(5) Friendship paradox in Facebook



The **friendship paradox**: a phenomenon first observed by the sociologist Scott L. Feld in 1991 that **most people have fewer friends than their friends have, on average.**

Facebook: 83.6% of users have less friends than the median friend count of their friends

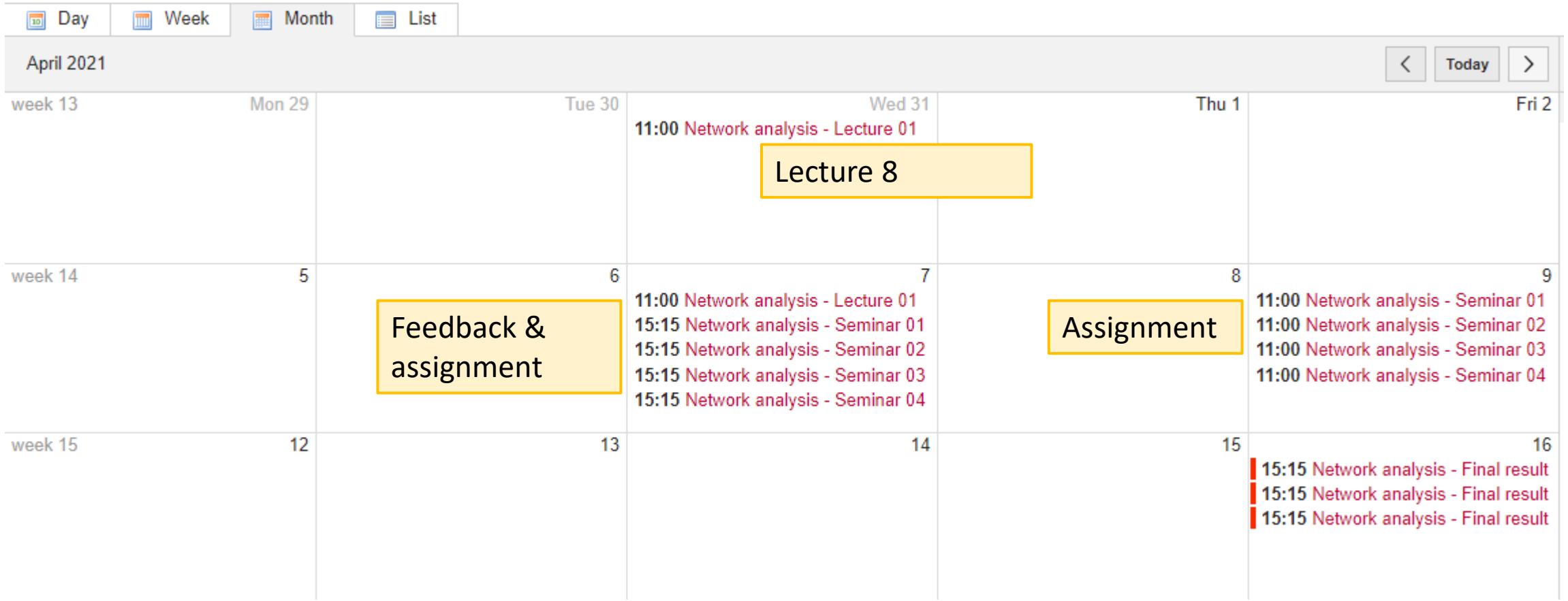
What might be the application for marketing strategy?

J. Ugander, B. Karrer, L. Backstrom, C. Marlow. The anatomy of the Facebook social graph.

Questions?

Practicalities of the courses

- Group assignments/ Computer labs
- Individual assignments
- Exam


Day
Week
Month
List
April 2021
Today
week 13
Mon 29
Tue 30
Wed 31
11:00 Network analysis - Lecture 01
Thu 1
Fri 2
Lecture 8
week 14
5
6
7
11:00 Network analysis - Lecture 01
15:15 Network analysis - Seminar 01
15:15 Network analysis - Seminar 02
15:15 Network analysis - Seminar 03
15:15 Network analysis - Seminar 04
8
9
11:00 Network analysis - Seminar 01
11:00 Network analysis - Seminar 02
11:00 Network analysis - Seminar 03
11:00 Network analysis - Seminar 04
Feedback & assignment
Assignment
week 15
12
13
14
15
16
15:15 Network analysis - Final result
15:15 Network analysis - Final result
15:15 Network analysis - Final result

Lecture 5: Social network and its measures

Ex1 in group assignment

Lecture 6: Network formation

Ex2 and bonus question in group assignment

Lecture 7: Simple and complex contagion

Lecture 8: Influence manipulation

Ex3 in group assignment