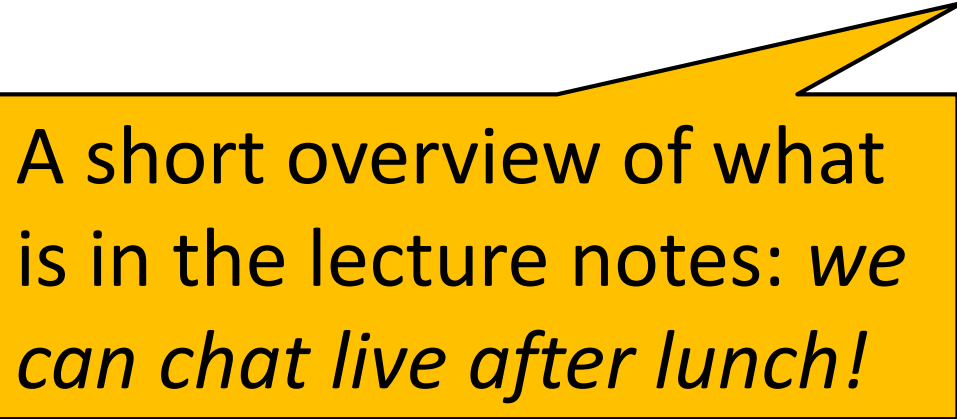


GIAN Course on Distributed Network Algorithms

Social Networks

GIAN Course on Distributed Network Algorithms

Social Networks

A yellow speech bubble with a black outline, pointing towards the top right. It contains text about the lecture notes and a live chat invitation.

A short overview of what is in the lecture notes: *we can chat live after lunch!*

Stanley Milgram's Experiment

The Experiment:

- Chose random people in US Midwest
- Tell them to send a letter to **random person** in Boston
- **Rule:** forward letter only to someone known on **first name basis**



Stanley Milgram's Experiment

The Experiment:

- Chose random people in US Midwest
- Tell them to send a letter to **random person** in Boston
- **Rule:** forward letter only to someone known on **first name basis**

The Outcome:



Stanley Milgram's Experiment

The Experiment:

- Chose random people in US Midwest
- Tell them to send a letter to **random person** in Boston
- **Rule:** forward letter only to someone known on **first name basis**

The Outcome:

- Many letters lost 😊
- But the ones that arrived:
average path length 5.5



Stanley Milgram's Experiment

The Experiment:

- Chose random people in US Midwest
- Tell them to send a letter to **random person** in Boston
- **Rule:** forward letter only to someone known on **first name basis**

The Outcome:

- Many letters lost 😊
- But the ones that arrived:
average path length 5.5

A small world!



Stanley Milgram's Experiment

The Experiment:

- Chose random people in US Midwest
- Tell them to send a letter to **random**
- **Rule:** forward letter only to someone

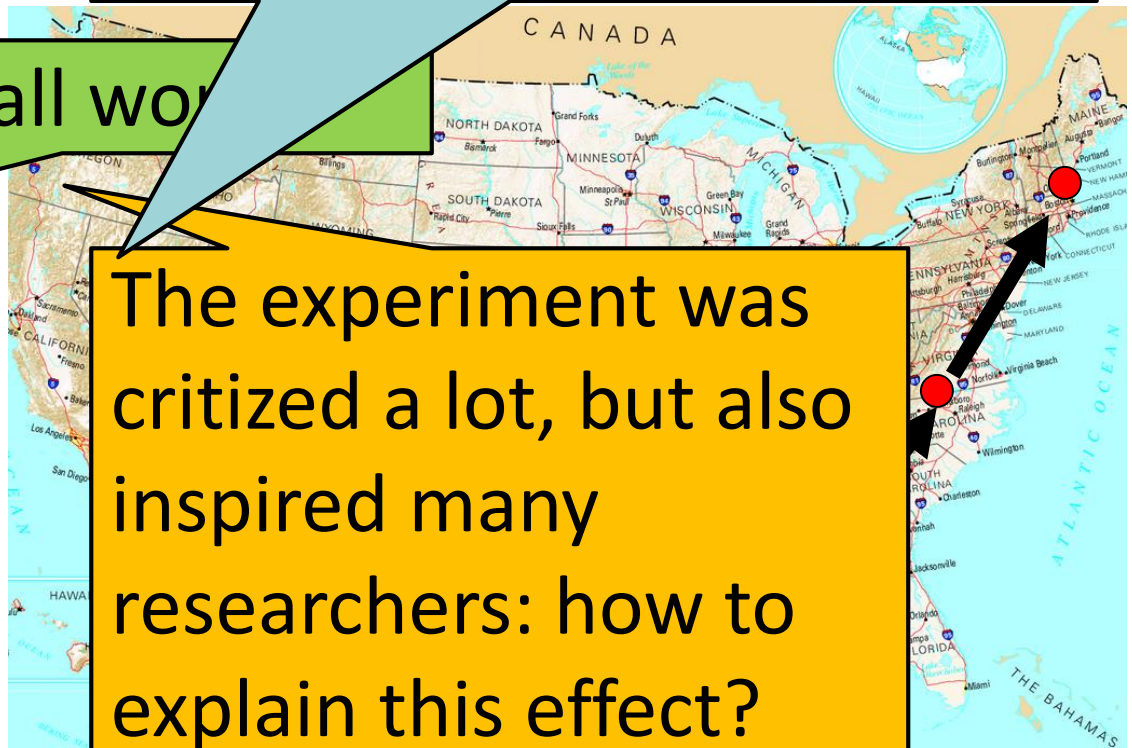
Researchers designed powerlaw graph models, Watts-Strogatz model, etc.

The Outcome:

- Many letters lost 😊
- But the ones that arrived:
average path length 5.5

A small world

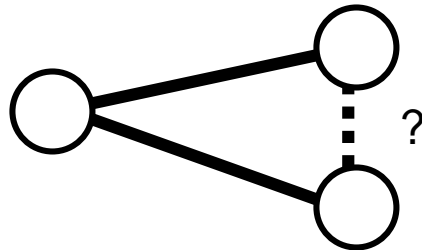
The experiment was criticized a lot, but also inspired many researchers: how to explain this effect?



Social Network =
Graph with large clustering
coefficient
+
random links

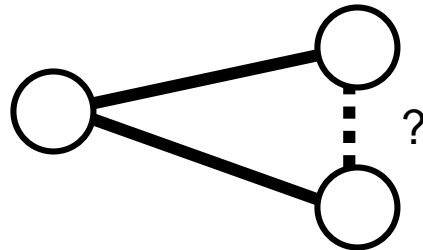
Cluster Coefficient

Likelihood that two friends of a node are also friends, summed up over all nodes.



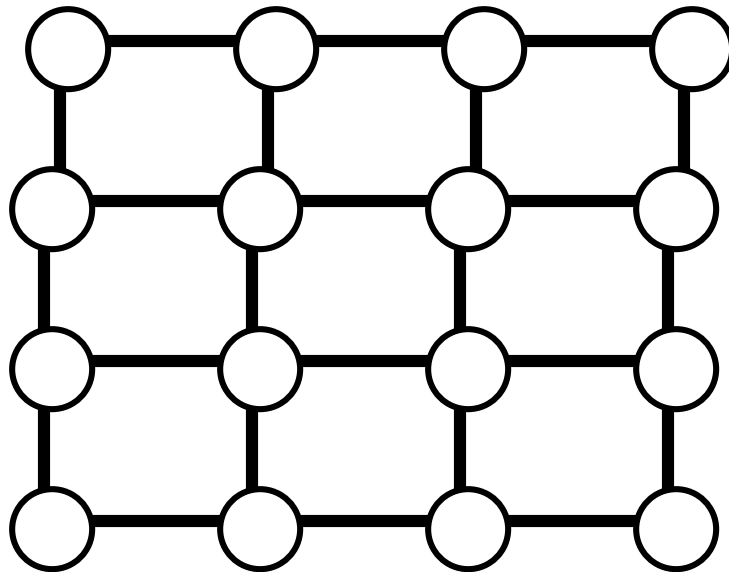
Cluster Coefficient

Likelihood that two friends of a node are also friends, summed up over all nodes.

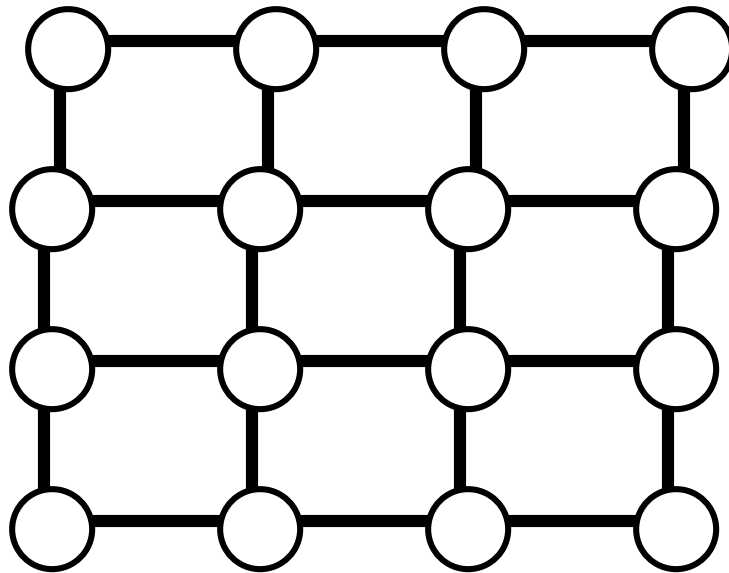


A good metric?

Clustering Coefficient?



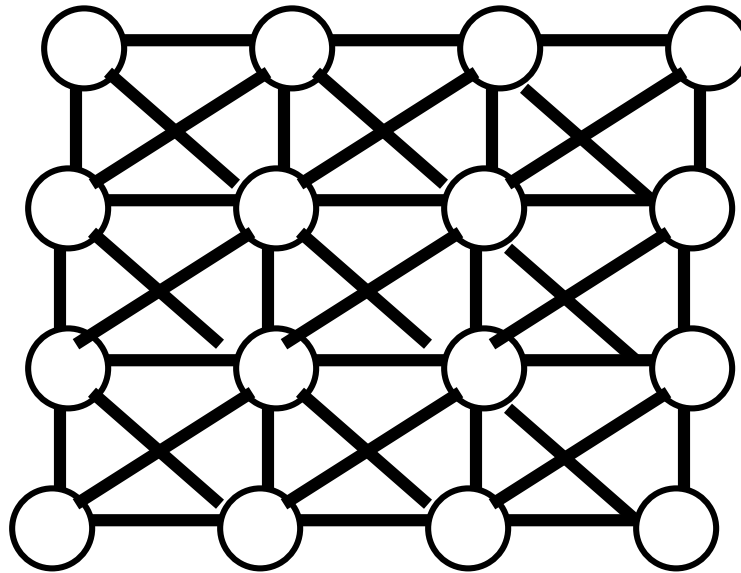
Clustering Coefficient?



My friends are never friends.

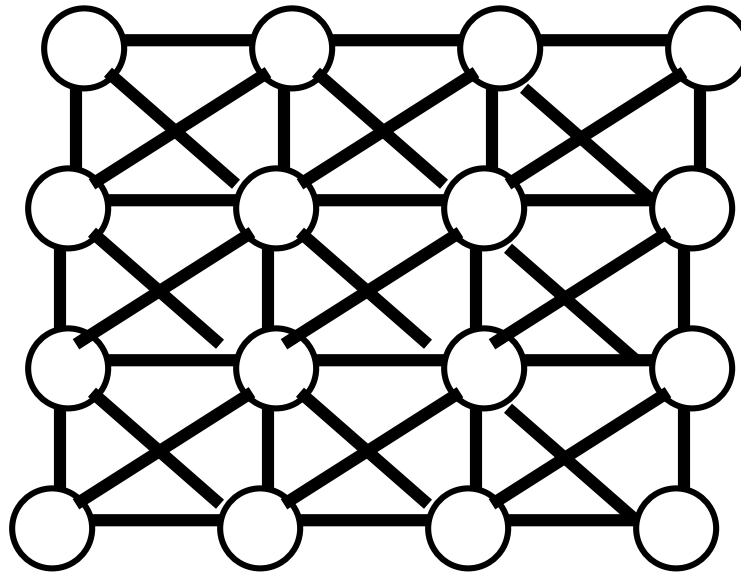
Coefficient: 0

Clustering Coefficient?



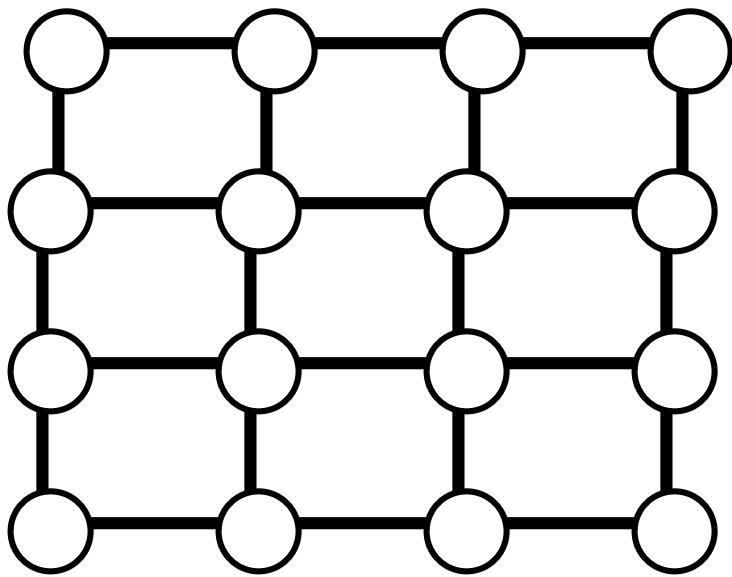
Coefficient: $3/7$

Clustering Coefficient?



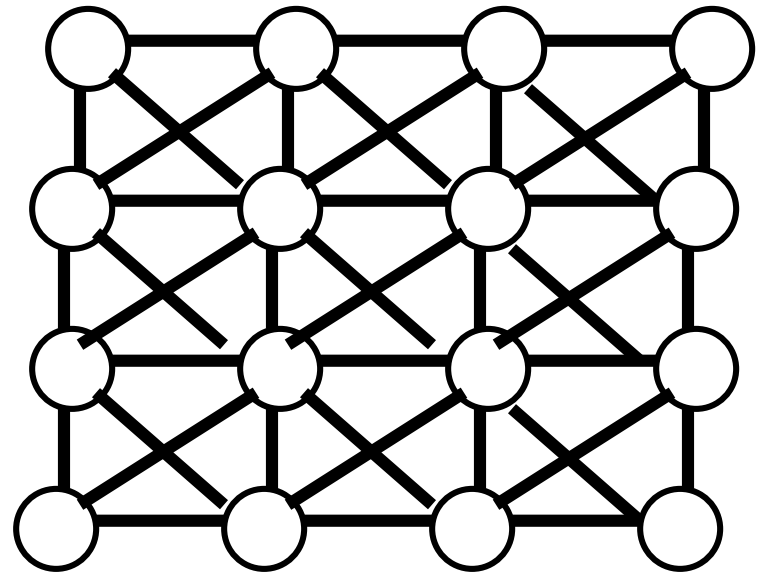
out of $\text{binom}(8,2)=28$
pairs of neighbors, 12
are neighbors

Coefficient: $3/7$



Coefficient: 0

VS



Coefficient: 3/7

Are these graphs really so different?
Is the CC a good metric?

It's about routing!

To understand Milgram's experiment, we do not only need to understand the network topology, but also the routing: How did the persons find these routes?



The network topology must also be navigable!

It's about routing!

To understand Milgram's experiment we do not only need to know the network topology, but also the routing protocol. How did the users choose their routes?

Routing was likely greedy: users could not be sure whether the next hop is really closer to the destination (in terms of number of hops).

The network topology must also be navigable!

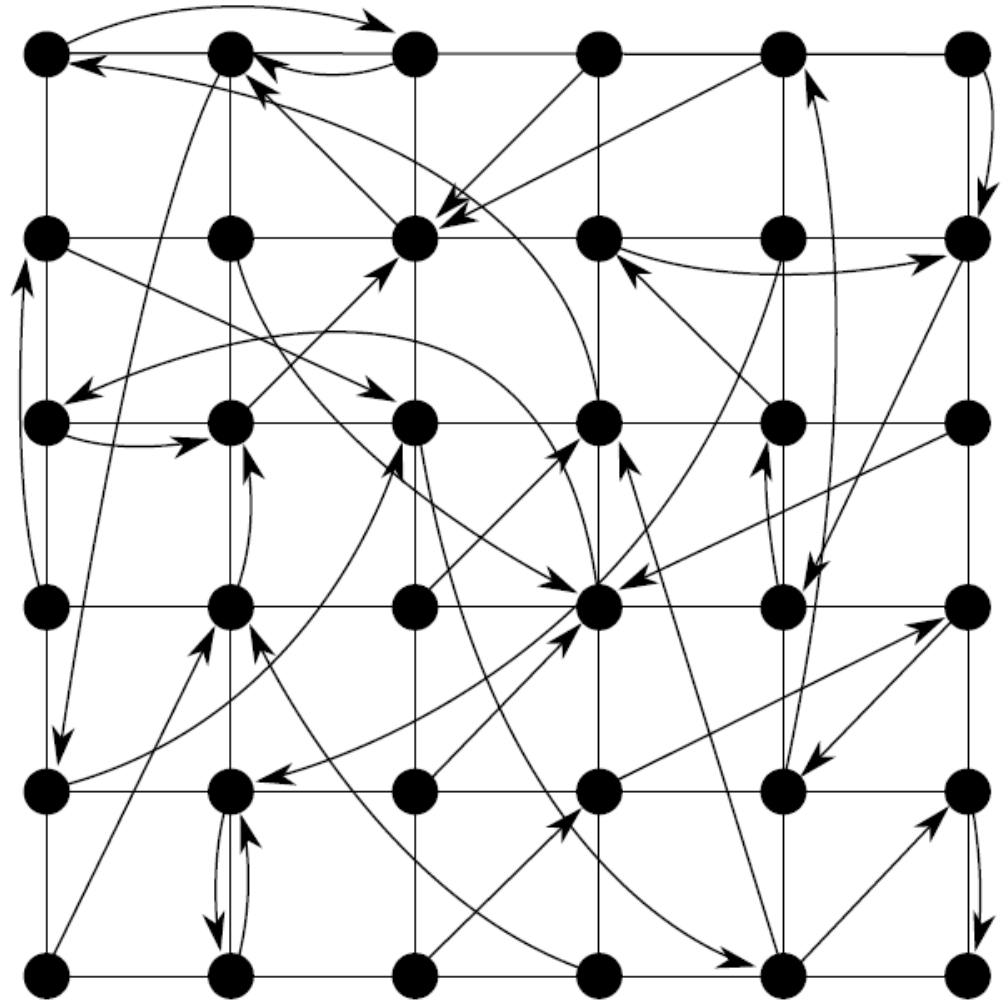
Kleinberg: Routing on Augmented Grid

Square grid (**small range contacts**) augmented with **long-range random links** chosen according to a probability distribution.

Node u points to node v with probability

$$d(u, v)^{-\alpha} / \sum_{w \in V \setminus \{u\}} d(u, w)^{-\alpha}$$

where $d()$ is the distance in the grid and α is a parameter.



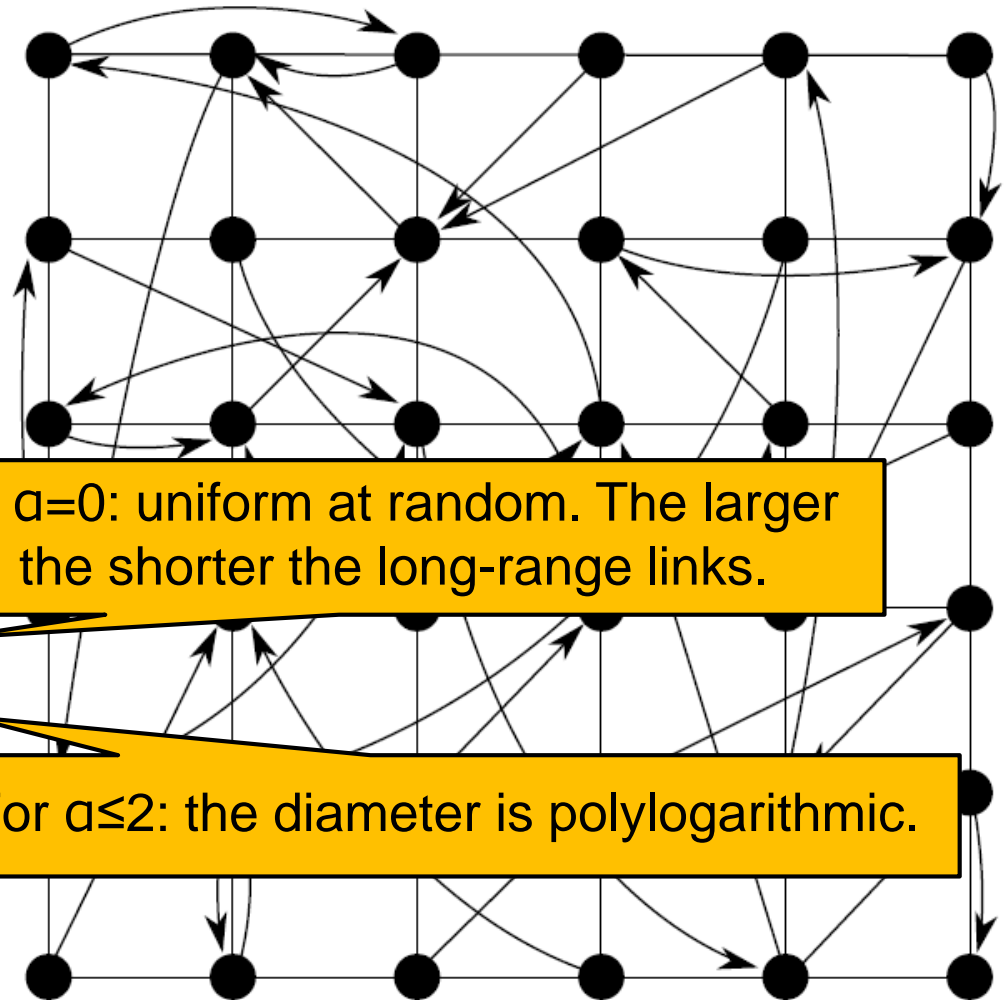
Kleinberg: Routing on Augmented Grid

Square grid (**small range contacts**) augmented with **long-range random links** chosen according to a probability distribution.

Node u points to node v with probability

$$d(u, v)^{-\alpha} / \sum_{w \in V \setminus \{u\}} d(u, w)^{-\alpha}$$

where $d()$ is the distance in the grid and α is a parameter.

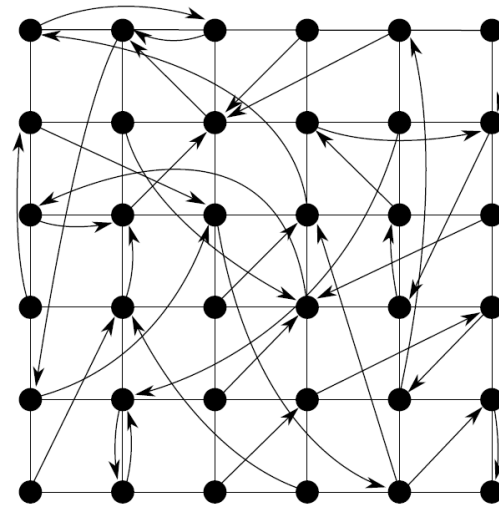


Kleinberg Routing

Kleinberg proposed the following greedy routing algorithm:

Greedy Routing

while (not at destination):
go to neighbor which is closest to destination
(considering **grid distance** only)



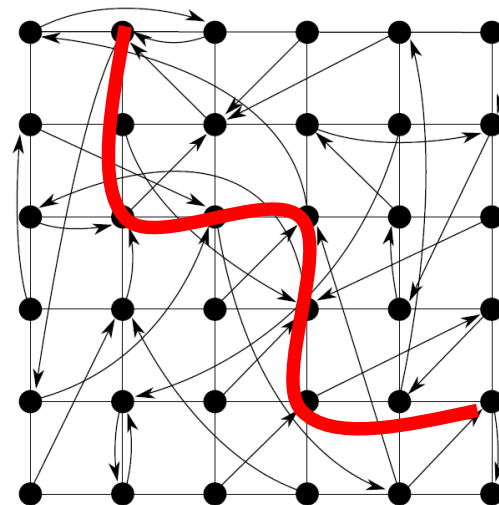
Kleinberg Routing

Kleinberg proposed the following greedy routing algorithm:

Greedy Routing

while (not at destination):
go to neighbor which is closest to destination
(considering **grid distance** only)

Note: in $O(\sqrt{n})$ we will definitely arrive at the destination: along grid!
But Milgram promises more...



Kleinberg Routing

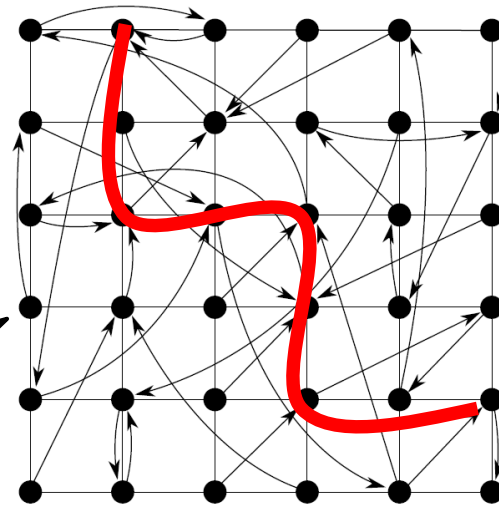
Kleinberg proposed the following greedy routing algorithm:

Greedy Routing

while (not at destination):
go to neighbor which is closest to destination
(considering **grid distance** only)

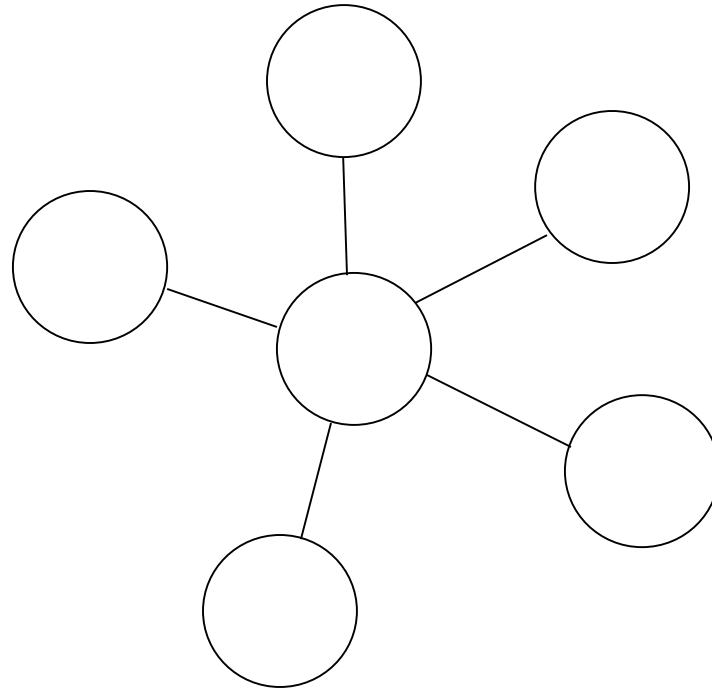
Note: in $O(\sqrt{n})$ we will definitely arrive at the destination: along grid!
But Milgram promises more...

In the lecture notes: proof that routing terminates in $O(\log^2 n)$ time.



Propagation Studies

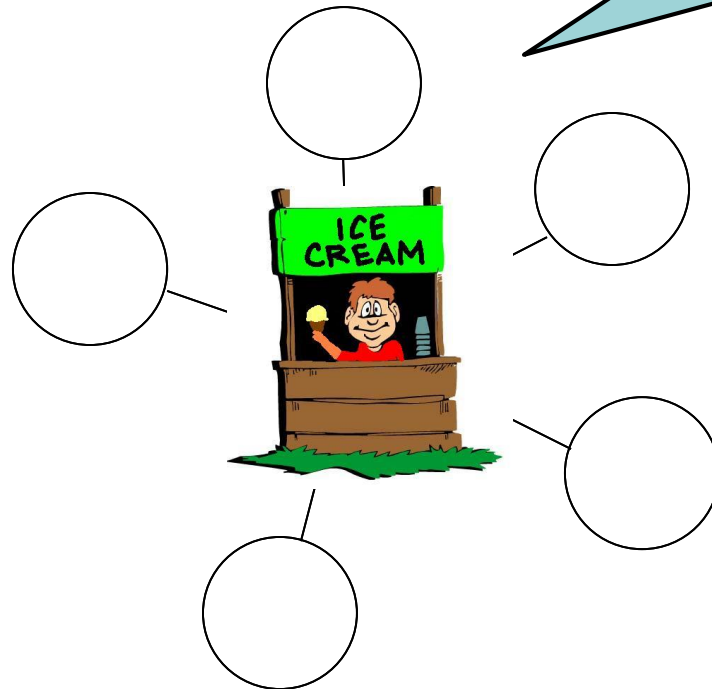
Where to place an ice-cream shop in a graph so I reach most customers?



Propagation Studies

Where to place an ice-cream shop in a graph so I reach most customers?

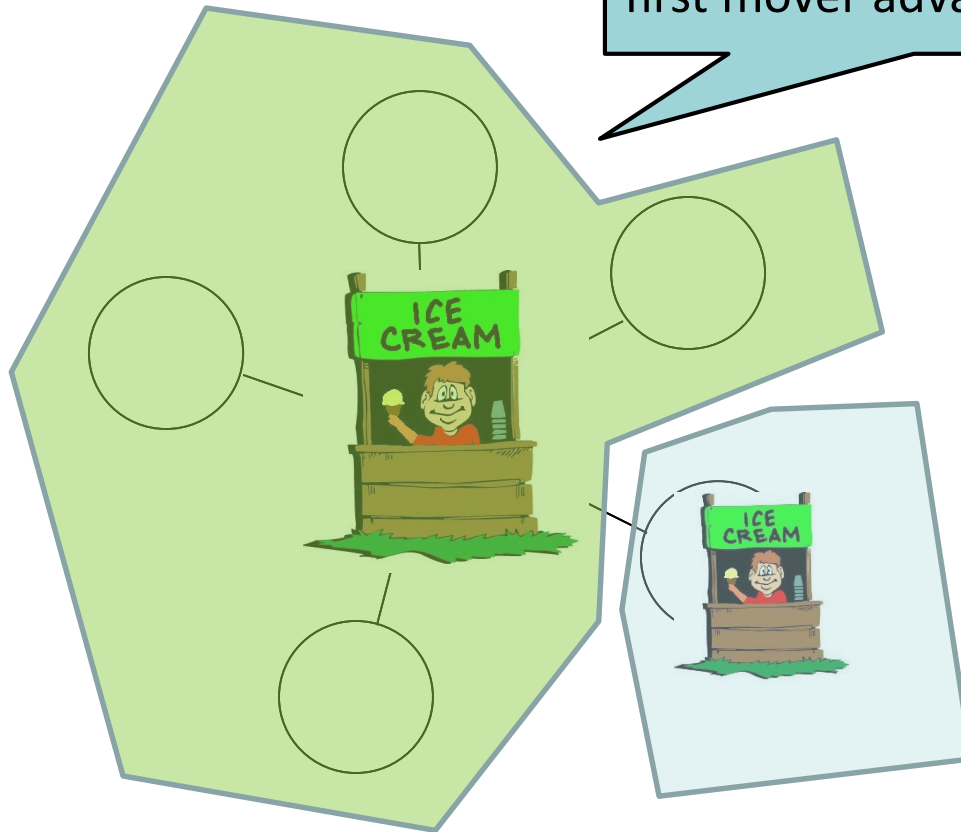
Placing the shop in the center:
first mover advantage!



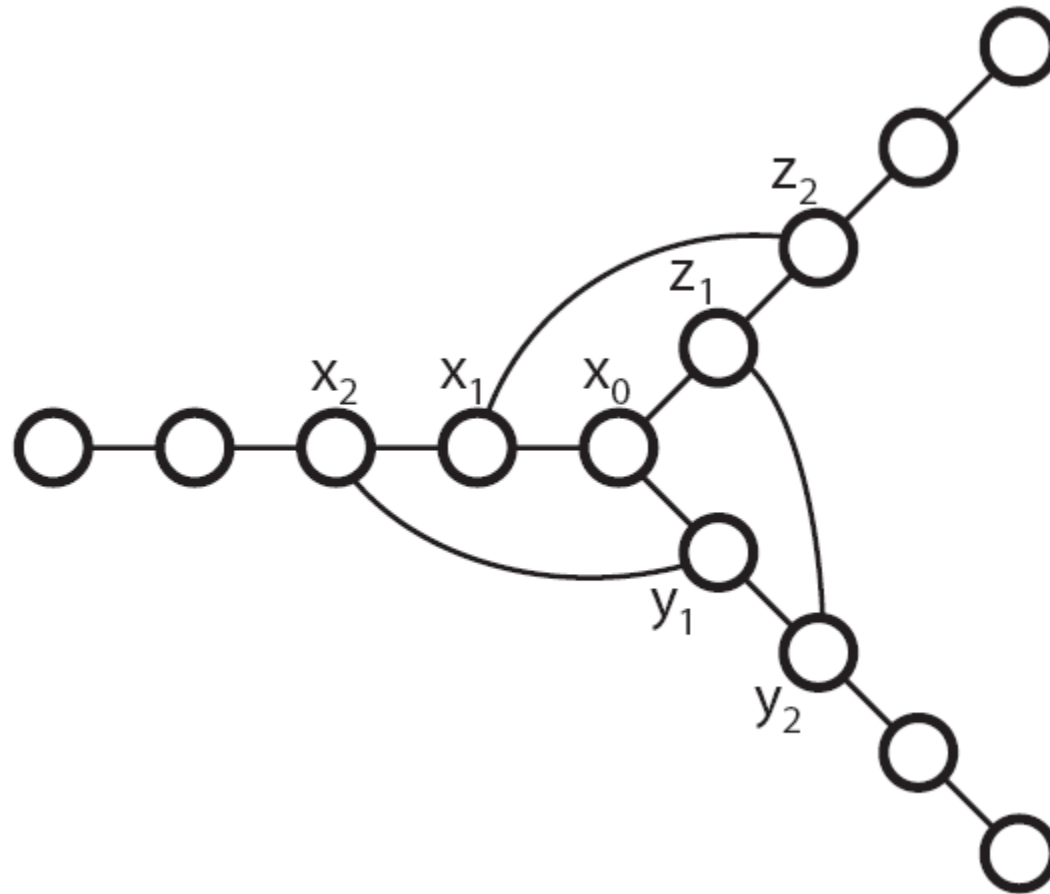
Propagation Studies

Where to place an ice-cream shop in a graph so I reach most customers?

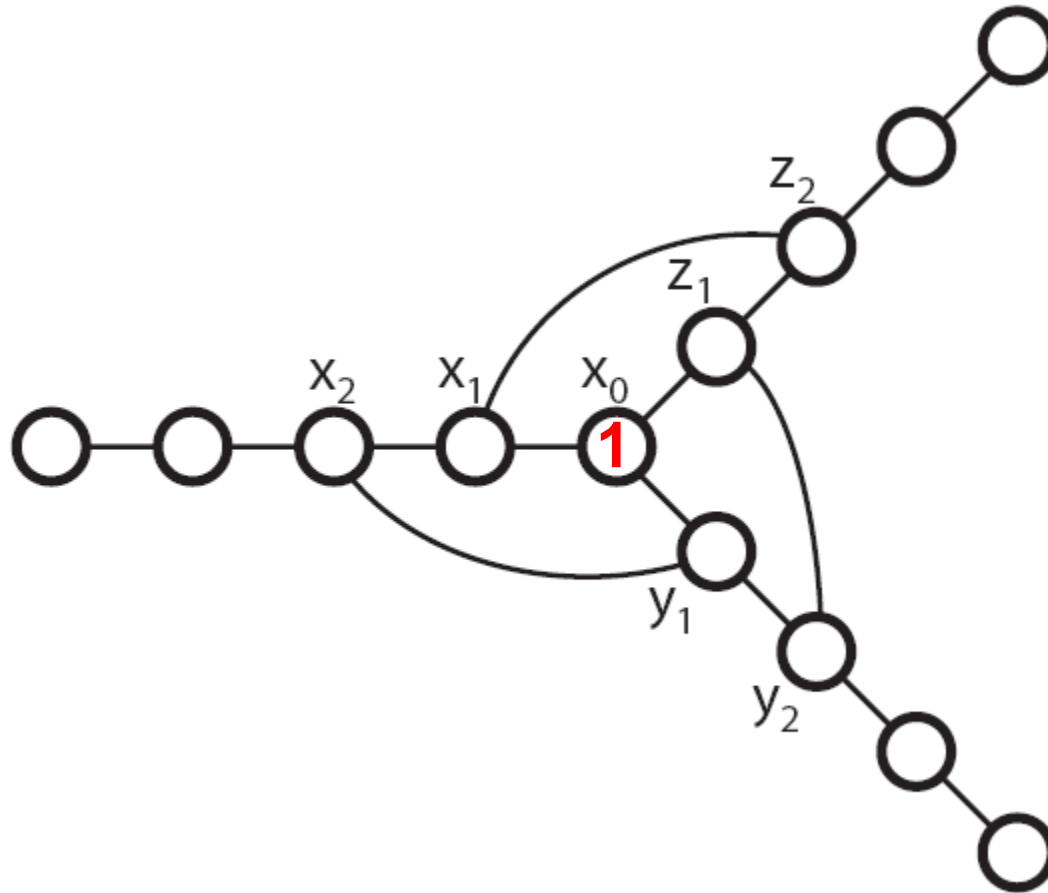
Placing the shop in the center:
first mover advantage!



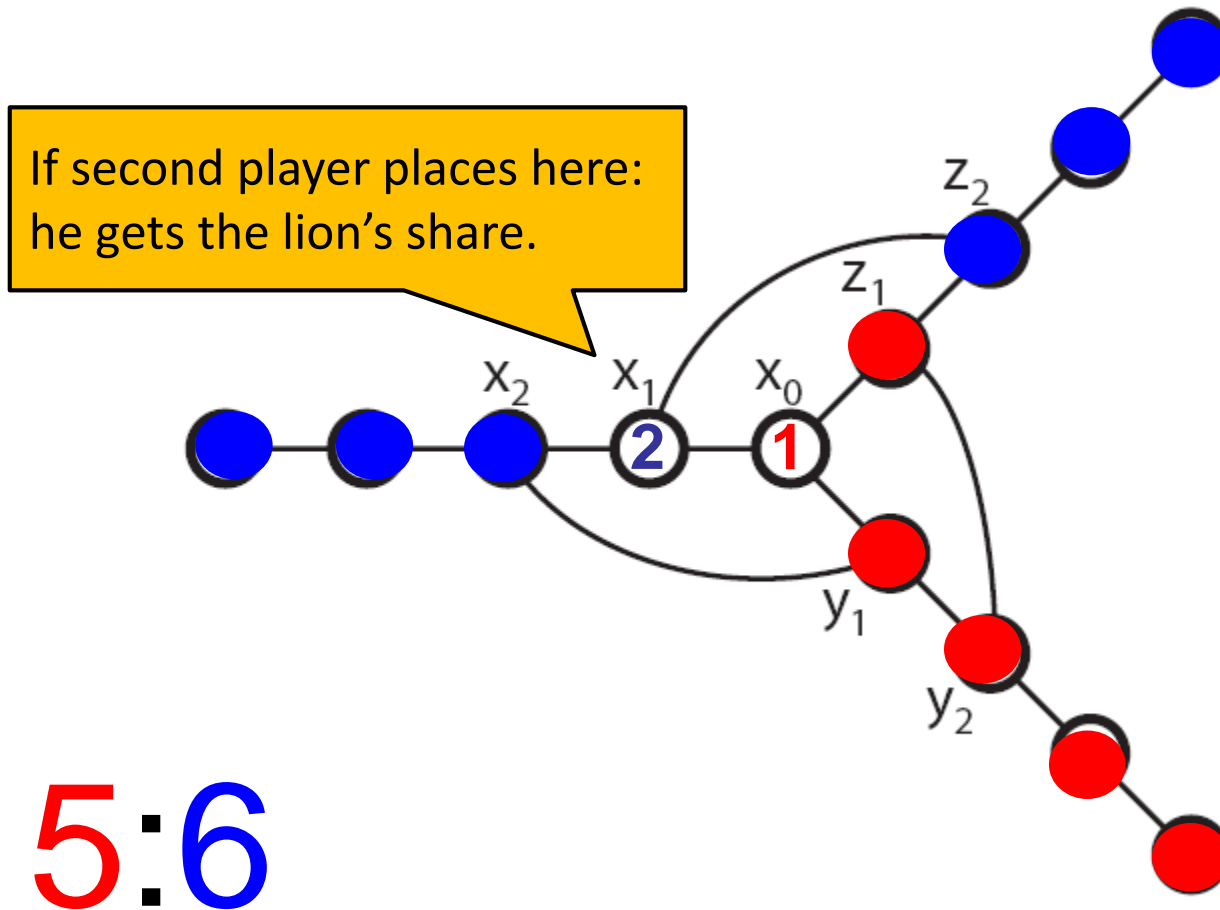
Always first mover advantage?



Always first mover advantage?

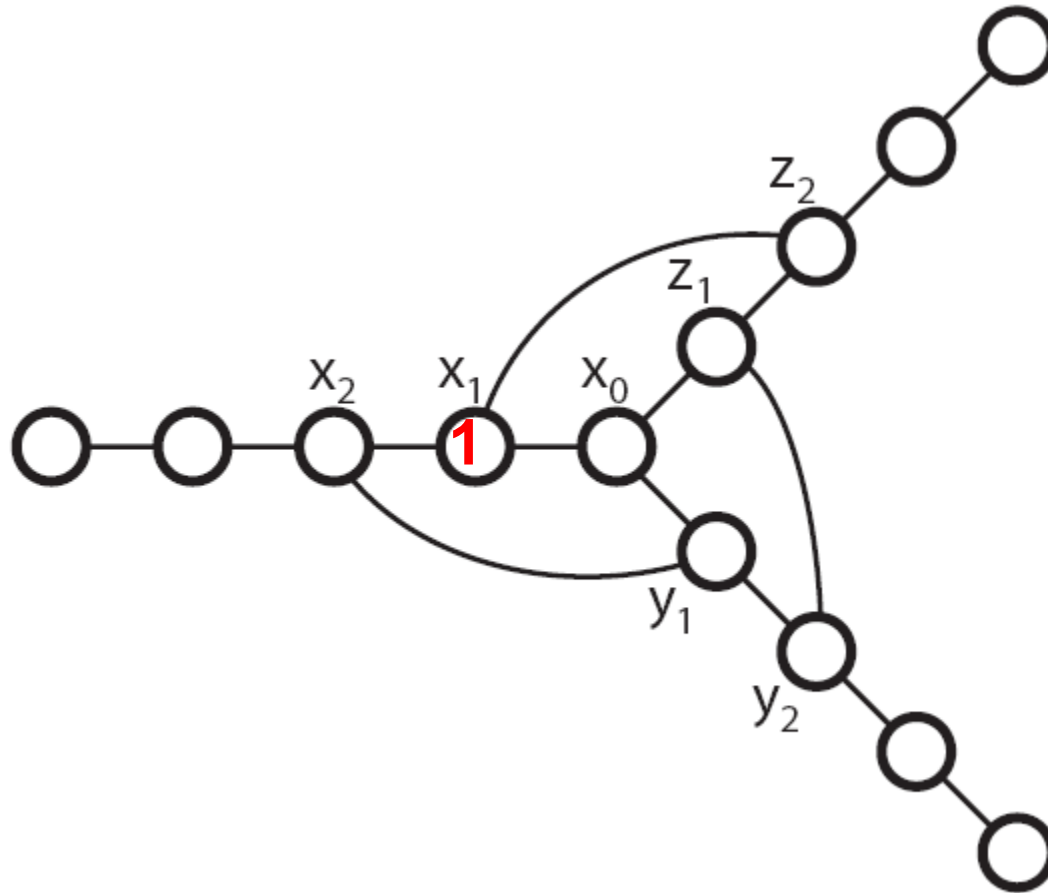


Always first mover advantage?

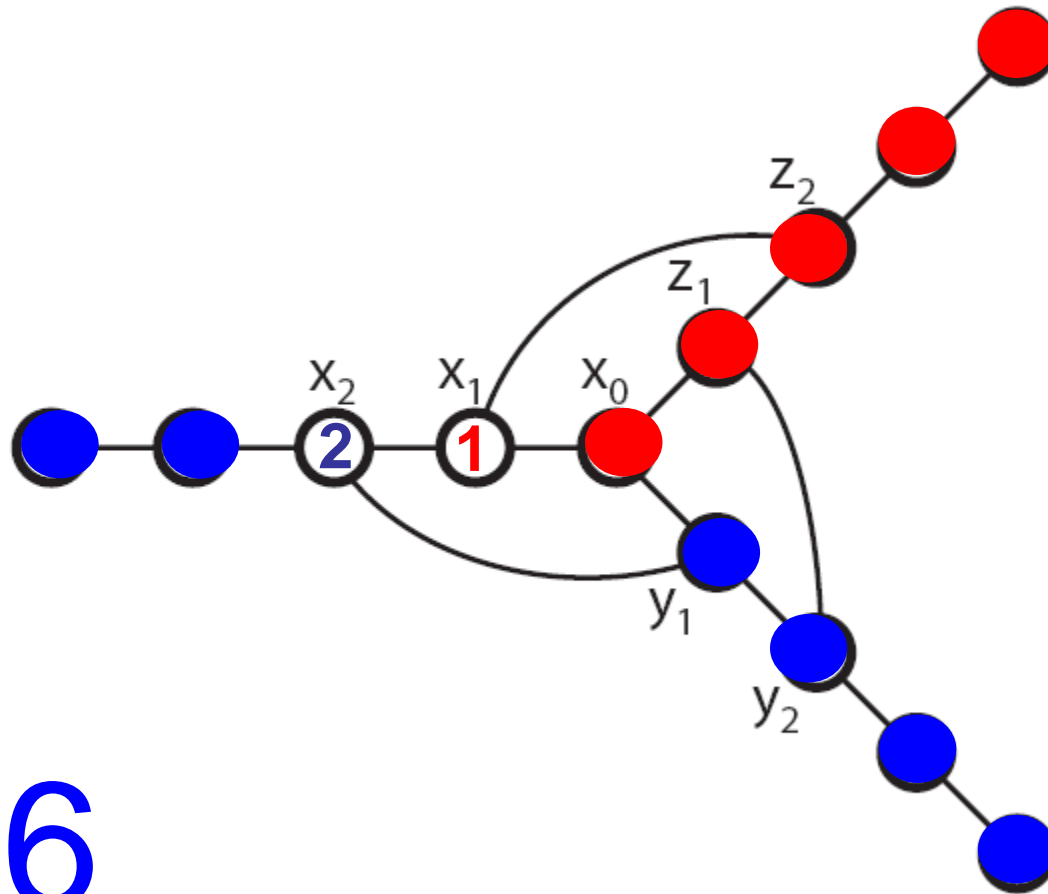


Always first mover advantage?

Where second?



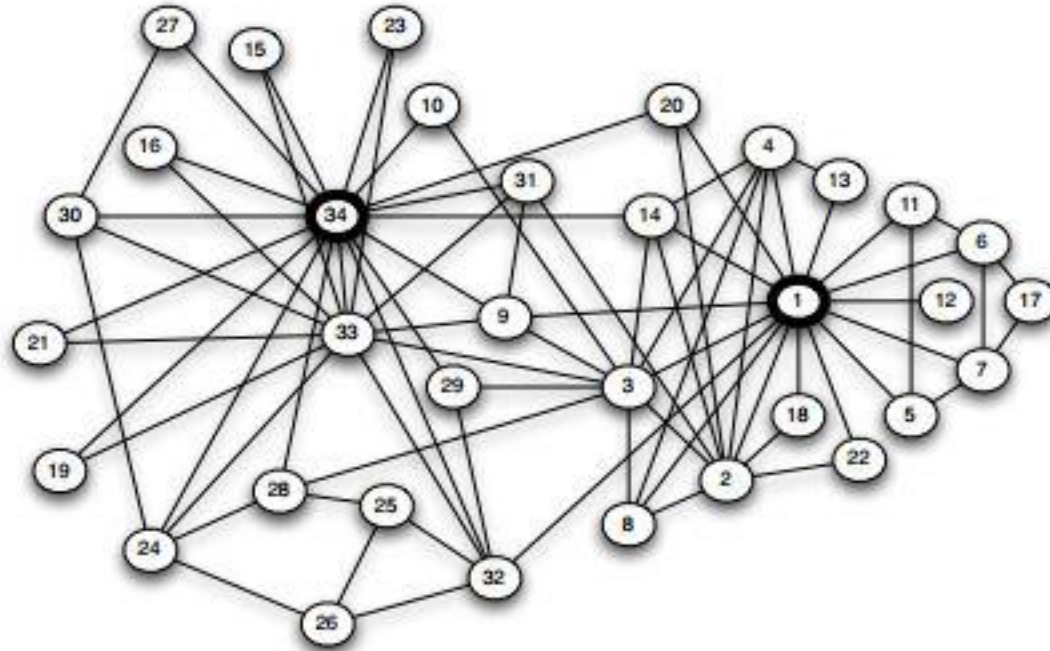
Always first mover advantage?



5:6

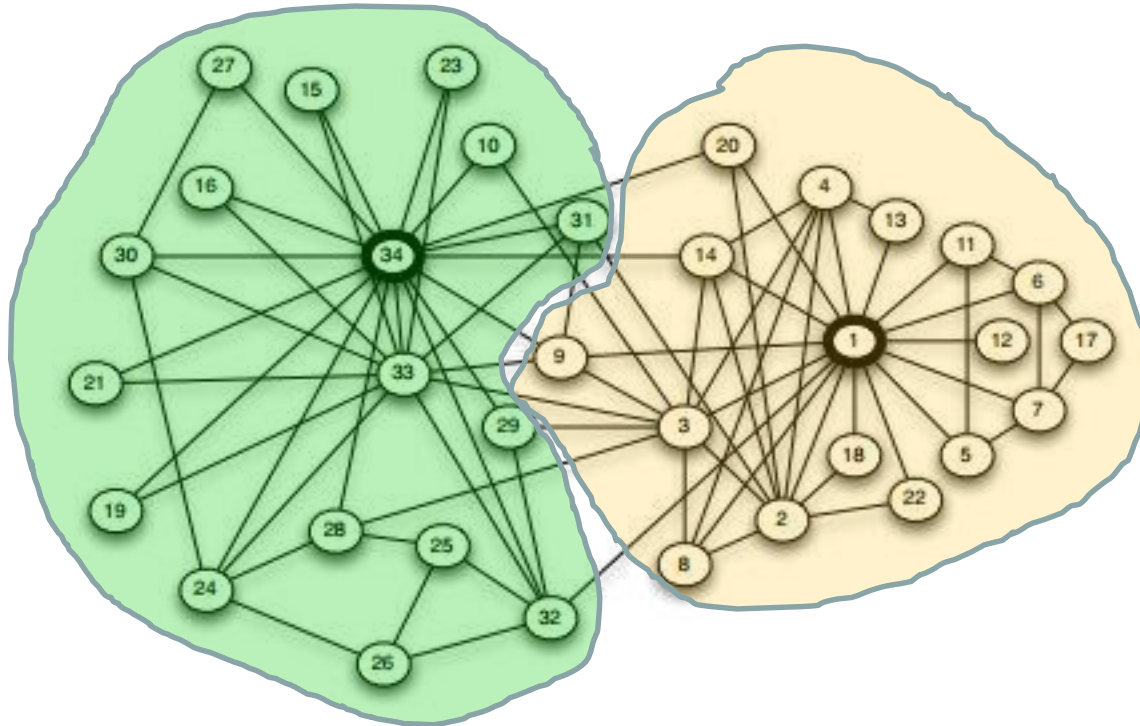
Community Detection

Social Network Analysis



Social network of a karate club: Dispute between instructor and president led the club to split into two rival clubs. Guess where!

Social Network Analysis

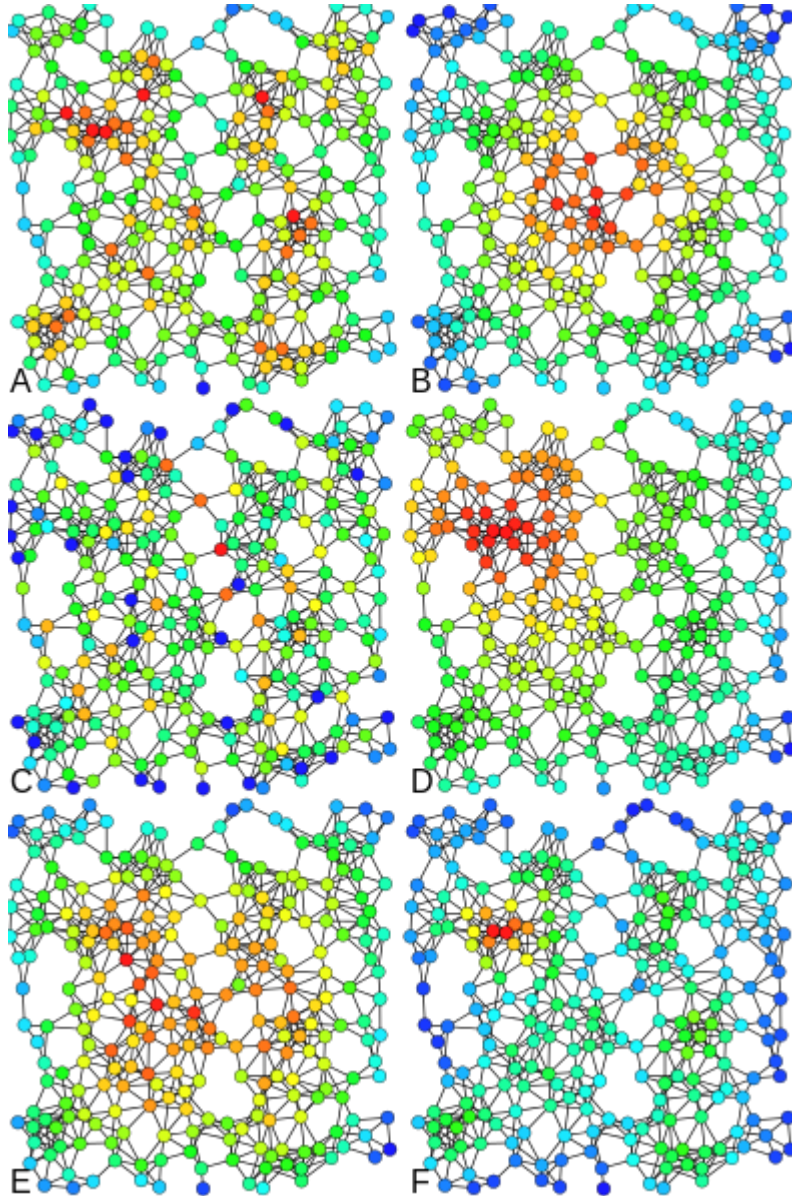


Social network of a karate club: Dispute between instructor and president led the club to split into two rival clubs. Guess where!

It's about **community detection!**

(Vertex) Centrality

How important is the node? Many definitions exist: degree/betweenness/closeness/... centrality.



A: Degree Centrality

Simply the degree of a node. (E.g., Karate Club president)

B: Closeness Centrality

Defined as the inverse of the sum of the distances to all other nodes.

$$C_C(v) = \sum_{t \in V \setminus v} 2^{-d_G(v,t)}.$$

C: Betweenness Centrality

Fraction of shortest paths that include v .

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

D: Eigenvector Centrality

E: Katz Centrality and Page Rank

F: Alpha Centrality

blue = min, red = max

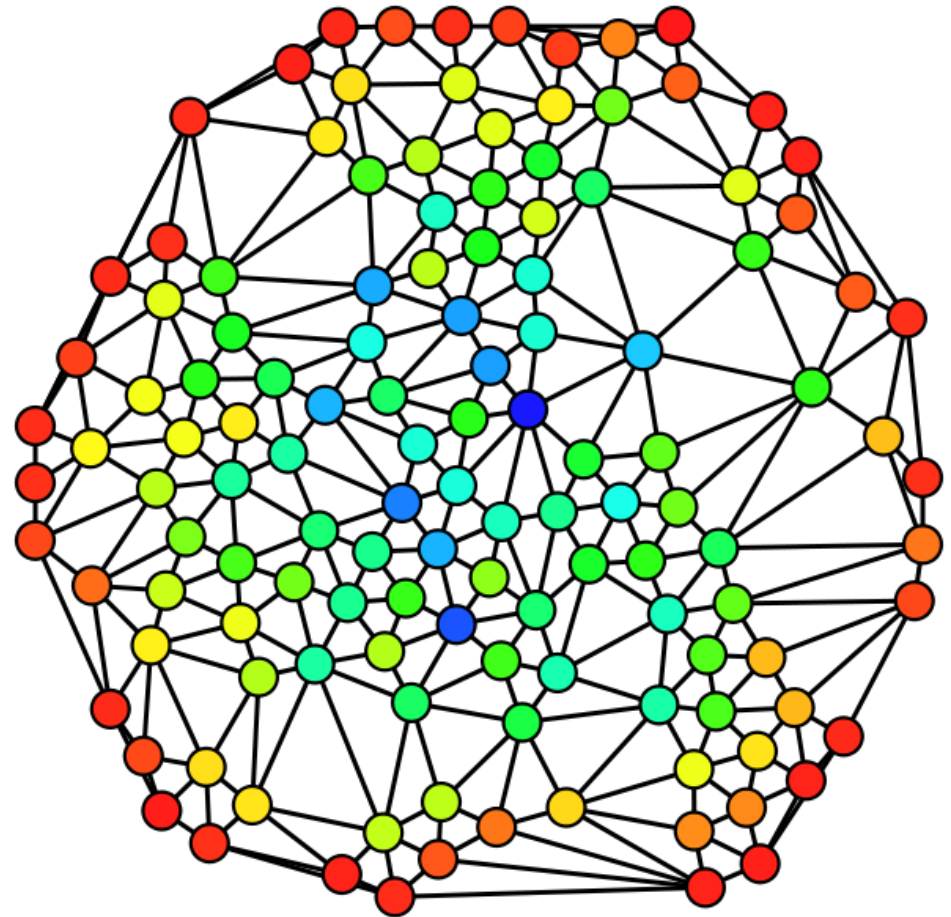
Betweenness Centrality

How to compute?

1. All-Pairs-Shortest-Paths
E.g., Floyd-Warshall
2. Better algos exist 😊

How to compute community from centralities?

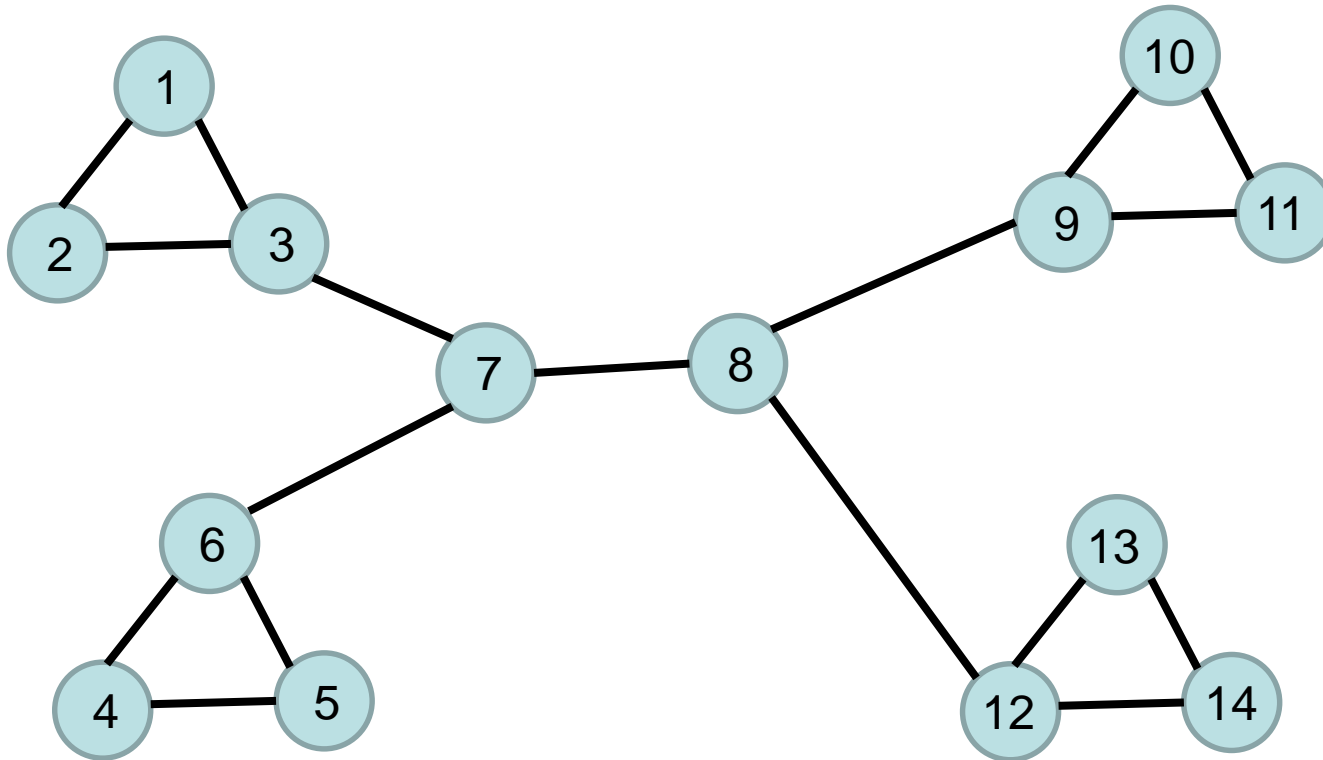
1. Idea: use edge-centrality
2. Remove edges of high centrality
(likely to connect different communities?)



red = min, blue = max

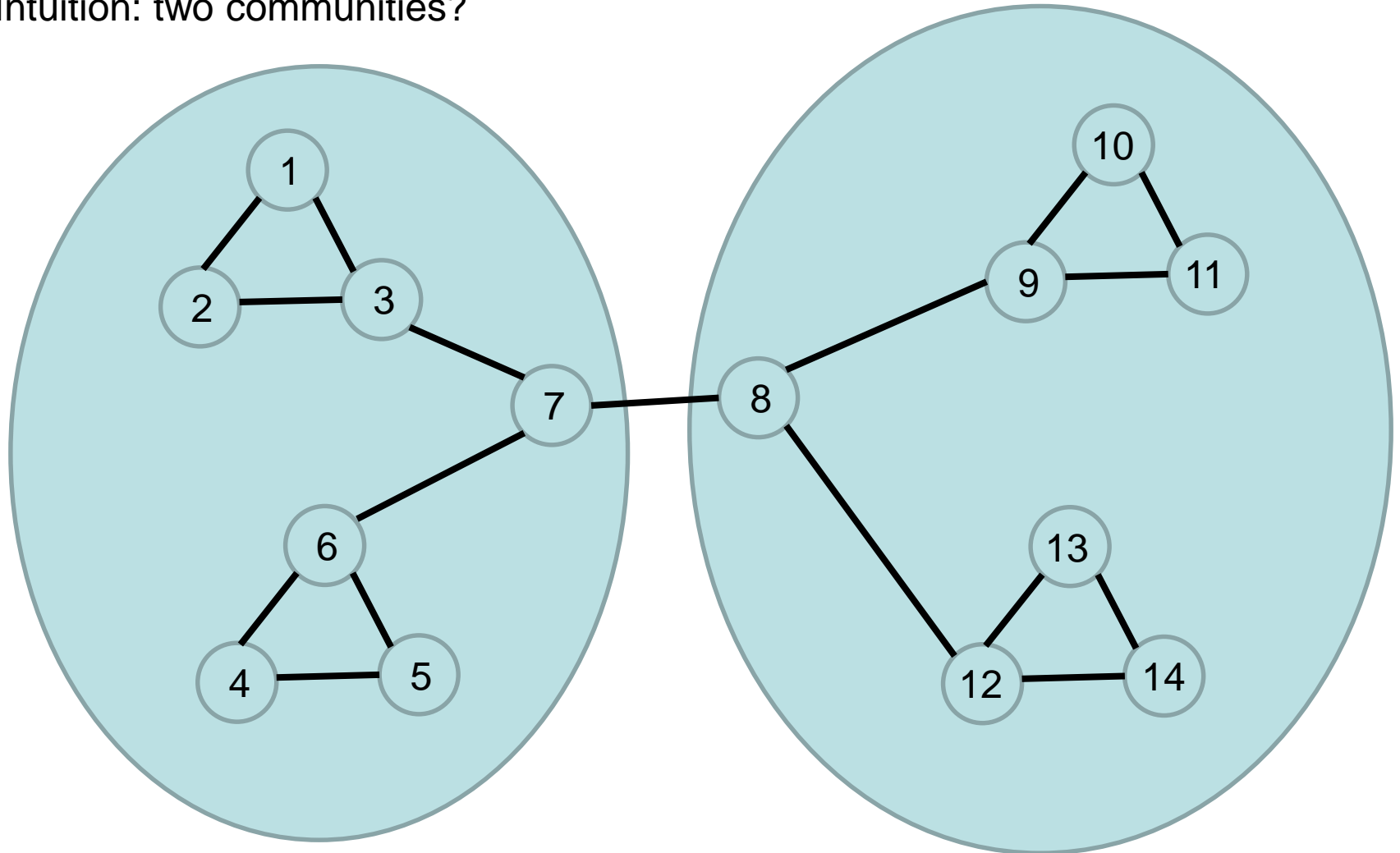
Betweenness Centrality

Intuition: two communities?



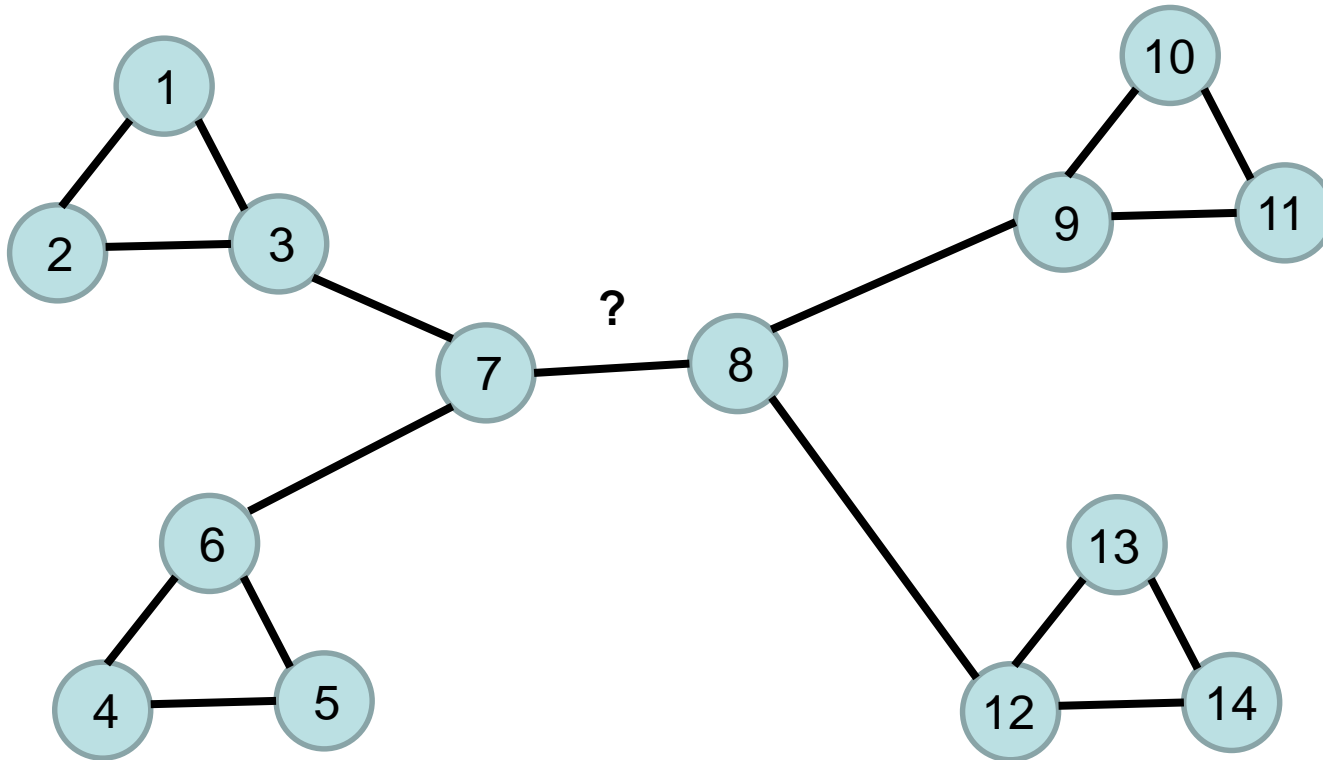
Betweenness Centrality

Intuition: two communities?



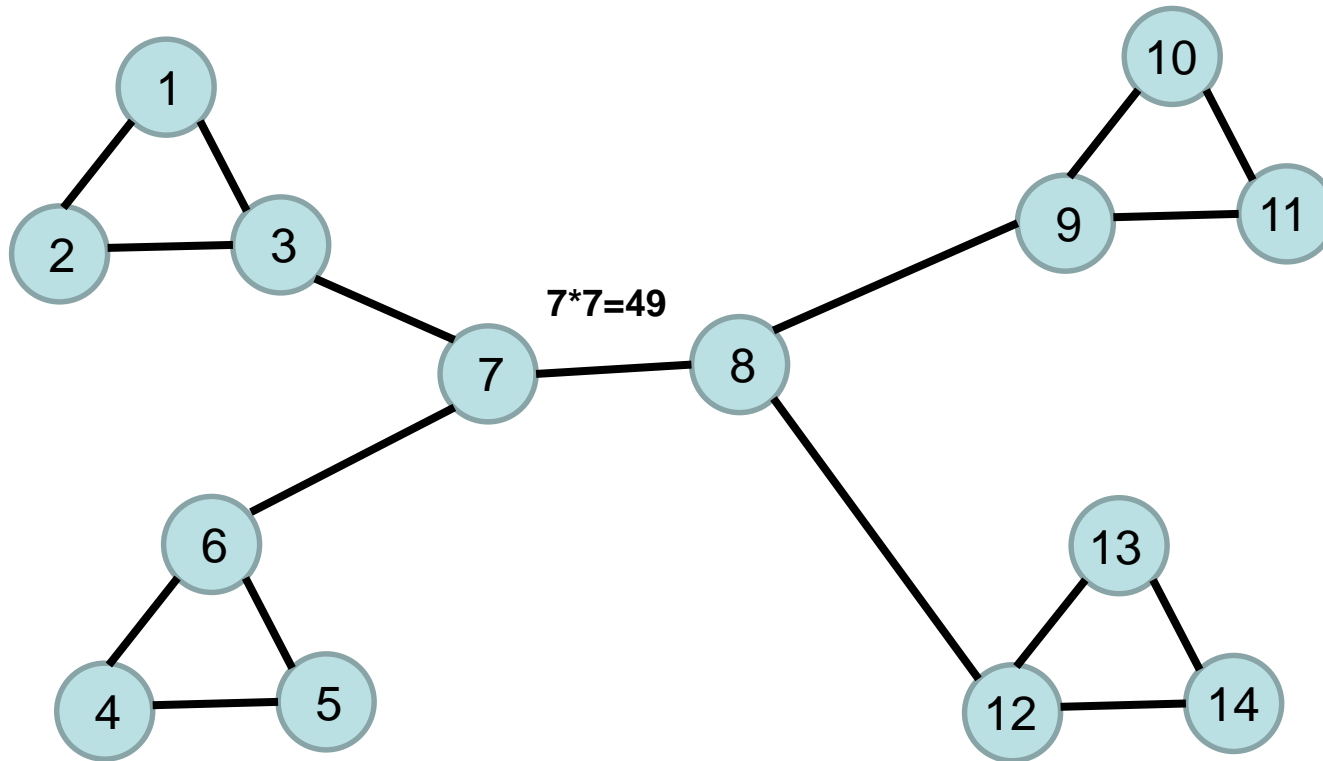
Betweenness Centrality

Betweenness?



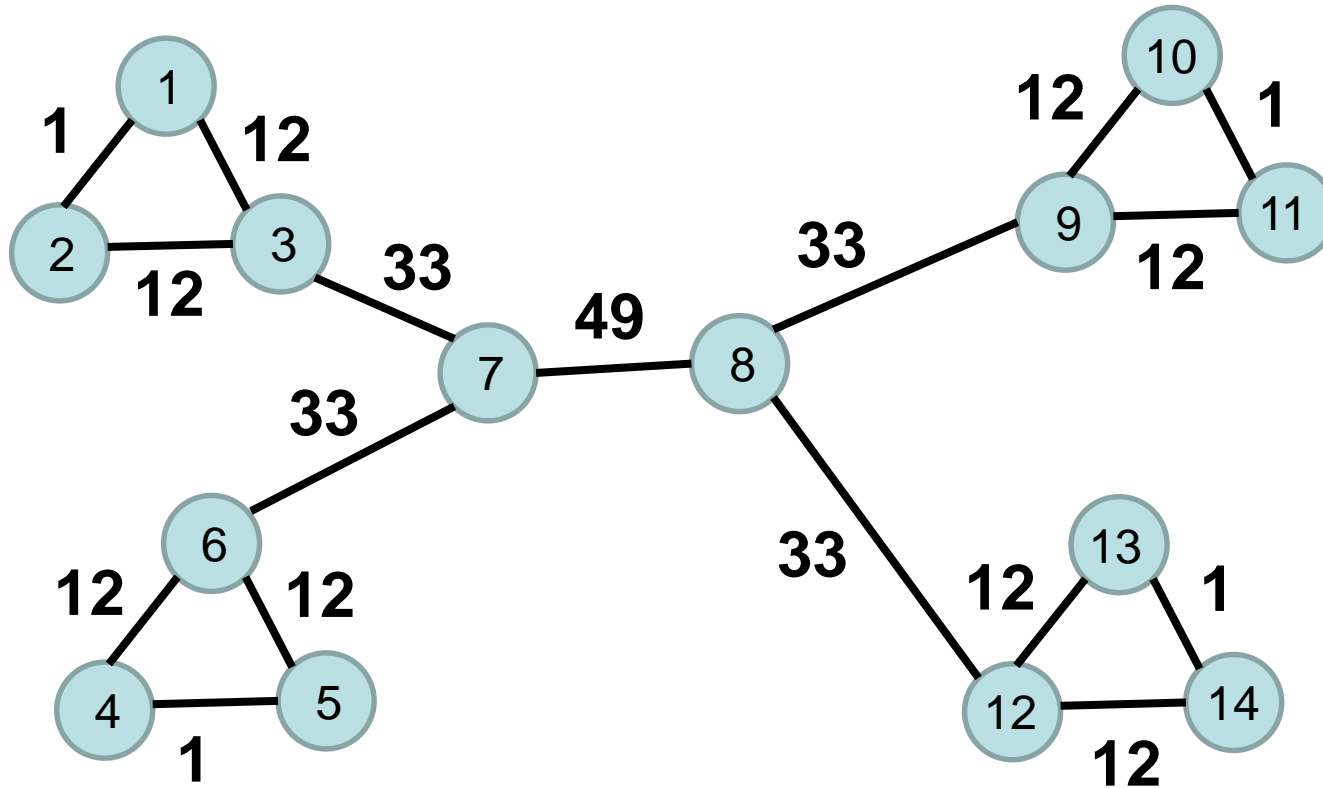
Betweenness Centrality

Betweenness?



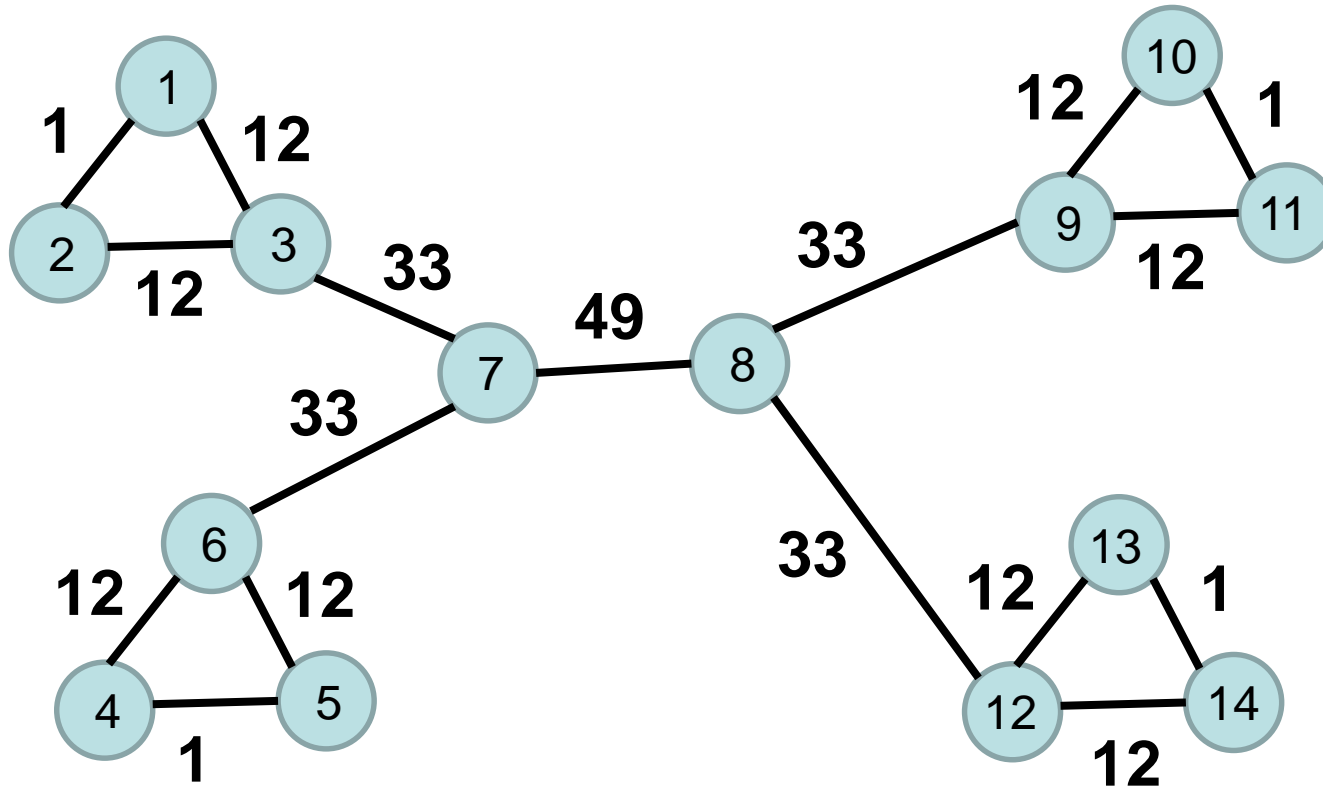
Betweenness Centrality

Betweenness?



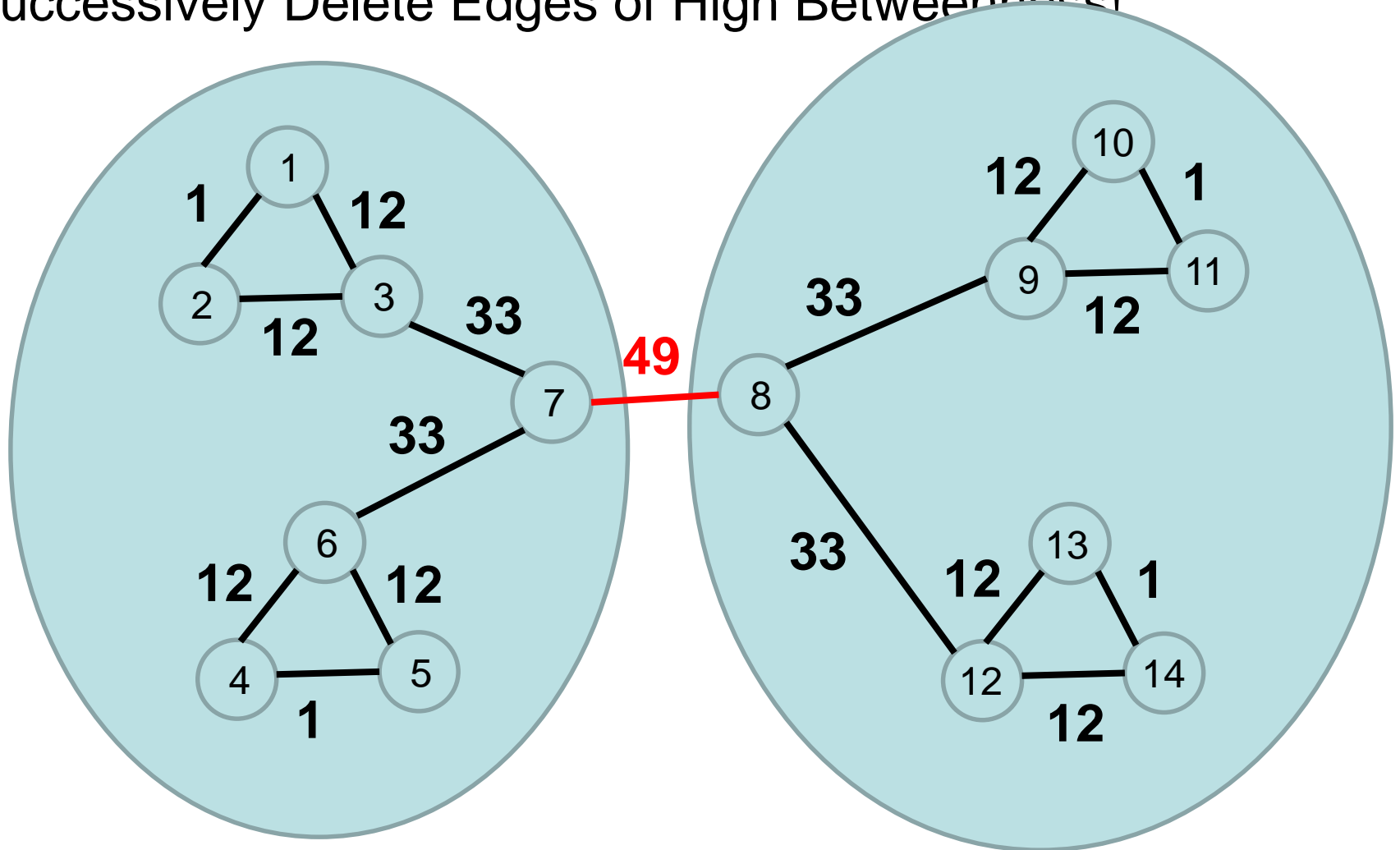
The Girvan-Newman Method

Successively Delete Edges of High Betweenness!



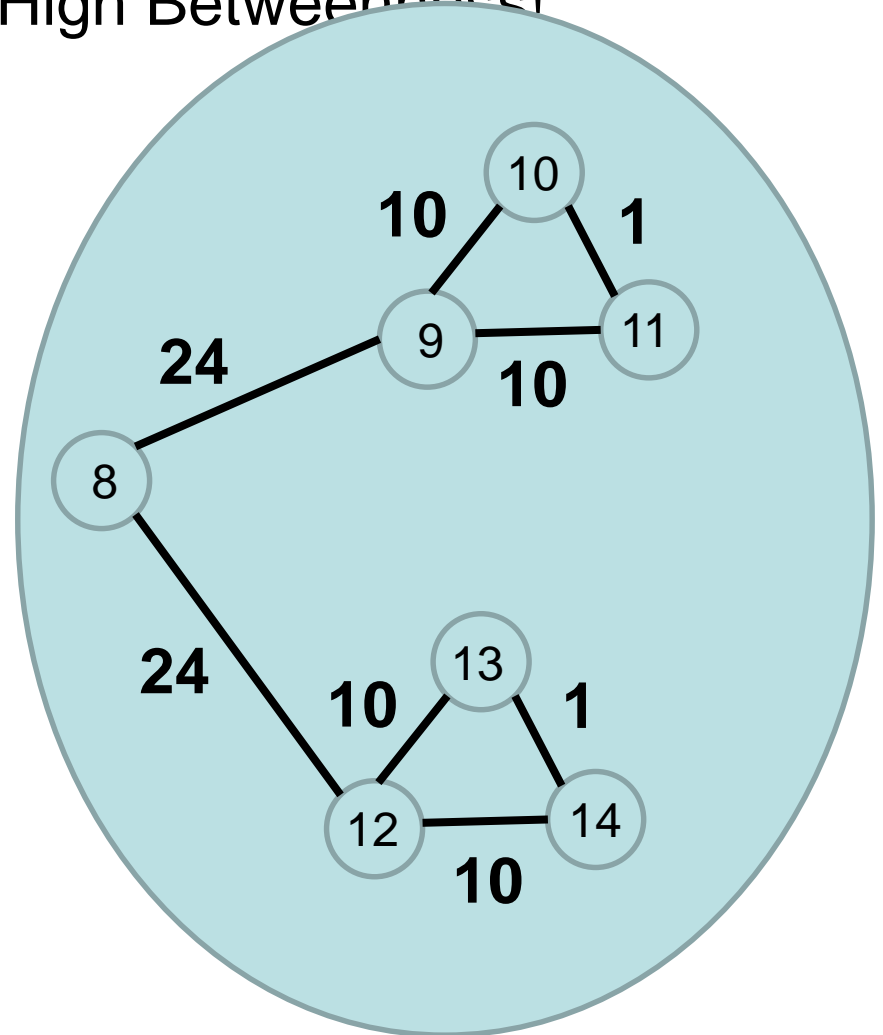
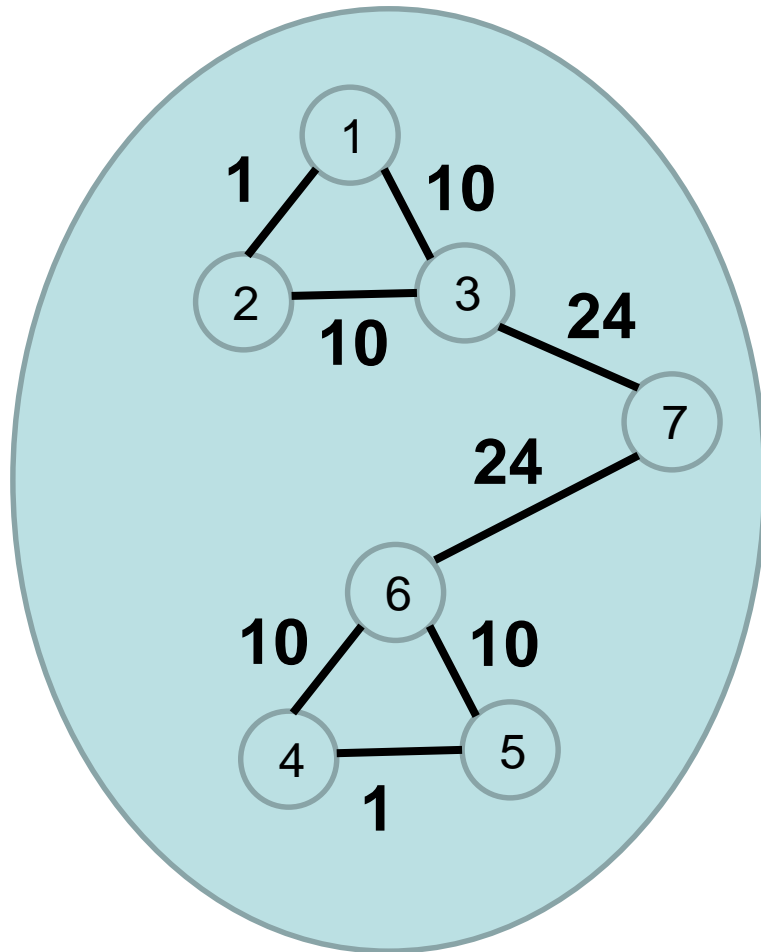
The Girvan-Newman Method

Successively Delete Edges of High Betweenness



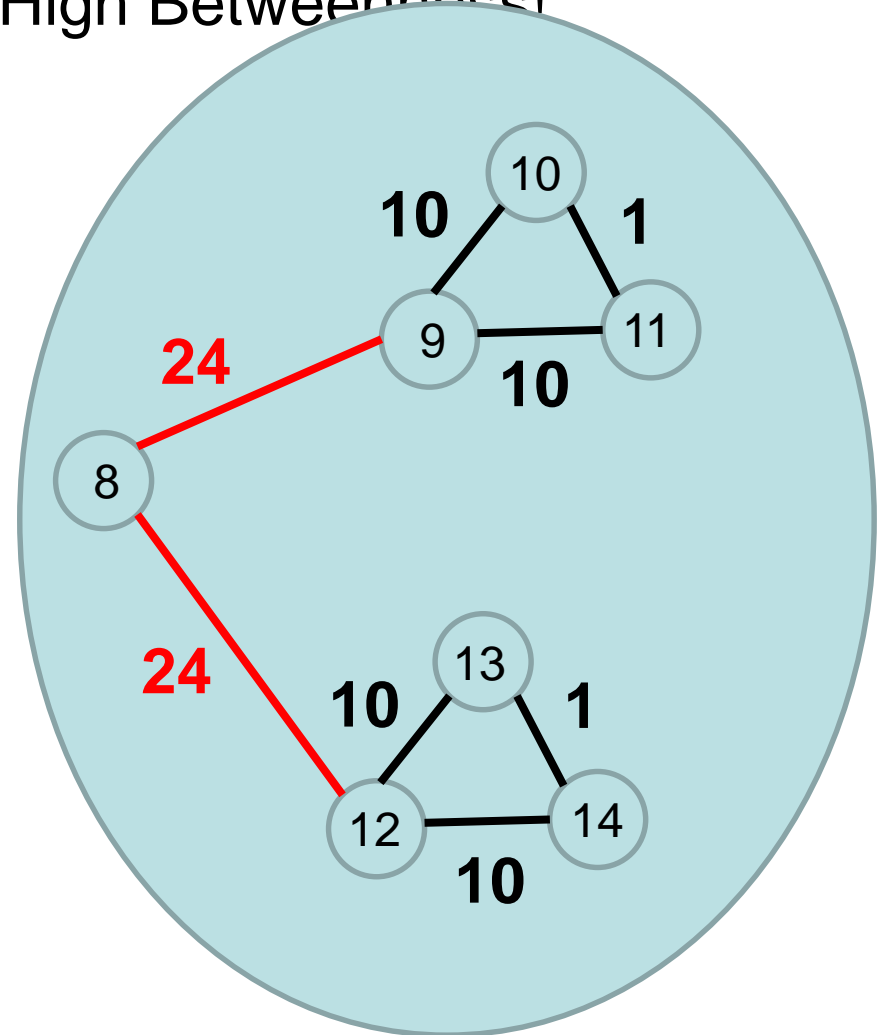
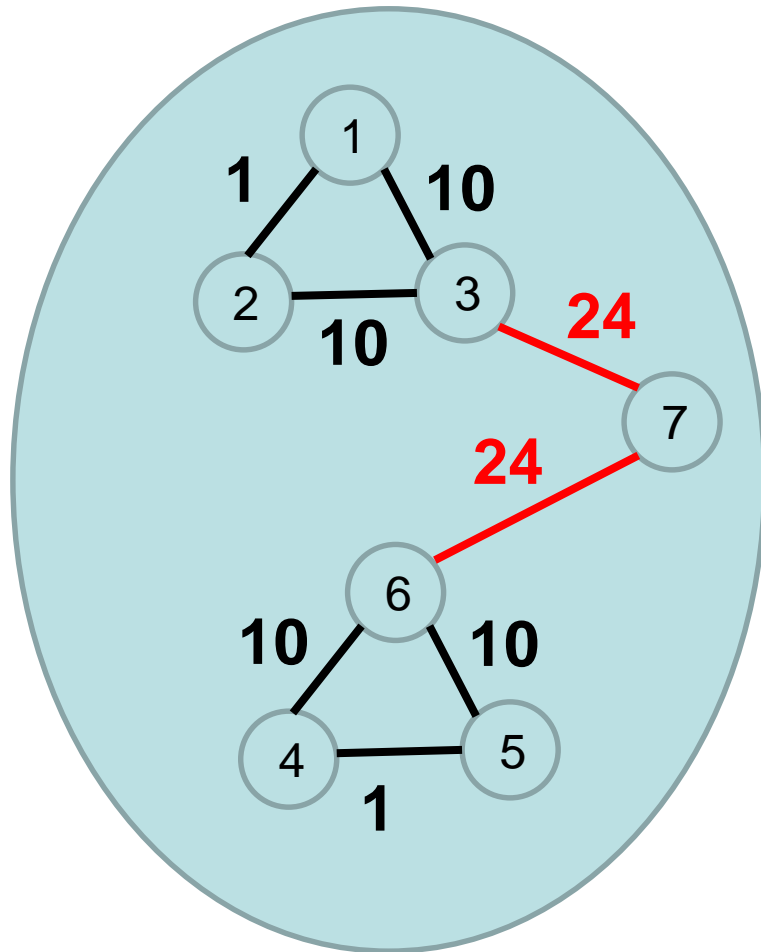
The Girvan-Newman Method

Successively Delete Edges of High Betweenness



The Girvan-Newman Method

Successively Delete Edges of High Betweenness



End of lecture