Web Science: Social Networks

(Part 1 - Graphs)

CS 432/532

Old Dominion University

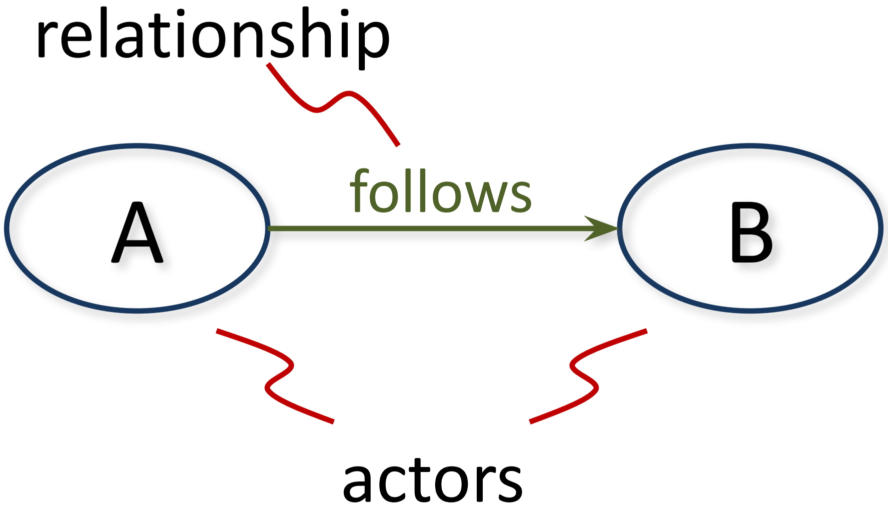
Most slides based on Ch 2 of *Networks, Crowds and Markets* by Easley & Kleinberg (2010)

(*abbreviated as EK10*)

Kleinberg is the HITS guy (hubs and authorities)

Social Network

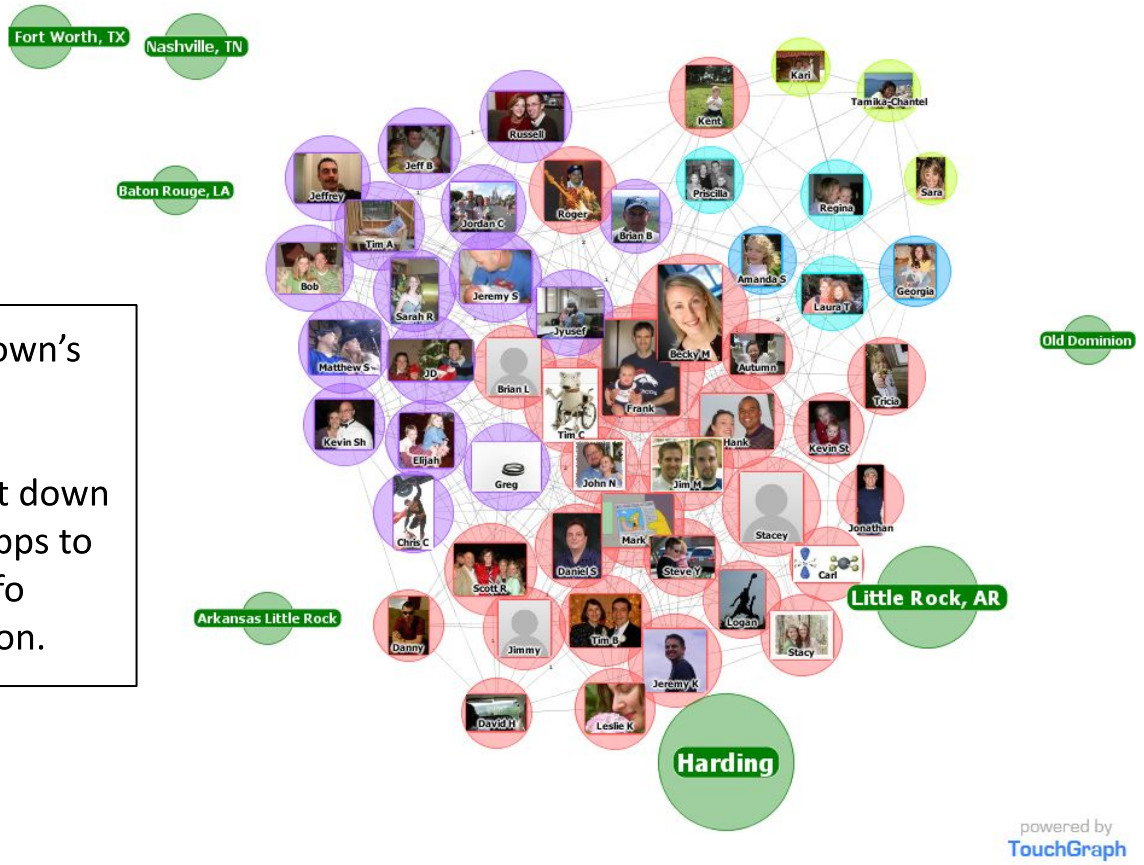
- Definition: Graph composed of *actors* (people, organizations, groups) that are tied by *social links*



Online Social Networks (OSN)

- Digital incarnation of a social network
- Created by social network sites, social networking services
- Web-based services that allows users to:[1]
  - construct (semi-)public profiles within the system
  - "connect" to other users
  - view and traverse lists of connections created by others

[1] Boyd & Ellison (2007) - Social Network Sites: Definition, History, and Scholarship

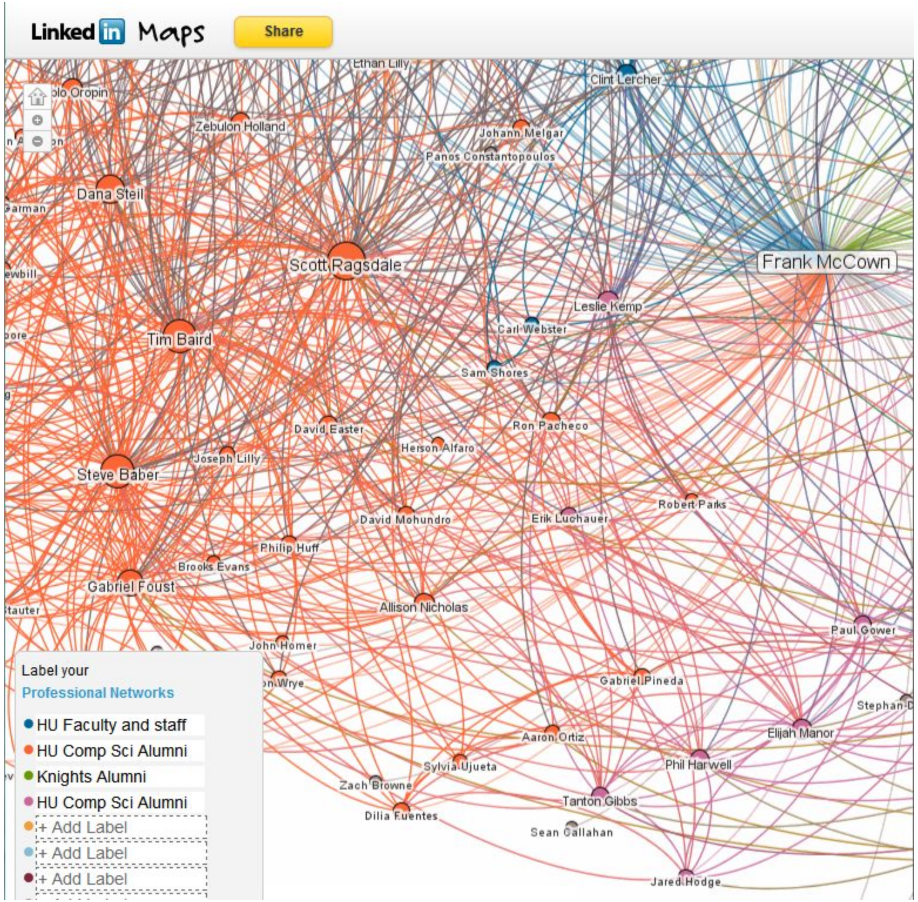


This is part of Dr. McCown's friend group pre-2014.

In 2015, Facebook shut down the API that allowed apps to access your friends' info without their permission.

Facebook Friends

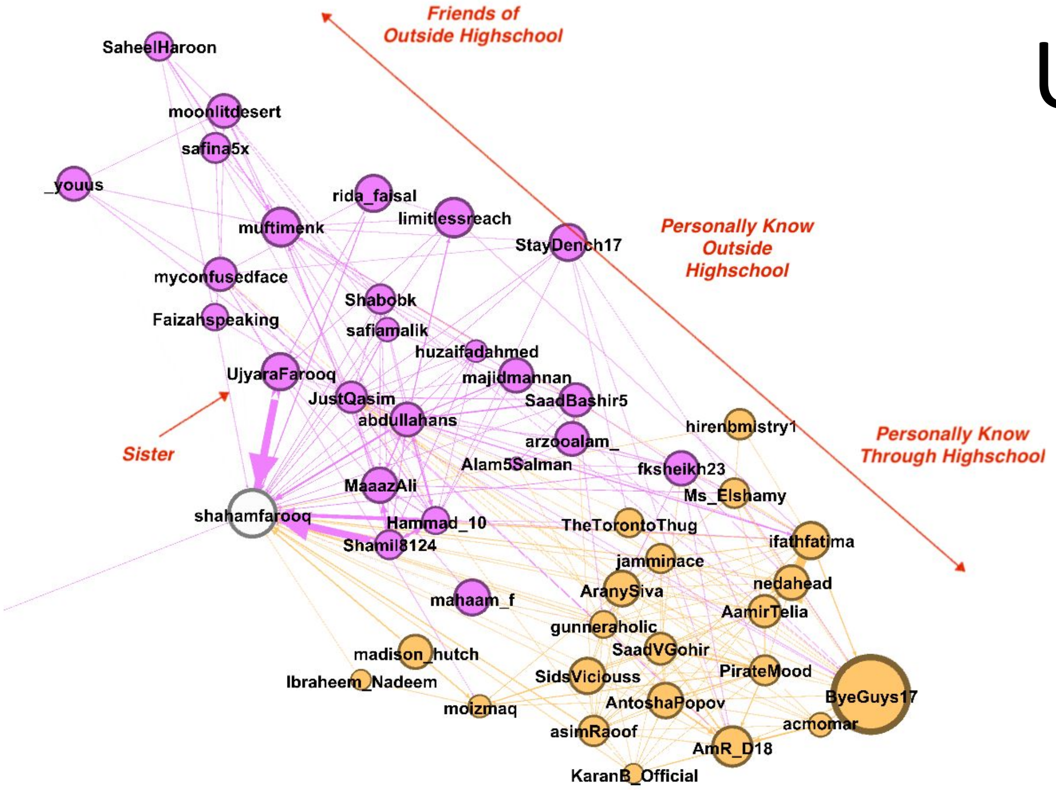
LinkedIn Maps



[Visualize your LinkedIn network with InMaps](), 2011

[LinkedIn Is Quietly Retiring Network Visualization Tool InMaps]() , 2014

User-generated Twitter Network



[Generating A Twitter Ego-Network & Detecting Communities | by Shaham]

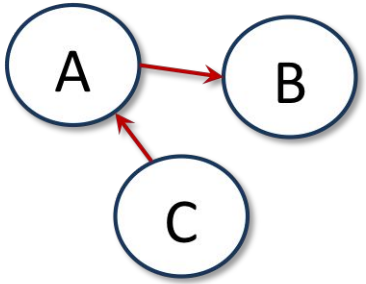
Why Study Social Networks?

- Social scientists use social networks to study how people interact and develop theories of social behavior. OSNs offer opportunities to study social networks at a larger scale.

- Understanding the structure of OSNs can lead to systems in the future that improve security, leverage trust, improve social interaction, etc.

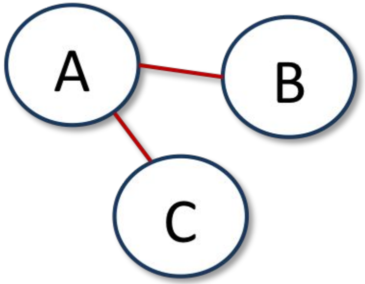
Graph Types

Directed



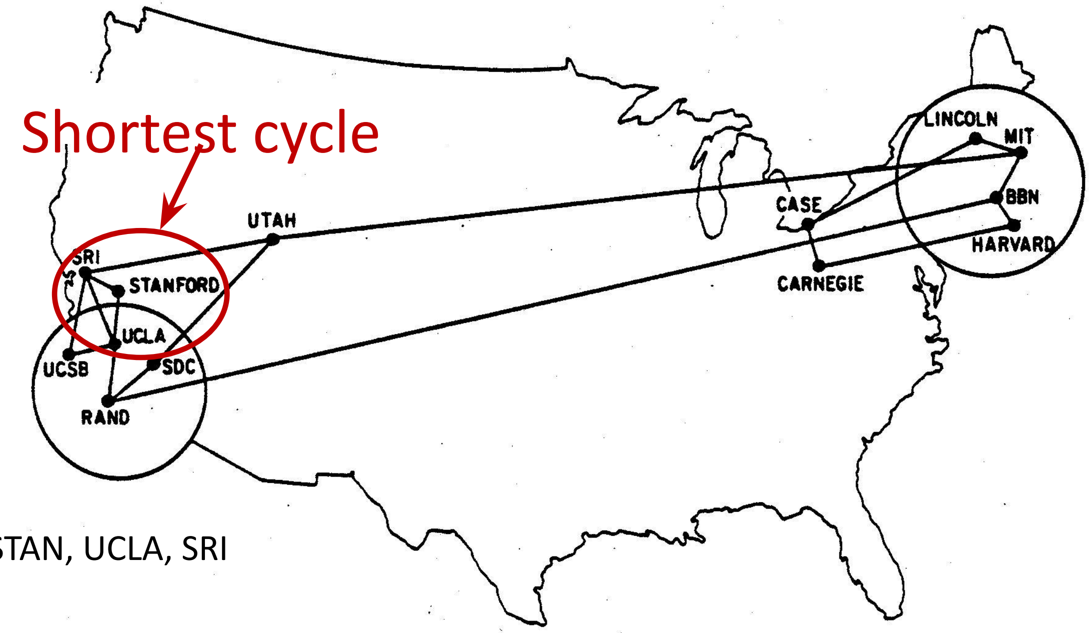
e.g., Twitter followers

Undirected



e.g., Facebook friends

Cycles



Shortest cycle is SRI, STAN, UCLA, SRI

[EK10] [Figure 2.2](): A network depicting the sites on the Internet, then known as the Arpanet, in December 1970. (Image from F. Heart, A. McKenzie, J. McQuillian, and D. Walden)

Components

Giant component

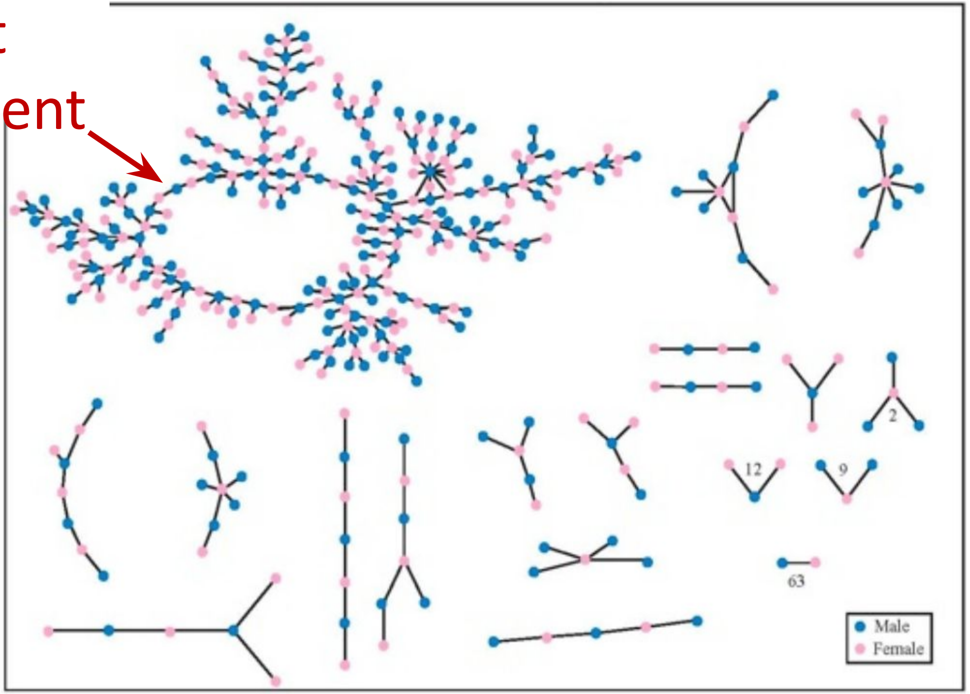


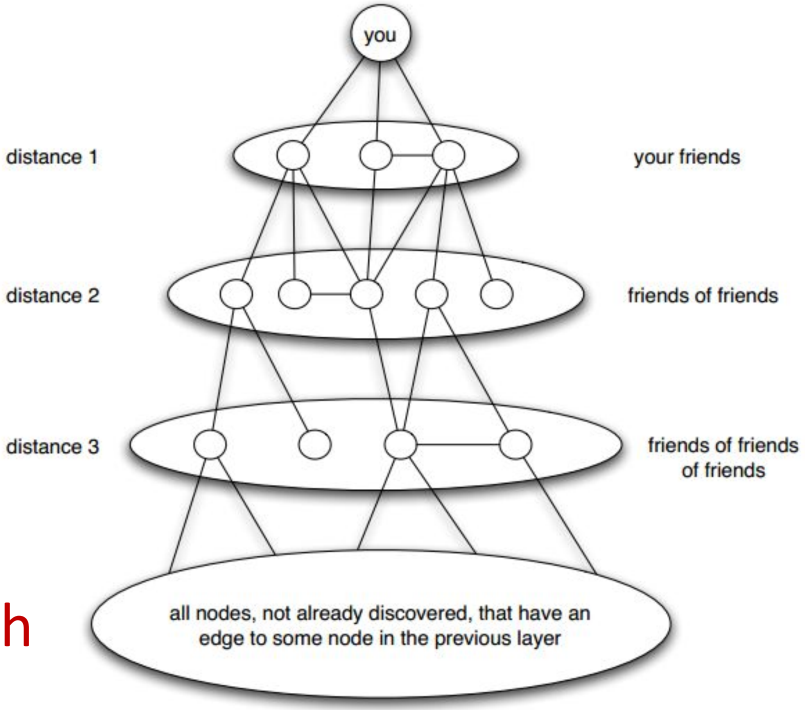
Figure 2.7: A network in which the nodes are students in a large American high school, and an edge joins two who had a romantic relationship at some point during the 18-month period in which the study was conducted [49].

[EK10]

Distance

Length of shortest path

Breadth-first search



[EK10] Figure 2.8: Breadth-first search discovers distances to nodes one “layer” at a time; each layer is built of nodes that have an edge to at least one node in the previous layer.

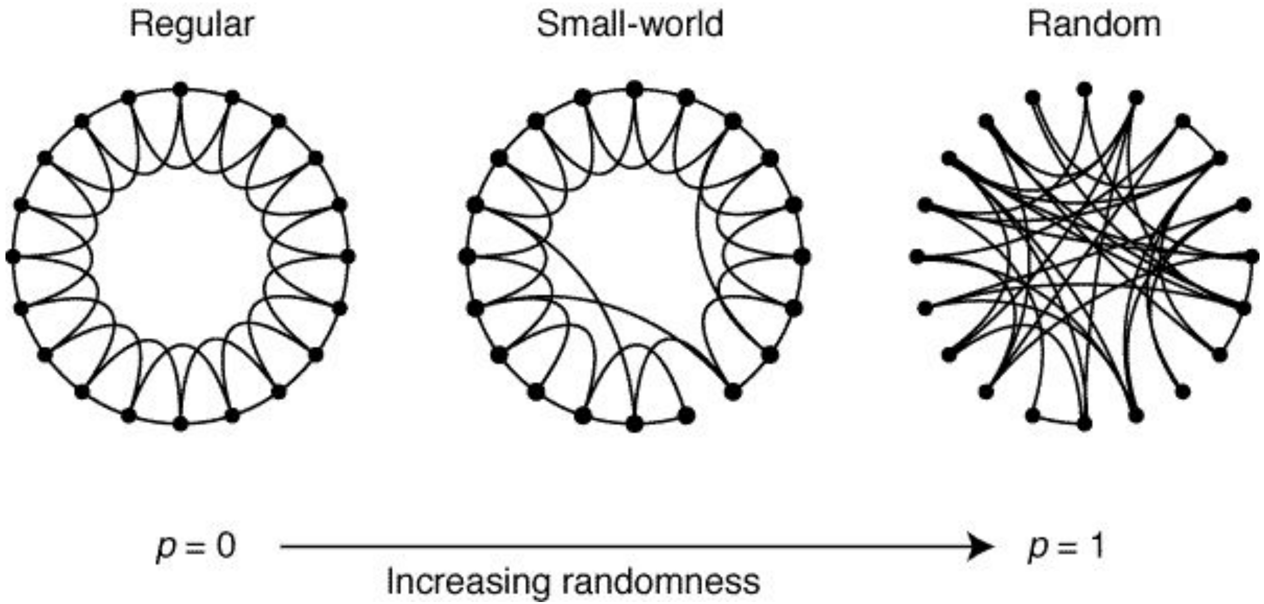
Small World Network

- Six degrees of separation
  - first tested by Stanley Milgram in 1960s who found median length of 6 between two individuals
- OSNs make studying this phenomena a little easier
- Six degrees of Kevin Bacon
- Erdos number

[Six Degrees of Kevin Bacon](#) (Wikipedia) / [The Oracle of Bacon](#)
[Erdős number](#) (Wikipedia)

[My Erdos number is 3](#)

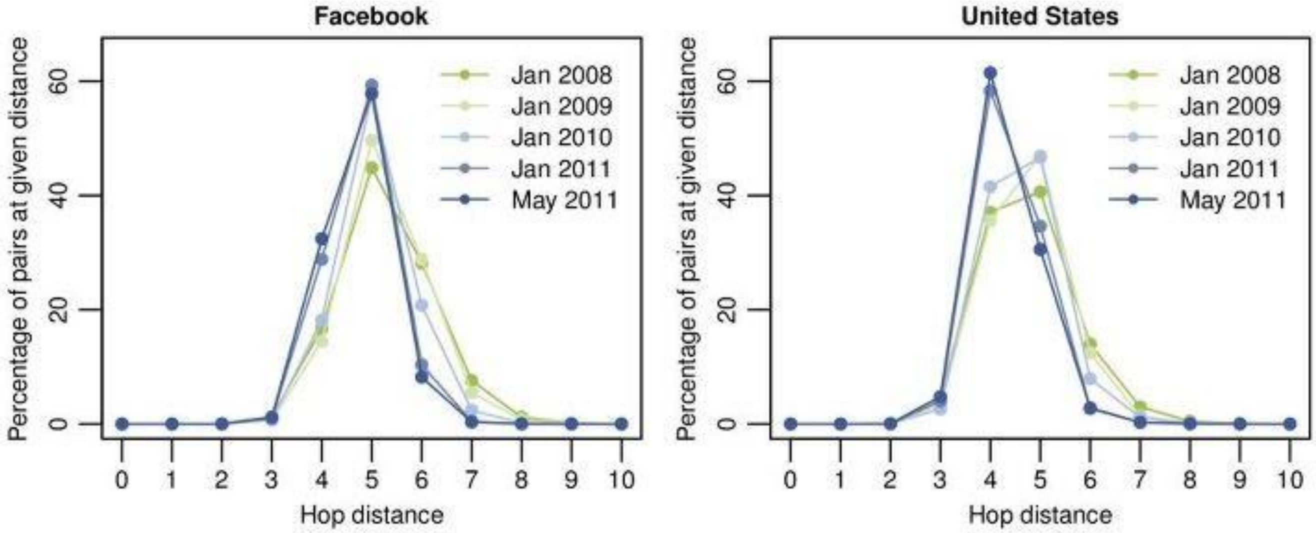
Regular → Small World → Random



Watts & Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, 1998 (pdf)

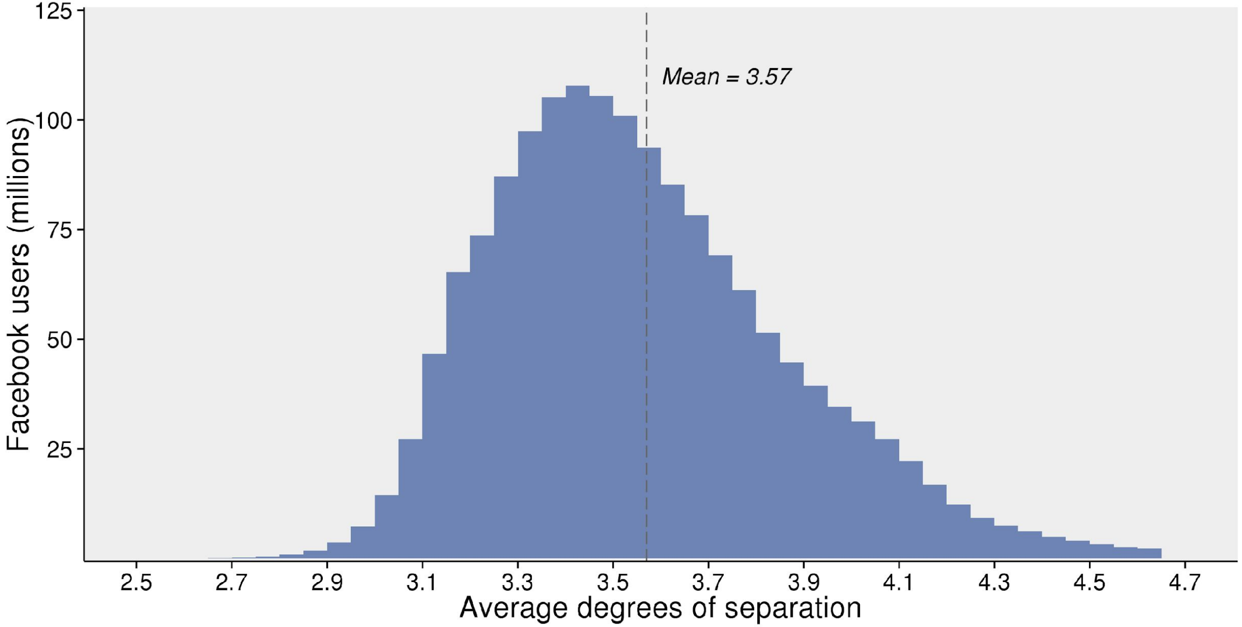
Four Degrees of Separation

Study of all Facebook users (721 million) in 2011 revealed avg distance of 4.74



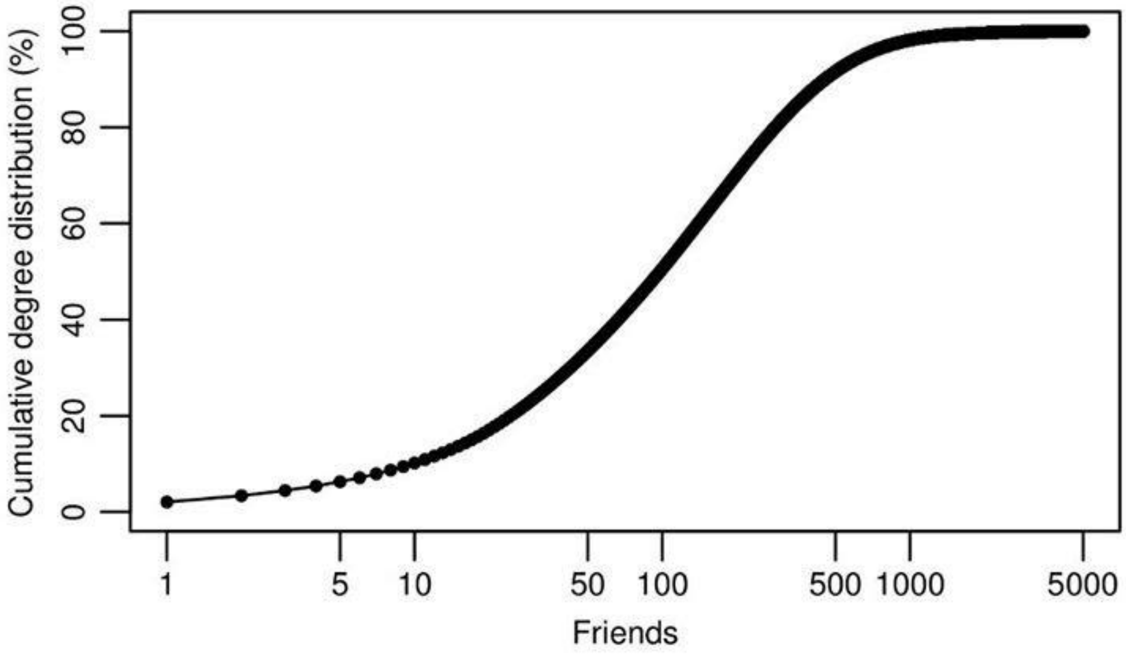
[Anatomy of Facebook]

2016 Update: 3.57 Degrees



[Three and a half degrees of separation]

How Many Friends?



[Anatomy of Facebook]

example of a CDF: [Cumulative distribution function] (Wikipedia)

"Friendship Paradox"

Fun fact: your friends have more friends than you do.

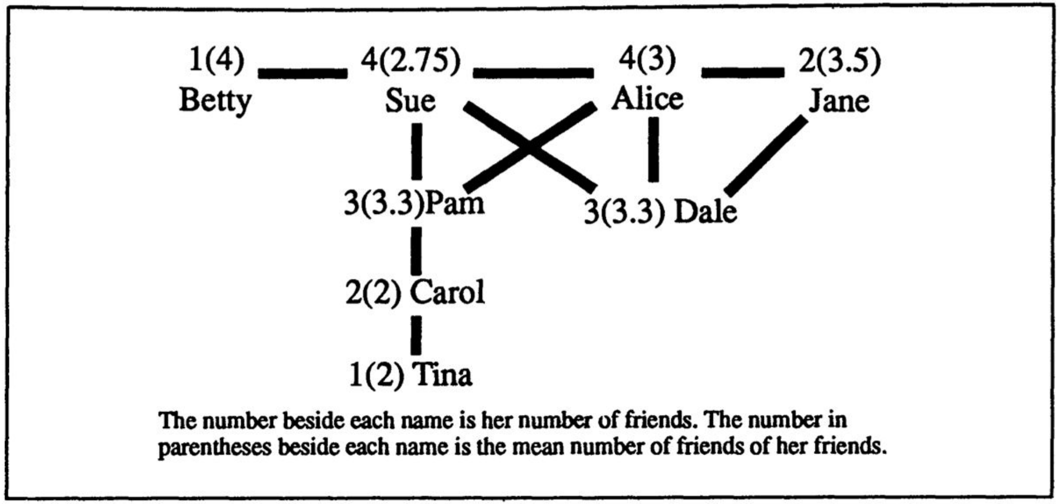


FIG. 1.—Friendships among eight girls at Marketville High School

Feld, "Why Your Friends Have More Friends Than You Do" (pdf)
Friendship paradox (Wikipedia)
Your friends are also happier than you: The happiness paradox: your friends are happier than you

Sue & Alice Rule Marketville

(Carol is doing ok)

TABLE 1

A SUMMARY OF THE NUMBERS OF FRIENDS AND THE MEAN NUMBERS OF FRIENDS OF FRIENDS FOR EACH OF THE GIRLS IN FIGURE 1

	Number of Friends (x_i)	Total Number of Friends of Her Friends (Σx_j)	Mean Number of Friends of Her Friends ($\Sigma x_j/x_i$)
Betty..............	1	4	4
Sue	4	11	2.75
Alice	4	12	3
Jane................	2	7	3.5
Pam................	3	10	3.3
Dale................	3	10	3.3
Carol	2	4	2
Tina	1	2	2
Total...........	20	60	23.92
Mean	2.5*	3†	2.99*

* For eight girls.
† For 20 friends.

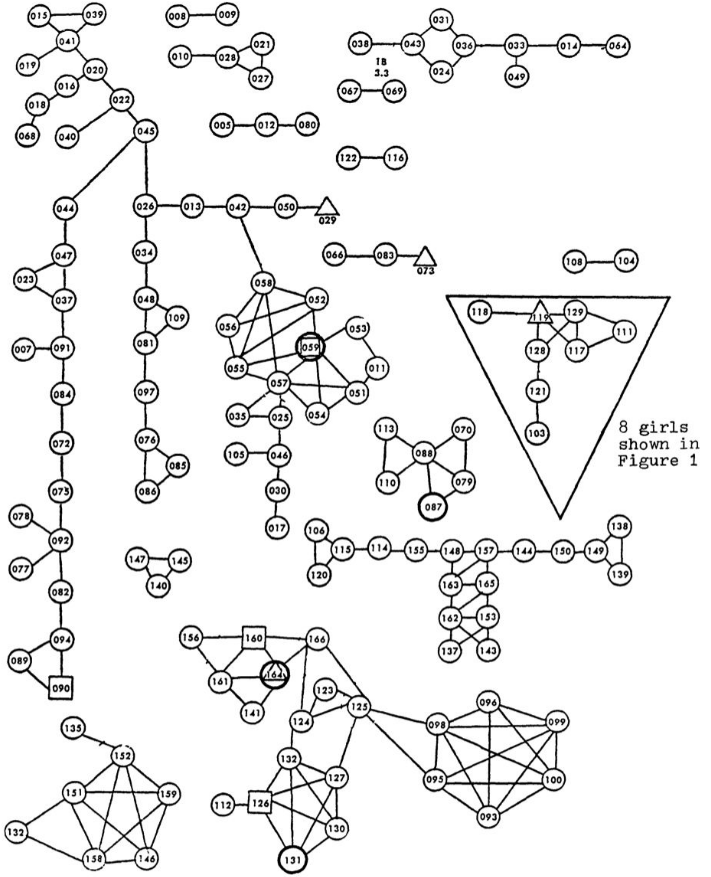
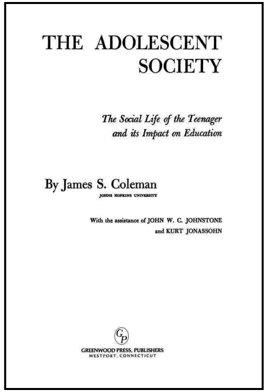


FIG. 2.—Network of reciprocated friendships among Marketville girls; the triangle at right indicates friendships illustrated in fig. 1. (From *The Adolescent Society* by James S. Coleman. © 1961 by the Free Press, a division of Macmillan, Inc. Used with permission.)

Complete Network of Marketville

In 1961 study, friendship is via bi-directional naming in a questionnaire (i.e., Facebook style).



Web Science: Social Networks

(Part 2 - Strong and Weak Ties)

CS 432/532

Old Dominion University

Most slides based on Ch 3 of *Networks, Crowds and Markets* by Easley & Kleinberg (2010)

(abbreviated as EK10)

Kleinberg is the HITS guy (hubs and authorities)

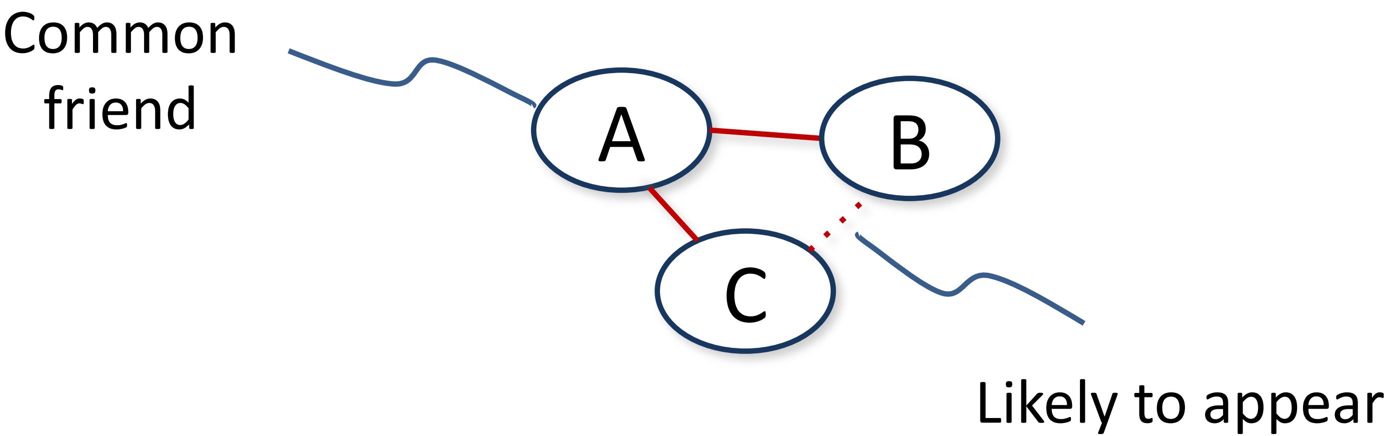
Network Evolution

- Social networks tend to change over time
- You make new friends, lose contact with others
- You change schools, jobs, etc.
- We're often interested in examining multiple snapshots of the social graph through time

"If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future."

\- Anatole Rapoport (1953)

Triadic Closure



Reasons for Triadic Closure

- *Opportunity* – A spending time with B & C increases chance they will meet
- *Trust* – Research shows that B & C are more likely to trust each other if they are aware of mutual friend A
- *Incentive* – Research shows that latent stress occurs when mutual friends are not friends with each other

Measuring Triadic Closure

- *Clustering coefficient* – probability that any two randomly selected friends of a node are also friends

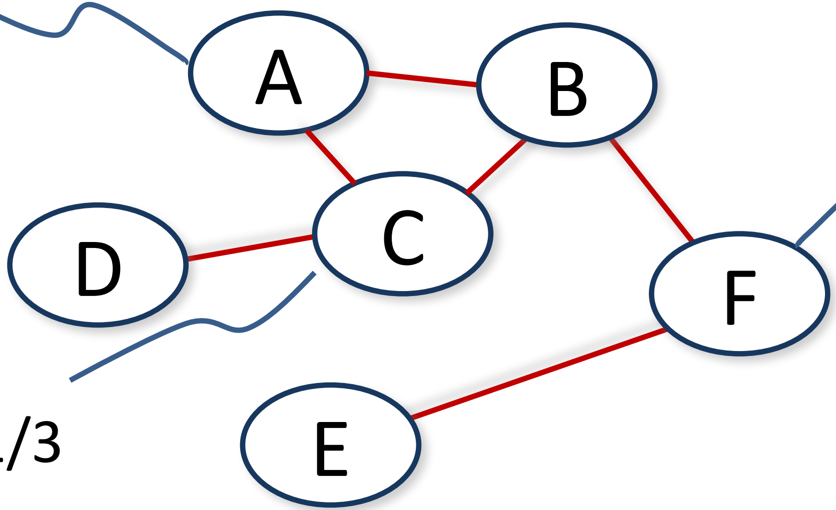
$$\textit{CC} \text{ for a node} = \frac{\text{\# of friend pairs}}{\text{\# of possible friend pairs}}$$

- *CC* = 0 when **none** of node's friends are friends with each other
- *CC* = 1 when **all** of node's friends are friends with each other

Clustering Coefficient

A's CC:
BC / BC = 1

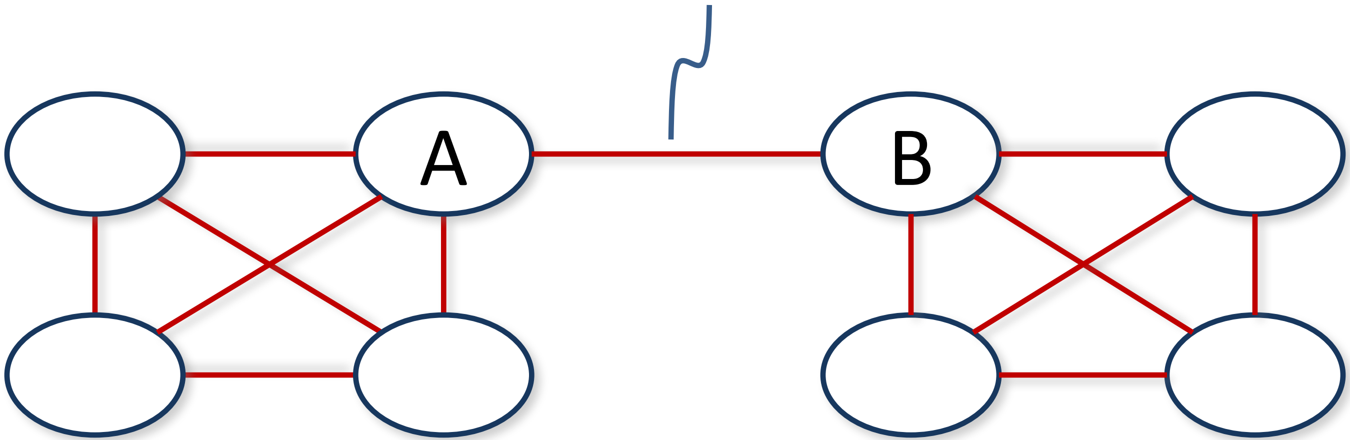
F's CC:
- / BE = 0



C's CC:
AB / AB, AD, BD = 1/3

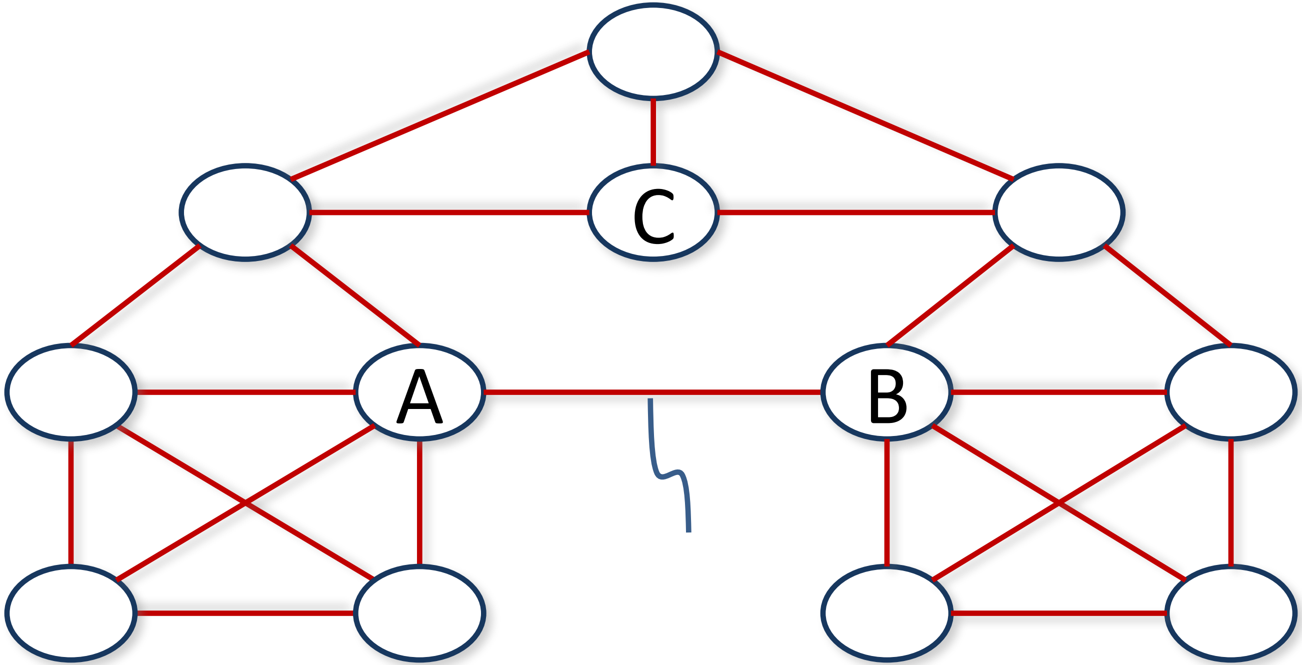
CC for a node = # of friend pairs / # of possible friend pairs

Bridges



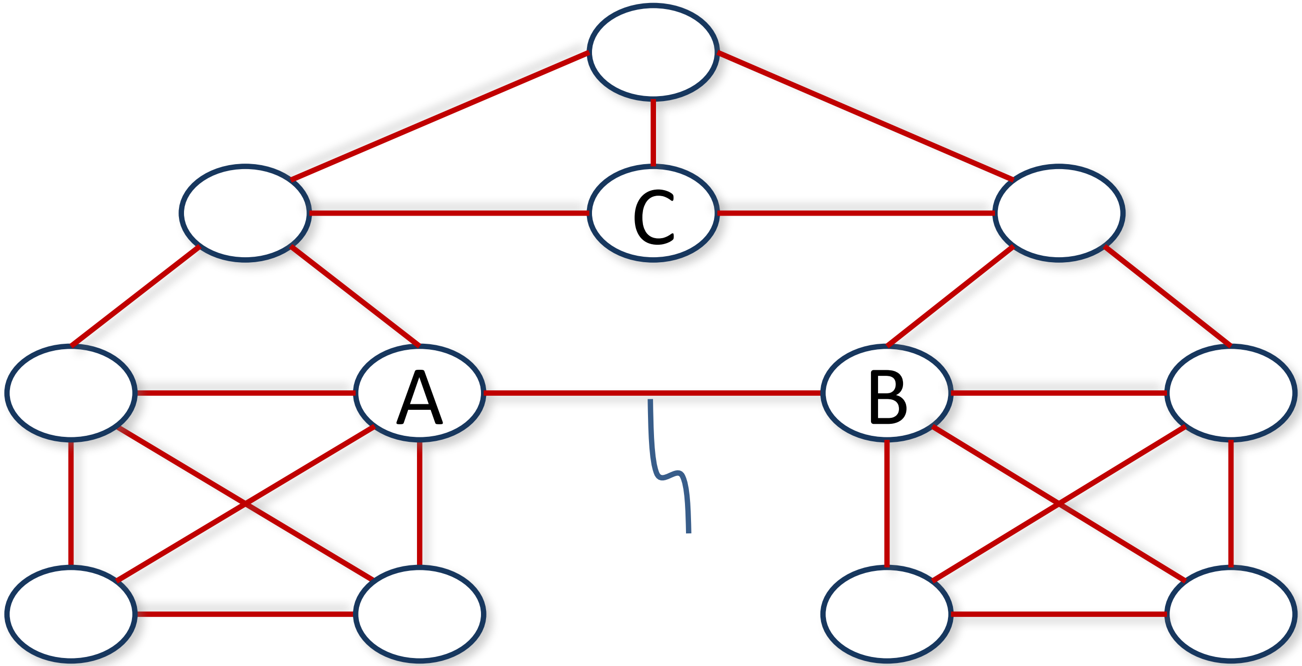
Edge is a **bridge** if removing the edge creates two components

Local Bridges



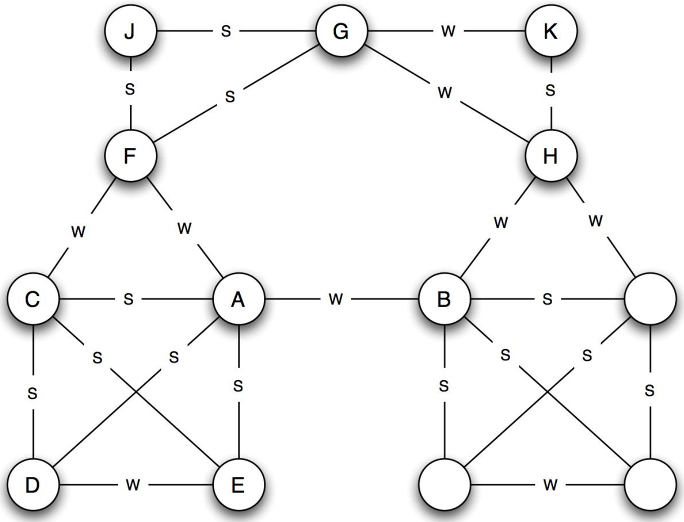
Edge is a **local bridge** if endpoints have no friends in common

Local Bridges



Removing a local bridge does **not** create two components!

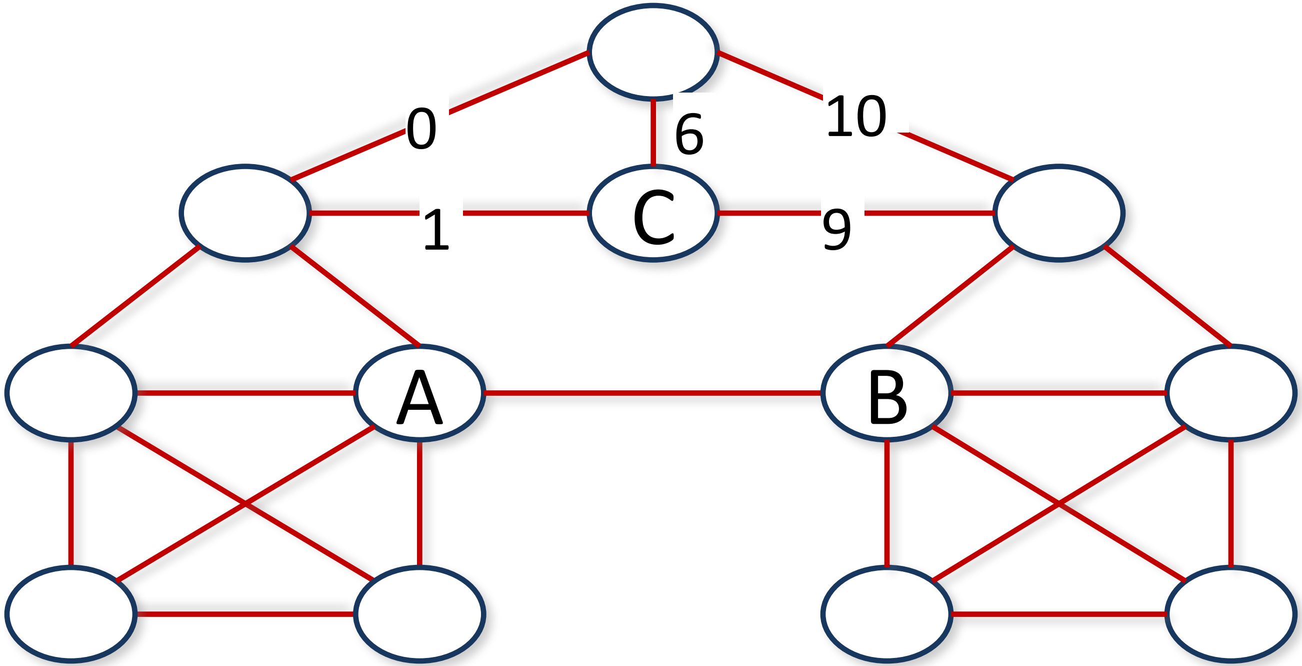
Strong and Weak Ties



[EK10] Figure 3.5: Each edge of the social network from Figure 3.4 is labeled here as either a *strong tie* (S) or a *weak tie* (W), to indicate the strength of the relationship. The labeling in the figure satisfies the Strong Triadic Closure Property at each node: if the node has strong ties to two neighbors, then these neighbors must have at least a weak tie between them.

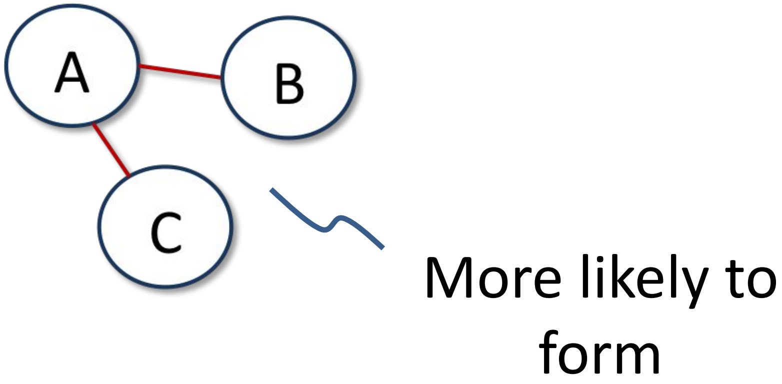
Not everyone is your BFF!

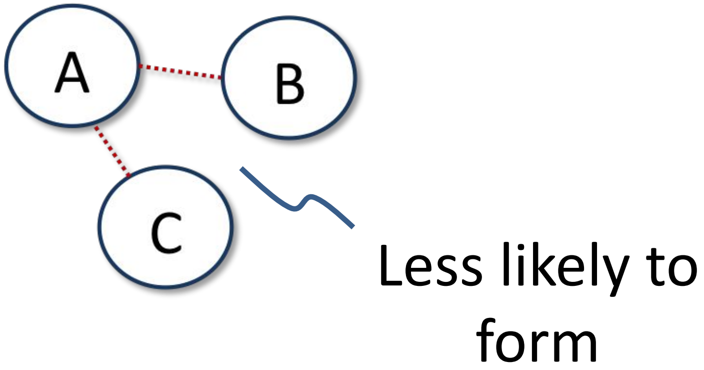
Variable Tie Strength



Can use more fine-grained methods to determine tie strength, like number of Facebook messages exchanged

Tie Strength and Triadic Closure



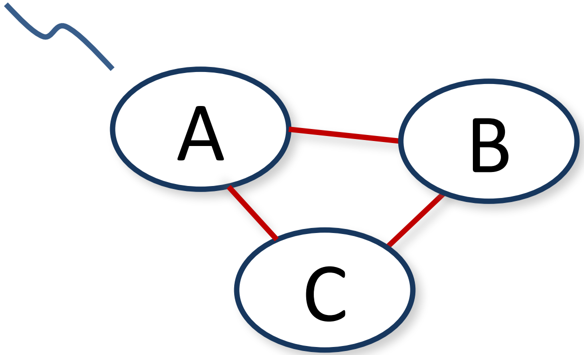


Triadic closure is more likely to form when initial edges are **strong**

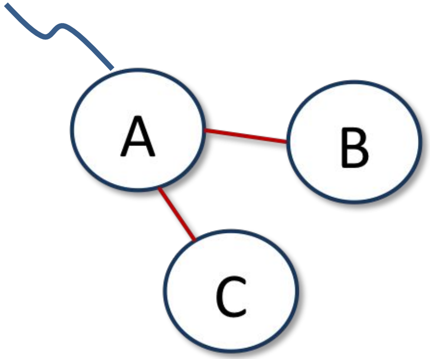
Strong Triadic Closure Property

Node A satisfies the **Strong Triadic Closure Property** if it has strong ties to two other nodes B and C, and there is some edge between B and C

satisfies

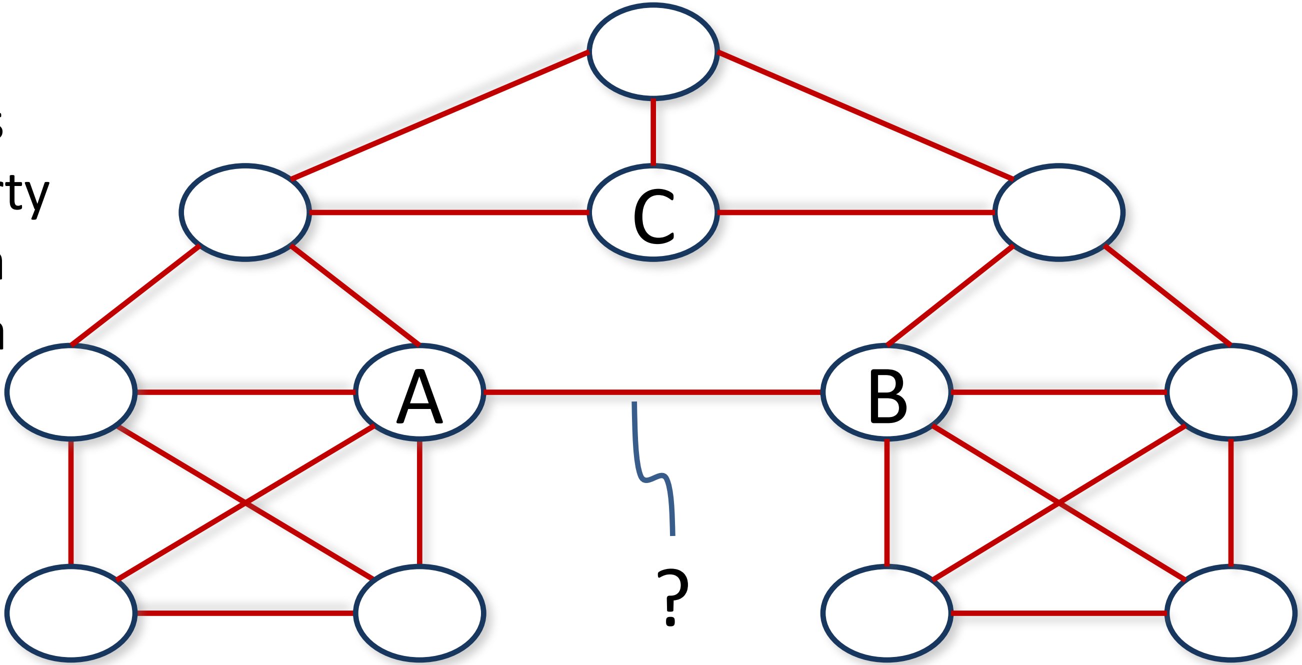


violates



Strong Triadic Closure and Bridges

Can a node with at least 2 strong edges satisfy the STC property and be involved in a local bridge that is a **strong tie**?



Answer: **No!** Why not?

Strong Triadic Closure and Bridges

Can a node with at least 2 strong edges satisfy the STC property and be involved in a local bridge that is a **strong tie**?

[EK10]

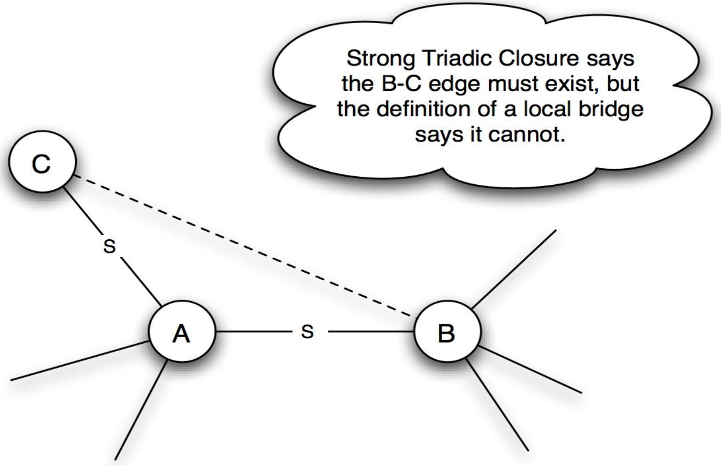


Figure 3.6: If a node satifies Strong Triadic Closure and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie. The figure illustrates the reason why: if the A-B edge is a strong tie, then there must also be an edge between B and C, meaning that the A-B edge cannot be a local bridge.

Answer: **No!** Why not?

Strong Triadic Closure and Bridges

- STC property says edge must form when two strong edges emanate from the same node
- Therefore a network that satisfies the STC property and has a sufficient number of strong ties will likely have *only weak* local bridges
- Does not always hold in real social networks, but useful theoretical framework in which to understand connections and test theories

Web Science: Social Networks

(Part 3 - Tie Strength and Embeddedness)

CS 432/532

Old Dominion University

Most slides based on Ch 3 of *Networks, Crowds and Markets* by Easley & Kleinberg (2010)

(abbreviated as EK10)

Kleinberg is the HITS guy (hubs and authorities)

Large-Scale Data Studies

- Put new concepts to work by examining several large-scale studies
- Examples:
  - Cell phone usage
  - Facebook
  - Twitter

Cell Phone Social Network

- 18 weeks of cell phone data covering 20% of country's population
- Node = cell phone user
- Edge = if two nodes called each other (both directions)
- Tie strength = duration of calls
  - rather than just “strong” or “weak”

Onnela et al., Structure and tie strengths in mobile communication networks (2007)

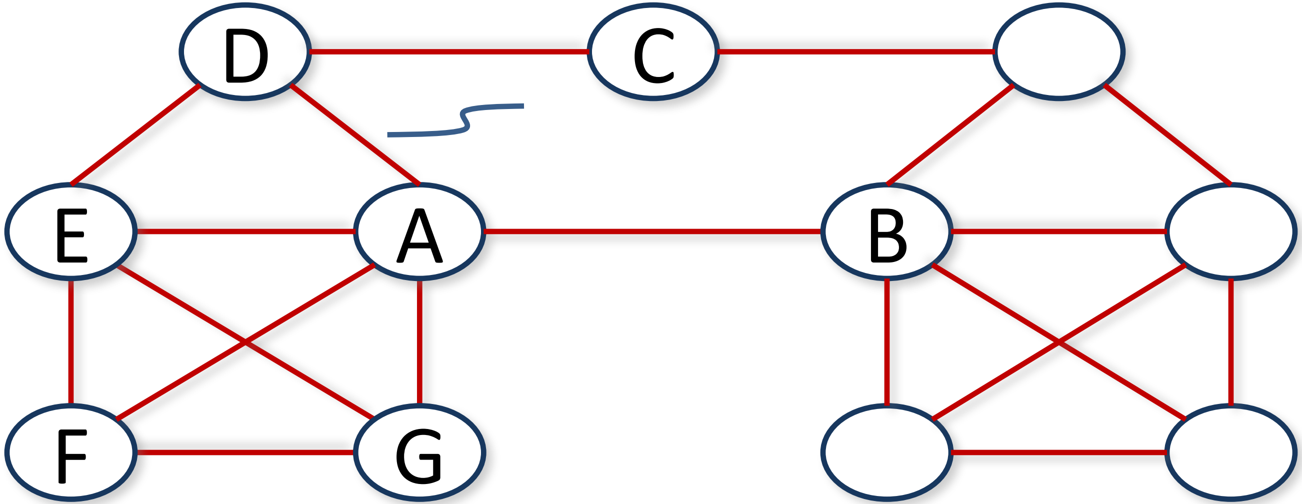
Cell Phone Social Network

- Instead of determining if edge is a local bridge or not, use neighborhood overlap

$$\text{Neighborhood overlap} = \frac{\text{neighbors of both A } \textit{and} \text{ B}}{\text{neighbors of A } \textit{or} \text{ B}}$$

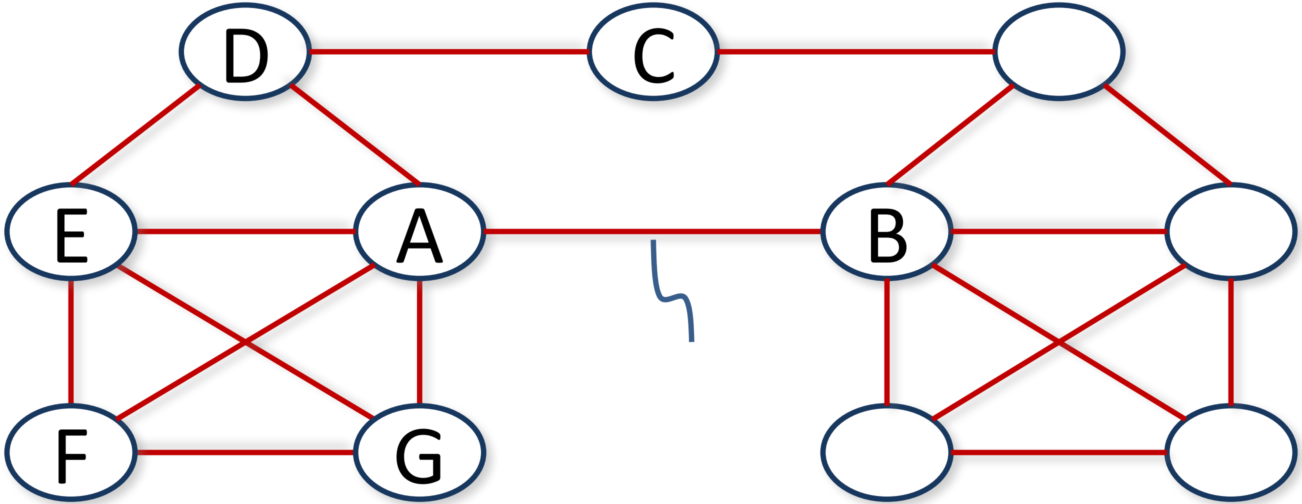
- = 1 if all neighbors of A and B are connected
- = 0 if no neighbors are connected (local bridge)
- Close to 0 is "almost" a local bridge

Neighborhood Overlap



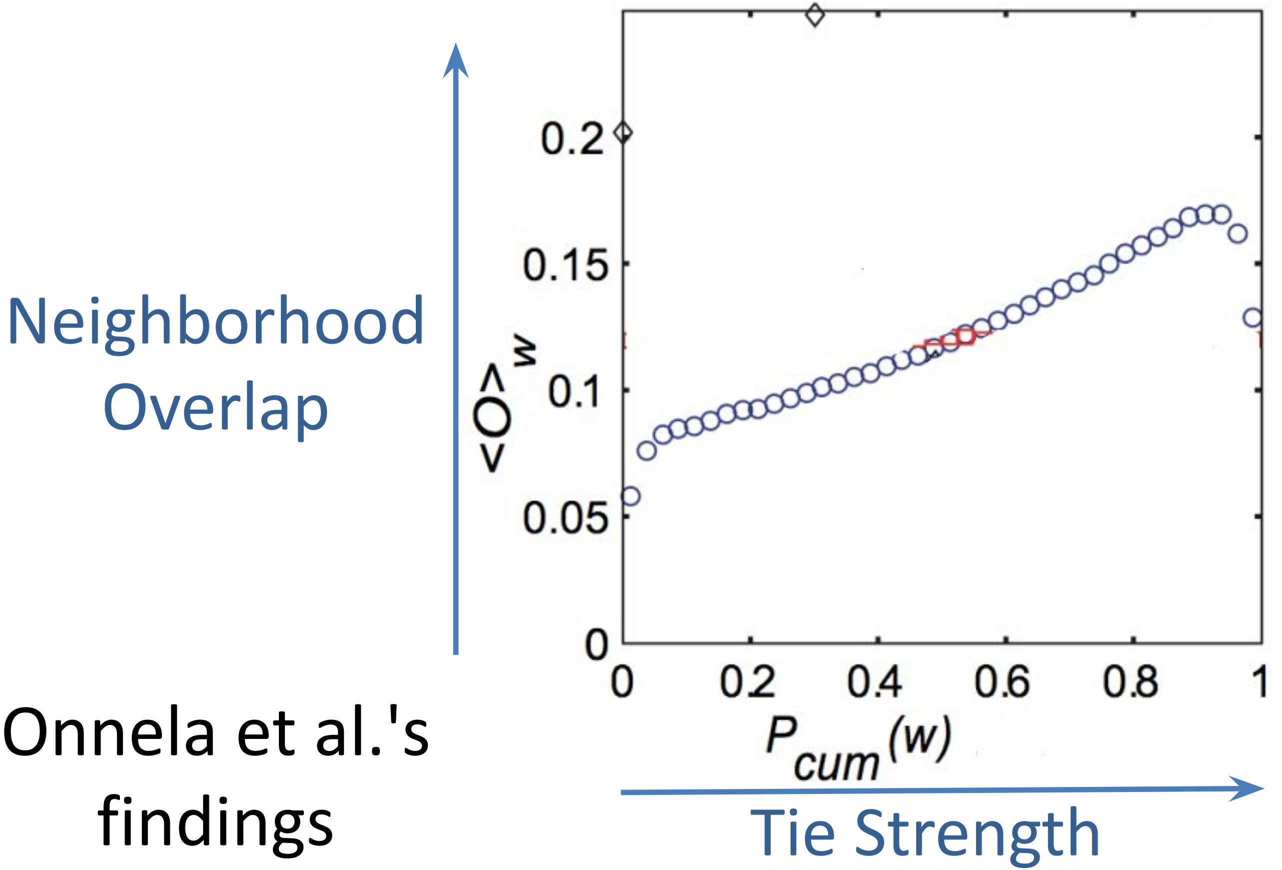
Neighborhood overlap of AD = $\frac{E}{C, E, F, G, B}$ = 1/5

Neighborhood Overlap



Neighborhood overlap of AB = $\dfrac{-}{\text{D, E, F, G, etc.}}$ = 0

Onnela et al.'s findings



Q: What's happening here?
A: We're not sure. Unusual phone usage?

[EK10] Fig 3.7. A plot of the neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength.

Removing Ties

- 84% of cell phone graph nodes lay in a single giant component
- Do weak ties link strongly connected components?
  - Removed edges one at a time starting with strongest edges → giant component slowly erodes
  - Removed edges starting with weakest edges → giant component shrank quicker, remnants broke apart quickly once critical weak edges removed

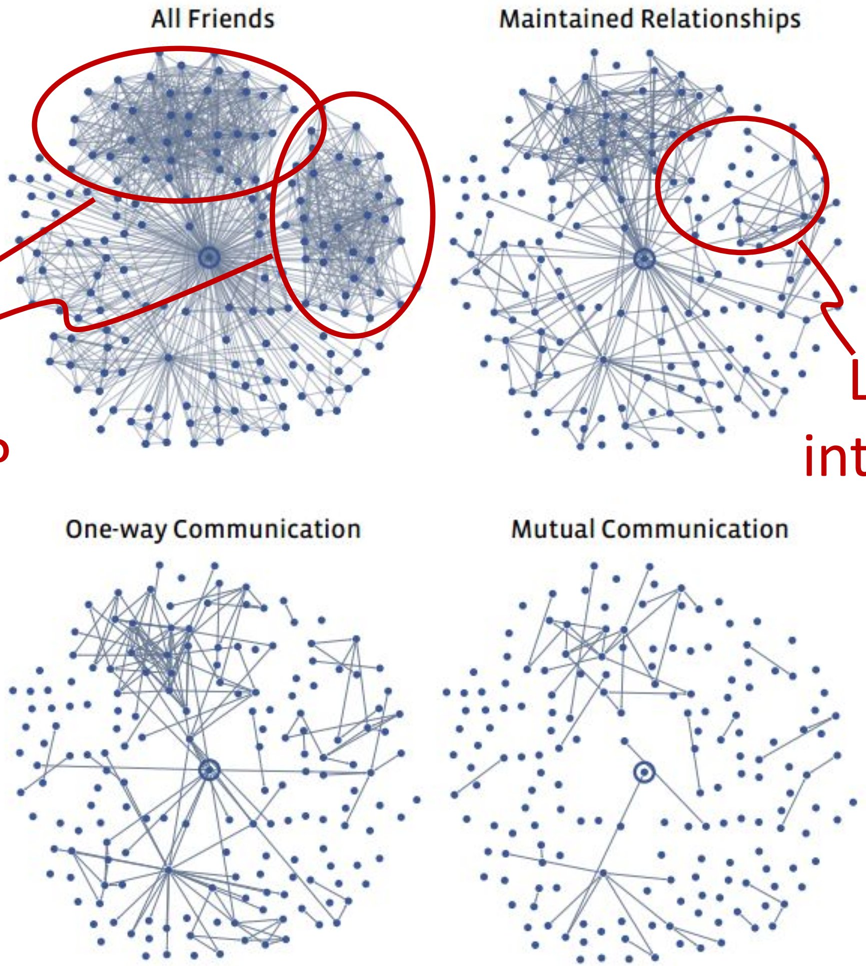
Tie Strength on Facebook

- Marlow et al., *Maintained relationships on Facebook*, 2009
- Analyzed tie strength of Facebook users using one month of interactions

Three Categories of Links

- *Reciprocal/Mutual* – messages sent both ways
- *One-way* – at least one message sent (friend may or may not have reciprocated)
- *Maintained* – user clicked on friend's content (from feed) or profile more than once

Typical User's Network



College & high school friends?

Less interest

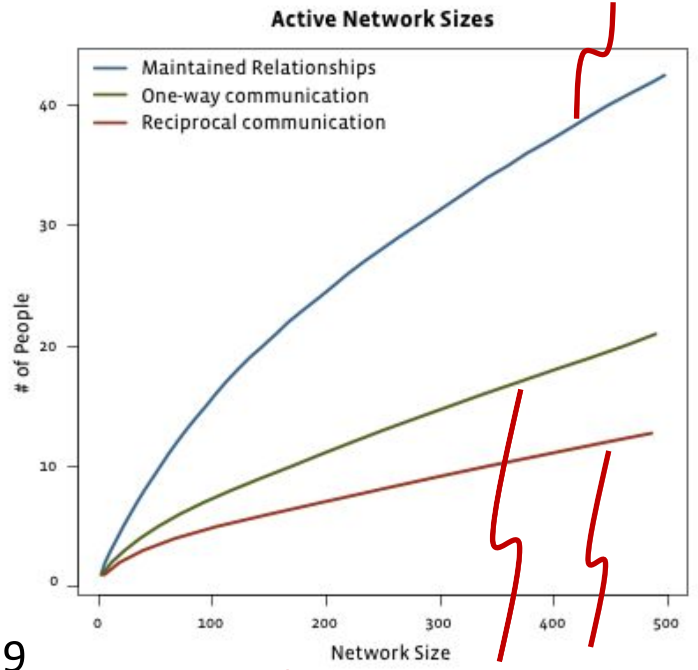
[EK10] Fig 3.8

[Maintained Relationships on Facebook](#)

Limits to Maintaining Relationships

- *Reciprocal* – messages sent both ways
- *One-way* – at least one message sent
- *Maintained* – user clicked on Friend's content (from feed) or profile more than once

passive engagement



[EK10] Fig 3.9

Can't keep up with everyone!

[Maintained Relationships on Facebook]

Tie Strength on Twitter

- Analyzed tie strength of 211,024 active Twitter users during observation period
- **Followees** – people being followed
- **Friend** – person who user has directed at least 2 tweets to (strong ties)

Huberman et al., Social networks that matter: Twitter under the microscope, *First Monday*, Jan 2009

Friends vs. Followees

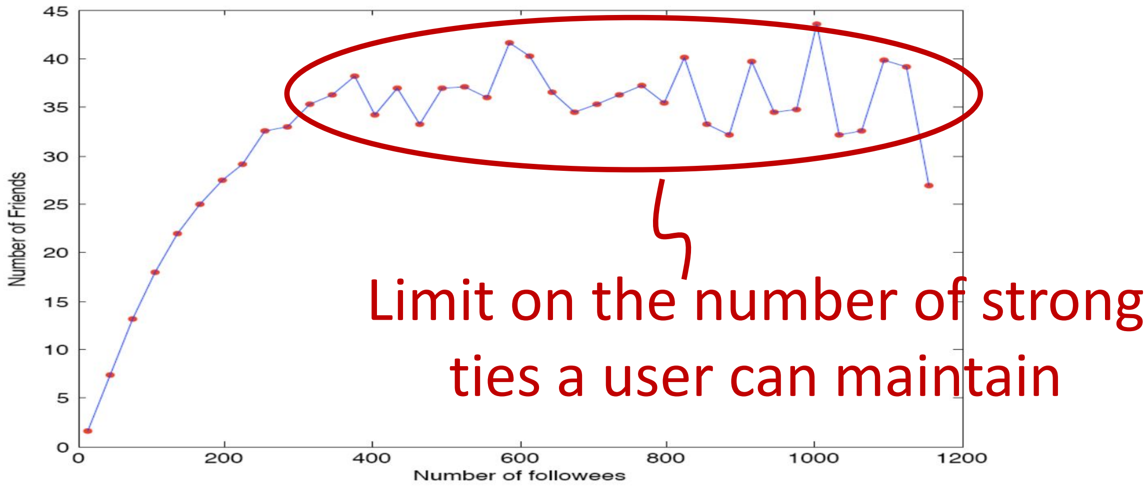


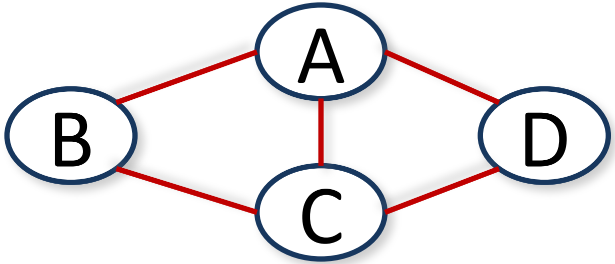
Figure 4: **Number of friends as a function of the number of followees. The total number of friends saturates while the number of followees keeps growing due to the minimal effort required to add a followee.**

Huberman et al., [Social networks that matter: Twitter under the microscope](), *First Monday*, Jan 2009

Now let's look at the edges to see which *nodes* are “important”…

Edge Embeddedness

- Defined as the number of common neighbors shared by two endpoints

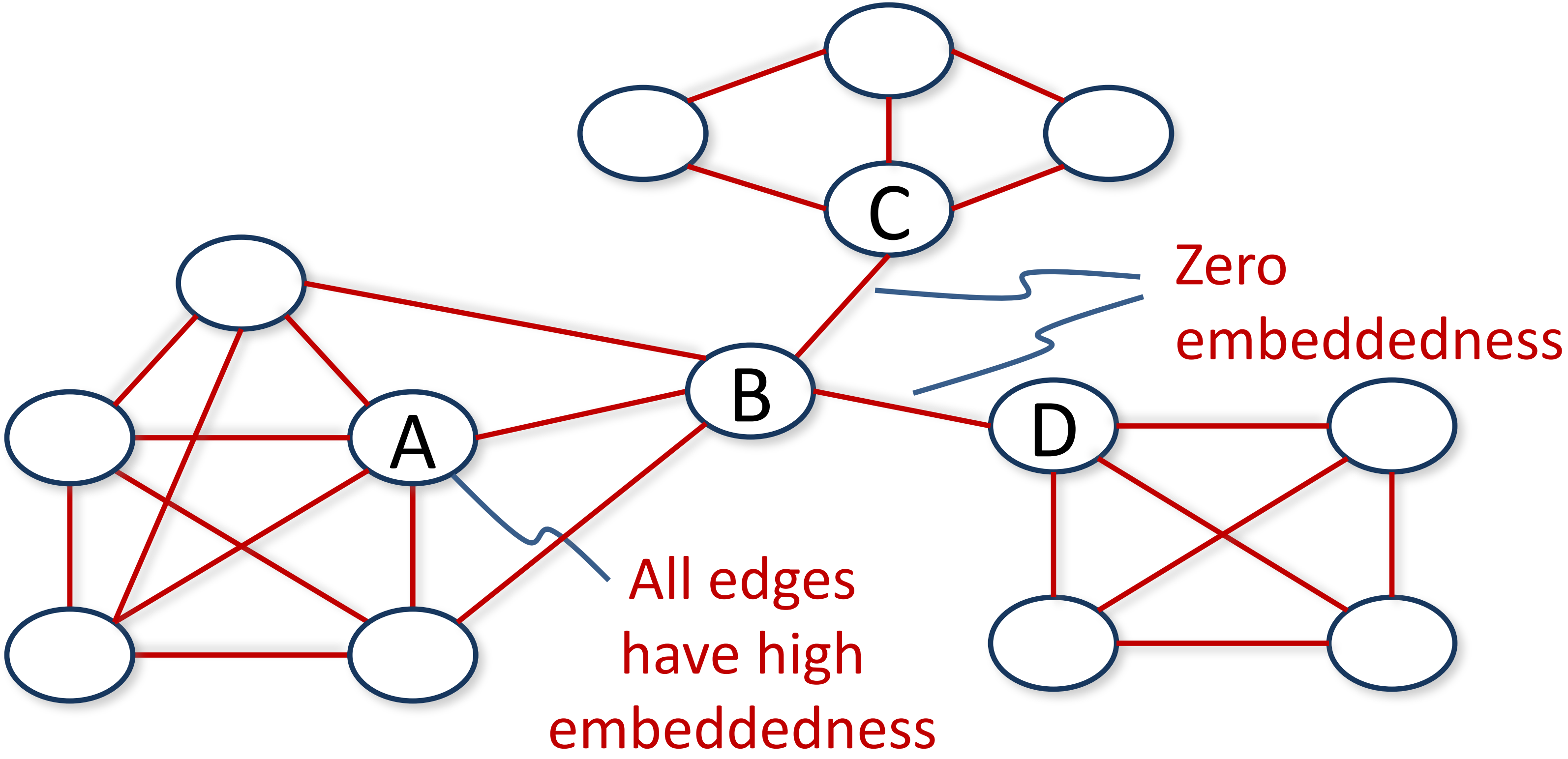


Embeddedness of A-B = 1

Embeddedness of A-C = 2

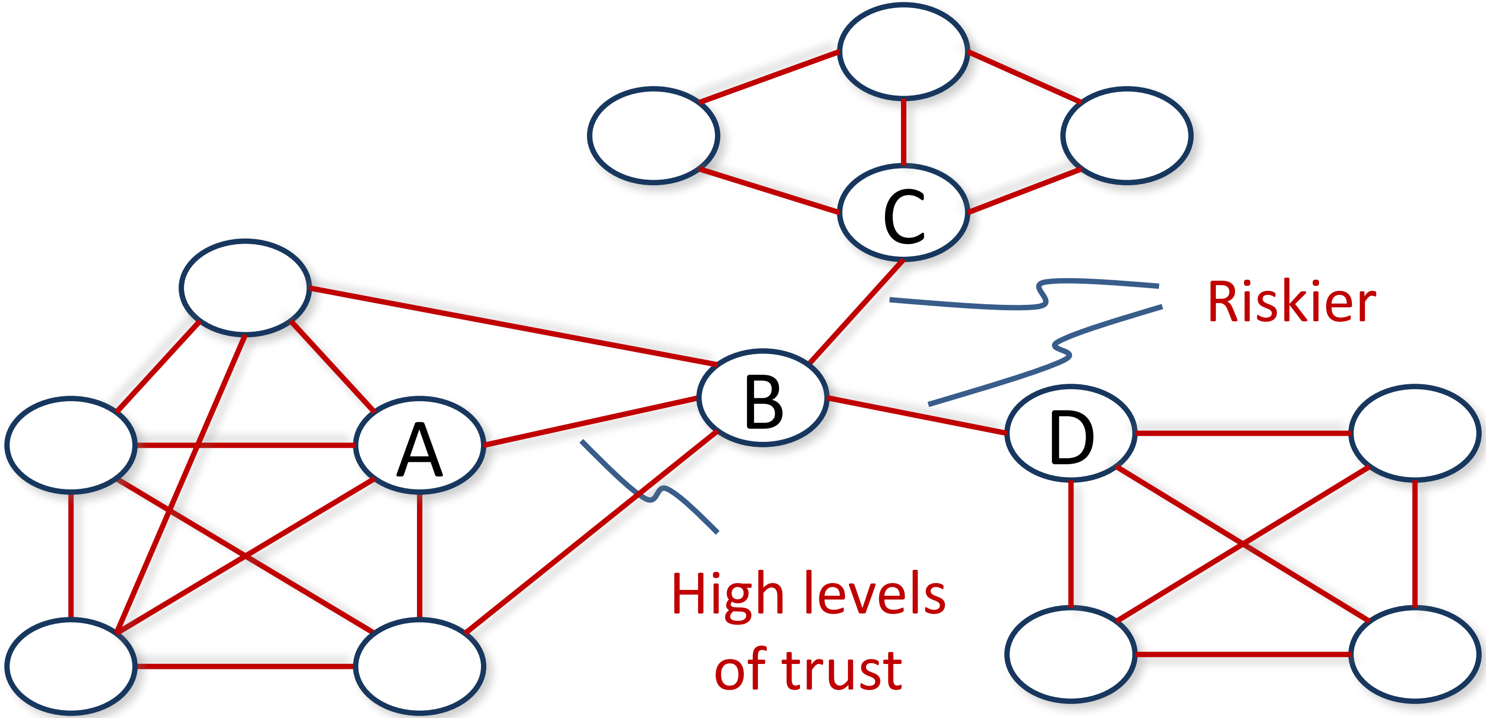
- Equal to numerator in neighborhood overlap
- Local bridges have embeddedness of zero

Edge Embeddedness Example

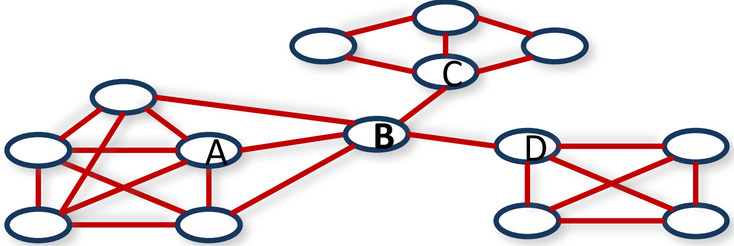


Embeddedness and Trust

- Research shows that if you and I are connected by an embedded edge (we share a mutual friend), there is more trust and confidence in the integrity of the transactions (social, economic, etc.) between us
- If I cheat you, our mutual friends are likely to find out, and they are less likely to trust me in the future!
- Threat is absent for edges of zero embeddedness; no mutual friends to punish me!

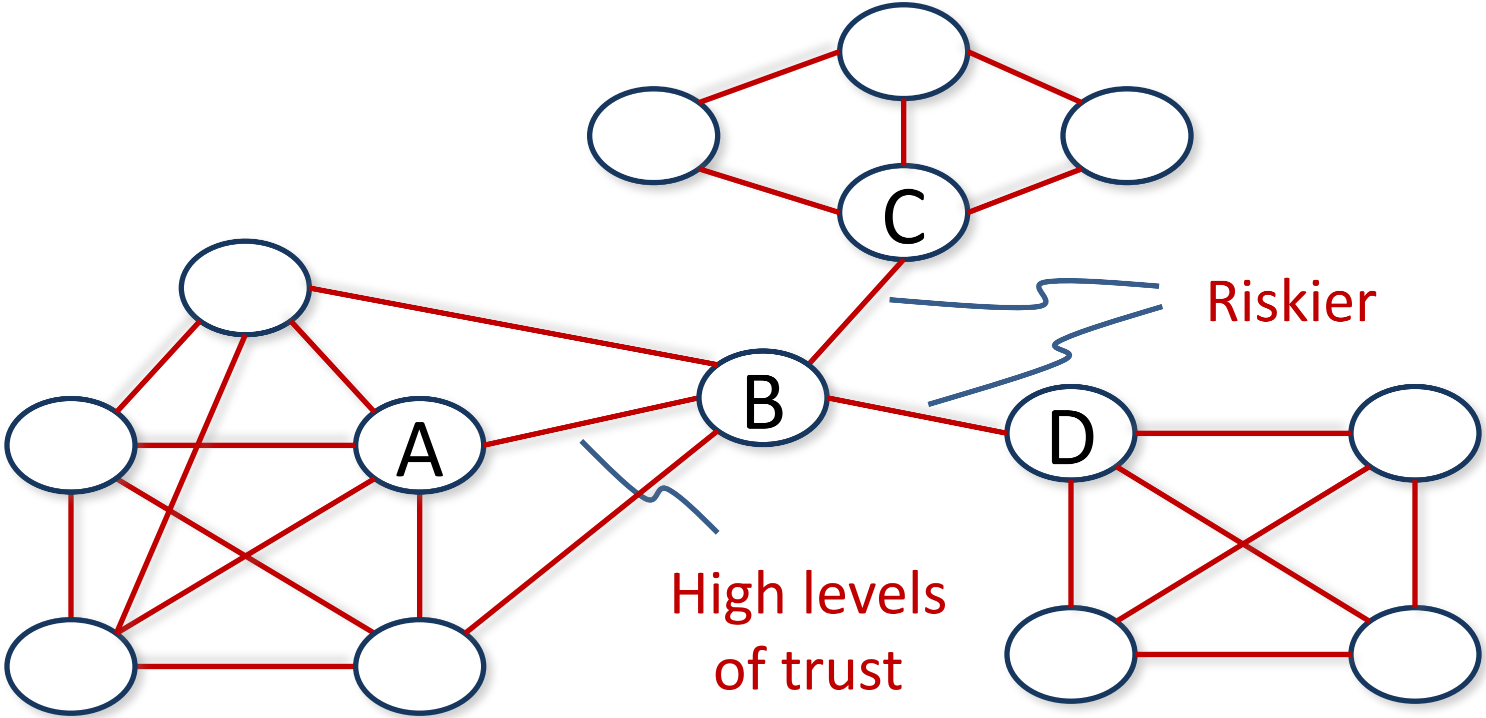


B also must deal with potentially contradictory norms and expectations from various groups



It's good to be B…

- Node B is at the end of multiple local bridges
- Node B spans a **structural hole** in the network – space between two components that do not interact closely
- If network represents a company or organization, some advantages for B:
  - B has early access to info that originates in various components
  - Having access to this info can amplify B's creativity by combining disparate info into novel ideas
  - B acts as "gatekeeper" and can control flow of information to various components
  - B seems “smart” to everybody else…



Who would you rather be? **A** with trusted relationships or risk-taker **B**?

Web Science: Social Networks

(Part 4 - Graph Partitioning)

CS 432/532

Old Dominion University

Most slides based on Ch 3 of *Networks, Crowds and Markets* by Easley & Kleinberg (2010)

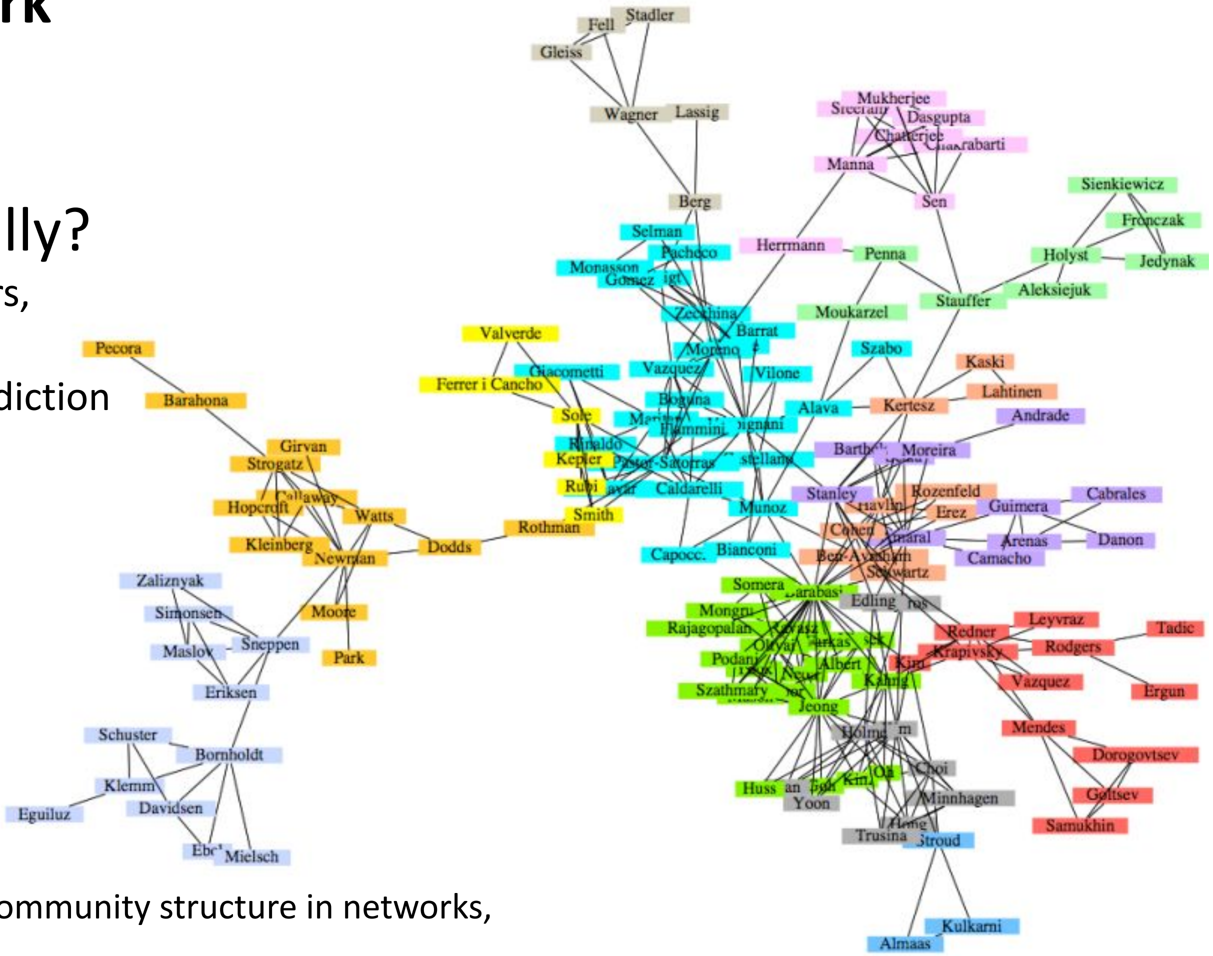
(*abbreviated as EK10)*

Kleinberg is the HITS guy (hubs and authorities)

Co-authorship network

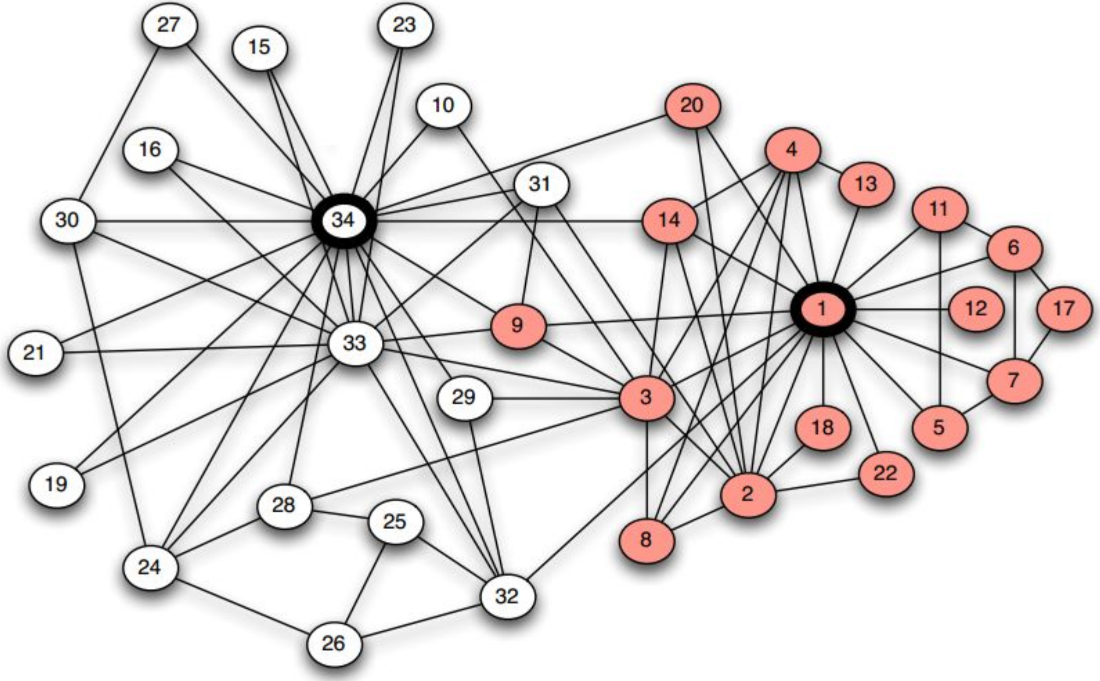
How can the tightly clustered groups be identified mathematically?

1. We can often "eyeball" clusters, important nodes
2. We often are interested in link prediction



Newman & Girvan, Finding and evaluating community structure in networks, *Physical Review E*, 69(2), 2004

Karate Club splits after a dispute. Can new clubs be identified based on network structure?



Zachary W. (1977). An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33, 452-473.

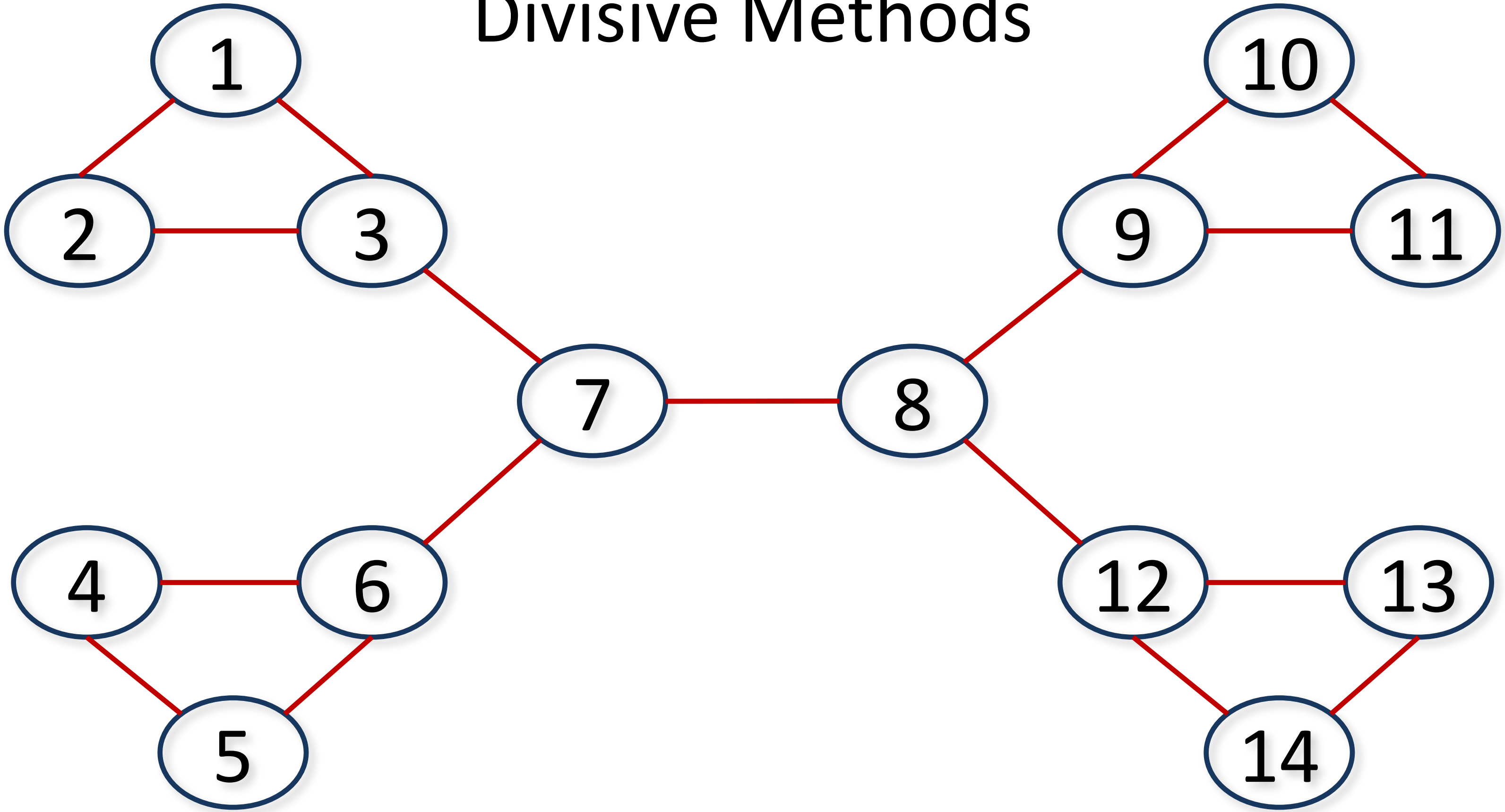
[Zachary's karate club](#) (Wikipedia) / see data set at [ZACHARY KARATE CLUB at UCINET IV Datasets](#) and [Zachary karate club at KONECT (archived)](#)

Graph Partitioning

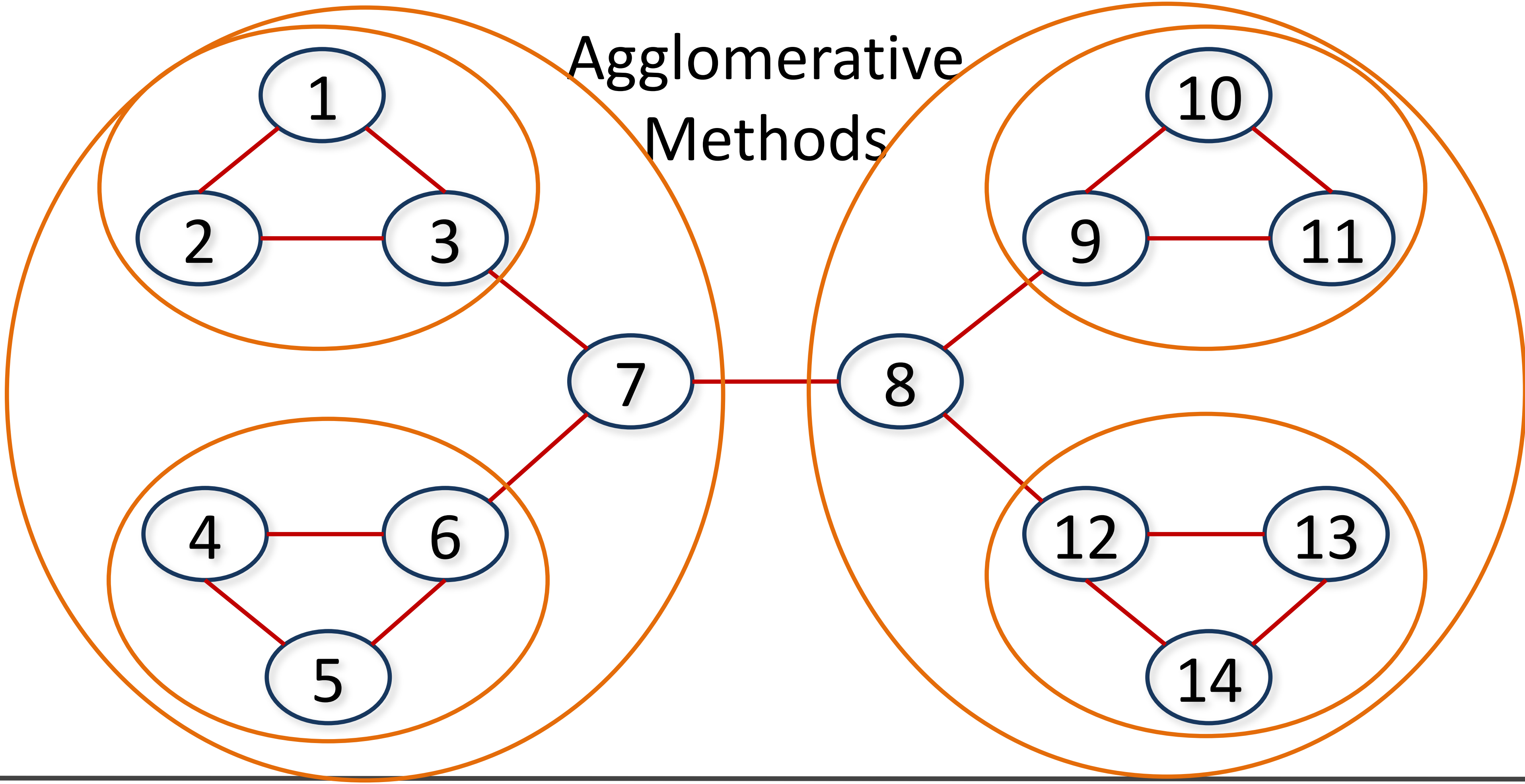
- Methods to break a network into sets of connected components called *regions*

- Many general approaches
  - **Divisive methods**: Repeatedly identify and remove edges connecting densely connected regions
  - **Agglomerative methods**: Repeatedly identify and merge nodes that likely belong in the same region

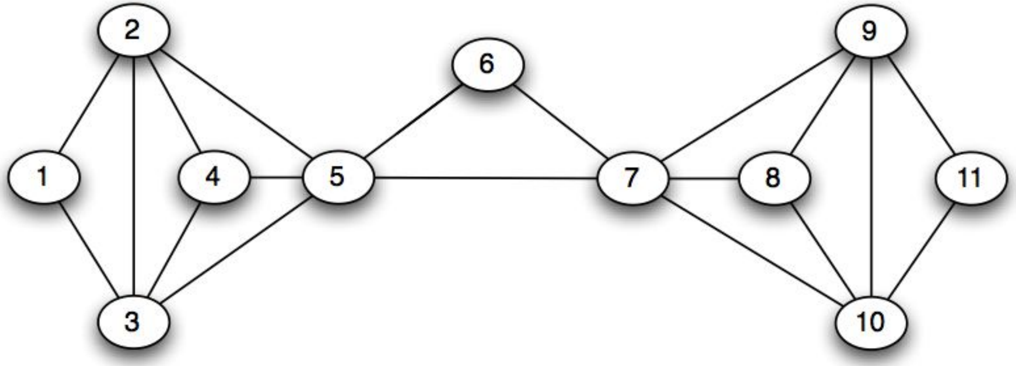
Divisive Methods



Agglomerative Methods



How to split/cluster?



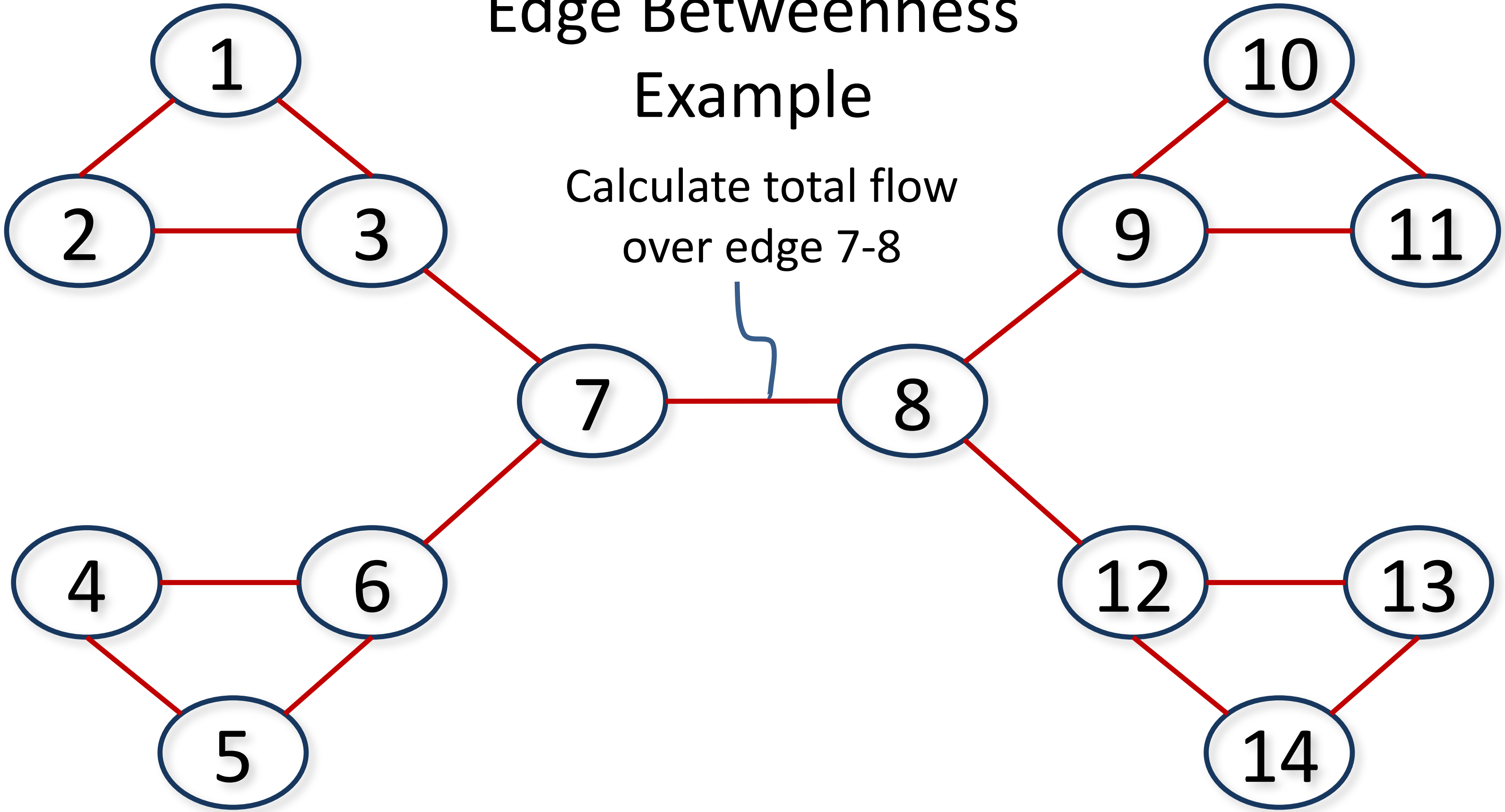
[EK10] Figure 3.15: A network can display tightly-knit regions even when there are no bridges or local bridges along which to separate it.

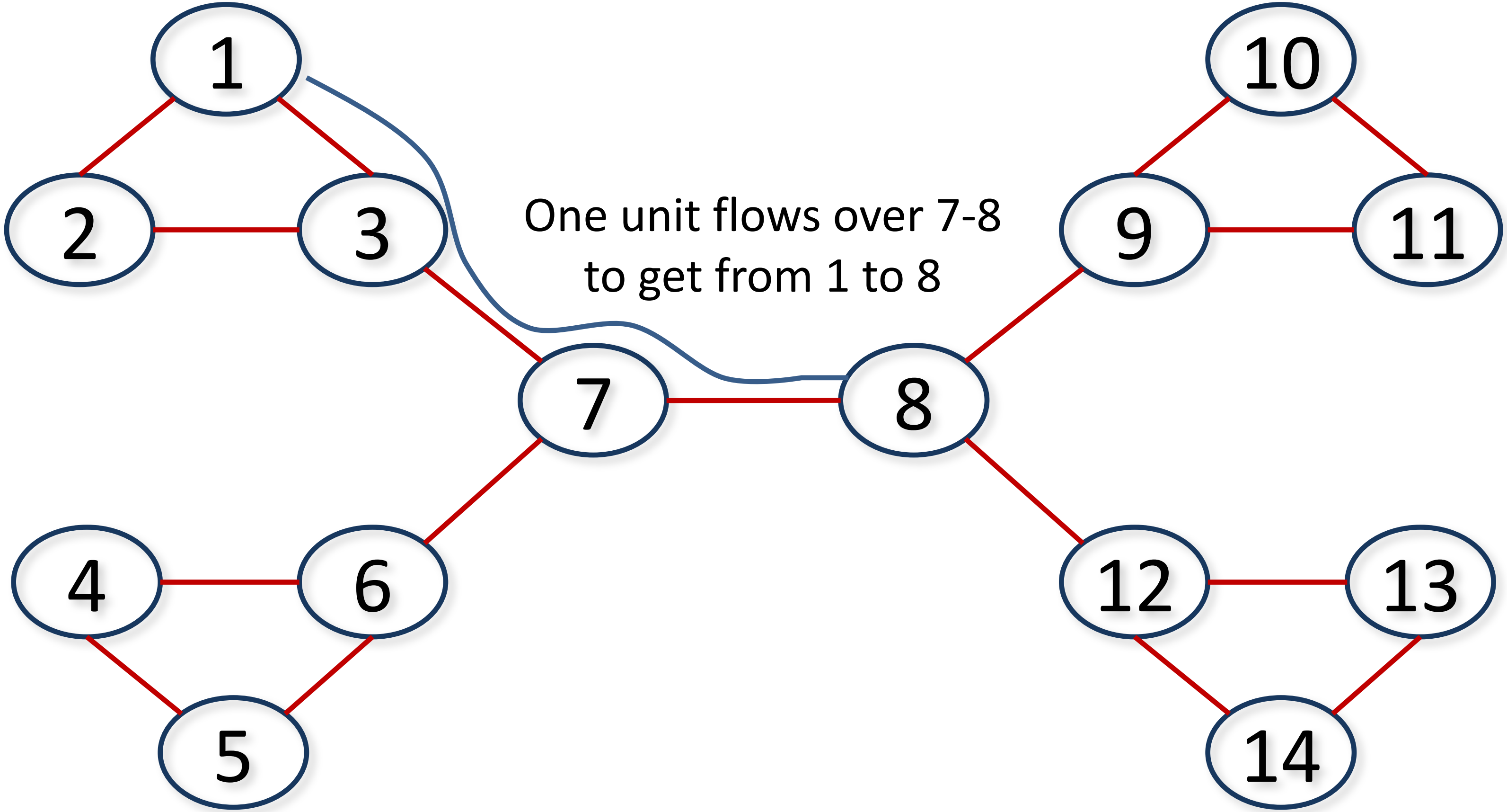
Girvan-Newman Algorithm

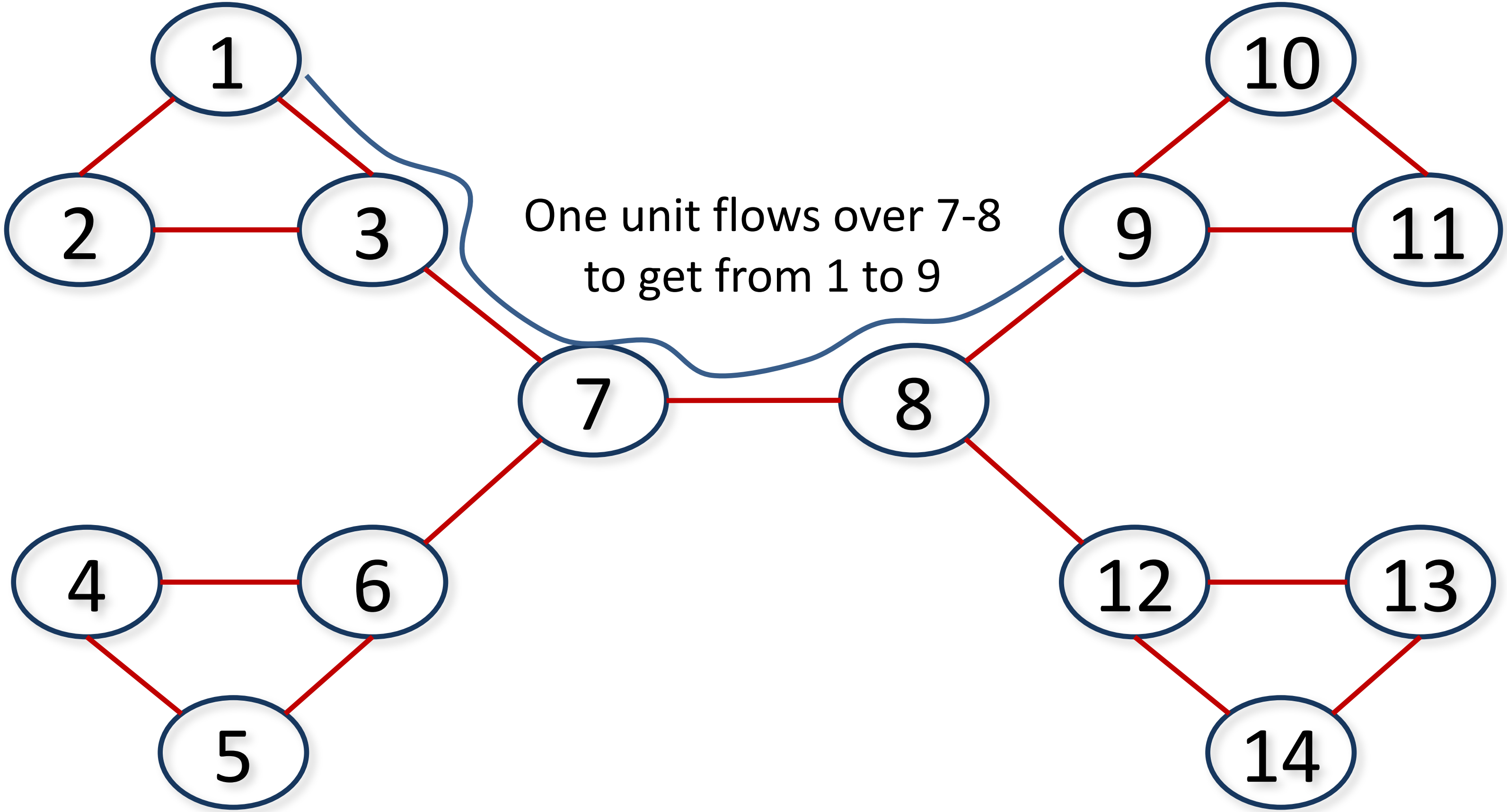
- Divisive method - Uses *edge betweenness* to identify edges to remove
- **Edge betweenness**: Total amount of "flow" an edge carries between all pairs of nodes
  - a single unit of flow between two nodes divides itself evenly among all shortest paths between the nodes (1/*k* units flow along each of *k* shortest paths)
- Betweenness is a measure of *centrality*, or importance, in a graph

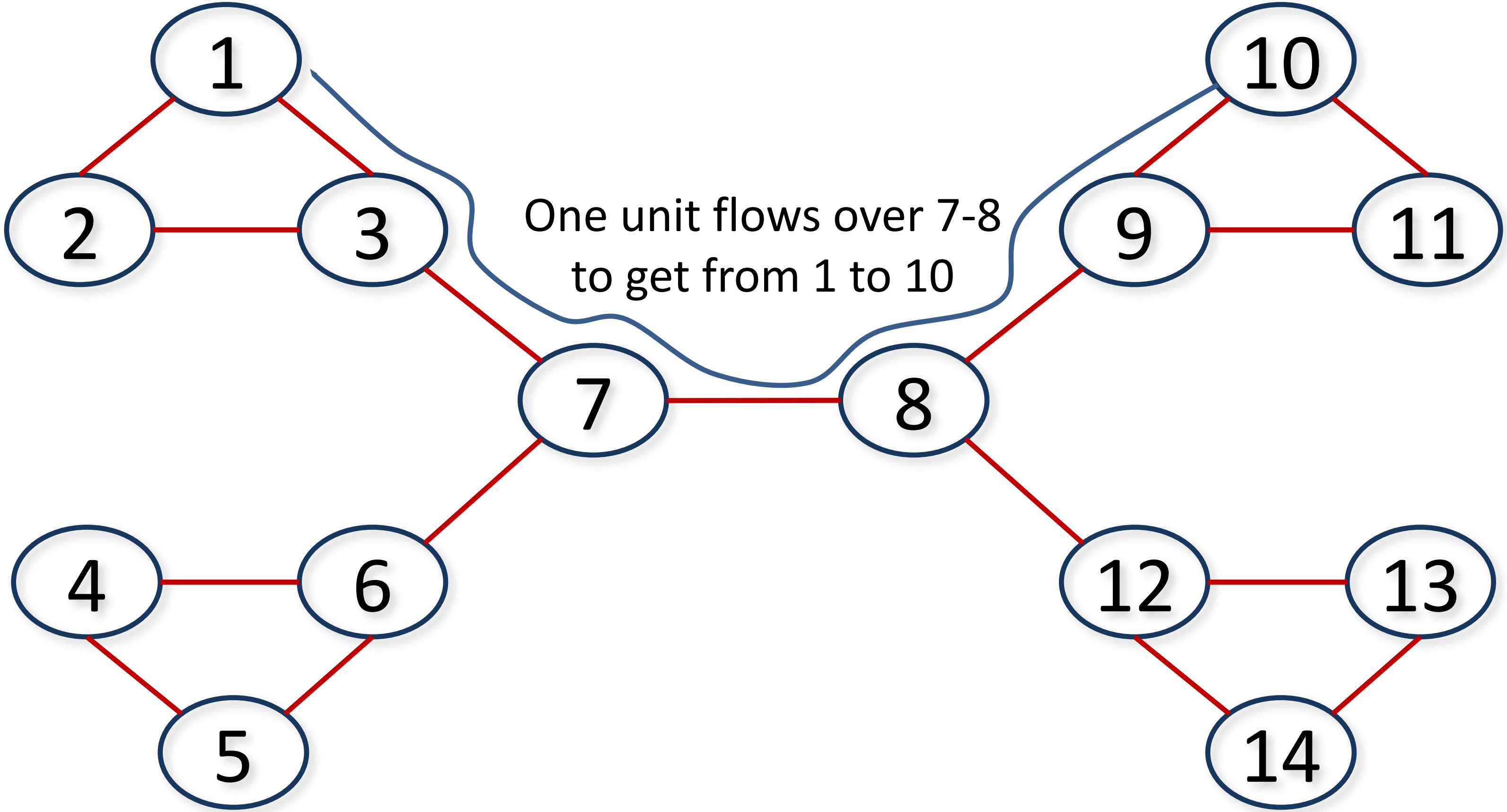
Girvan M. and Newman M. E. J., "Community structure in social and biological networks", Proc. Natl. Acad. Sci. USA **99**, 7821–7826 (2002)
Centrality (Wikipedia)

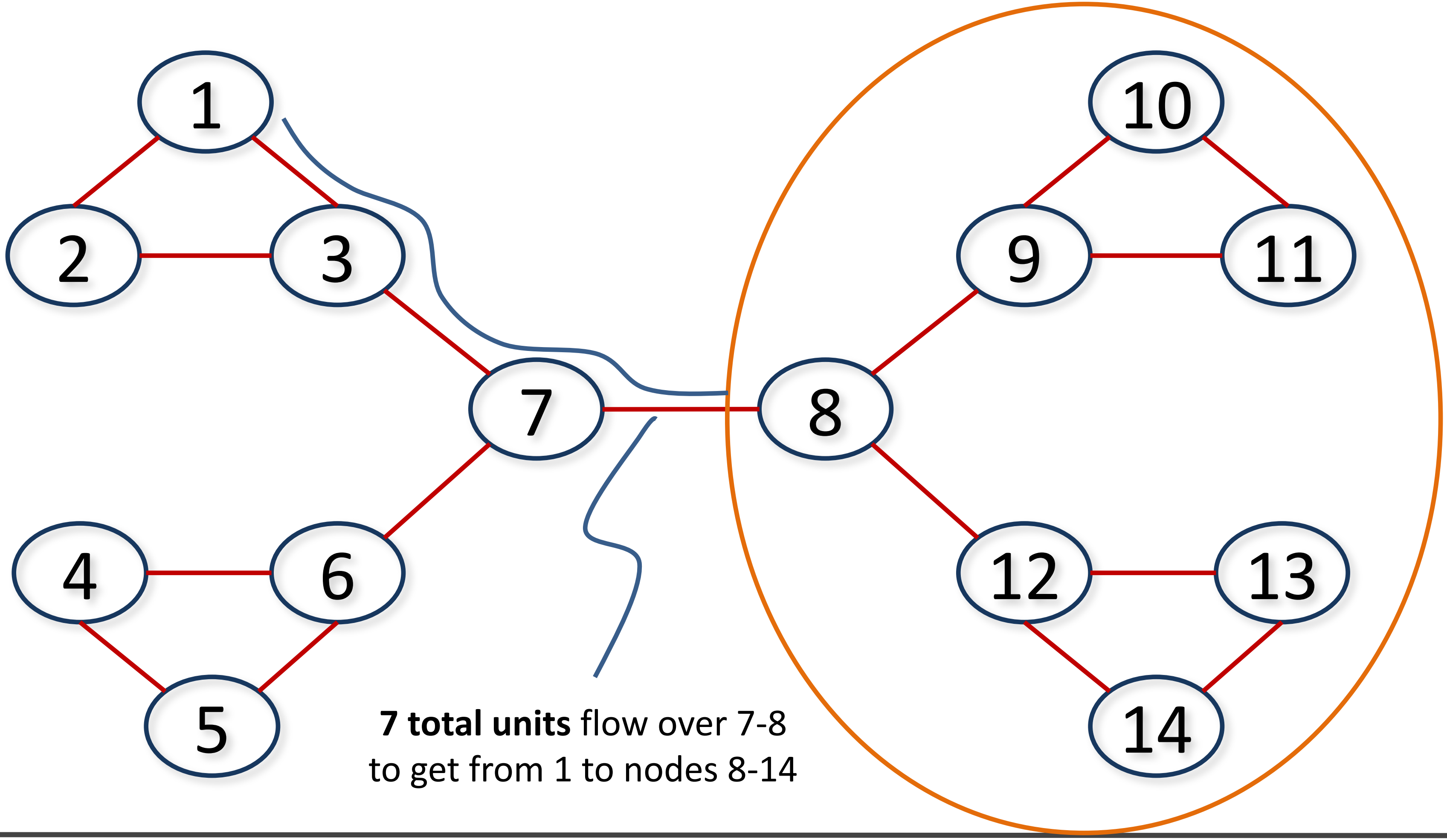
Edge Betweenness Example



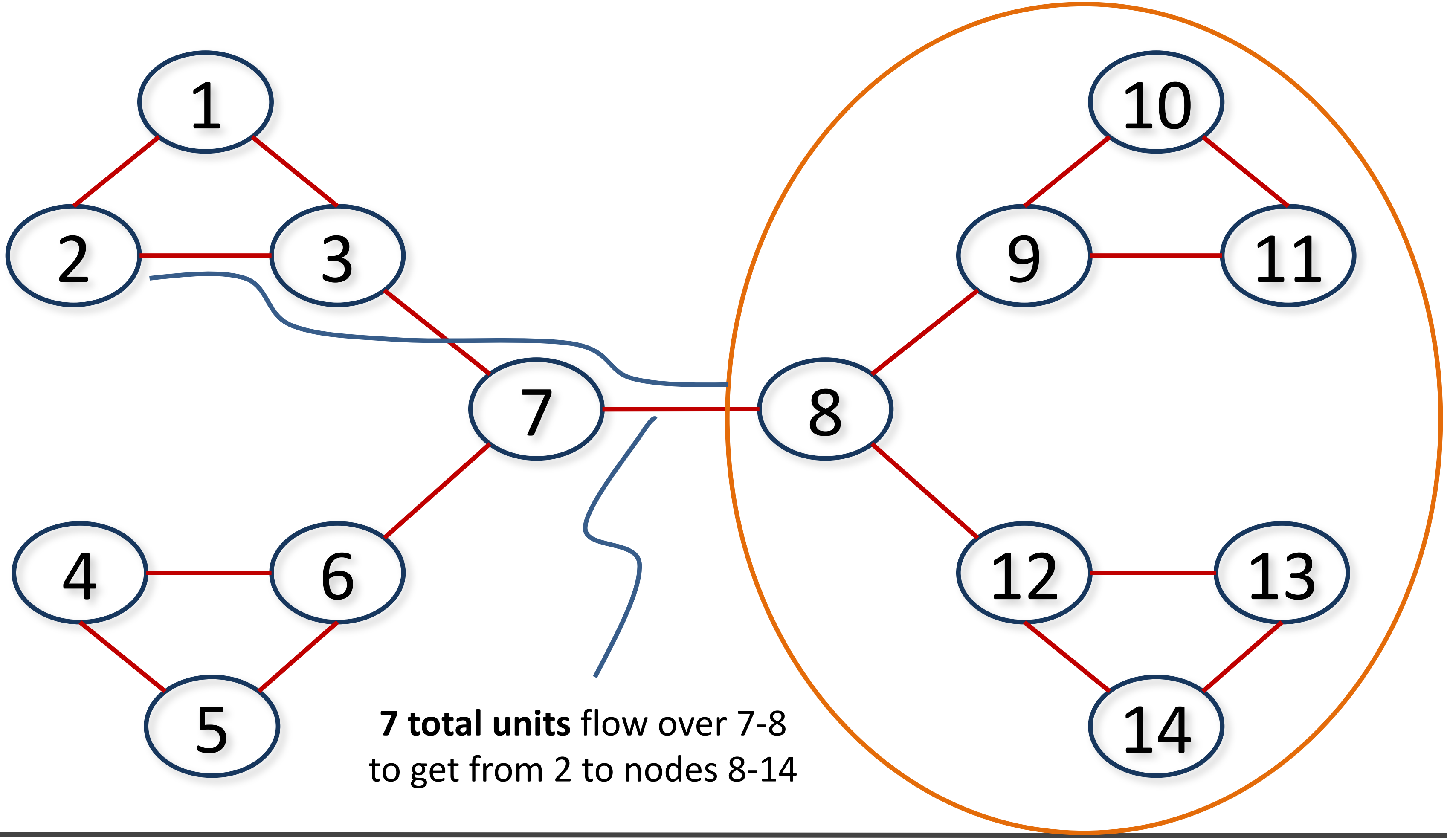

1
2
3
One unit flows over 7-8
to get from 1 to 8
7
8
9
10
11
4
5
6
12
13
14


1
10
2
3
One unit flows over 7-8
to get from 1 to 9
9
11
7
8
4
6
12
13
5
14

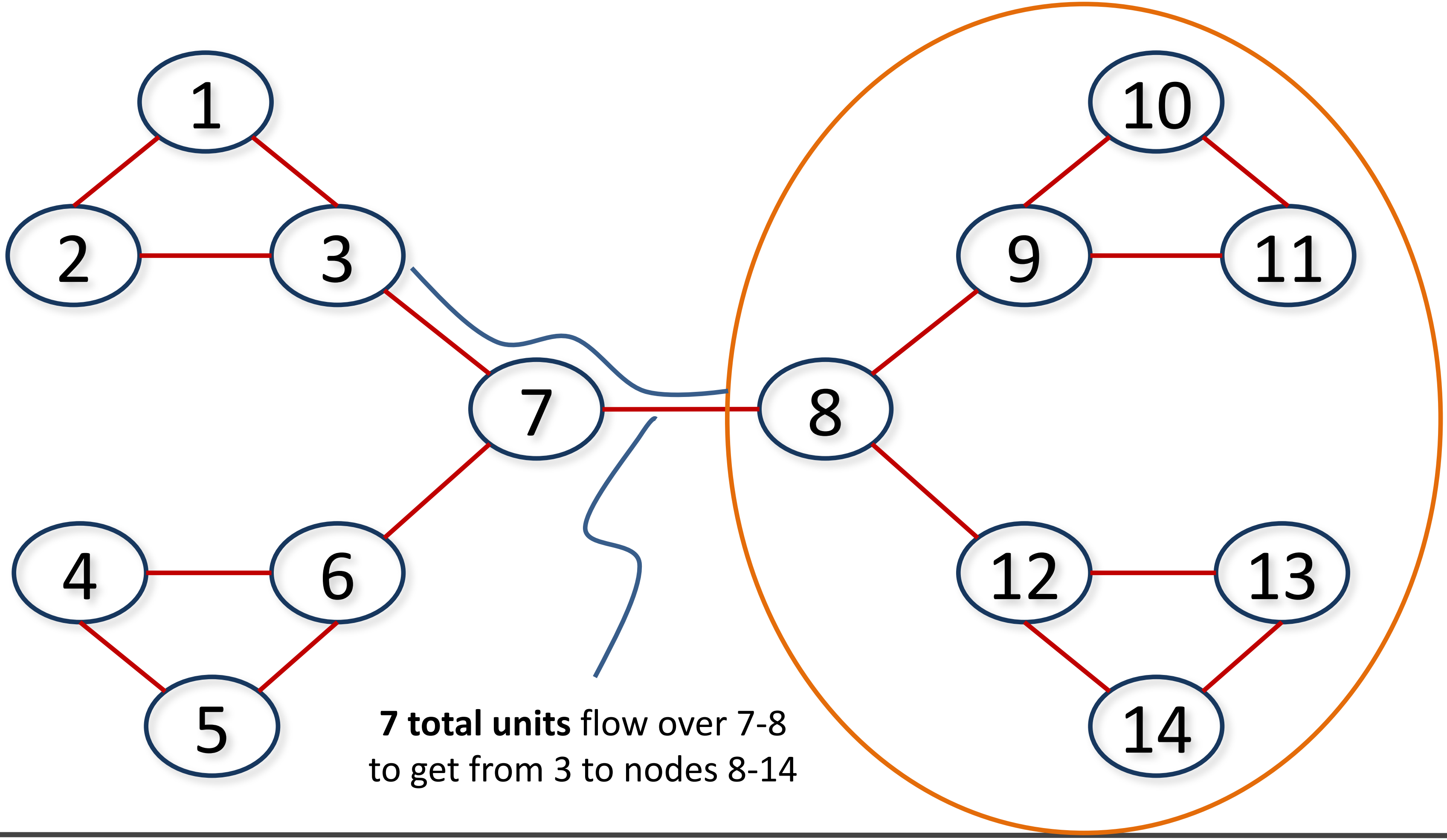

1
2
3
4
5
6
7
8
9
10
11
12
13
14
One unit flows over 7-8
to get from 1 to 10



7 total units flow over 7-8
to get from 1 to nodes 8-14

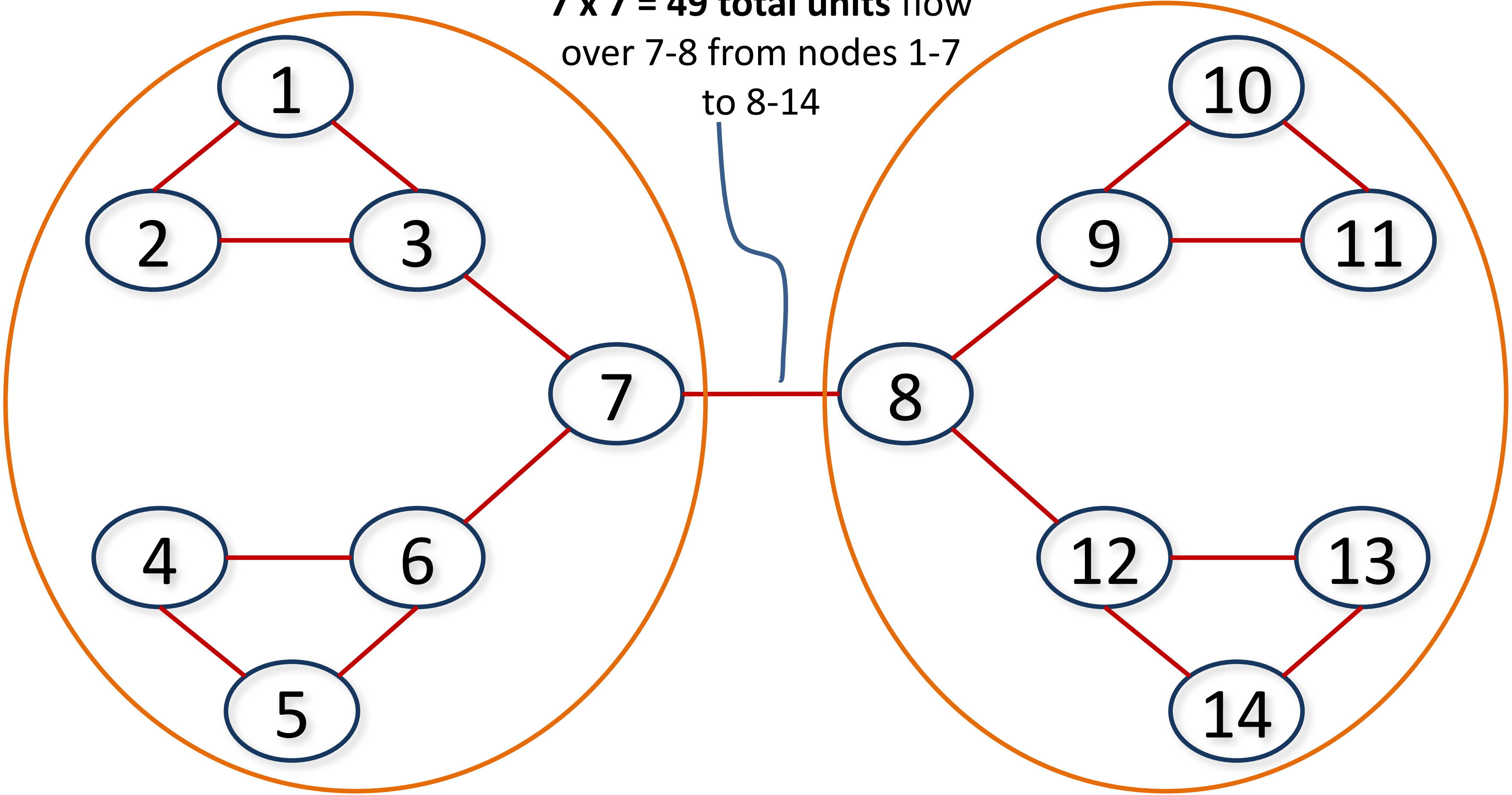


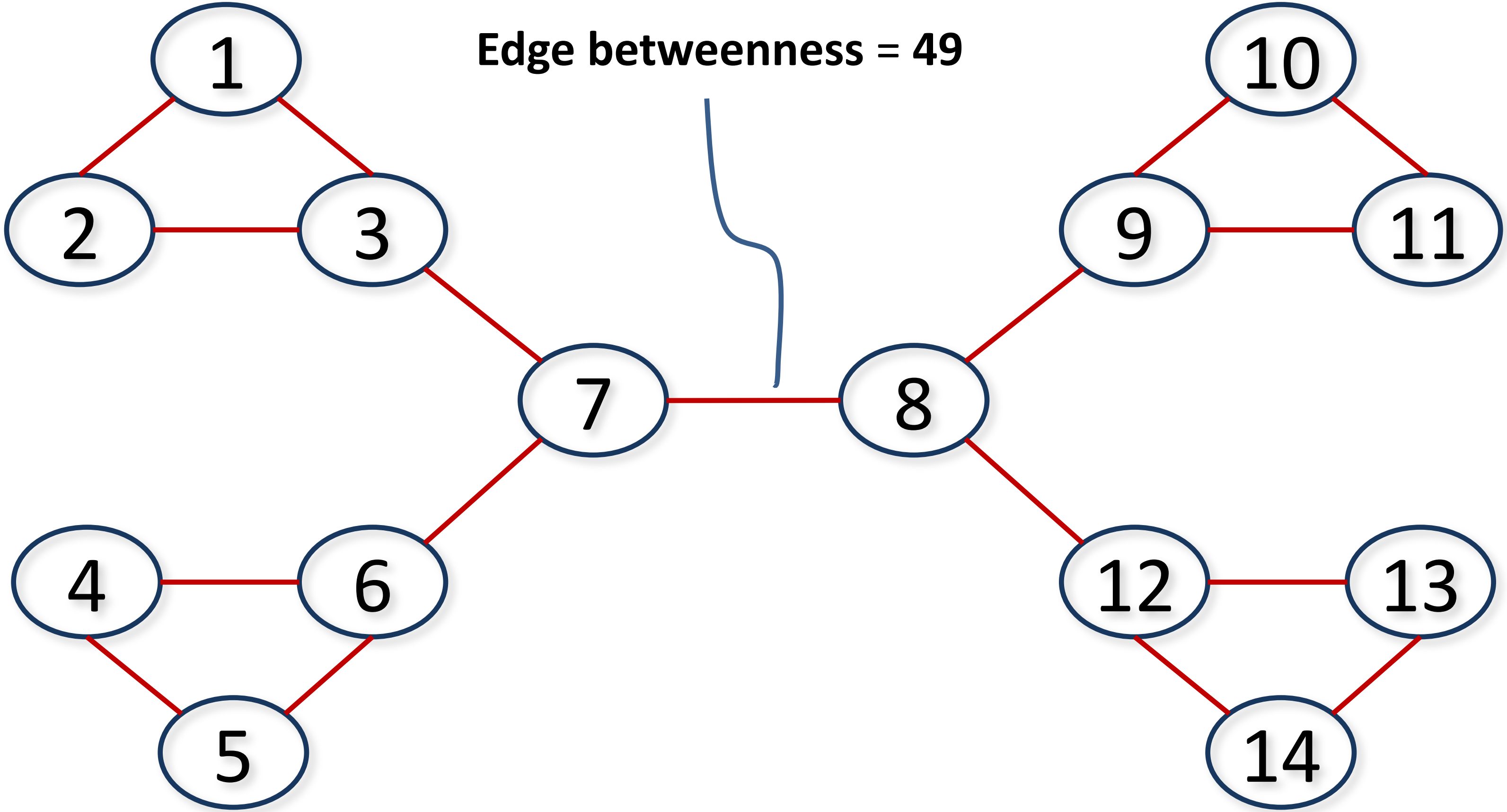
7 total units flow over 7-8
to get from 2 to nodes 8-14

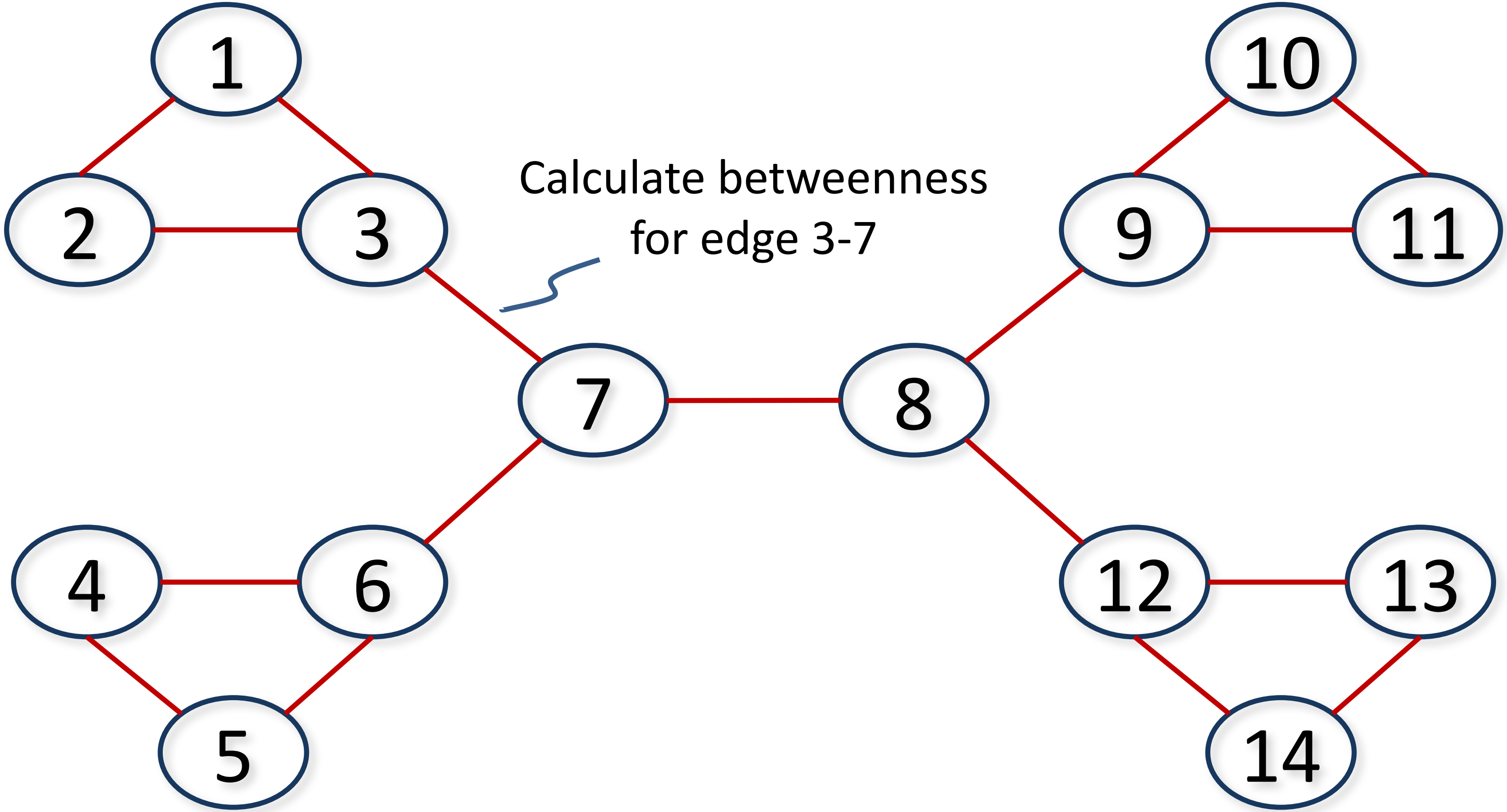


7 total units flow over 7-8
to get from 3 to nodes 8-14

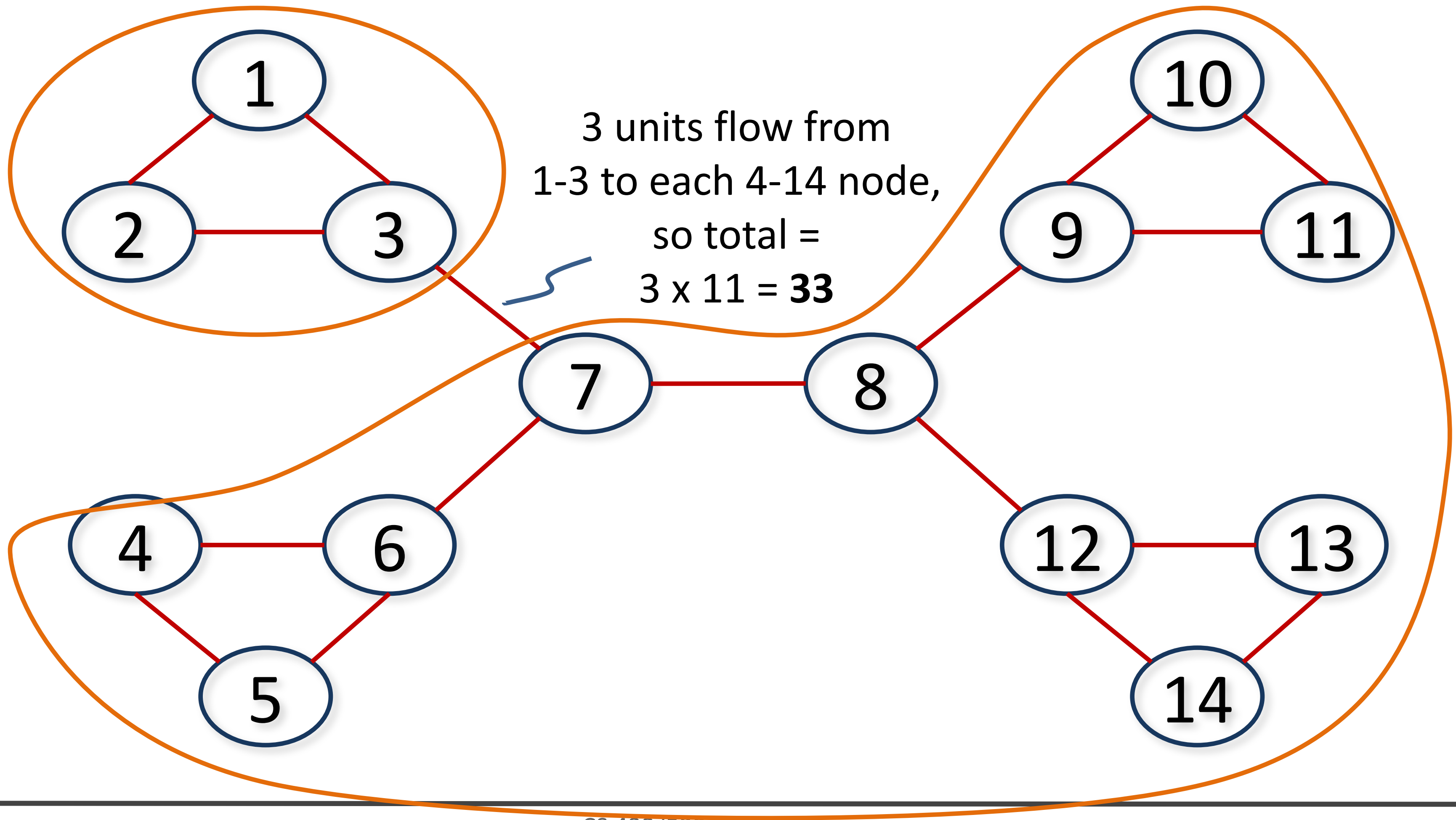
7 x 7 = 49 total units flow over 7-8 from nodes 1-7 to 8-14



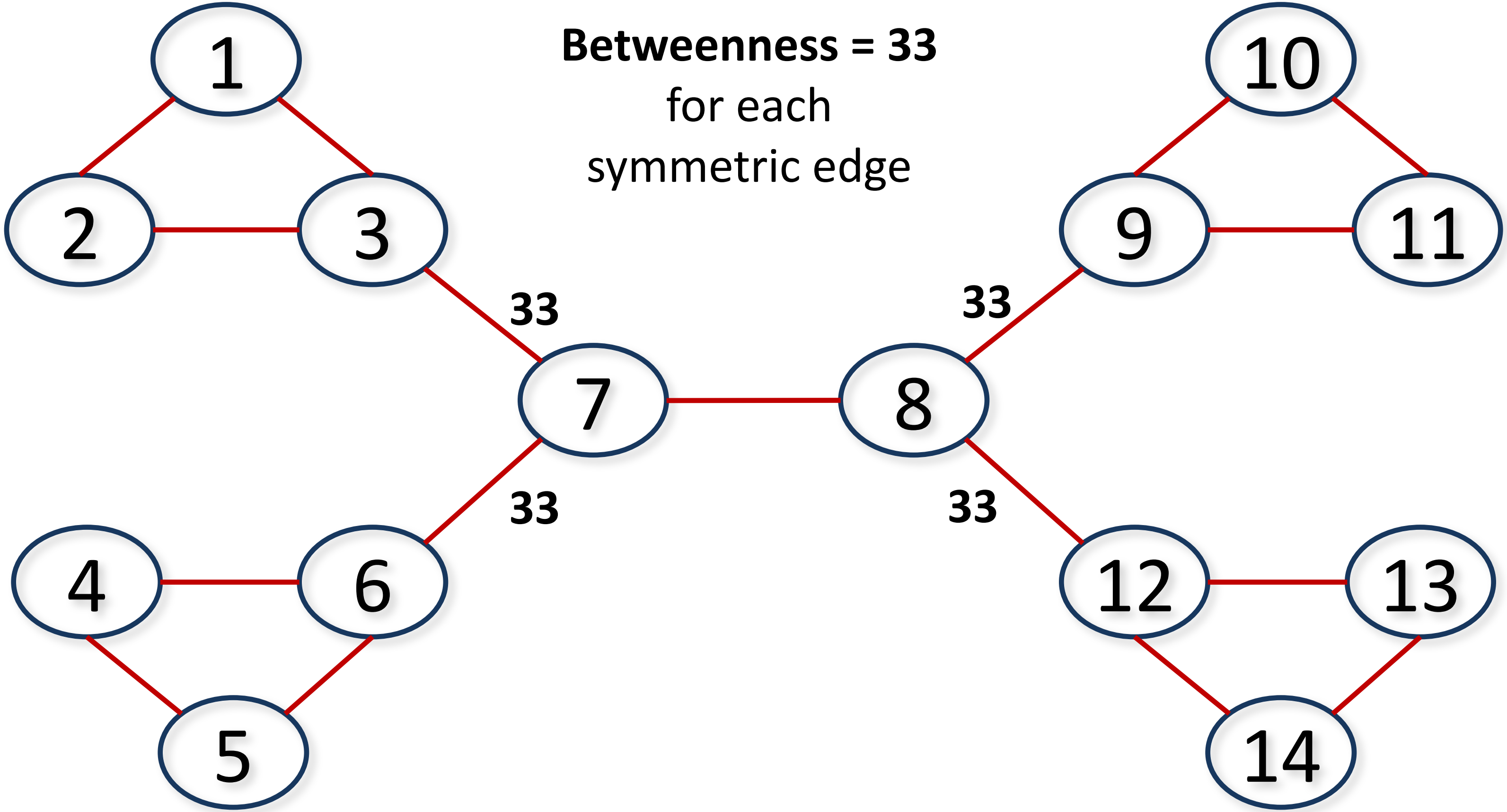

Edge betweenness = 49
1
2
3
4
5
6
7
8
9
10
11
12
13
14

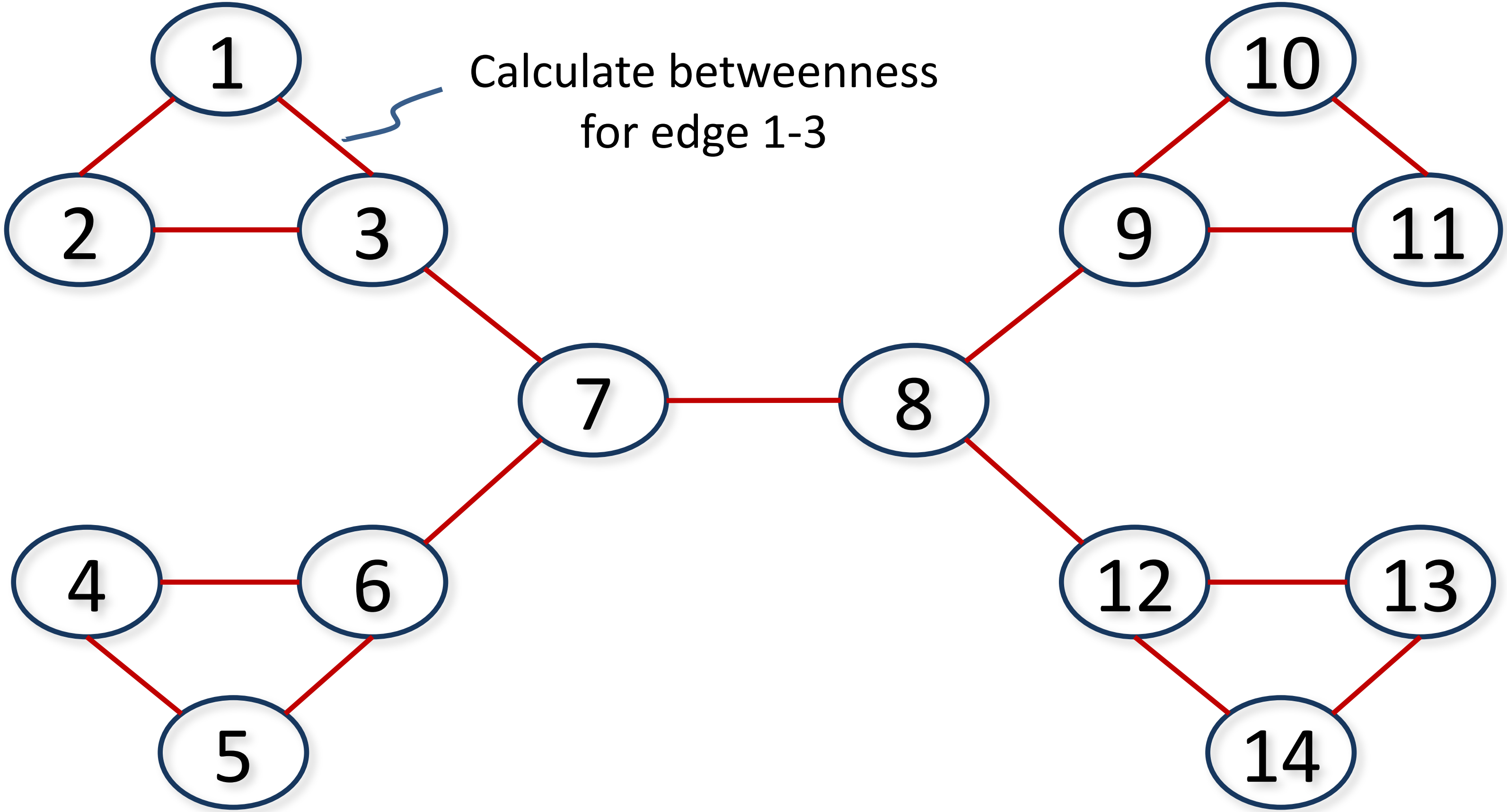


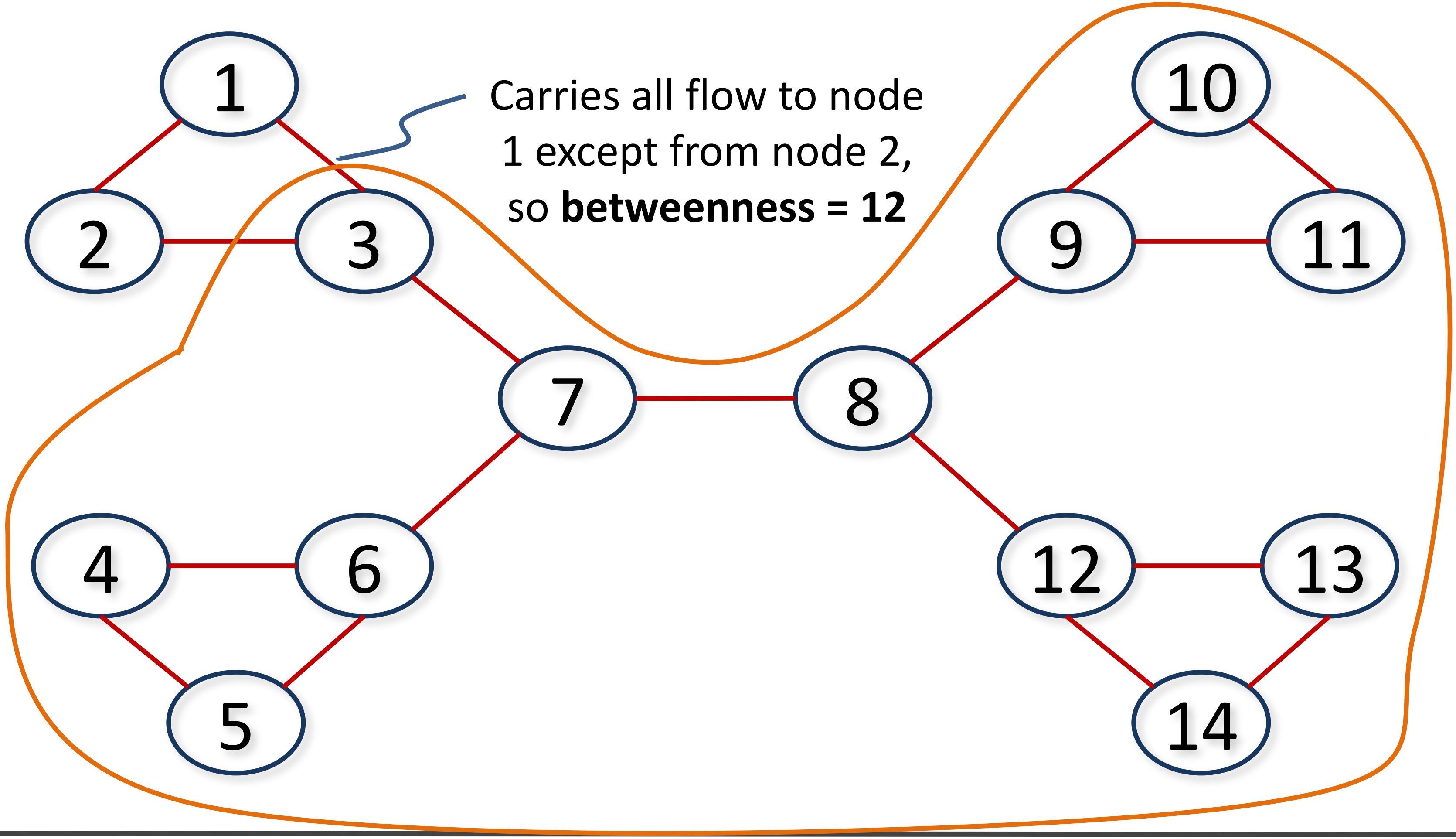
Calculate betweenness for edge 3-7

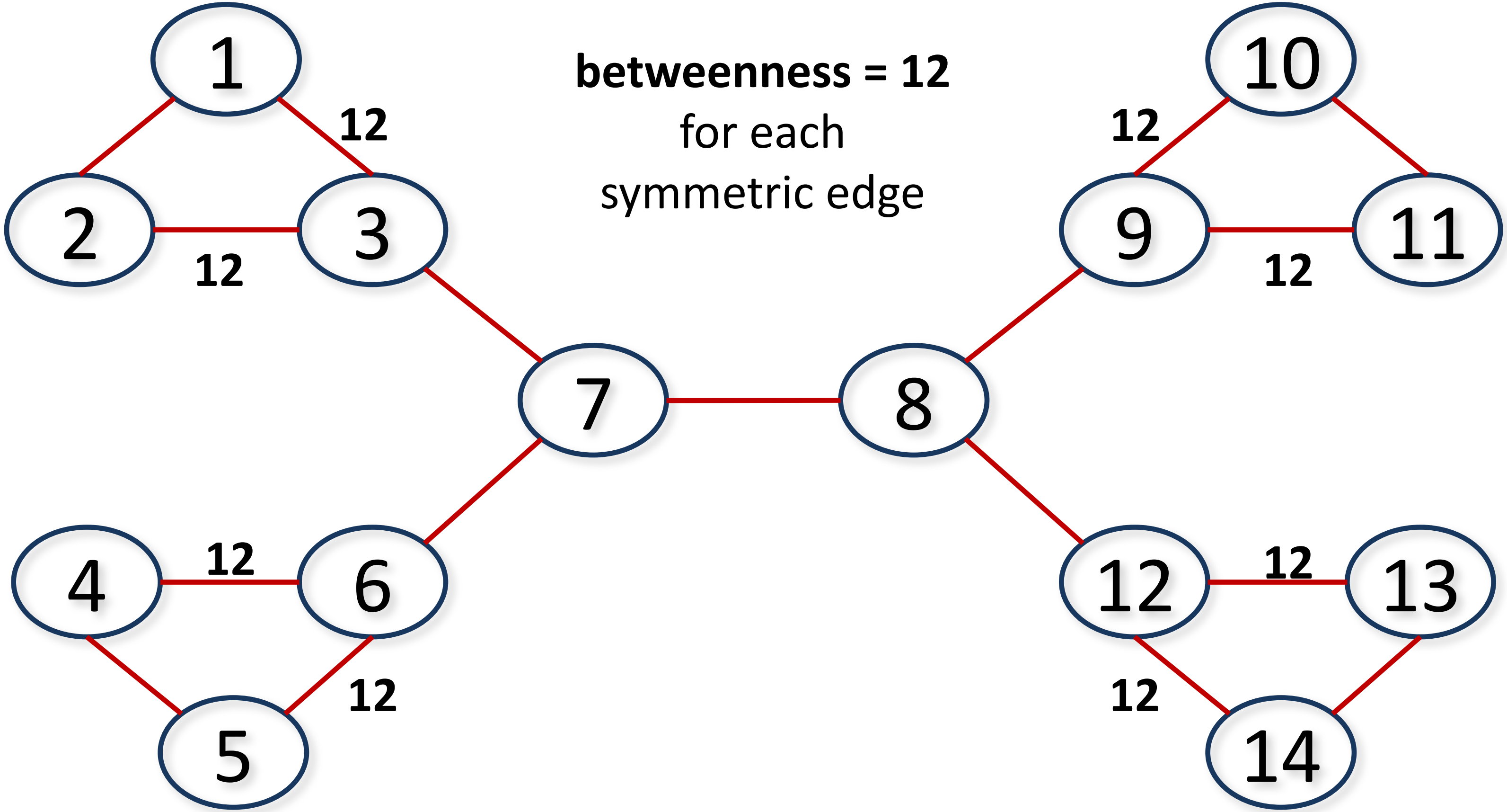

1
2
3
3 units flow from
1-3 to each 4-14 node,
so total =
3 x 11 = **33**
4
5
6
7
8
9
10
11
12
13
14

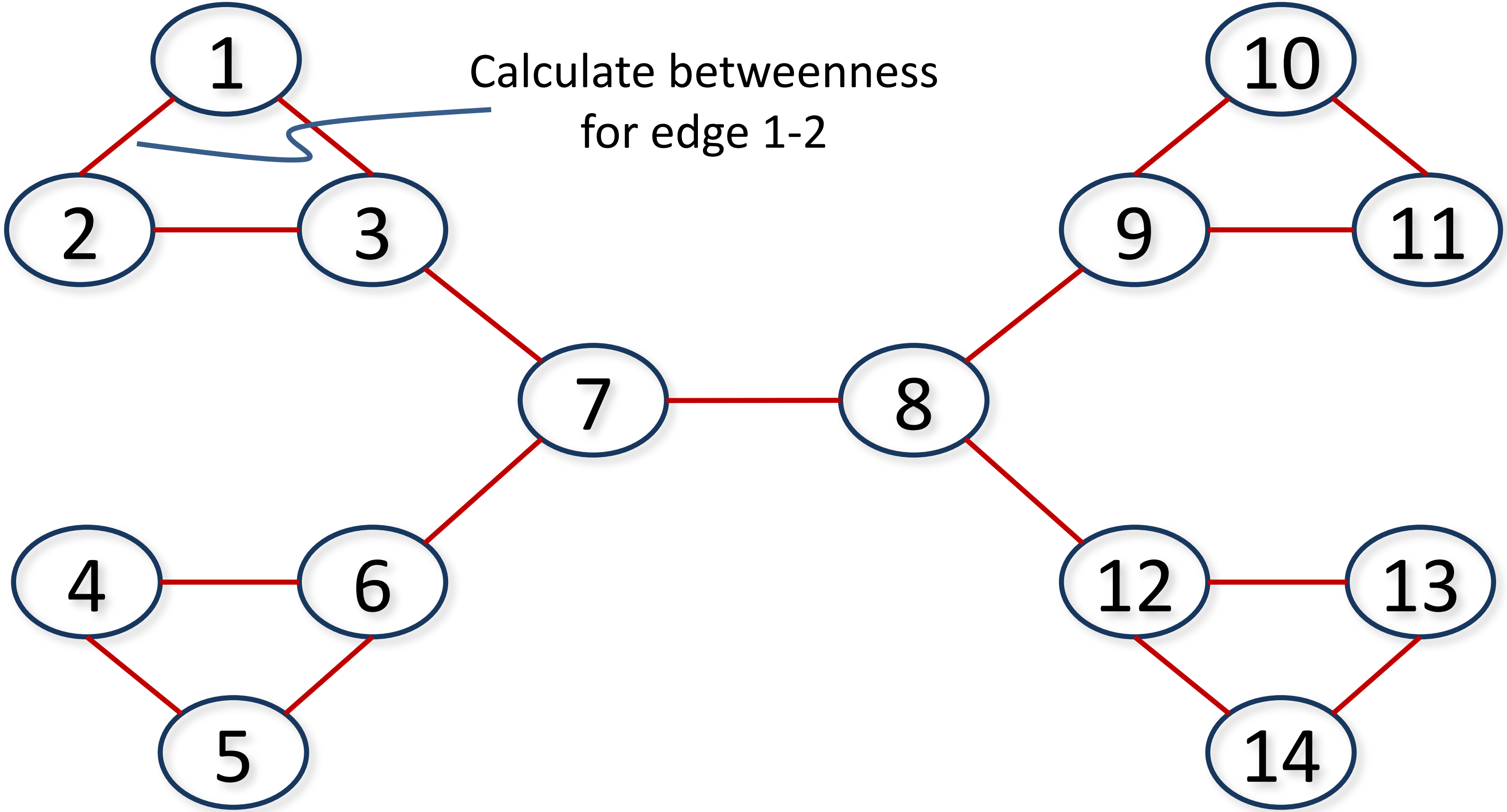
Betweenness = 33
for each
symmetric edge

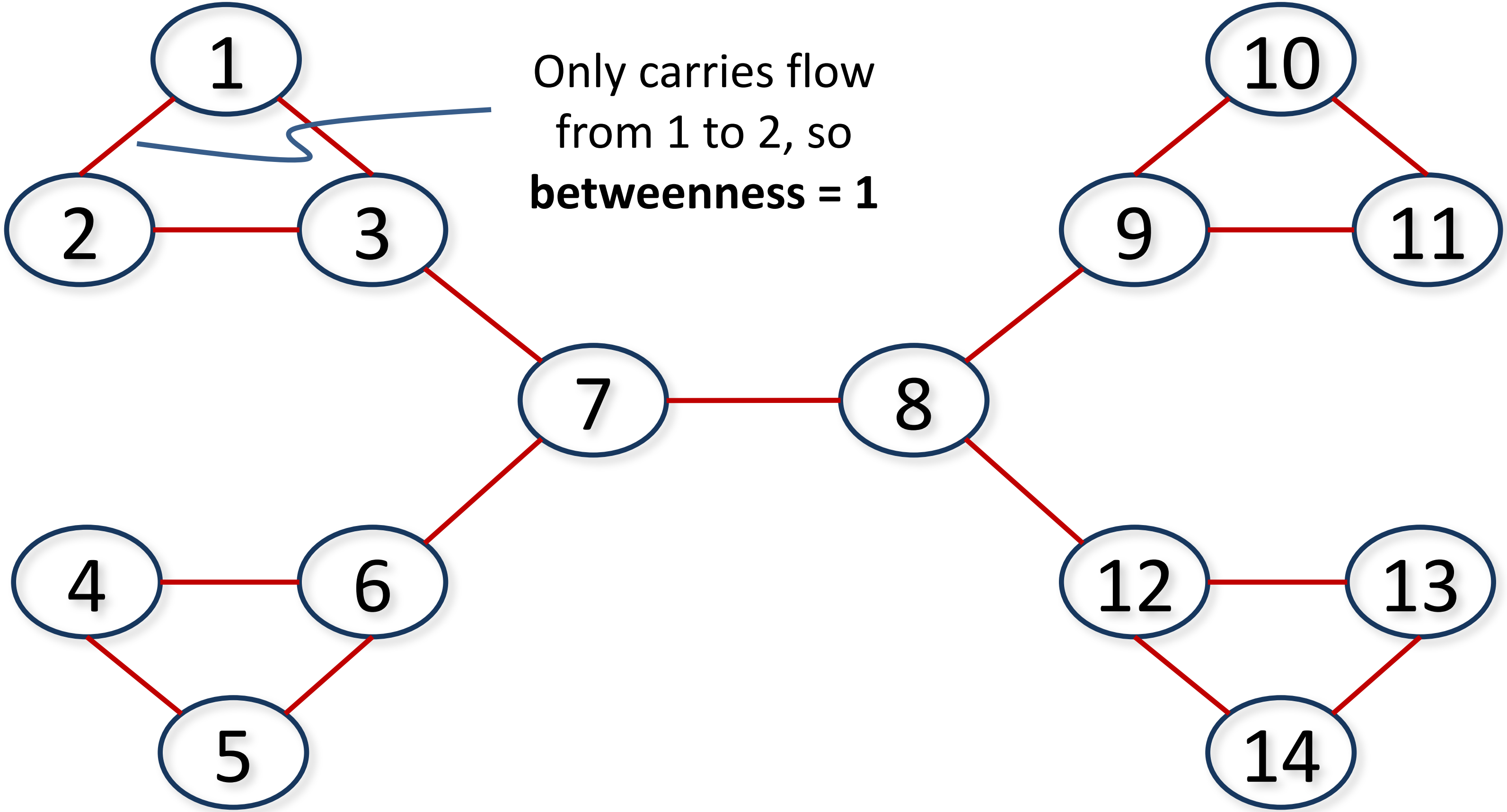



1
2
3
4
5
6
7
8
9
10
11
12
13
14
Calculate betweenness
for edge 1-3


Carries all flow to node 1 except from node 2, so **betweenness = 12**
1
2
3
4
5
6
7
8
9
10
11
12
13
14

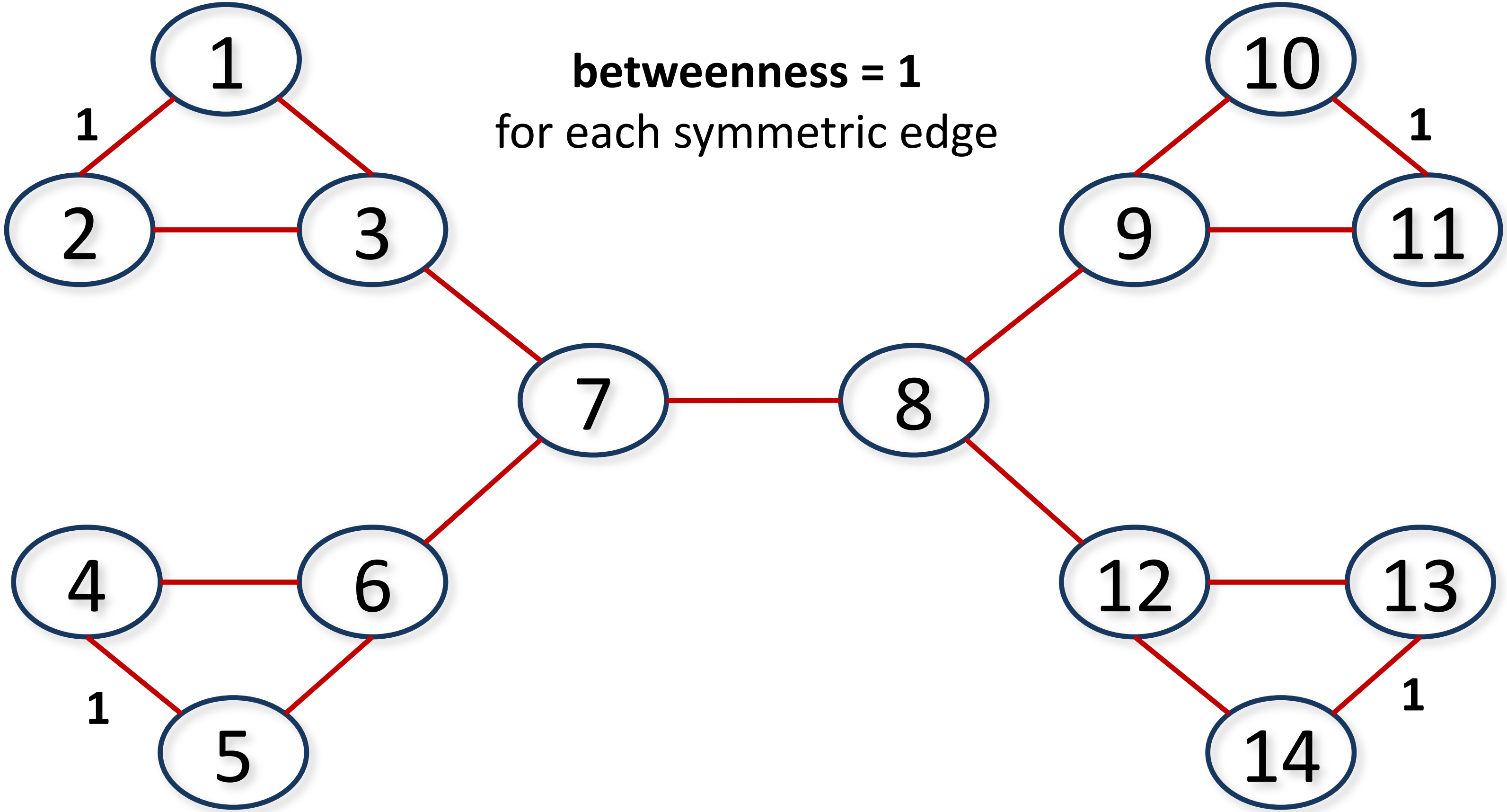

betweenness = 12
for each
symmetric edge
1
2
3
4
5
6
7
8
9
10
11
12
13
14
12
12
12
12
12
12
12
12

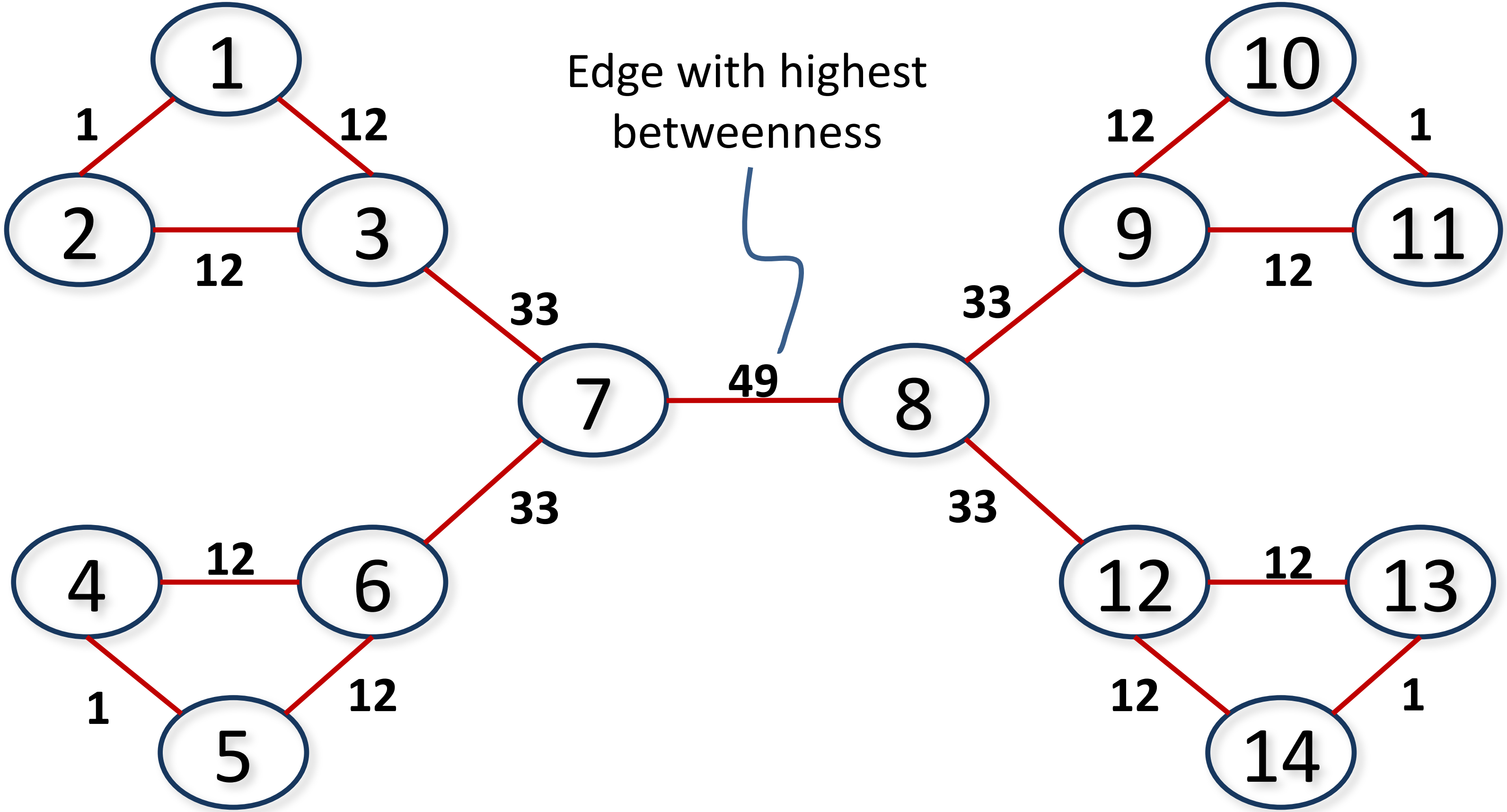

Calculate betweenness
for edge 1-2
1
2
3
4
5
6
7
8
9
10
11
12
13
14



Only carries flow
from 1 to 2, so
betweenness = 1

betweenness = 1
for each symmetric edge




Edge with highest betweenness
1
2
3
4
5
6
7
8
9
10
11
12
13
14
1
12
12
33
33
12
1
12
49
33
33
12
1
12
12
12
1

Girvan-Newman Algorithm

1. Calculate betweenness of all edges

2. Remove the edge(s) with highest betweenness

3. Repeat steps 1 and 2 until graph is partitioned into as many regions as desired

Girvan-Newman Pseudocode

Step 3

Step 1

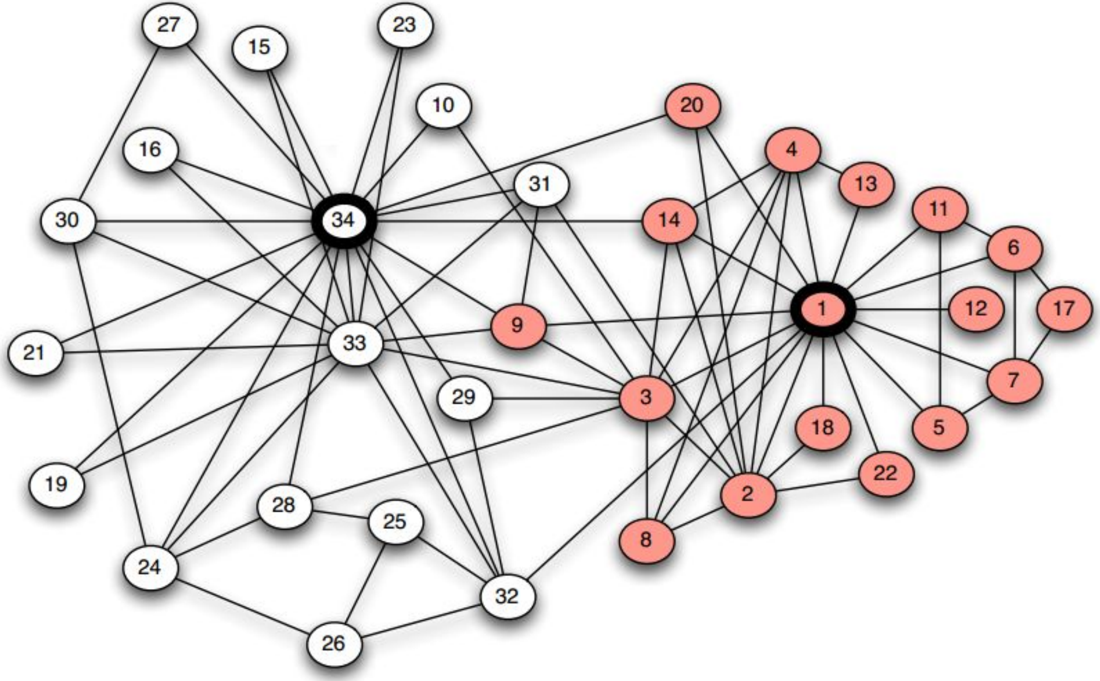
Step 2

```
Given graph G,
while( edgeCount > 0 and clusterCount < maximumClusterThreshold )
{
    allEdgesBetweennessList = G.getBetweennessValues()
    maximumBetweennessValue = getMaximum(allEdgeBetweennessList)

    edgeWithMaximumBetweennessValue =
    getEgdeWithMaximumBetweennessValue(maximumBetweennessValue)

    G.remove(edgeWithMaximumBetweennessValue)
}
```

Karate Club splits after a dispute. Can new clubs be identified based on network structure?

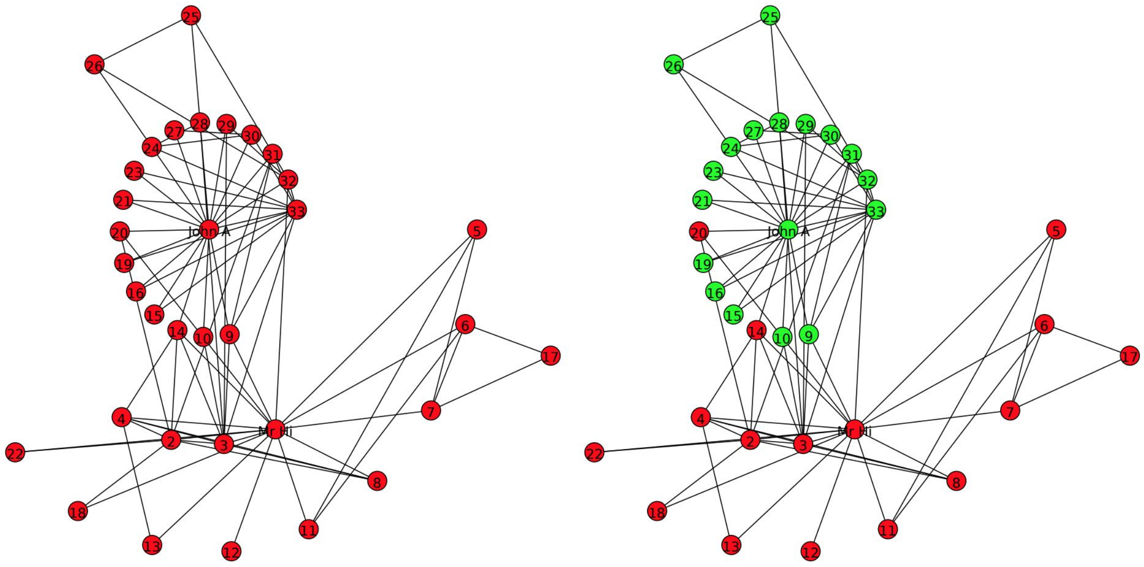


Zachary W. (1977). An information flow model for conflict and fission in small groups. Journal of Anthropological Research, 33, 452-473.

[Zachary's karate club](#) (Wikipedia) / see data set at [ZACHARY KARATE CLUB at UCINET IV Datasets](#) and [Zachary karate club at KONECT (archived)](#)

Girvan-Newman run on Karate-Club graph

Figure 1: Zachary's Karate Club Graphs

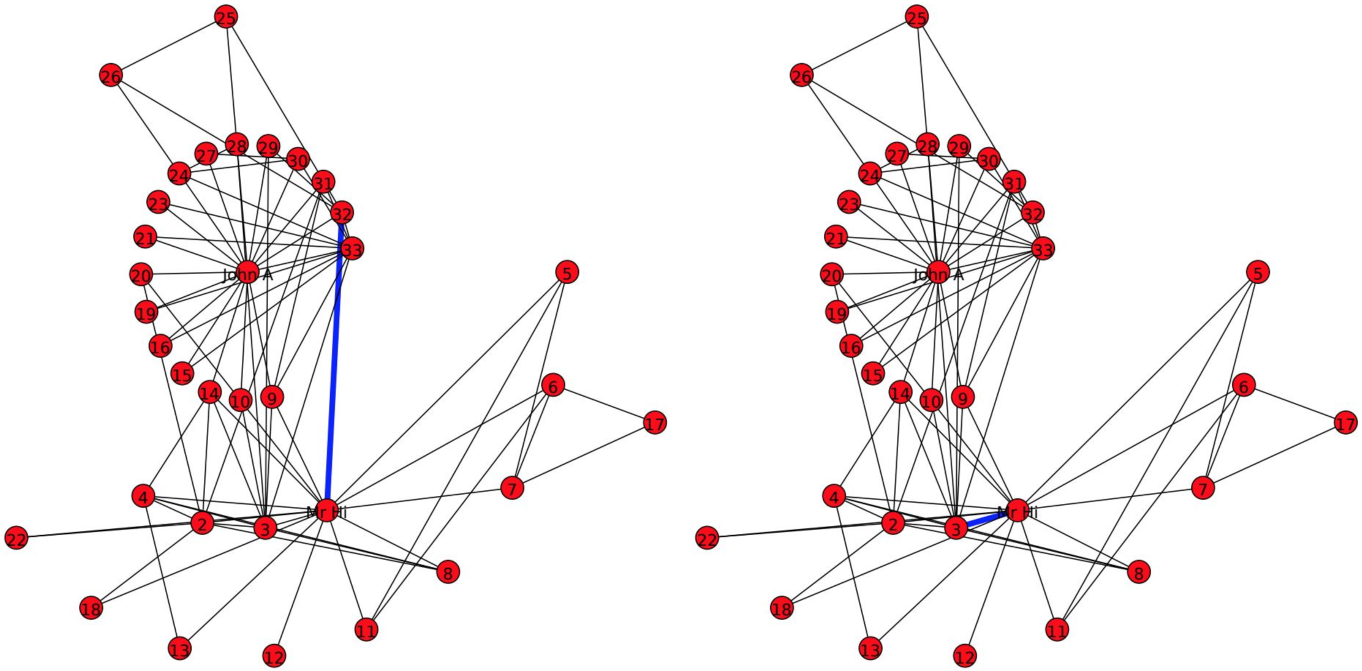


(a) Original Graph

(b) Result of Split Graph

Graphs generated by Alexander Nwala, Fall 2014

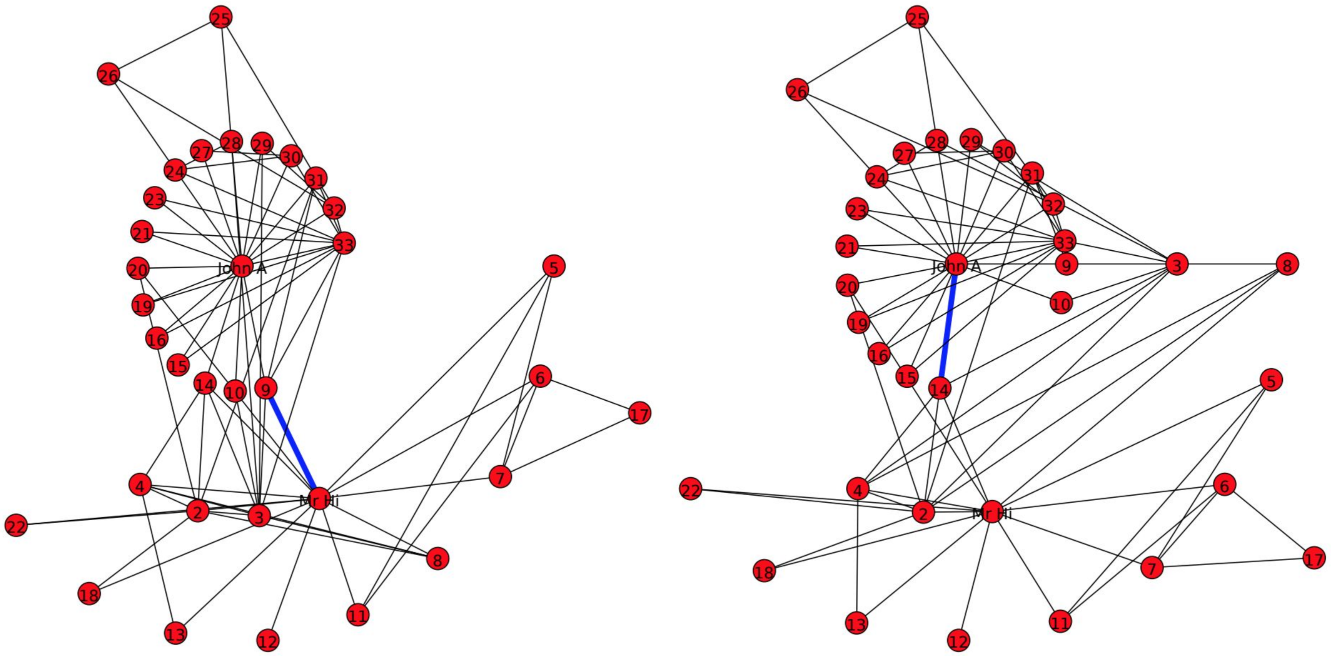
Girvan-Newman run on Karate-Club graph



(a) Iteration: 0, Clusters: 1

(b) Iteration: 1, Clusters: 1

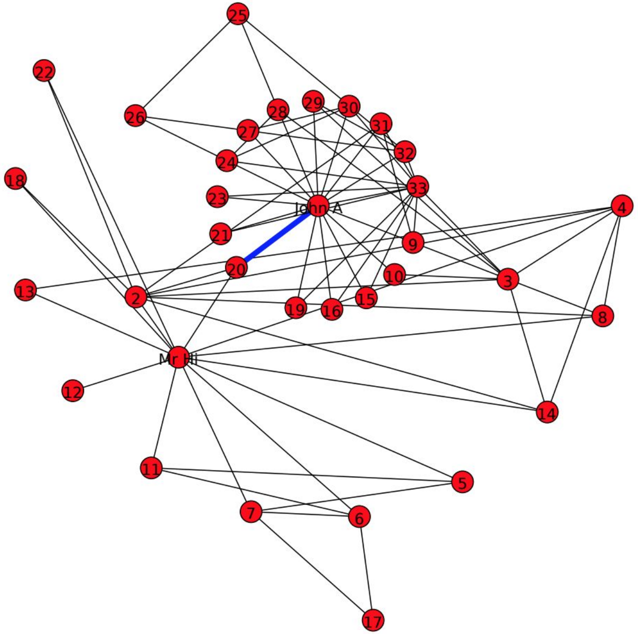
Girvan-Newman run on Karate-Club graph



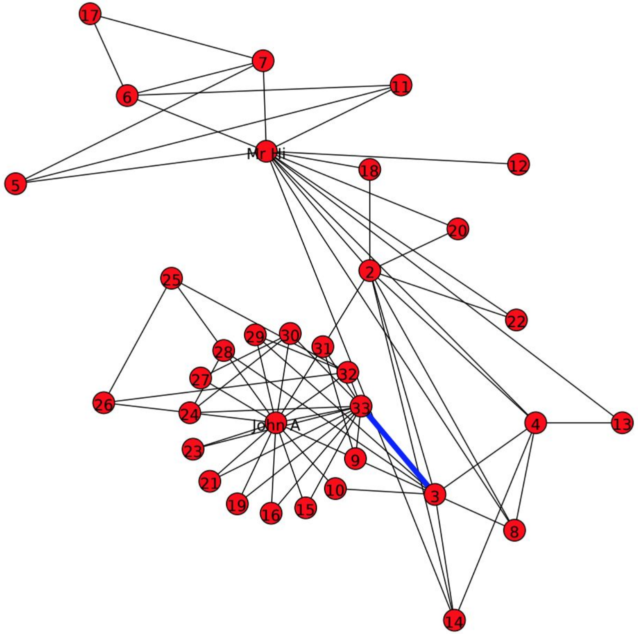
(c) Iteration: 2, Clusters: 1

(d) Iteration: 3, Clusters: 1

Girvan-Newman run on Karate-Club graph

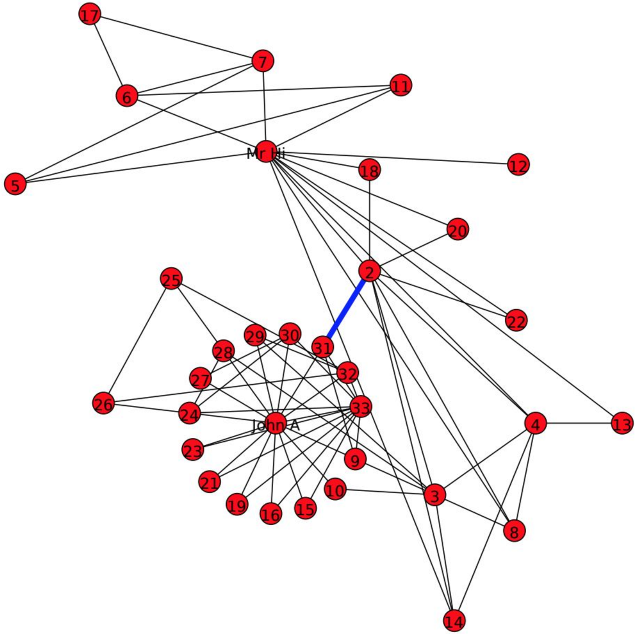


(a) Iteration: 4, Clusters: 1

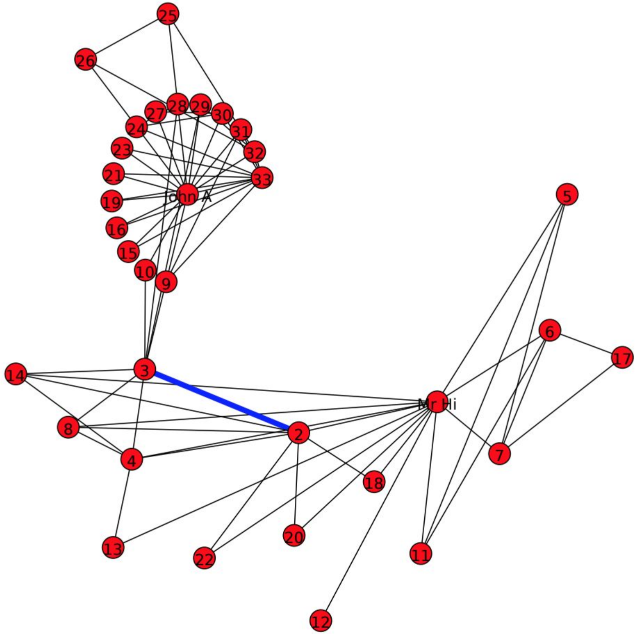


(b) Iteration: 5, Clusters: 1

Girvan-Newman run on Karate-Club graph

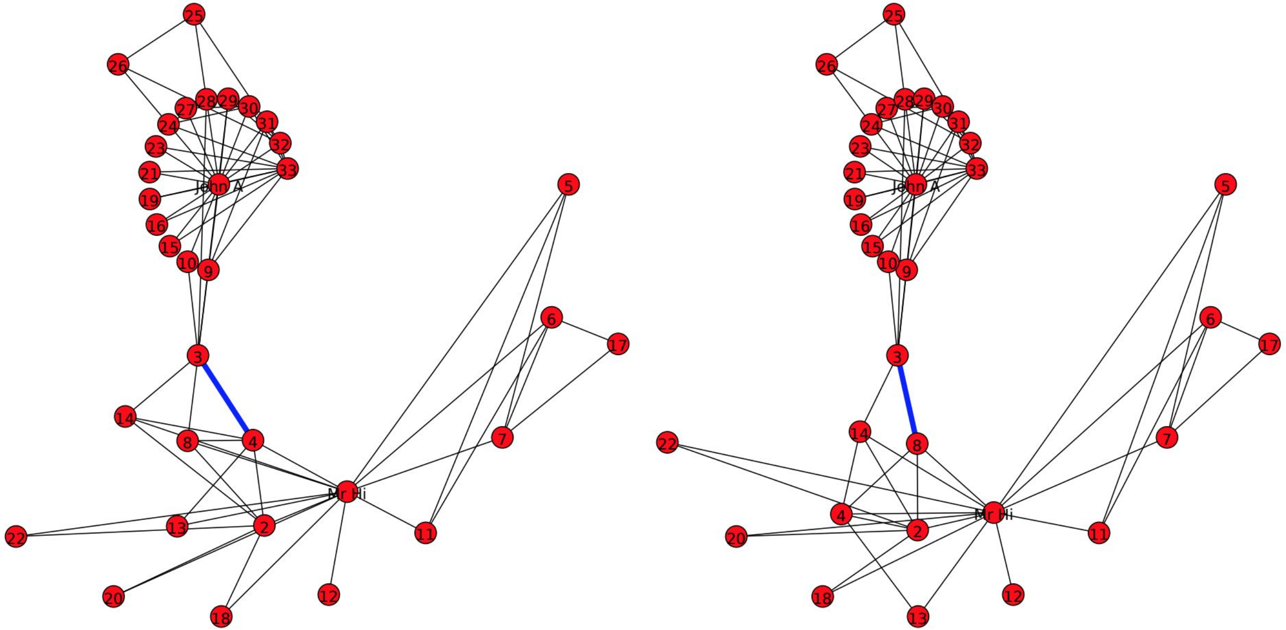


(c) Iteration: 6, Clusters: 1



(d) Iteration: 7, Clusters: 1

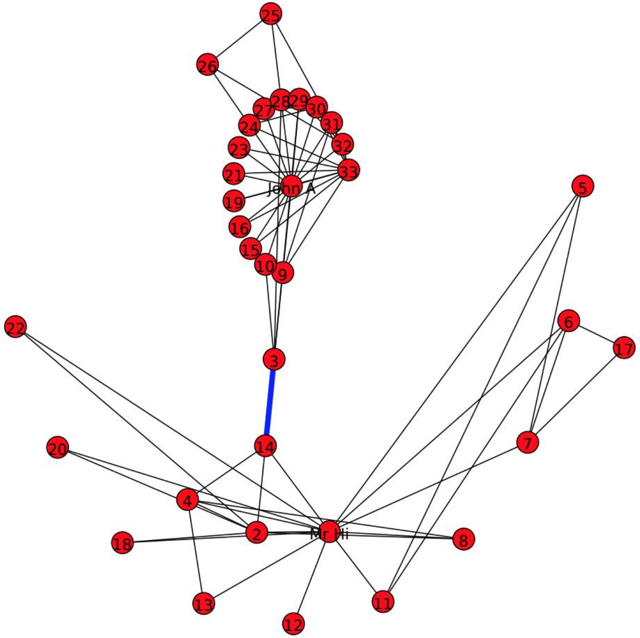
Girvan-Newman run on Karate-Club graph



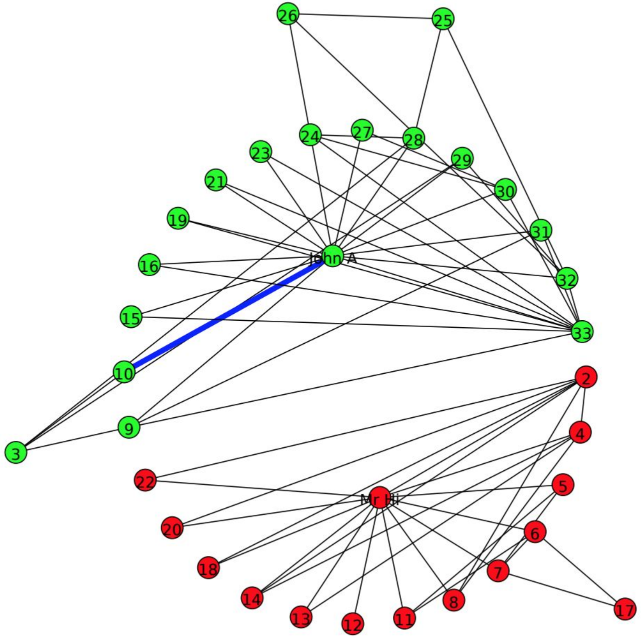
(a) Iteration: 8, Clusters: 1

(b) Iteration: 9, Clusters: 1

Girvan-Newman run on Karate-Club graph



(c) Iteration: 10, Clusters: 1



(d) Iteration: 11, Clusters: 2

Objectives

- Describe the friendship paradox.
- Explain the triadic closure in social networks.
- Explain how edge embeddedness affects trust.
- Explain the edge betweenness property.
- Explain the steps in the Girvan-Newman graph partitioning algorithm.