PolNet 2024

Network Modeling

Sarah Shugars
they/them/theirs

Assistant Professor
Rutgers University

Welcome!

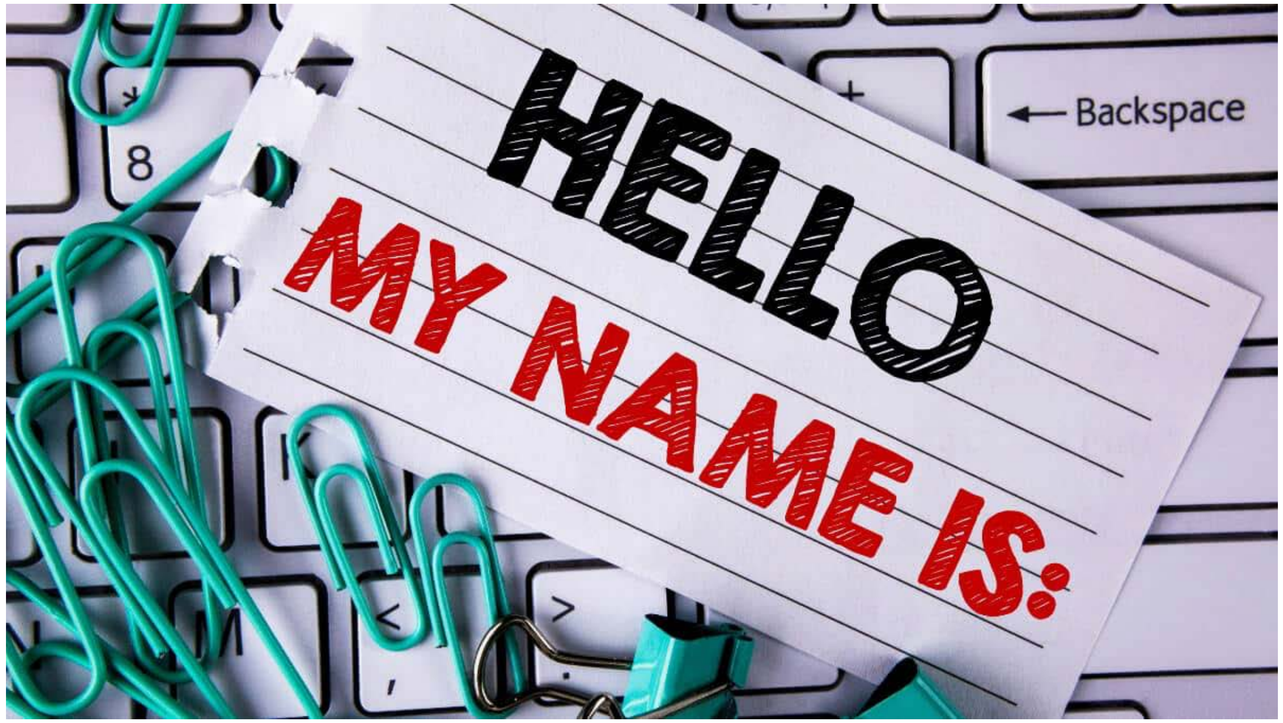
Dr. Sarah Shugars
sarah.shugars@rutgers.edu
they/them/theirs

- First gen to college
- PolNet regular since 2017
- PhD in Network Science; Northeastern 2020
- Faculty Fellow at NYU's Center for Data Science 2020-2022
- Now: Assistant Professor of Communication, Rutgers University
- *Research areas*: Political communication, computational methods

Introductions

Please share your:

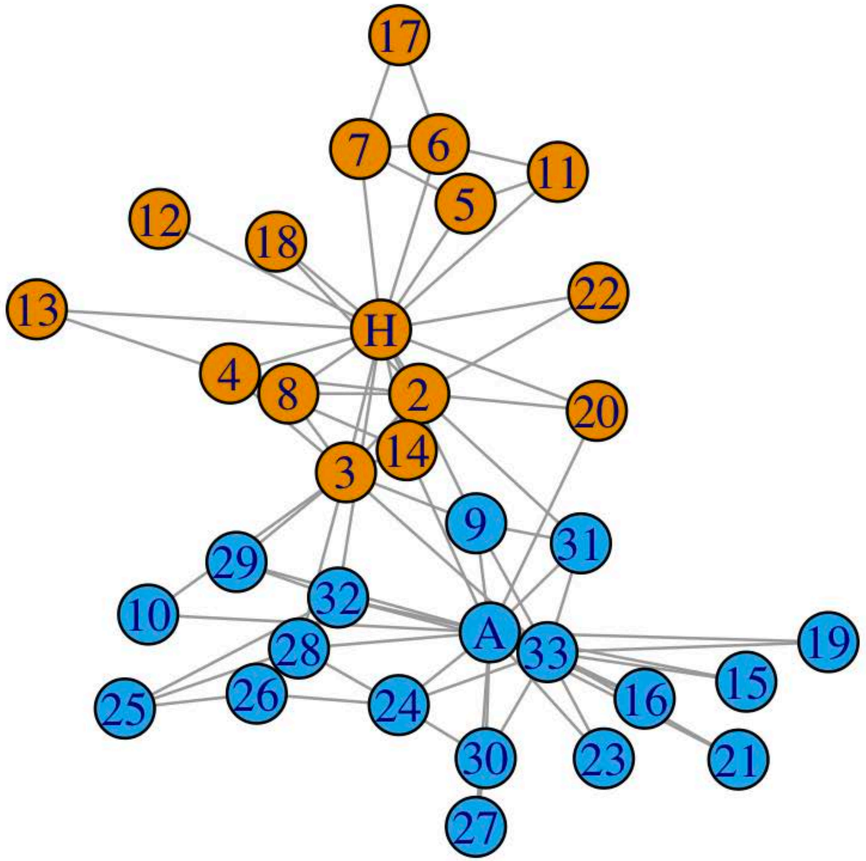
- ✦ Name
- ✦ Institution and role (ie, PhD student)
- ✦ Interest in networks:
  - ➡ Do you have a network you are studying or planning to study?



Workshop Motivation

	V1	V2
1	Mr Hi	Actor 2
2	Mr Hi	Actor 3
3	Mr Hi	Actor 4
4	Mr Hi	Actor 5
5	Mr Hi	Actor 6
6	Mr Hi	Actor 7
7	Mr Hi	Actor 8
8	Mr Hi	Actor 9
9	Mr Hi	Actor 11
10	Mr Hi	Actor 12
11	Mr Hi	Actor 13
12	Mr Hi	Actor 14
13	Mr Hi	Actor 18
14	Mr Hi	Actor 20
15	Mr Hi	Actor 22
16	Mr Hi	Actor 32
17	Actor 2	Actor 3
18	Actor 2	Actor 4
19	Actor 2	Actor 8
20	Actor 2	Actor 14
21	Actor 2	Actor 18
22	Actor 2	Actor 20
23	Actor 2	Actor 22
24	Actor 2	Actor 31
25	Actor 3	Actor 4
26	Actor 3	Actor 8
27	Actor 3	Actor 9
28	Actor 3	Actor 10
29	Actor 3	Actor 14
30	Actor 3	Actor 28

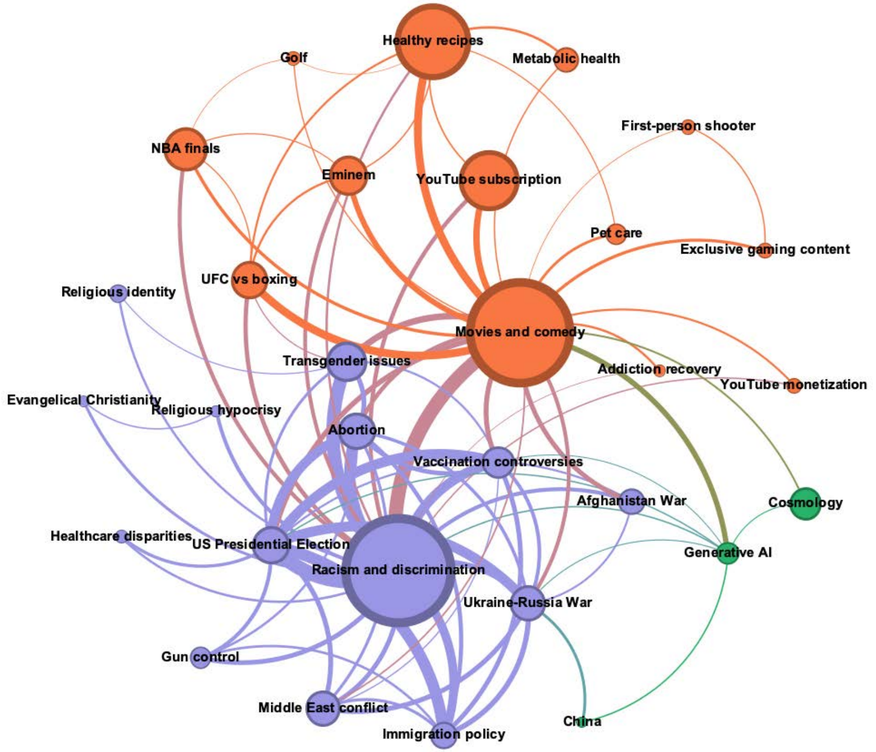
Network data doesn't magically show up looking like this...



Workshop Motivation

How do we go from

To

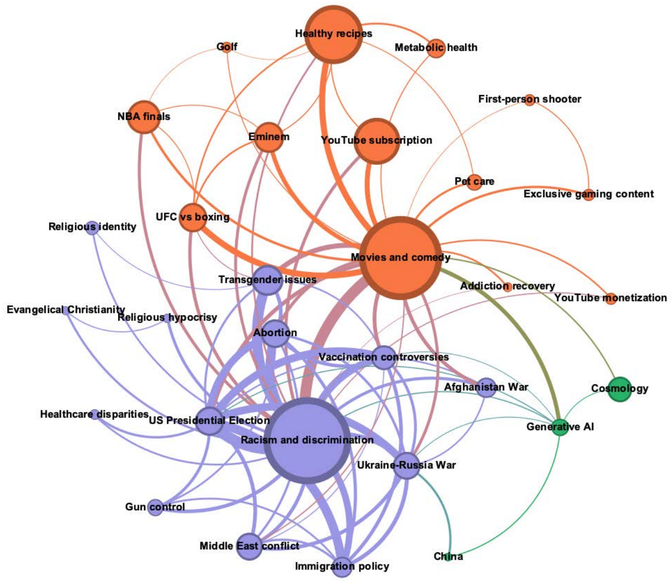


Observing something in the world

Building and analysing a network model?

Short Answer: It Depends

- Research always involves making and justifying **choices**
  - ➡ What choices are best will always depend on your **research questions**
  - ➡ …And, probably **practical considerations** as well



This workshop will reflect on (some of) the choices you will make during network modeling

- Will focus on conceptual discussion
- Plus some examples in R

Agenda

Part 1:

- Example: Zachary's Karate Club
  - ➡ Dealing with temporality in networks
- Example: Ethnic Networks
  - ➡ Aggregating edge measures
- Overview of network inference
  - ➡ Network comparison (QAP)

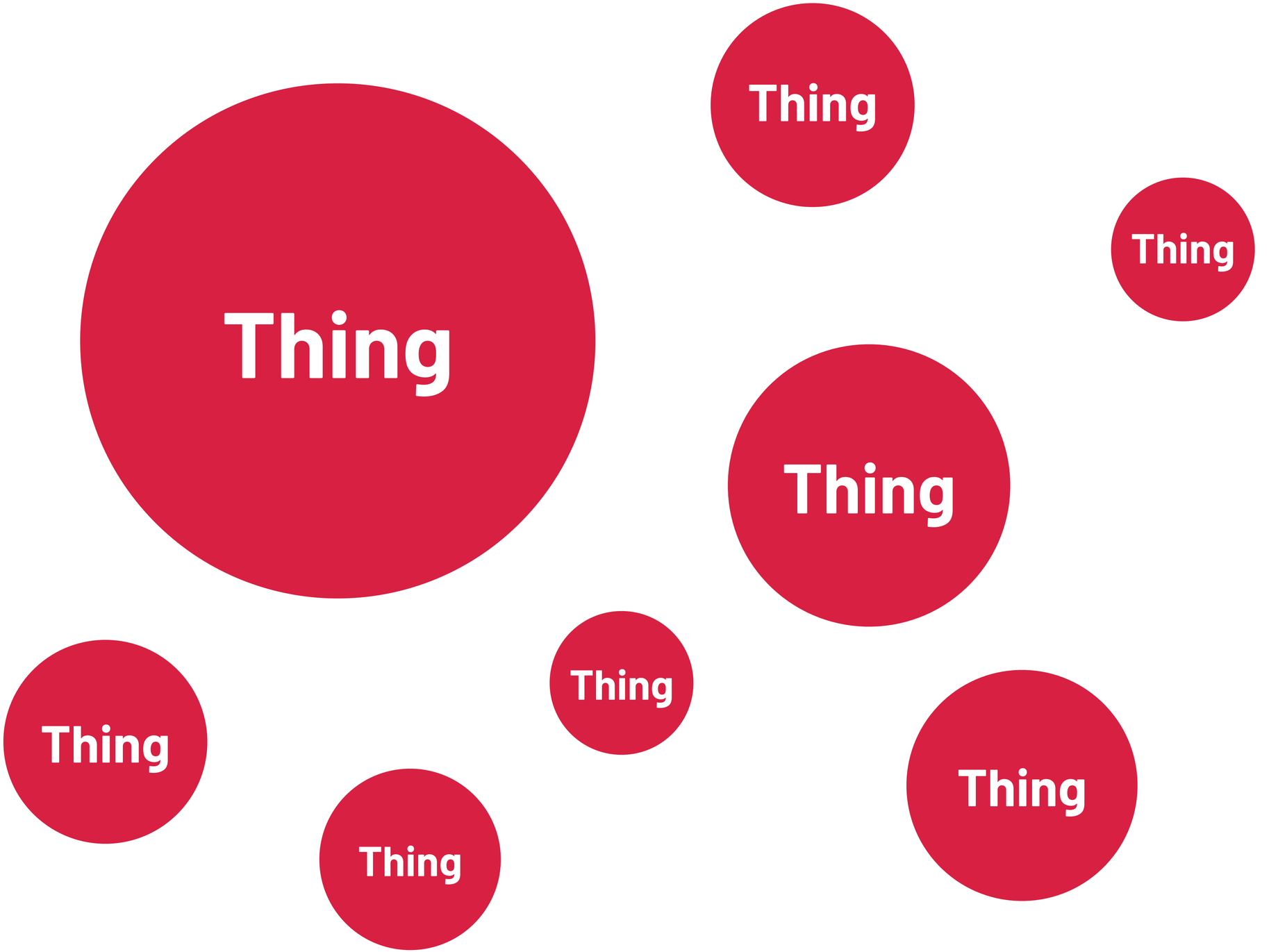
Break

Part 2:

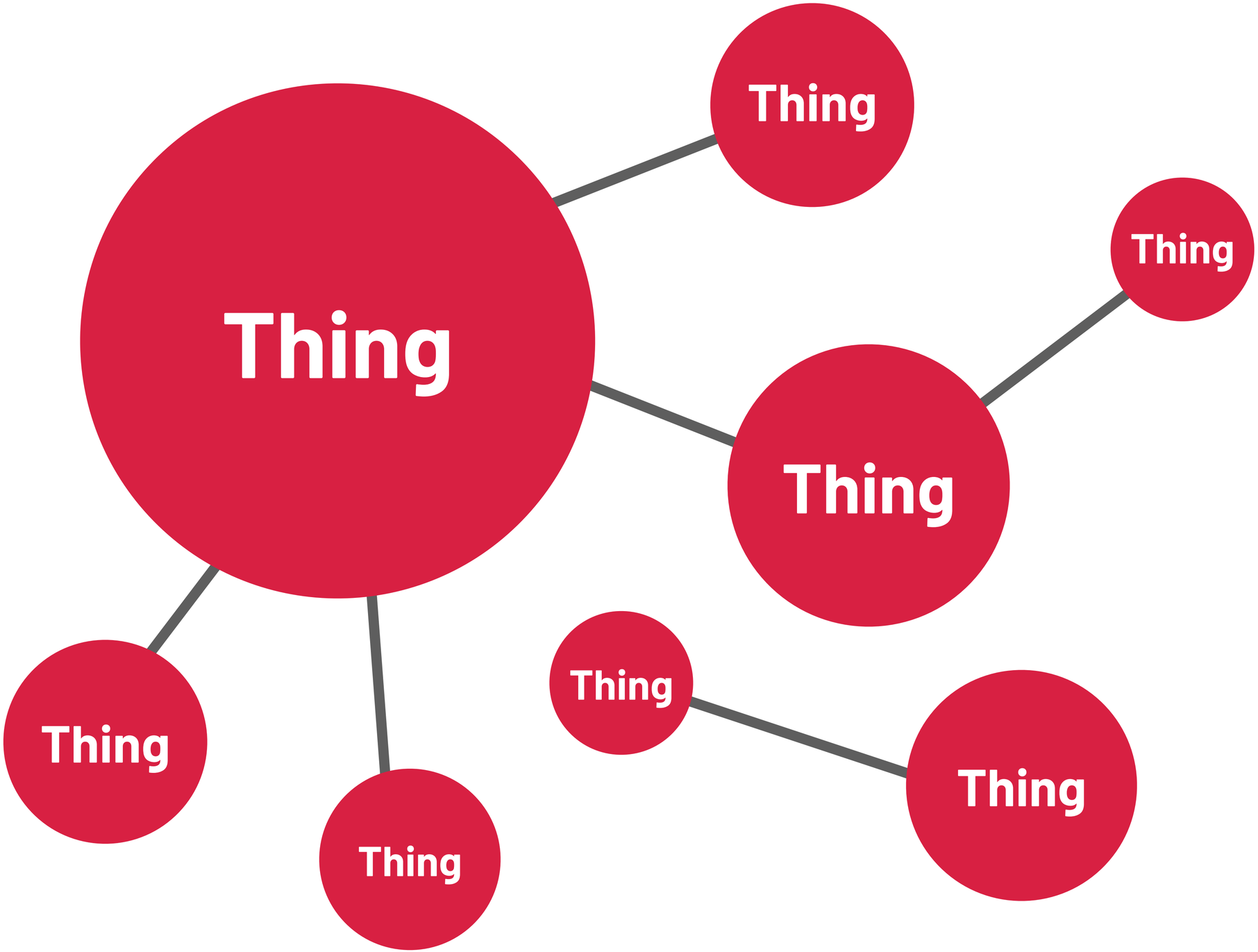
- American National Election Survey (ANES)
  - ➡ Whole networks v. Ego-networks
  - ➡ Regressing responsibly
- Topical co-occurance
  - ➡ Pruning and backboning

This workshop will reflect on (some of) the choices you will make during network modeling

PSA: Networks are *models* that imperfectly capture a complex world

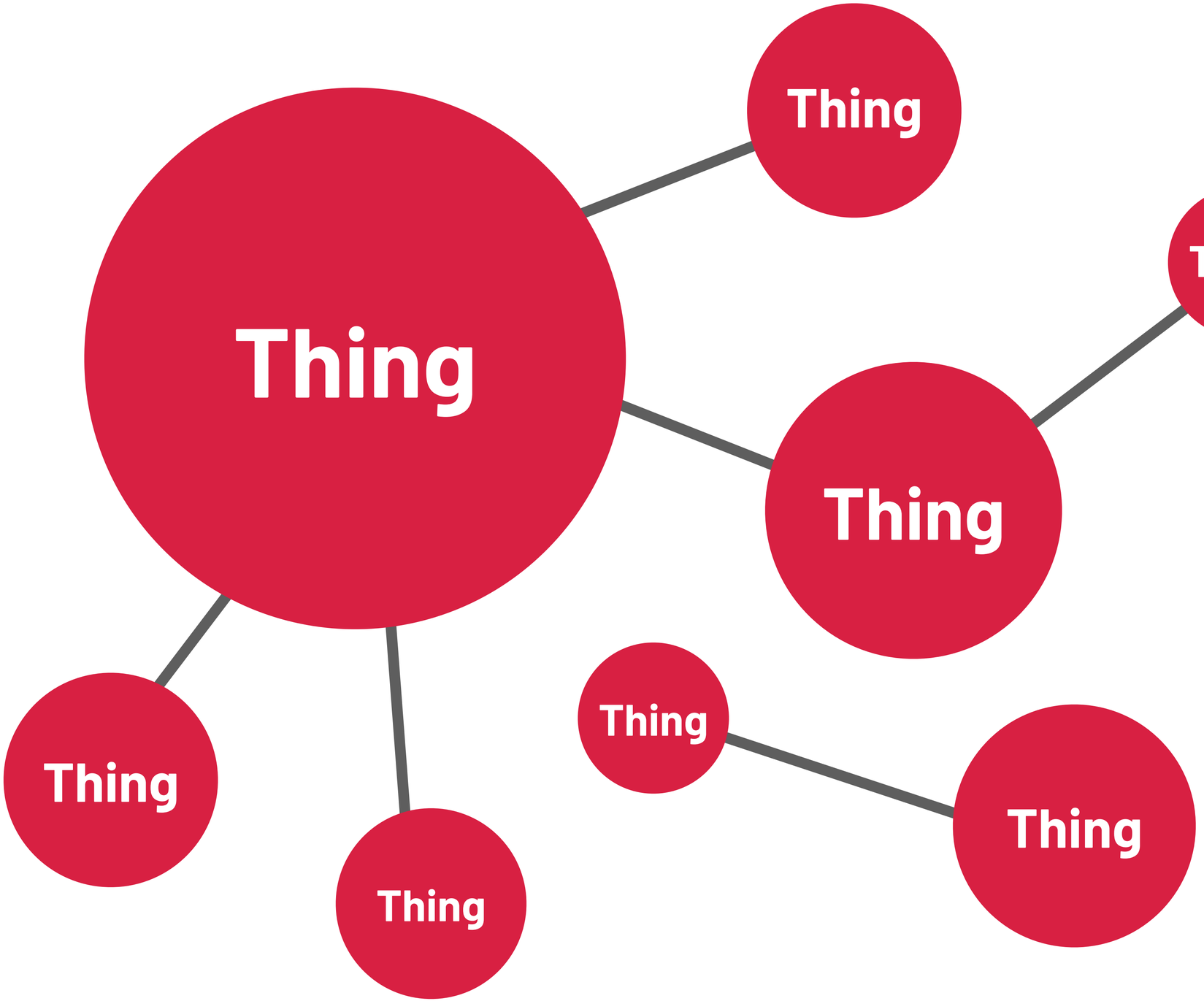


A network is a collection of things



A network is a collection of things

...Connected to other things



A network is a collection of things

...Connected to other things

We care about networks when the <u>connections</u> between things are as important as the things themselves!

What kinds of “things” might we care about?

- People
- Countries
- Organizations
- Dolphins
- Words
- Concepts
- …And more!

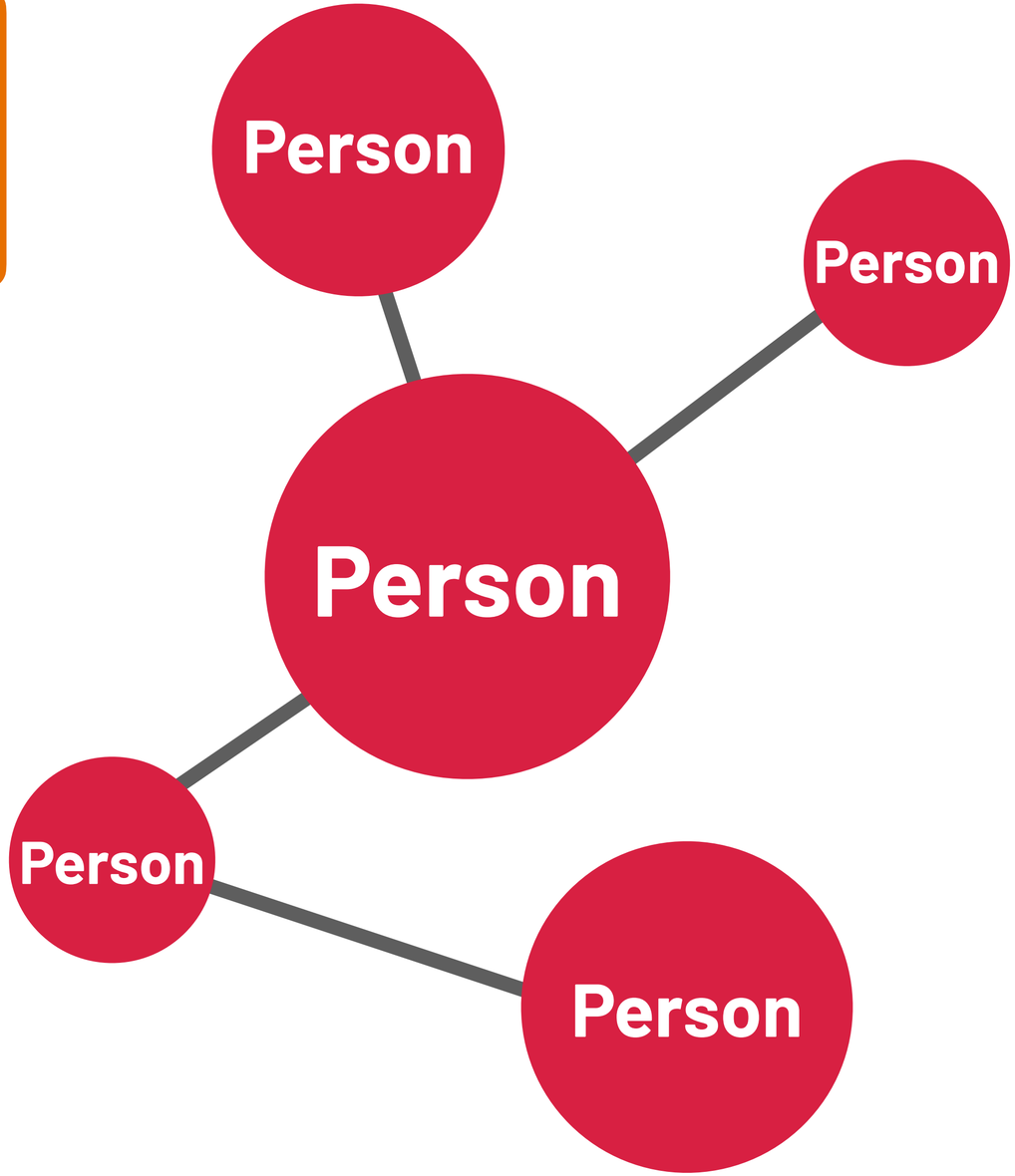
In the social sciences, we’re generally interested connections between in **social agents** OR their **artifacts**

What kinds of *connections* might we care about?

Imagine a “social network” where are nodes are individual people

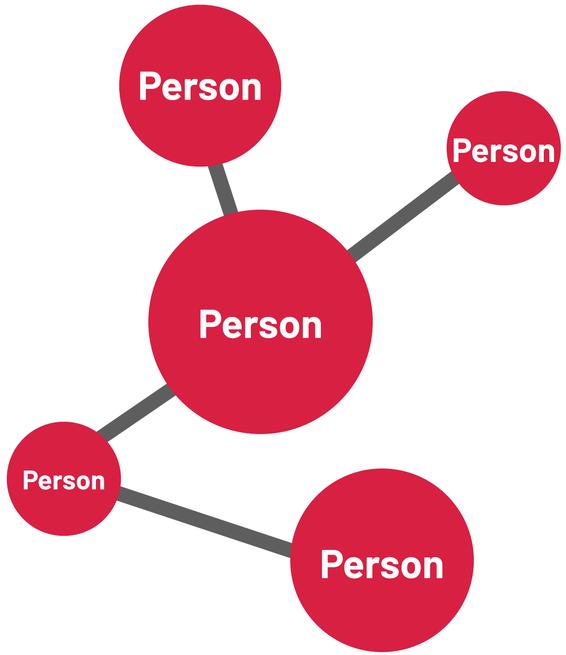
What might the connections between people represent?

- ➡ Sharing physical space together
- ➡ Giving/receiving advice
- ➡ Giving/asking for money
- ➡ People who would help you move
- ➡ People who follow you online
- ➡ People who repost your content
- ➡ ...and more!



What kinds of *connections* might we care about?

Imagine a “social network” where are nodes are individual people



What might the connections between people represent?

- ➡ Sharing physical space together
- ➡ Giving/receiving advice
- ➡ Giving/asking for money
- ➡ People who would help you move
- ➡ People who follow you online
- ➡ People who repost your content
- ➡ ...and more!

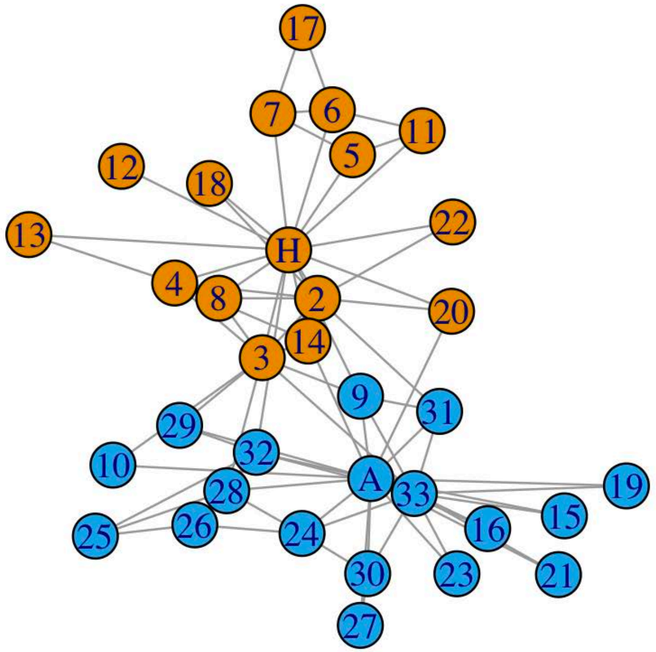
These connections tell us something meaningful about the relationship between people

Example 1:
Zachary's Karate Club

Example 1: Zachary's Karate Club

Always start by asking:

What are the **nodes** and what are the **edges**?



THE UNIVERSITY OF CHICAGO PRESS JOURNALS

An Information Flow Model for Conflict and Fission in Small Groups

Author(s): Wayne W. Zachary

Source: *Journal of Anthropological Research*, Winter, 1977, Vol. 33, No. 4 (Winter, 1977), pp. 452-473

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/3629752

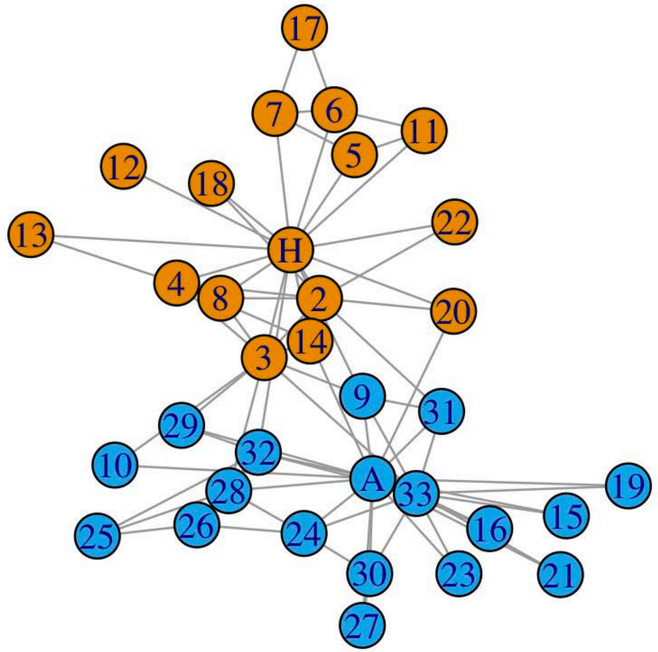
Example 1: Zachary's Karate Club

Nodes:

Members of a university-based Karate group

➡ Membership observed between 1970-1972 **Time-bound snapshot!**

➡ "Between 50-100" all together, only 34 included in paper! **Dropped isolates**



THE UNIVERSITY OF CHICAGO PRESS JOURNALS

An Information Flow Model for Conflict and Fission in Small Groups

Author(s): Wayne W. Zachary

Source: *Journal of Anthropological Research*, Winter, 1977, Vol. 33, No. 4 (Winter, 1977), pp. 452-473

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/3629752

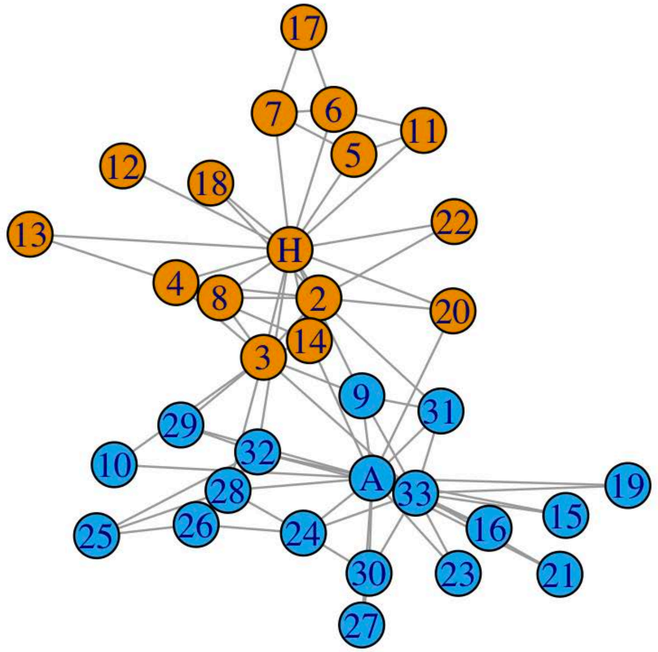
Example 1: Zachary's Karate Club

Edges:

"Friendship" between members

➡ Interacted in at least ONE of the following contexts:

(1) Association in and between academic classes at the university.
(2) Membership in Mr. Hi's private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
(3) Membership in Mr. Hi's private karate studio on the east side of the city, where many of his supporters worked out on weekends.
(4) Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their sutdents.
(5) Interaction at the university rathskeller, located in the same basement as the karate club's workout area.
(6) Interaction at a student-oriented bar located across the street from the university campus.
(7) Attendance at open karate tournaments held through the area at private karate studios.
(8) Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.



NOT treated as a weighted network (Though it could be!)

Example 1: Zachary's Karate Club

Nodes:

Members of a university-based Karate group

- ➡ Membership observed between 1970-1972
- ➡ "Between 50-100" all together, only 34 included in paper!

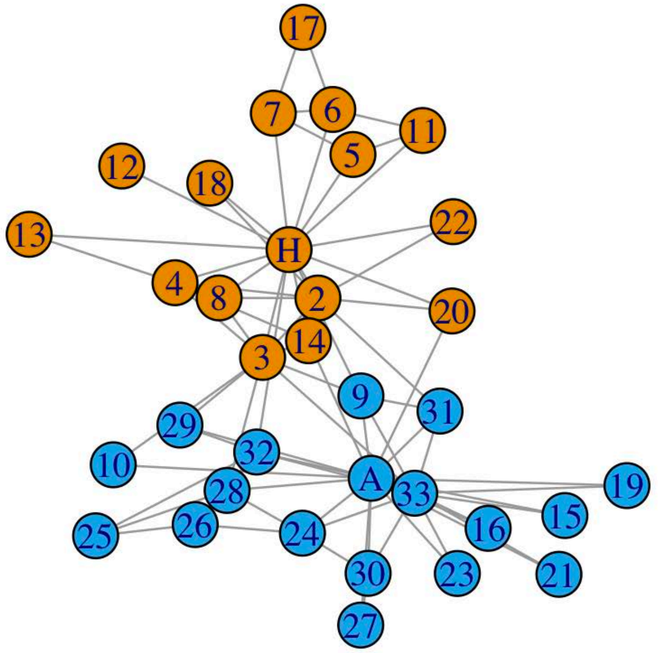
Edges:

- ➡ Interacted in at least 1 of 8 relevant contexts

Are these choices justifiable?

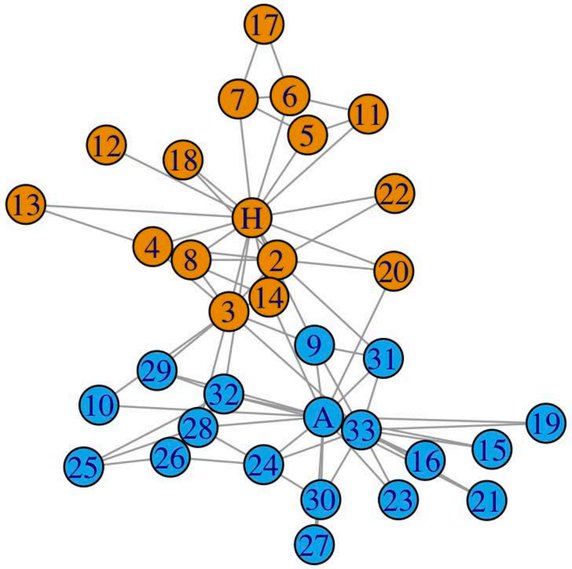
Sure? Could definately "Reviewer 2" this, but it's not *un*reasonable

Remember: EVERY model has limitations. Networks are models



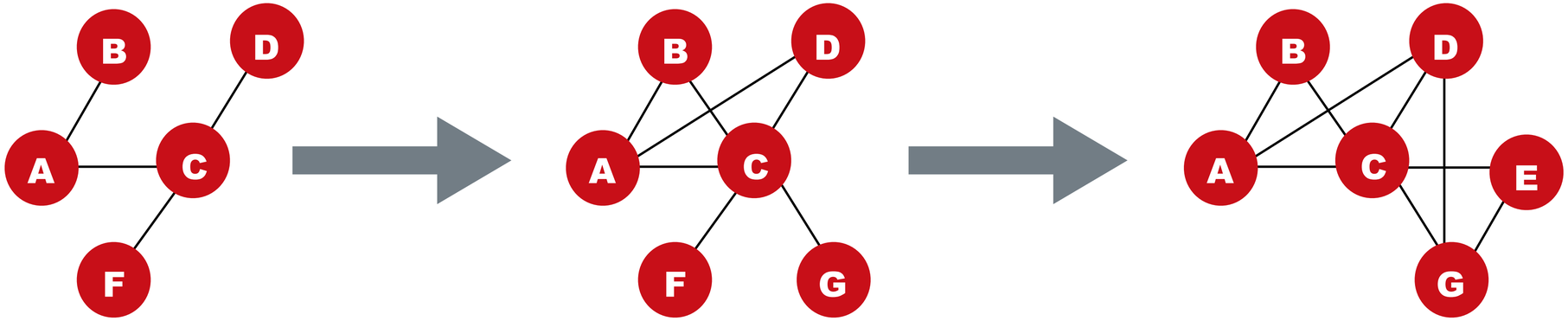
Choices: Temporal

Nodes defined as "people who were members at any point during a 3-year period of time."



What are some other possible choices?

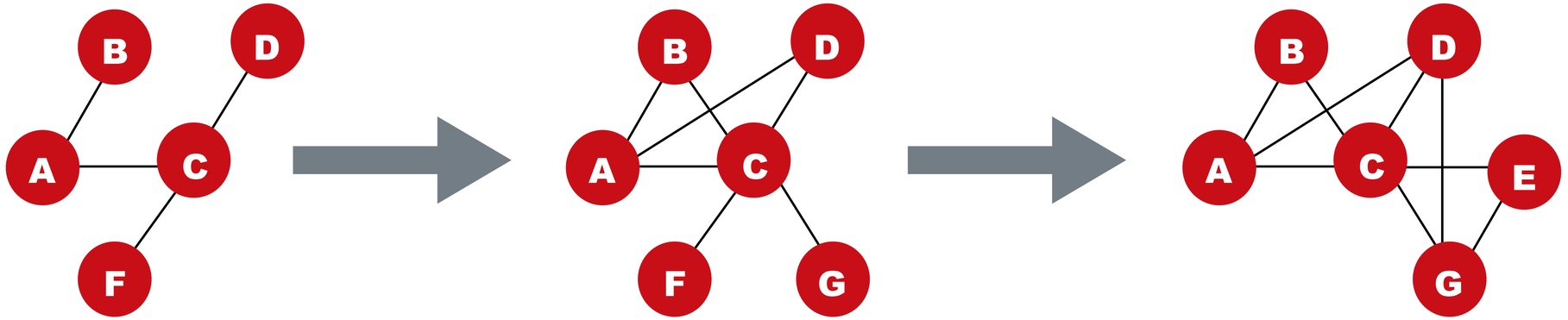
- ➡ Study a longer/shorter period of time
- ➡ Study temporal networks (ie, snapshots over time)



1970 1971 1972

Choices: Temporal

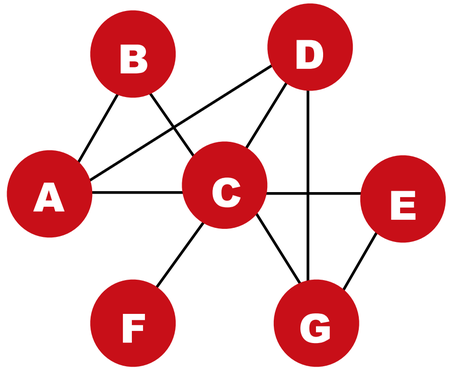
When is this:



1970 1971 1972

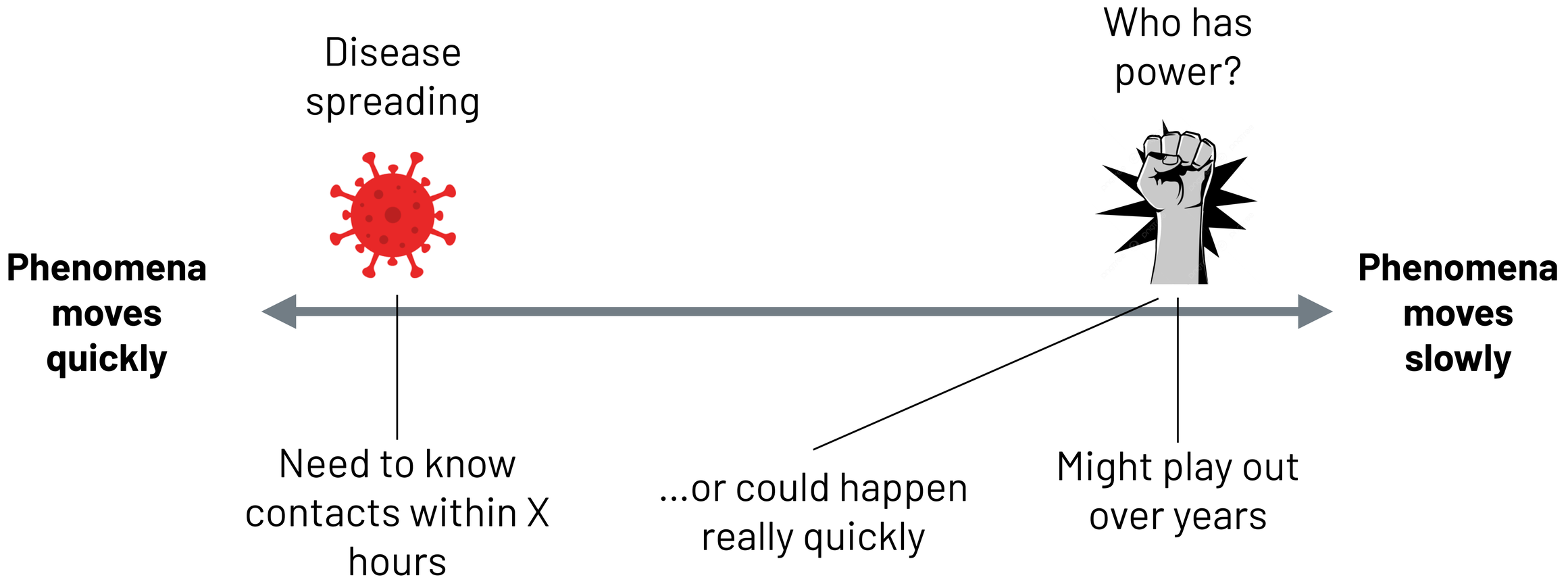
Better (more useful than) this:

1970-1972



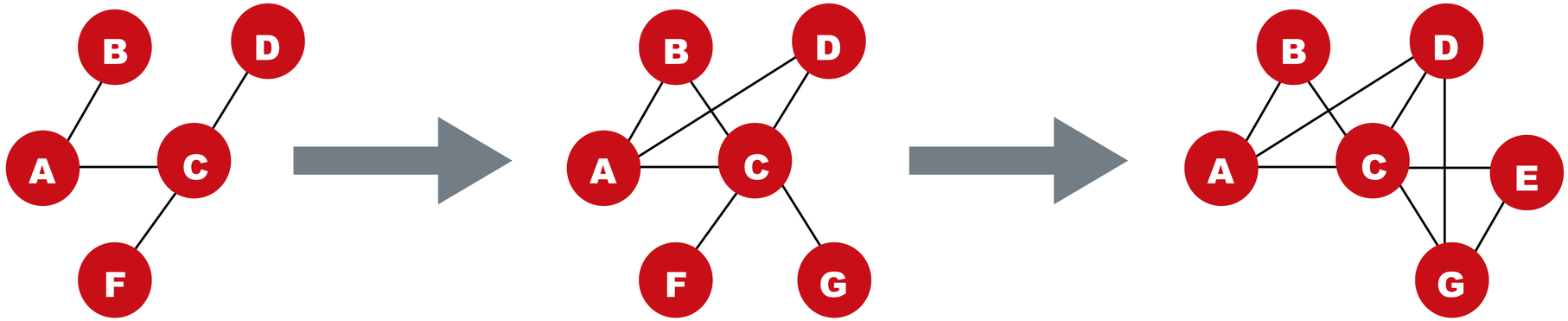
Temporal properties of your network should match the temporal tendencies of your phenoma of interest

Choices: Temporal



Choices: Temporal

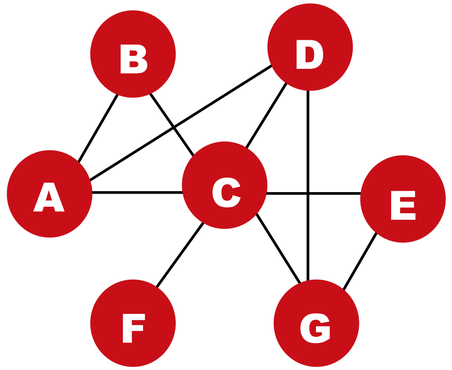
When is this:



1970 1971 1972

Better (more useful than) this:

1970-1972

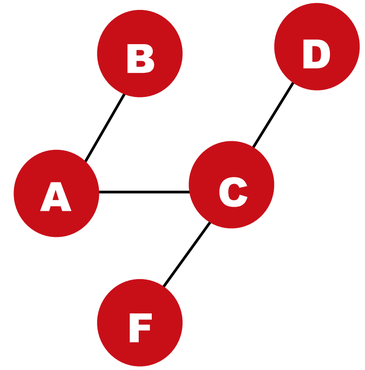


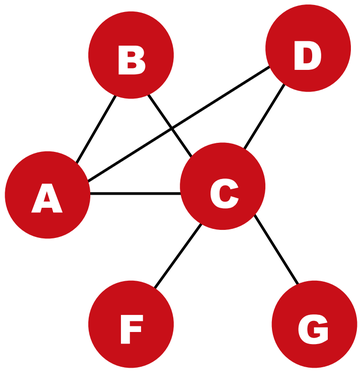
Ask yourself:

- How quickly does the phenomena happen?
- How quickly does the network change?
- Do you need to know how the network changes within a time frame?

Reflection: Your Network

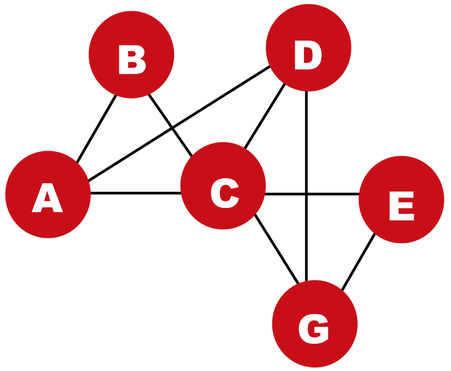
- What time scale is appropriate for a network you are studying?
- Do you need to compare your network over time or is aggregation okay for your research question?





Remember:

- All networks are aggregated to some extent (you're not recording every microsecond!)
- You want a model that is as simple as possible while capturing <u>needed</u> complexity



Example 2:
Ethnic Networks

Example 2: Ethnic Networks

Research question:
How does information spread in a homogenous v. ethnically diverse community?

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

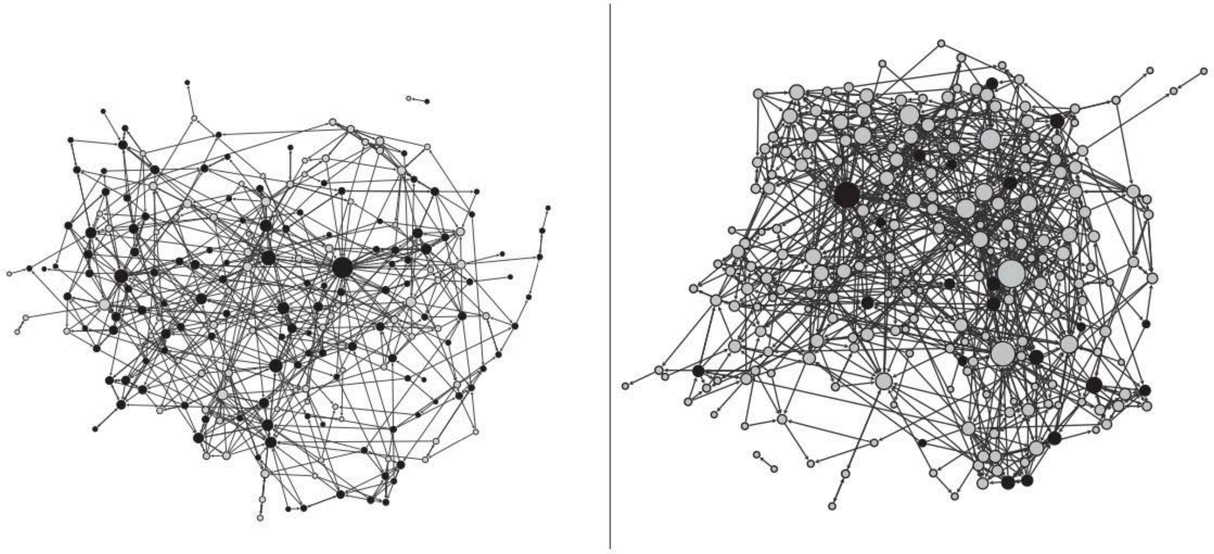
Ethnic Networks

Jennifer M. Larson New York University
Janet I. Lewis U.S. Naval Academy

Abstract: *Active research on a wide range of political contexts centers on ethnicity's role in collective action. Many theories posit that information flows more easily in ethnically homogeneous areas, facilitating collective action, because social networks among coethnics are denser. Although this characterization is ubiquitous, little empirical work assesses it. Through a novel field experiment in a matched pair of villages in rural Uganda, this article directly examines word-of-mouth information spread and its relationship to ethnic diversity and networks. As expected, information spread more widely in the homogeneous village. However, unexpectedly, the more diverse village's network is significantly denser. Using unusually detailed network data, we offer an explanation for why network density may hamper information dissemination in heterogeneous areas, showing why even slight hesitation to share information with people from other groups can have large aggregate effects.*

- Seeded information: "In three days an event would be held at which all adults in attendance would receive a valuable block of soap in exchange for taking a survey."
- Collected network information at event
- After event, conducted door-to-door interviews of households collecting additional network information

FIGURE 1 Abalang (left) and Mugana (right) Social Networks, Visualized with Force-Directed Layout



Note: Those who heard are shaded black; node size is proportional to degree.

Example 2: Ethnic Networks

Nodes:

Residents in two Ugandan villages

➡ Single network constructed for each village, collected over about a week

Time-bound snapshot

➡ "At least one individual in all households"

(Close to) Whole network data

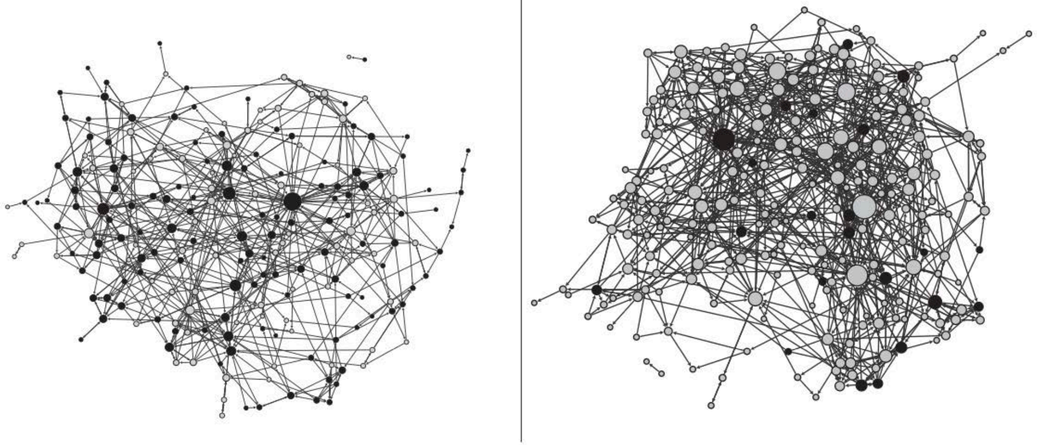
AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Ethnic Networks

Jennifer M. Larson New York University
Janet I. Lewis U.S. Naval Academy

Abstract: *Active research on a wide range of political contexts centers on ethnicity's role in collective action. Many theories posit that information flows more easily in ethnically homogeneous areas, facilitating collective action, because social networks among coethnics are denser. Although this characterization is ubiquitous, little empirical work assesses it. Through a novel field experiment in a matched pair of villages in rural Uganda, this article directly examines word-of-mouth information spread and its relationship to ethnic diversity and networks. As expected, information spread more widely in the homogeneous village. However, unexpectedly, the more diverse village's network is significantly denser. Using unusually detailed network data, we offer an explanation for why network density may hamper information dissemination in heterogeneous areas, showing why even slight hesitation to share information with people from other groups can have large aggregate effects.*

FIGURE 1 Abalang (left) and Mugana (right) Social Networks, Visualized with Force-Directed Layout



Note: Those who heard are shaded black; node size is proportional to degree.

Example 2: Ethnic Networks

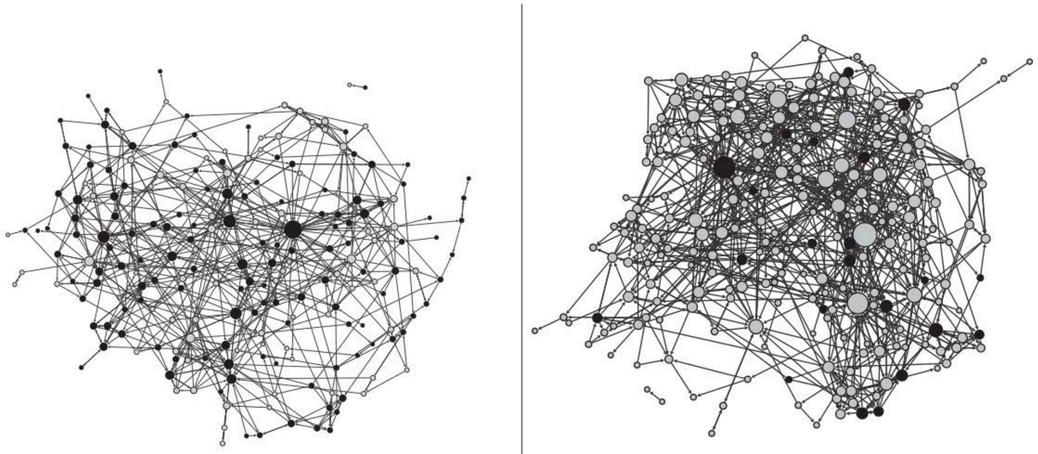
Edges

Respondents asked to name up to 5 people (for each category), whom they:

1. Visit
2. Discuss politics
3. Discuss religion
4. Spend time
5. Share meals
6. Share secrets
7. Speak on the phone

Final network:
Union of sub-networks,
edges treated as **directed**

FIGURE 1 Abalang (left) and Mugana (right) Social Networks, Visualized with Force-Directed Layout



Note: Those who heard are shaded black; node size is proportional to degree.

Example 2: Ethnic Networks

Nodes:

Residents in two Ugandan villages, nearly all residents captured as nodes

Edges:

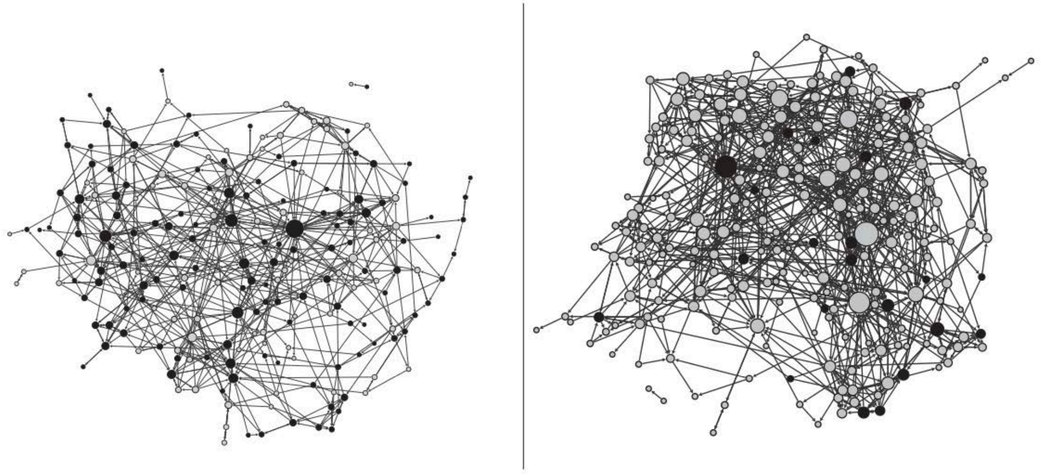
- ➡ Included in at least 1 of 7 name generators
- ➡ Directed
- ➡ Unweighted

Are these choices justifiable?

Sure? Could definately "Reviewer 2" this, but it's not *un*reasonable

Remember: EVERY model has limitations. Networks are models

FIGURE 1 **Abalang (left) and Mugana (right) Social Networks, Visualized with Force-Directed Layout**



Note: Those who heard are shaded black; node size is proportional to degree.

Choices: Measuring Edges

Aggregating networks

Both these examples used multiple prompts to construct a single, unweighted network edge

Why?

- Each prompt measures a slightly different relationship, BUT
- Each relationship conceptually tied to the same phenoma of interest
- Single measure will be noisy, multiple measures more robust

Note: this is very similar to survey batteries which are collapsed into a single measure

Edges

People whom they:

1. Visit
2. Discuss politics
3. Discuss religion
4. Spend time
5. Share meals
6. Share secrets
7. Speak on the phone

(1) Association in and between academic classes at the university.
(2) Membership in Mr. Hi's private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
(3) Membership in Mr. Hi's private karate studio on the east side of the city, where many of his supporters worked out on weekends.
(4) Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their sutdents.
(5) Interaction at the university rathskeller, located in the same basement as the karate club's workout area.
(6) Interaction at a student-oriented bar located across the street from the university campus.
(7) Attendance at open karate tournaments held through the area at private karate studios.
(8) Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.

Choices: Measuring Edges

Zachary's Karate Club

Who are you likely to side with during a power struggle between club leaders?

(1) Association in and between academic classes at the university.
(2) Membership in Mr. Hi's private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
(3) Membership in Mr. Hi's private karate studio on the east side of the city, where many of his supporters worked out on weekends.
(4) Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their sutdents.
(5) Interaction at the university rathskeller, located in the same basement as the karate club's workout area.
(6) Interaction at a student-oriented bar located across the street from the university campus.
(7) Attendance at open karate tournaments held through the area at private karate studios.
(8) Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.

Ethnic Networks

Who are you likely to share information with?

Edges

People whom they:

1. Visit
2. Discuss politics
3. Discuss religion
4. Spend time
5. Share meals
6. Share secrets
7. Speak on the phone

Choices: Measuring Edges

Zachary's Karate Club

Who are you likely to side with during a power struggle between club leaders?

(1) Association in and between academic classes at the university.
(2) Membership in Mr. Hi's private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
(3) Membership in Mr. Hi's private karate studio on the east side of the city, where many of his supporters worked out on weekends.
(4) Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their sutdents.
(5) Interaction at the university rathskeller, located in the same basement as the karate club's workout area.
(6) Interaction at a student-oriented bar located across the street from the university campus.
(7) Attendance at open karate tournaments held through the area at private karate studios.
(8) Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.

Ethnic Networks

Who are you likely to share information with?

Edges

People whom they:

1. Visit
2. Discuss politics
3. Discuss religion
4. Spend time
5. Share meals
6. Share secrets
7. Speak on the phone

Both networks *could* be treated as **weighted**, but that assumes:

- Each measure isn't too noisy
- The the # of connections significantly influences outcomes

SIMPLE MODELS ARE OFTEN BETTER

Network Similarity

It typically makes sense to aggregate edge measurements into a single, unweighted network if sub-networks are relatively “similar”

Common measures of similarity include:

- **Correlation:** Do nodes tend to share the same neighbors in both networks?
- **Distance:** How “far apart” are the networks in some similarity space?

Problem: With only two observations (ie, two networks), how can you tell if networks are *significantly* similar?

(1) Association in and between academic classes at the university.
(2) Membership in Mr. Hi's private karate studio on the east side of the city where Mr. Hi taught nights as a part-time instructor.
(3) Membership in Mr. Hi's private karate studio on the east side of the city, where many of his supporters worked out on weekends.
(4) Student teaching at the east-side karate studio referred to in (2). This is different from (2) in that student teachers interacted with each other, but were prohibited from interacting with their sutdents.
(5) Interaction at the university rathskeller, located in the same basement as the karate club's workout area.
(6) Interaction at a student-oriented bar located across the street from the university campus.
(7) Attendance at open karate tournaments held through the area at private karate studios.
(8) Attendance at intercollegiate karate tournaments held at local universities. Since both open and intercollegiate tournaments were held on Saturdays, attendance at both was impossible.

Edges

People whom they:

1. Visit
2. Discuss politics
3. Discuss religion
4. Spend time
5. Share meals
6. Share secrets
7. Speak on the phone

Network Inference

Landscape of Network Inference

Generative models

- Create a network using a defined set of rules
- Example: “Nodes connect with probability *p*”
- Useful when you don’t have empirical data (often used in Agent Based Models)
- Examples: ER, BA, Watts-Strogatz

Permutation Tests

- Shuffle your data according to a set of rules
- Example: “Randomly re-wire edges”
- Useful for testing a network-level measure while controlling for node inter-dependence
- Examples: CUG, QAP

Network Regression (!)

- Useful if your DEPENDENT VARIABLE (DV) is NOT a network feature
- NOT appropriate for estimating a network/ network feature as an output variable
- Be cautious: Network IVs are correlated!

ERGMs and Latent Space Models

- Estimate probability of observing your network as a function of specified network features
- Estimate probability of observing a tie while controlling for network features
- Generate coefficients estimating the effects of specified network features

Permutation Tests

Permutation Tests

- Shuffle your data according to a set of rules
- Example: “Randomly re-wire edges”
- Useful when you want to test a network-level measure while controlling for node inter-dependence

Types of Permutation Tests

- Configuration model
- Conditional Uniform Graph (CUG)
- Quadratic Assignment Procedure (QAP)

	A	B	C
A	0	1	1
B	1	0	0
C	0	1	0

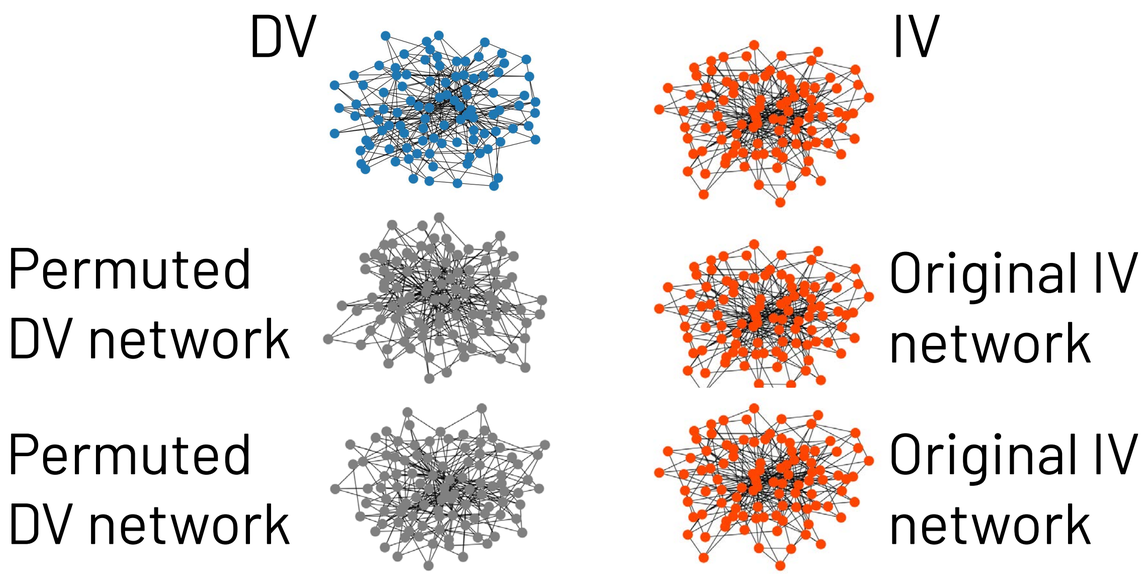
Input network

	B	C	A
B	0	0	1
C	1	0	0
A	1	1	0

A shuffled network

Quadratic Assignment Procedure (QAP)

Test the observed <u>similarity between networks</u> against a null model of simulated networks conditioning on researcher-defined network properties



You might see a permutation test as the main statistical test in a paper, but it's used more commonly as a suplementry analysis.

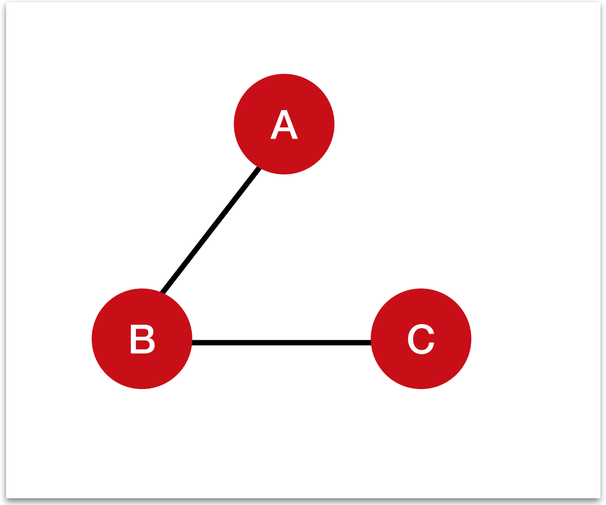
For example, a robustness check to justify merging "edge" measures.

Compare similarities of networks across permutations of a DV network

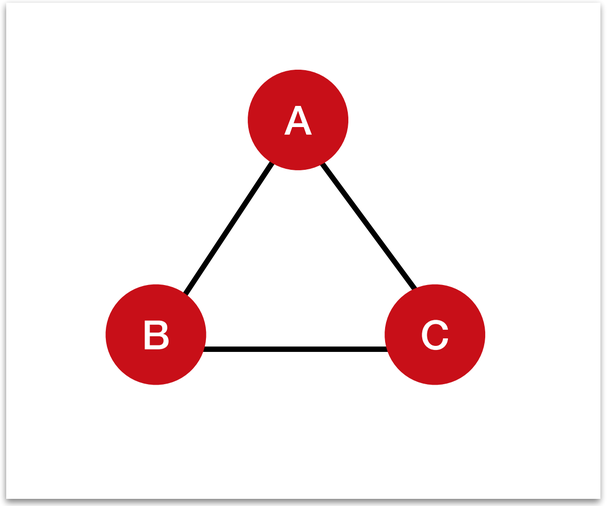
QAP — Simple Example

Imagine we observe two networks on the same set of nodes

Graph 1



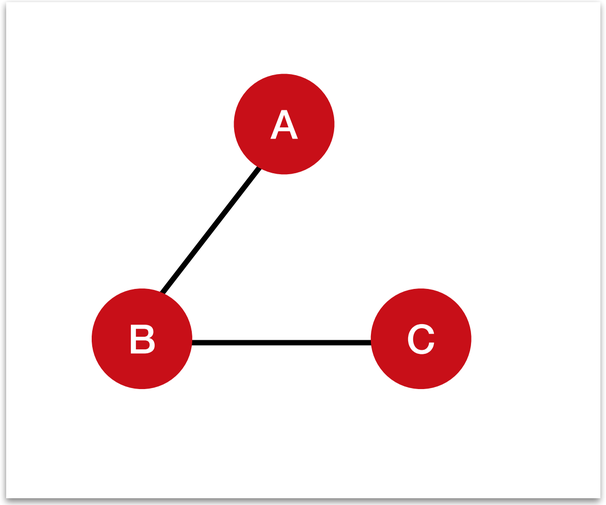
Graph 2



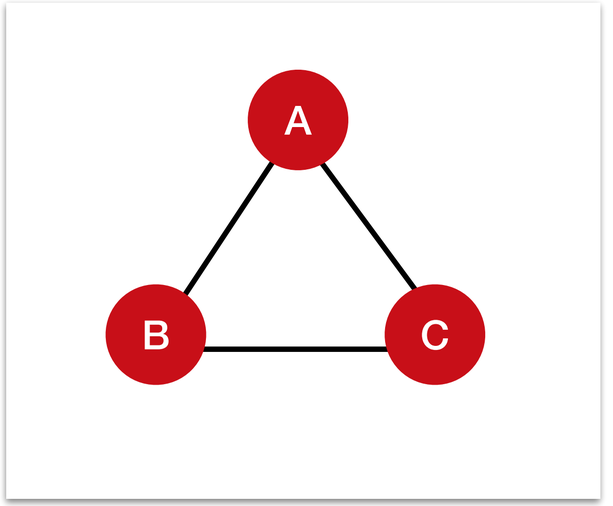
QAP – Simple Example

Imagine we observe two networks on the same set of nodes

Graph 1



Graph 2



Adjacency matrix representation

	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

QAP – Simple Example

We can use correlation to measure the similarity between the two networks:

cor(

	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0

,

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

) = .63

QAP – Simple Example

We can use correlation to measure the similarity between the two networks:

cor(

	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0

,

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

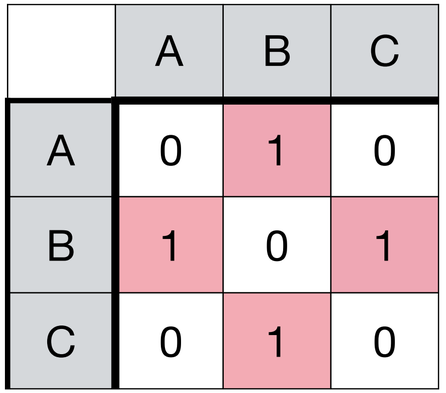
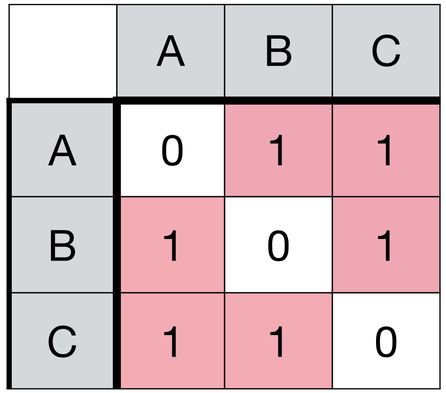
) = .63

One way to calculate the correlation between two matrices is to flatten the matrix into vector representations, ie:

= cor(0 1 0 1 0 1 0 1 0 , 0 1 1 1 0 1 1 1 0) = .63

QAP – Simple Example

These graphs are fairly similar

cor( , ) = .63

	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0

Graph 1

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

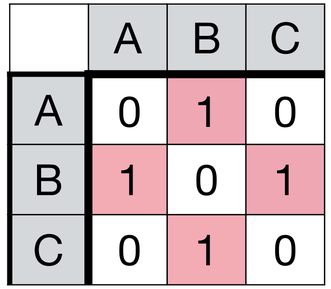
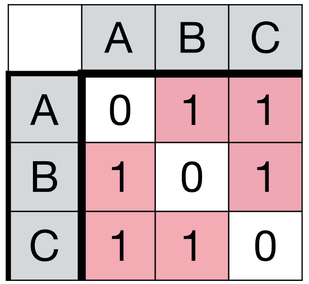
Graph 2

Is that what we would expect at random?

To test this, we'll shuffle the rows and columns of Graph 1 and test the correlation of the permuted graph against Graph 2

QAP – Simple Example

Observed relationship

cor( , ) = .63

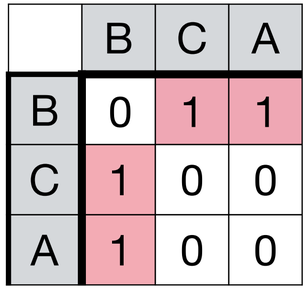
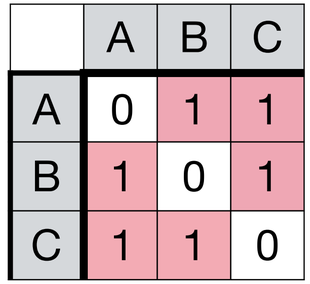
	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0

Graph 1

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

Graph 2

Permuted relationship

cor( , ) = ??

	B	C	A
B	0	1	1
C	1	0	0
A	1	0	0

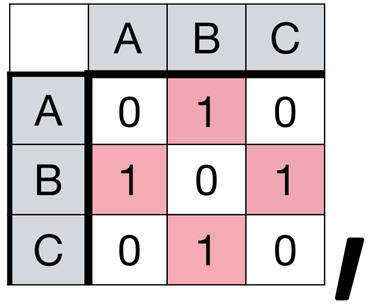
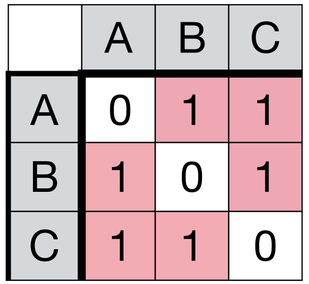
Graph 1 (Permuted)

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

Graph 2

QAP – Simple Example

Observed relationship

cor( , ) = .63

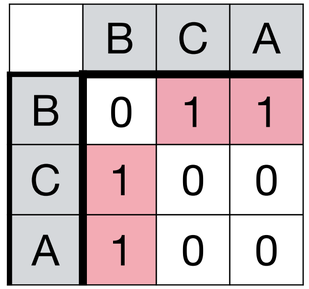
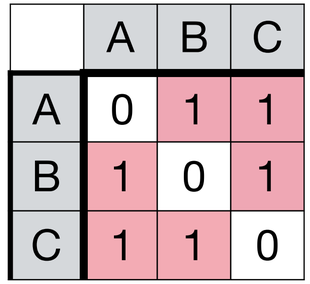
Graph 1

	A	B	C
A	0	1	0
B	1	0	1
C	0	1	0

Graph 2

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

Permuted relationship

cor( , ) = ?

Graph 1

	B	C	A
B	0	1	1
C	1	0	0
A	1	0	0

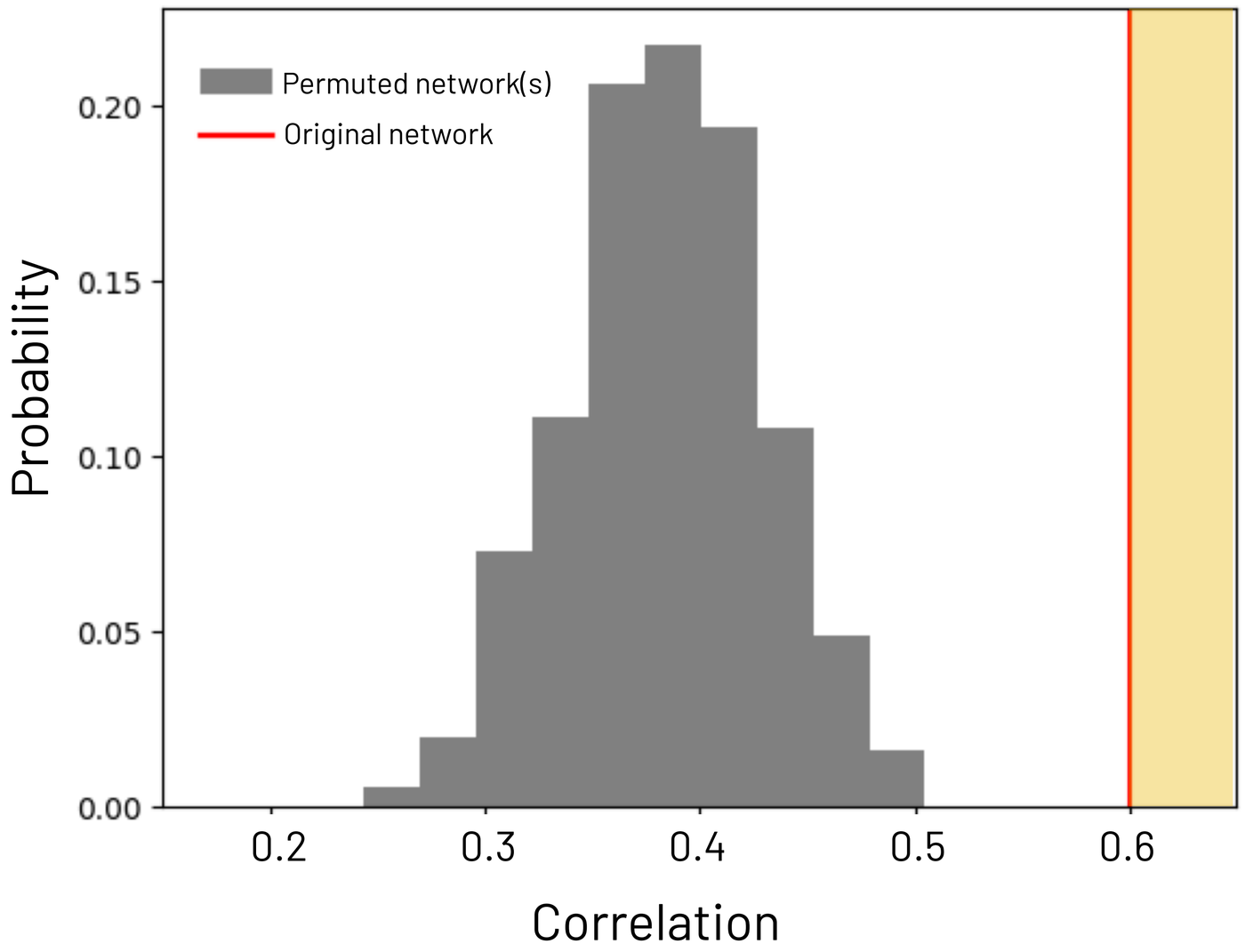
Graph 2

	A	B	C
A	0	1	1
B	1	0	1
C	1	1	0

Only permute Graph 1

Do this a large number of times, create a null distribution

Simulating a Null Distribution



The **observed** correlation between our two networks is corr = 0.63

When we permute one network, there are **0** observations with a correlation >= 0.63

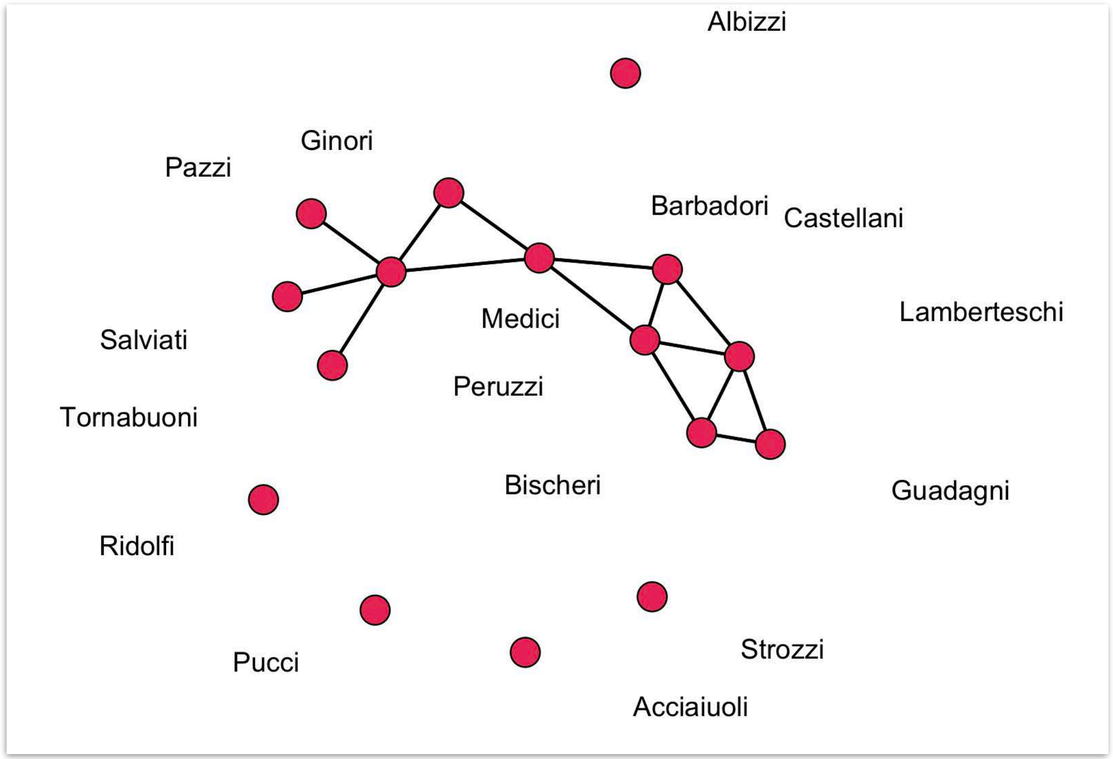
p(x >= obs) = 0

p-value!

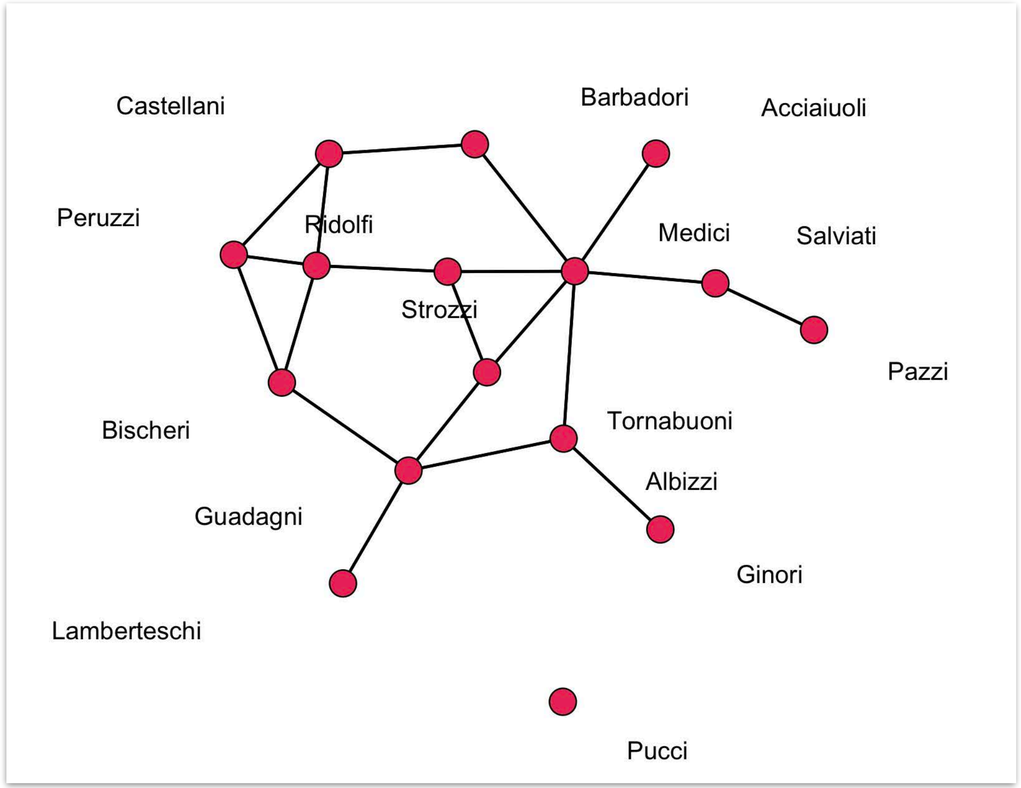
QAP – Empirical Example

```
library(statnet)
data(florentine)
```

This dataset includes two networks: one on business connections (flobusiness) and one with marriage connections (flomarriage)



flobusiness



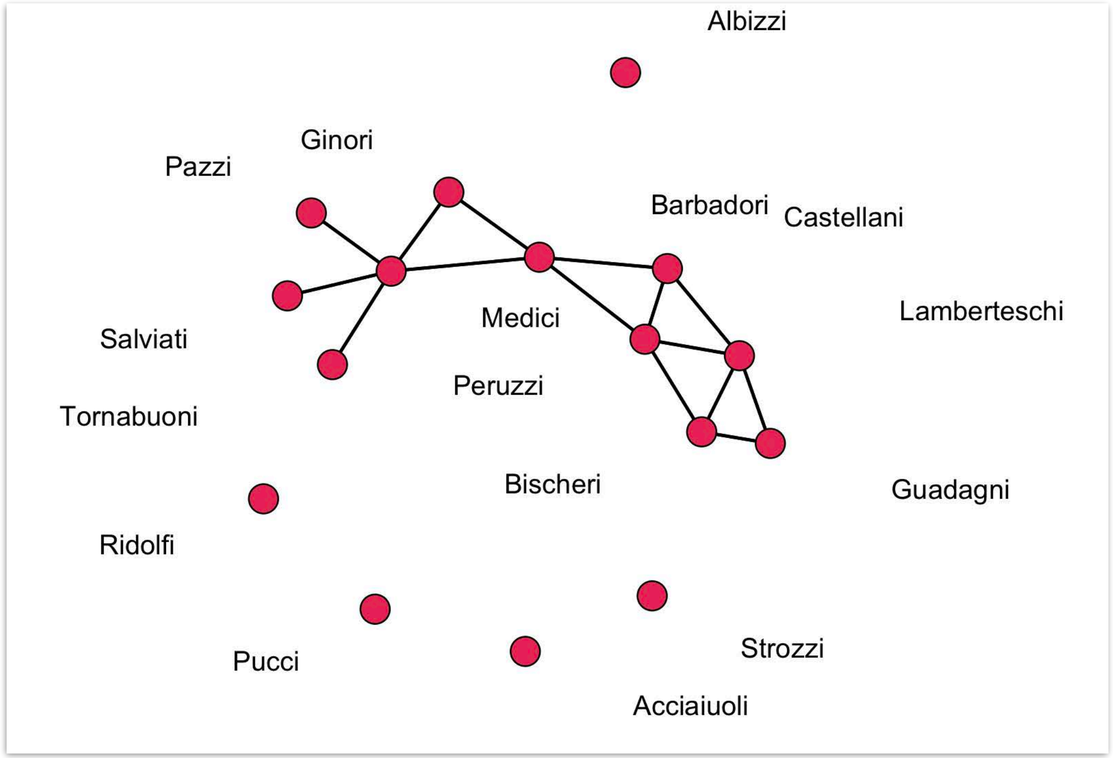
flomarriage

QAP – Empirical Example

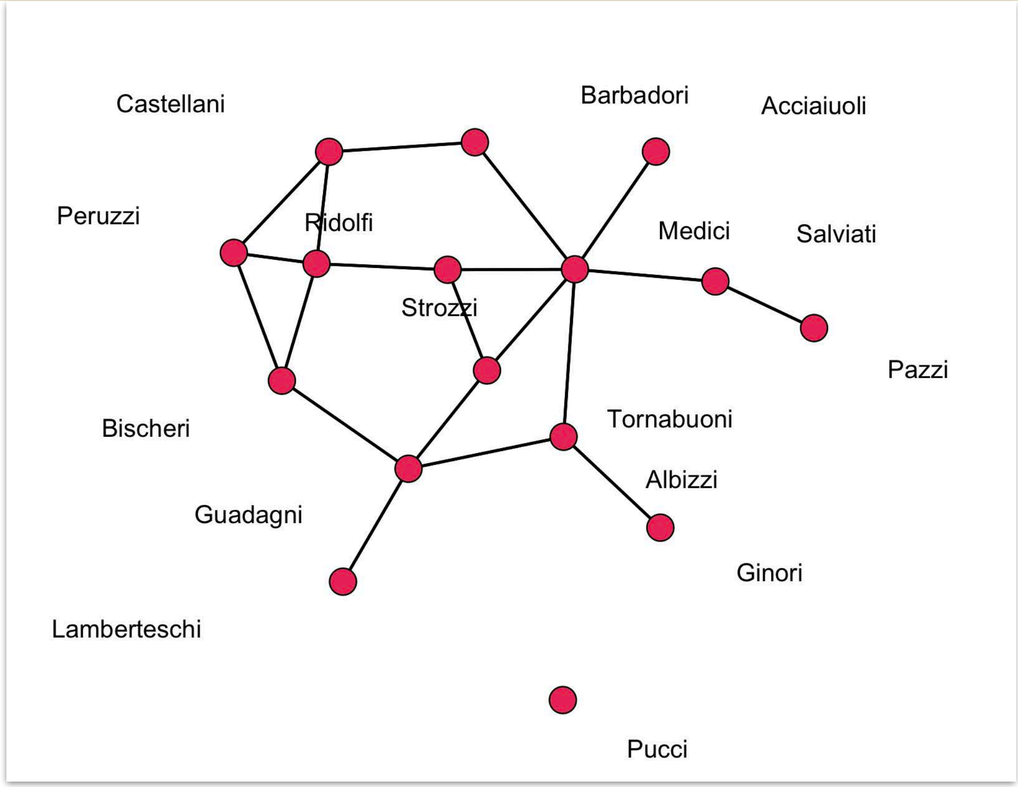
```
gcor(flobusiness, flomarriage)
```

```
> 0.37
```

Is this what we would expect at random given the underlying structure of the data?



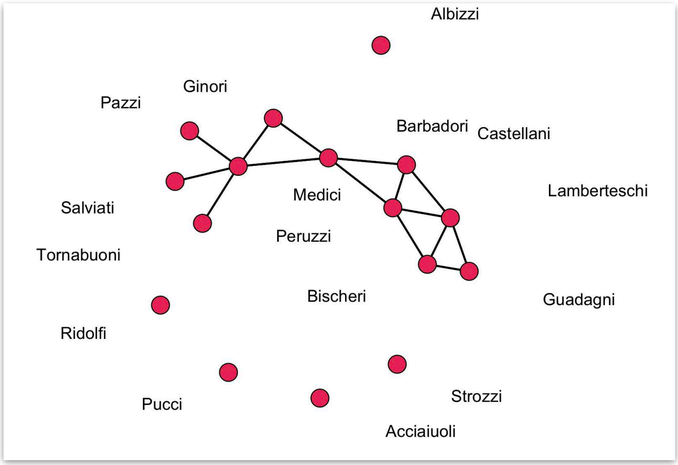
flobusiness



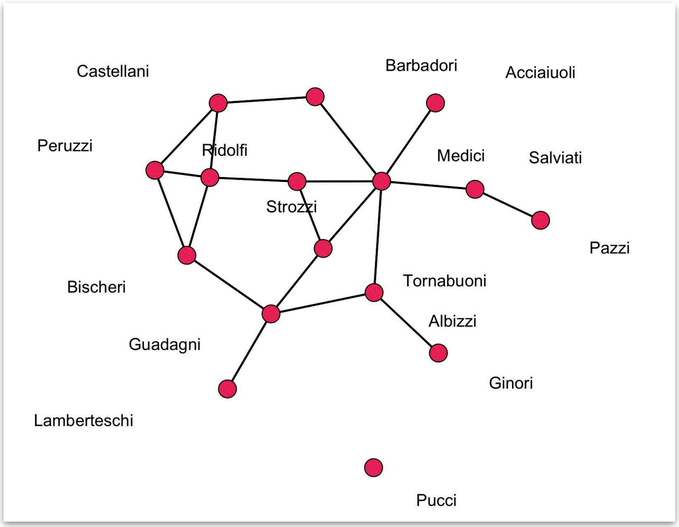
flomarriage

QAP – Empirical Example

```
qt <- qaptest(list(flobusiness,flomarriage),
                gcor,
                g1=1, g2=2
                )
```



flobusiness



flomarriage

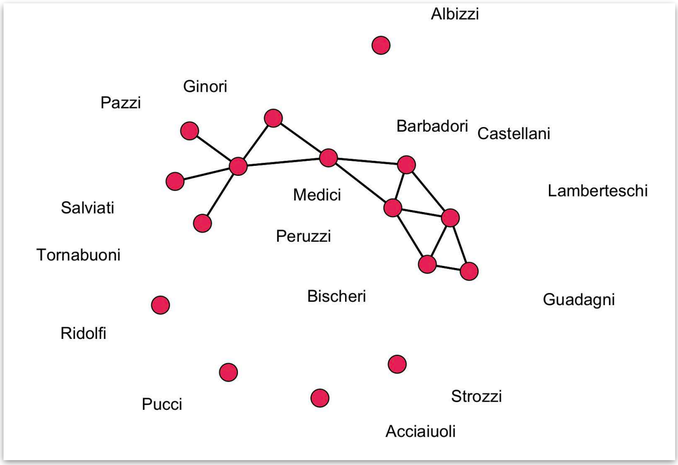
QAP – Empirical Example

List of networks to compare

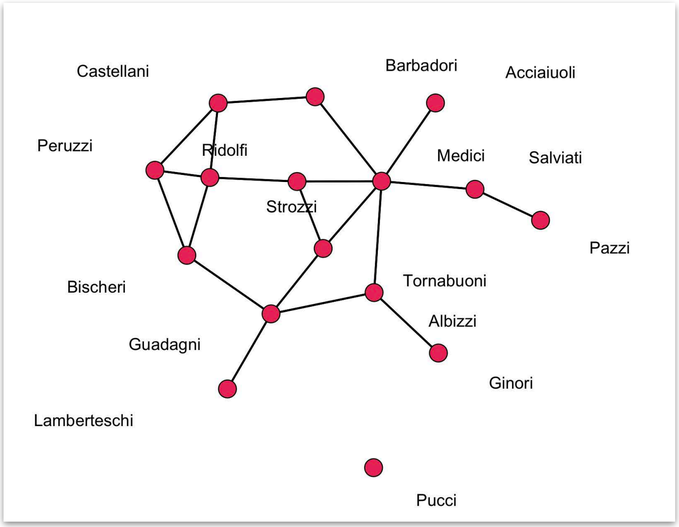
```
qt <- qaptest(list(flobusiness,flomarriage),
              gcor,
              g1=1, g2=2
             )
```

Function for comparison metric

g1: the index of the network to be permuted
g2: the index of the network to compare (not permuted)



flobusiness



flomarriage

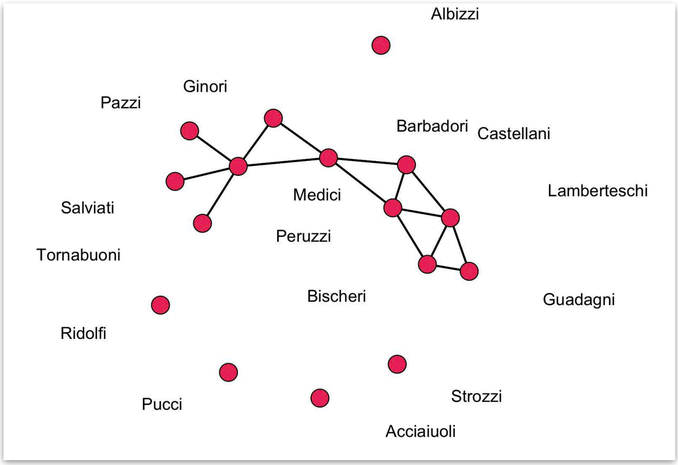
QAP – Empirical Example

List of networks to compare

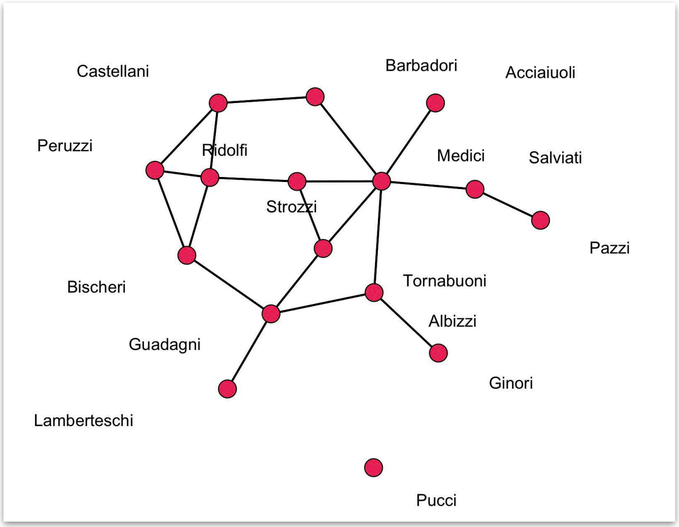
```
qt <- qaptest(list(flobusiness,flomarriage),
              gcor,
              g1=1, g2=2
              )
```

```
list(flobusiness,flomarriage)
```

Index	1	2
Item	flobusiness	flomarriage



flobusiness



flomarriage

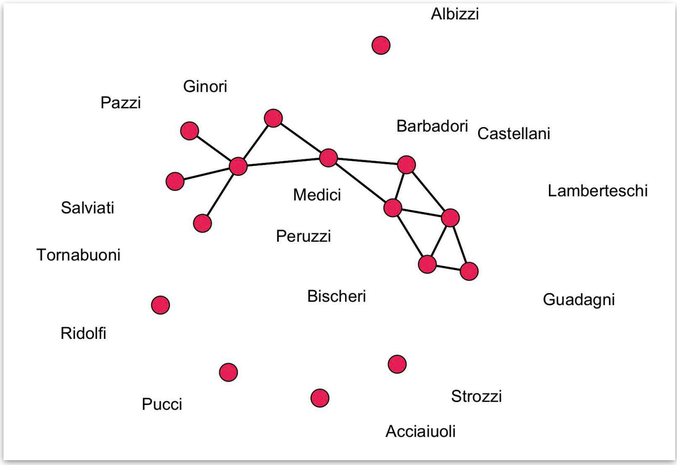
QAP – Empirical Example

```
qt <- qaptest(list(flobusiness,flomarriage),
                    gcor, g1=1, g2=2)
summary(qt)
```

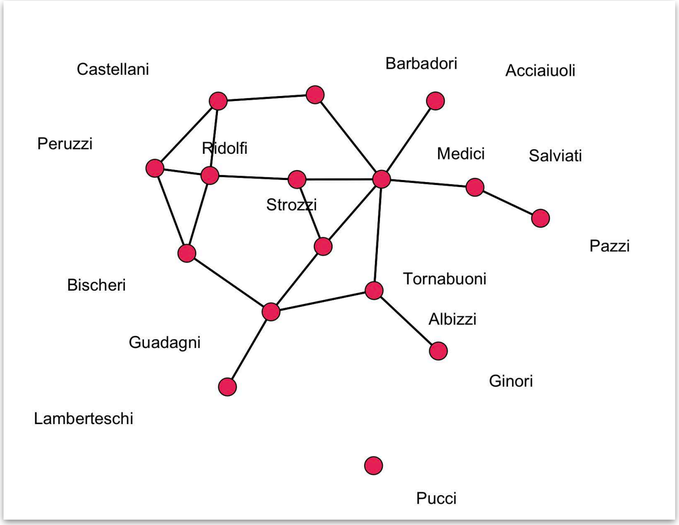
QAP Test Results

Estimated p-values:
p(f(perm) >= f(d)): 0
p(f(perm) <= f(d)): 1

Test Diagnostics:
Test Value (f(d)): 0.3718679
Replications: 1000



flobusiness



flomarriage

QAP – Empirical Example

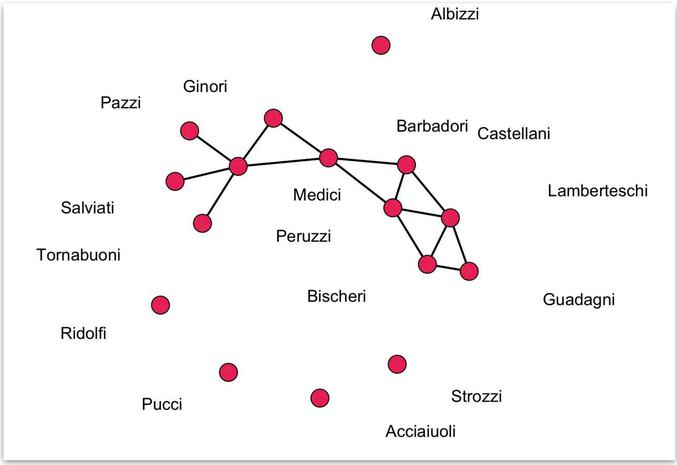
```
qt <- qaptest(list(flobusiness,flomarriage),
                    gcor, g1=1, g2=2)
summary(qt)
```

QAP Test Results

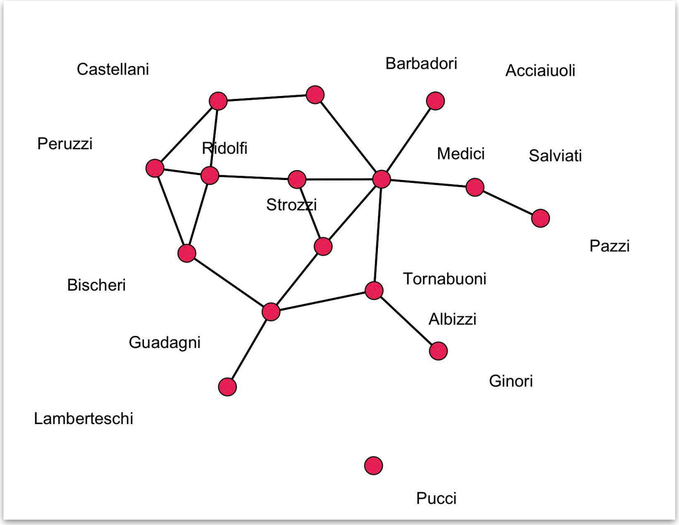
Estimated p-values:
p(f(perm) >= f(d)): 0
p(f(perm) <= f(d)): 1

Test Diagnostics:
Test Value (f(d)): 0.3718679
Replications: 1000

Observed correlation between networks



flobusiness



flomarriage

QAP – Empirical Example

```
qt <- qaptest(list(flobusiness,flomarriage),
                    gcor, g1=1, g2=2)
summary(qt)
```

QAP Test Results

Estimated p-values:

p(f(perm) >= f(d)): 0

p(f(perm) <= f(d)): 1

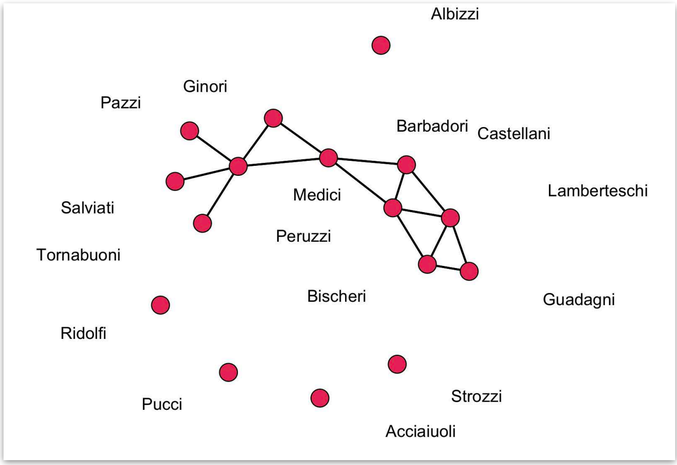
Test Diagnostics:

Test Value (f(d)): 0.3718679

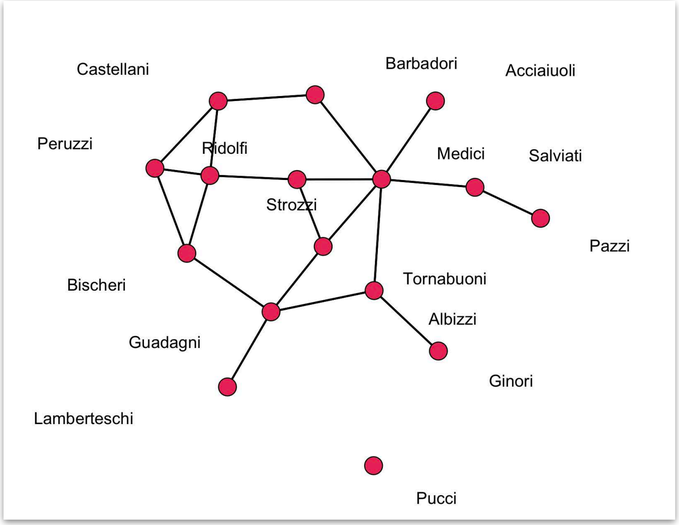
Replications: 1000

Very unlikely to see a correlation this high by random chance

Observed correlation between networks



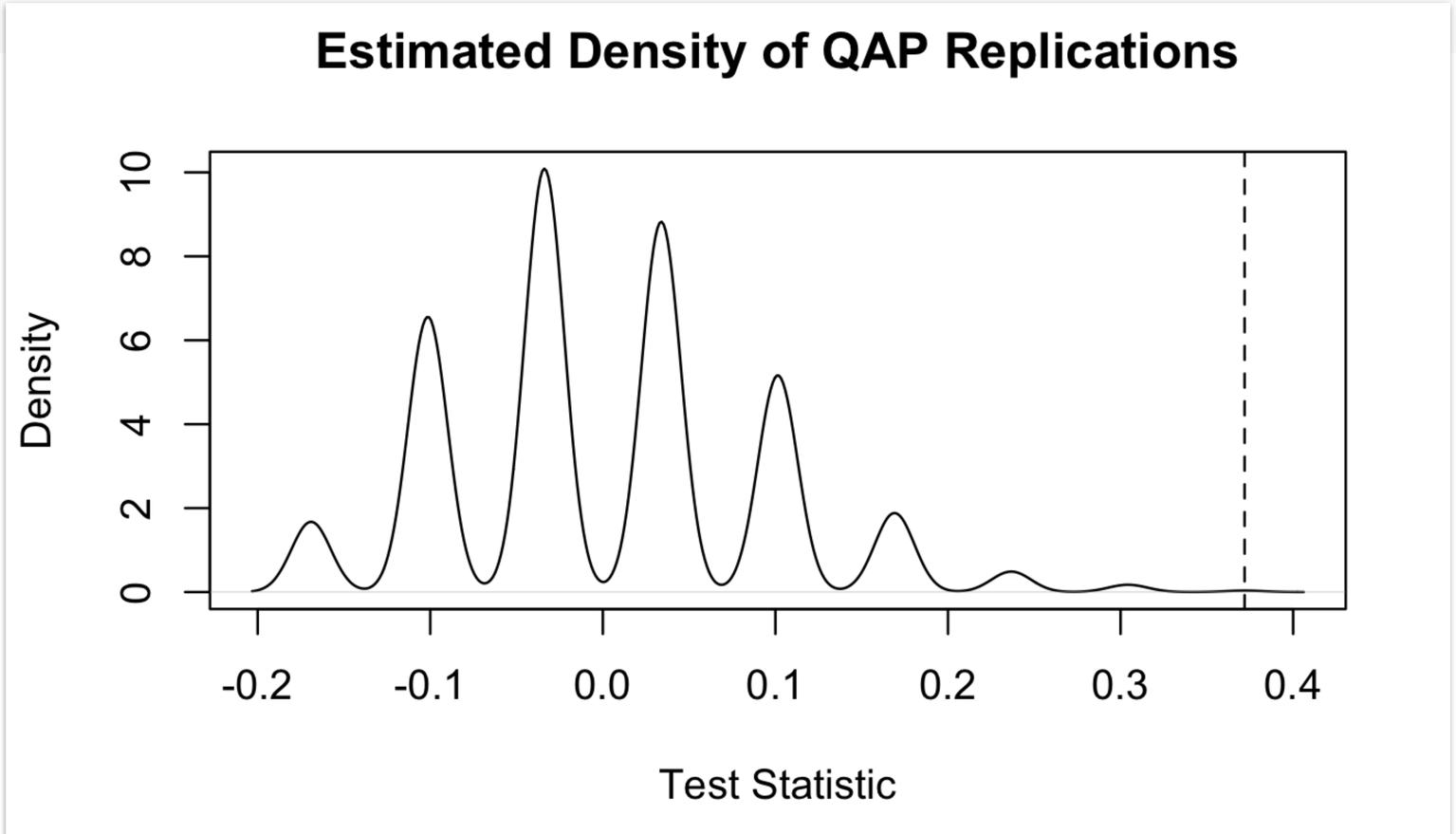
flobusiness



flomarriage

QAP – Empirical Example

```
qt <- qaptest(list(flobusiness,flomarriage),
              gcor, g1=1, g2=2)
plot(qt)
```

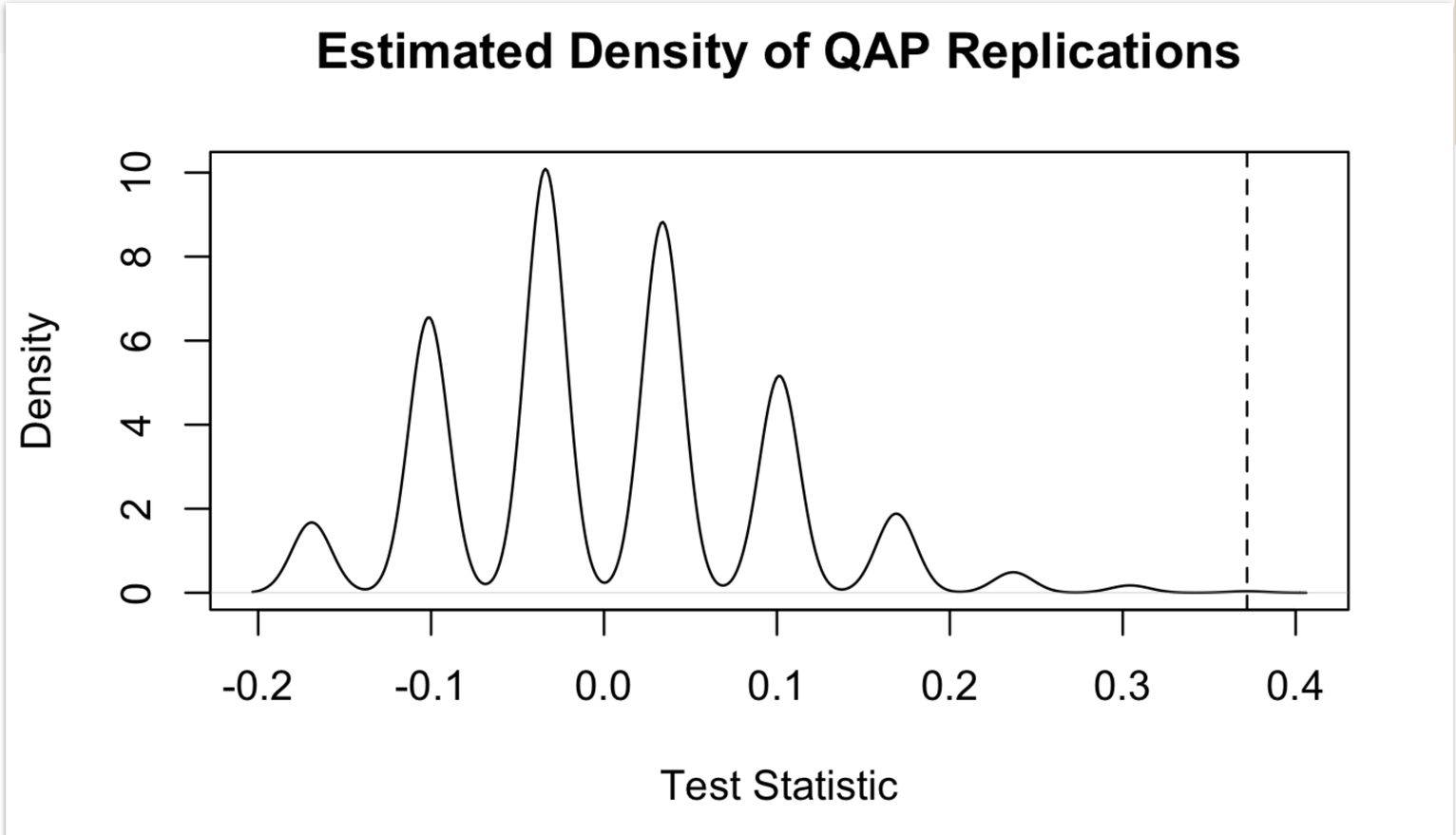


Estimated p-values:
p(f(perm) >= f(d)): 0
p(f(perm) <= f(d)): 1

Reject null hypothesis. It is very unlikely we would see a correlation this high as a result of random chance

QAP – Empirical Example

```
qt <- qaptest(list(flobusiness,flomarriage),
              gcor, g1=1, g2=2)
plot(qt)
```



With this type of simulation test, it's common to get these "weird" p=0 type results

Estimated p-values:
p(f(perm) >= f(d)): 0
p(f(perm) <= f(d)): 1

Reject null hypothesis. It is very unlikely we would see a correlation this high as a result of random chance

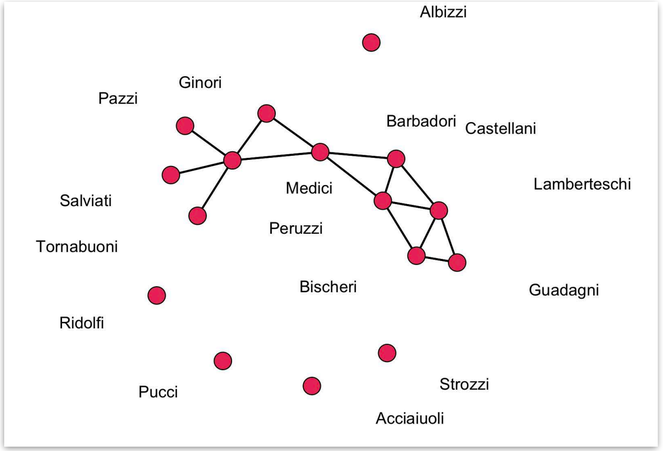
QAP – Empirical Example

How this might appear in a paper:

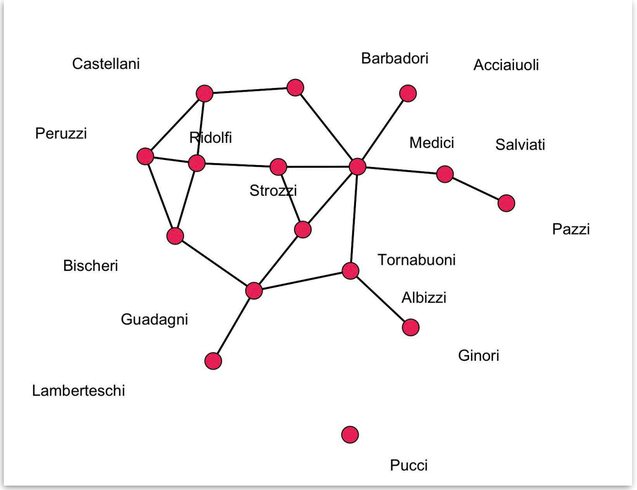
"There is a low correlation (0.37, $p < 0.01$) between the business and marriage networks.

This suggests these networks represent different types of relationships."

If there were high correlation, this could be used as justification for aggregating these measures into a single network



flobusiness



flomarriage

Agenda

Part 1:

- Example: Zachary's Karate Club
  - ➡ Dealing with temporality in networks
- Example: Ethnic Networks
  - ➡ Aggregating edge measures
- Overview of network inference
  - ➡ Network comparison (QAP)

Break

Part 2:

- American National Election Survey (ANES)
  - ➡ Whole networks v. Ego-networks
  - ➡ Regressing responsibly
- Topical co-occurance
  - ➡ Pruning and backboning

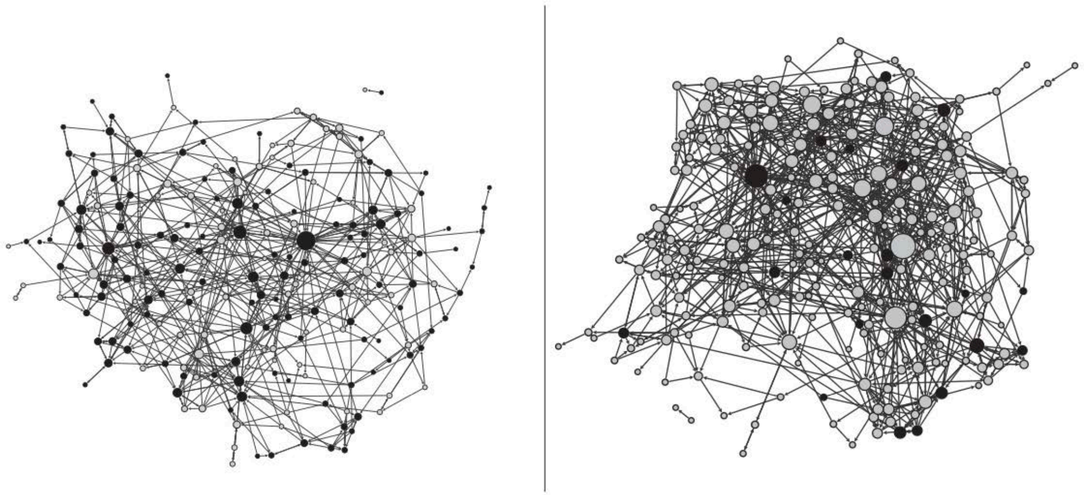
This workshop will reflect on (some of) the choices you will make during network modeling

PSA: Networks are *models* that imperfectly capture a complex world

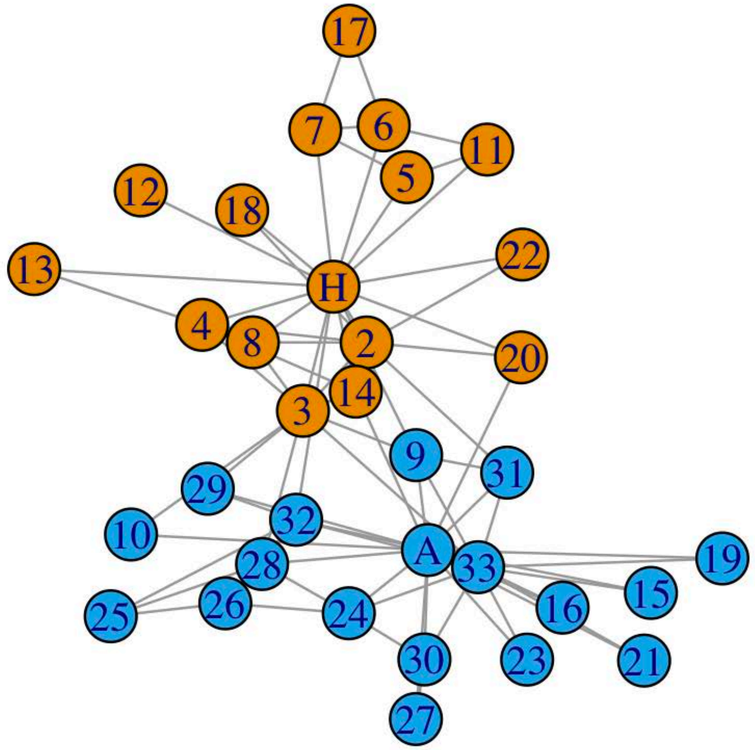
Recap: Defining your nodes and edges

- A network is the result of **choices** you make about how to define nodes and edges
- These choices must be explained and justified!
- There are wrong choices, but there are no “right” choices

FIGURE 1 Abalang (left) and Mugana (right) Social Networks, Visualized with Force-Directed Layout



Note: Those who heard are shaded black; node size is proportional to degree.



Example 3:
ANES

Example 3: ANES

The ANES (American National Election Survey) asks:

2000 | V001699 | Z1. Person R discusses politics with: na

Z1.
Now let's shift our attention to another area.
From time to time, people discuss government, elections and politics with other people. I'd like to ask you about the people with whom you discuss these matters.
These people might or might not be relatives.
Can you think of anyone? R IDENTIFIES NAME 1

Example 3: ANES

The ANES (American National Election Survey) asks:

2000 V001699 Z1. Person R discusses politics with: na

Z1.
Now let's shift our attention to another area.
From time to time, peopl
with other people. I'd like
discuss these matters.
These people might or m
Can you think of anyone

2000	V001708	Z10. When you talk with [NAME 1], do you **DISCUSS** political matters...often, sometimes, rarely, or never?
2000	V001716	Z14. When you talk with [NAME 2], do you **DISCUSS** political matters...often, sometimes, rarely, or never?
2000	V001724	Z18. When you talk with [NAME 3], do you **DISCUSS** political matters...often, sometimes, rarely, or never?
2000	V001732	Z22. When you talk with [NAME 4], do you **DISCUSS** political matters...often, sometimes, rarely, or never?

Asking in this way can help insure respondents are thinking of specific people

Example 3: ANES

2000 | V001699 | Z1. Person R discusses politics with: na

Z1.
Now let's shift our attention to another area.
From time to time, people discuss government, elections and politics with other people. I'd like to ask you about the people with whom you discuss these matters.
These people might or might not be relatives.
Can you think of anyone? R IDENTIFIES NAME 1

Nodes:

- ➡ Survey respondents

- ➡ AND people named by survey respondents

We know a lot about the ego — we have their full survey response!

We know very little about the alter

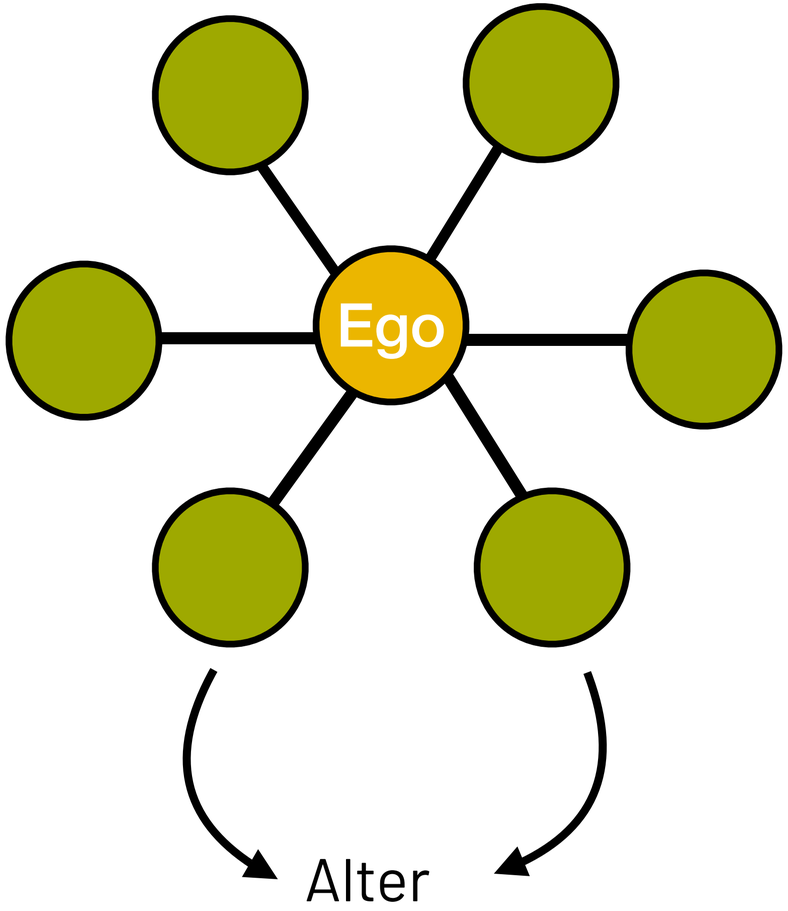
Edges:

- ➡ “Discuss important matters” together

In this case, the generated networks can be treated as **independent**. It is unlikely that the networks reported by respondents overlap (include the same people)

Egocentric Networks

An **ego-net** is one person's local perspective of a network.

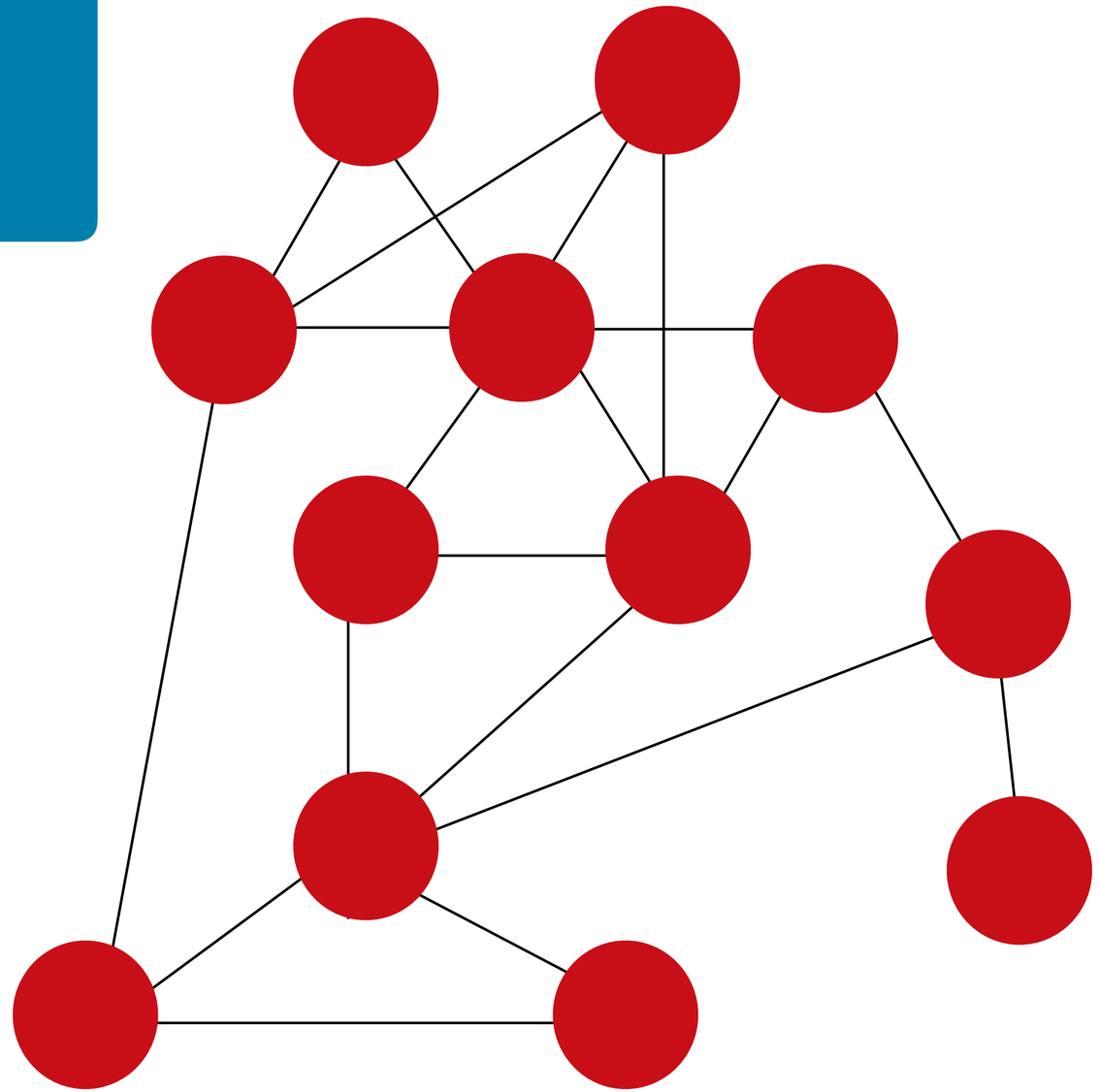


Comparison: Whole Network Data

"**Whole network data**" means that we know ALL nodes and edges of our network

We **rarely** have whole network data, but sometimes we're "close enough"

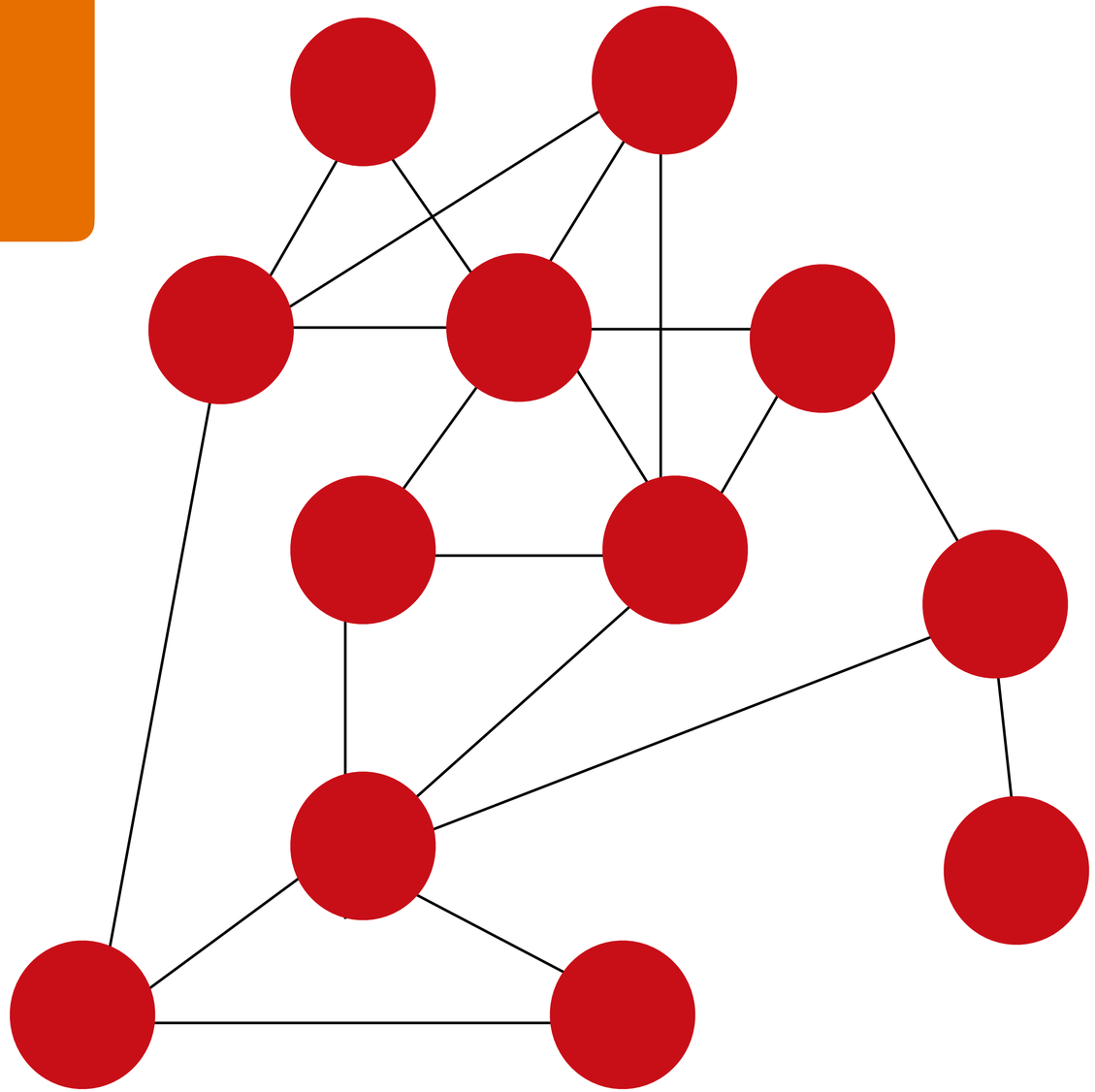
For example, the networks collected for the Ethnic Networks paper included at least one respondent from each household. We can consider that to be a "whole network"



Egocentric Networks

An **ego-net** is one person's local perspective of a network.

Imagine that there is a whole network which exists, but don't have access to all the information



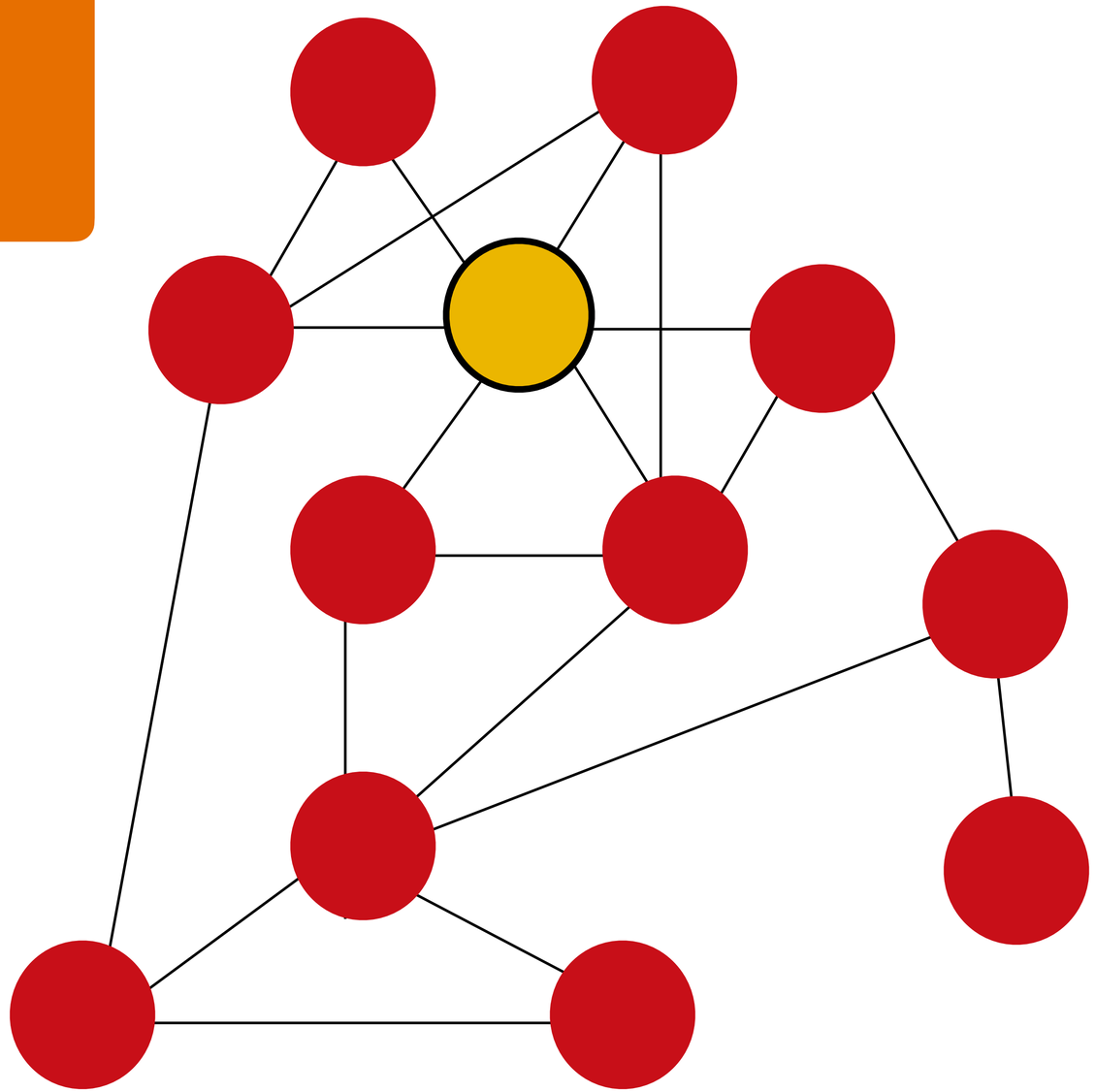
Egocentric Networks

An **ego-net** is one person's local perspective of a network.

Imagine that there is a whole network which exists, but don't have access to all the information

We survey one respondent

"Who are your friends?"



Egocentric Networks

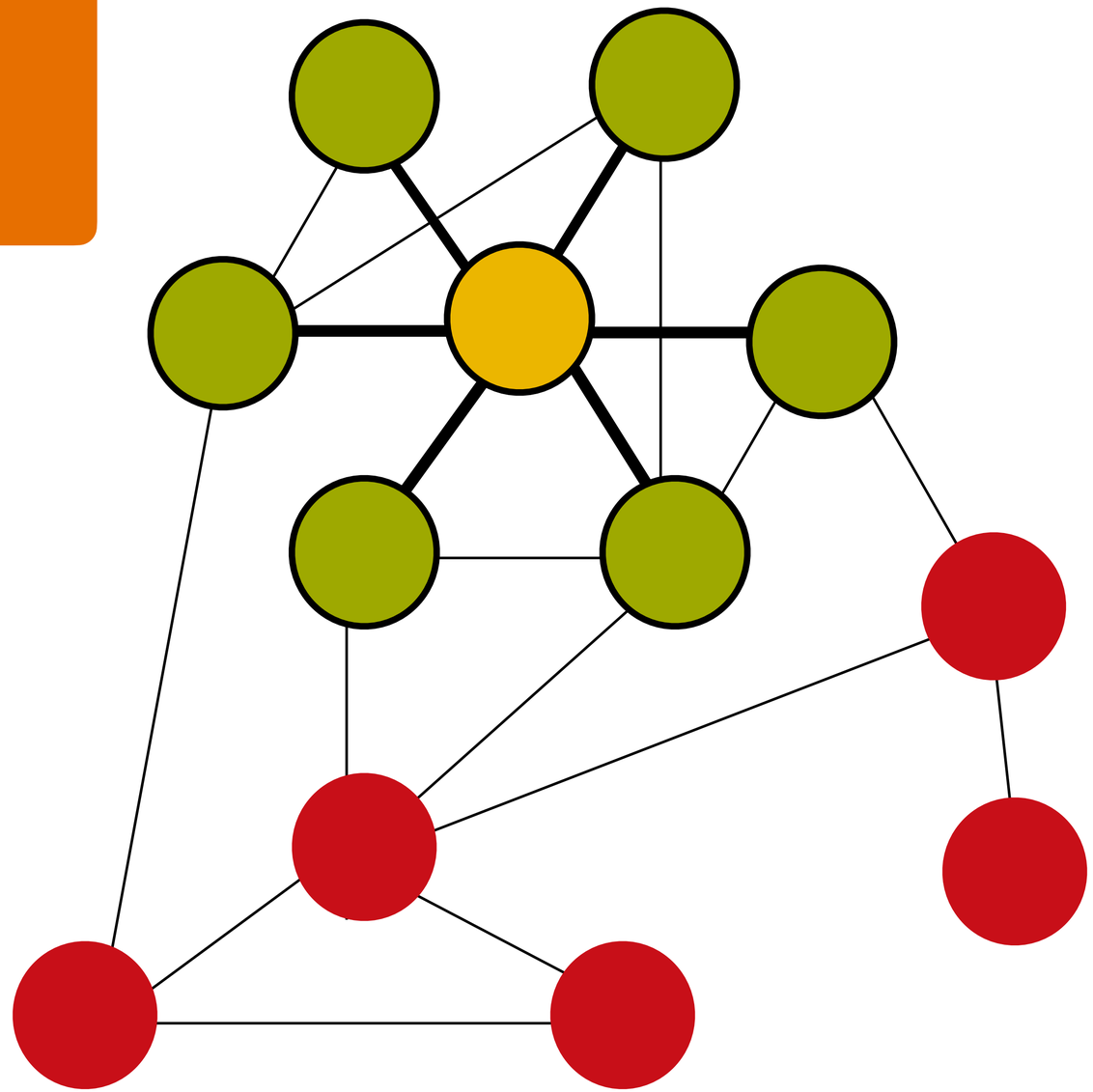
An **ego-net** is one person's local perspective of a network.

Imagine that there is a whole network which exists, but don't have access to all the information

We survey one respondent

"Who are your friends?"

Now we know who they're connected with ("who their alters are")



Egocentric Networks

An **ego-net** is one person's local perspective of a network.

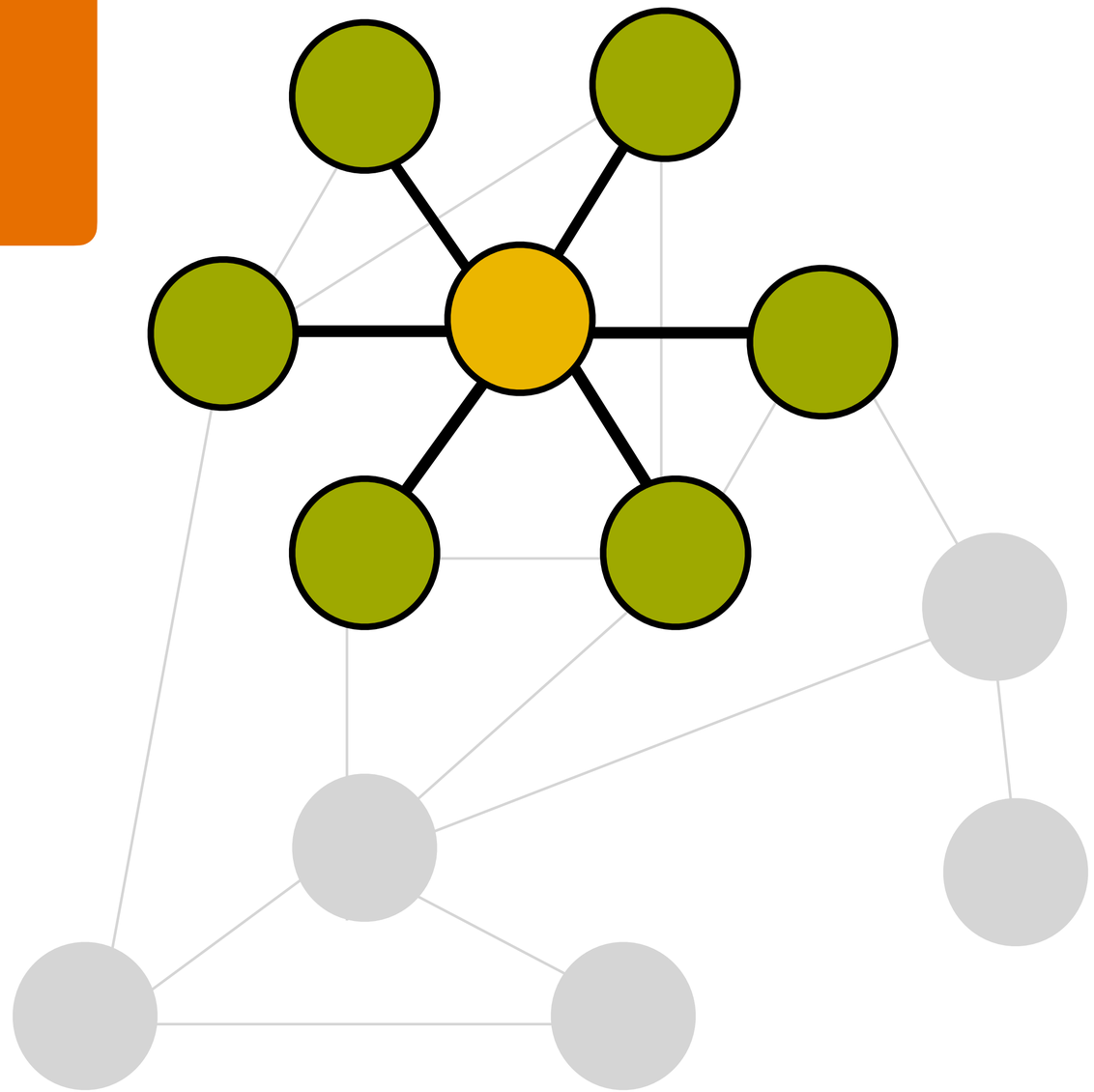
Imagine that there is a whole network which exists, but don't have access to all the information

We survey one respondent

"Who are your friends?"

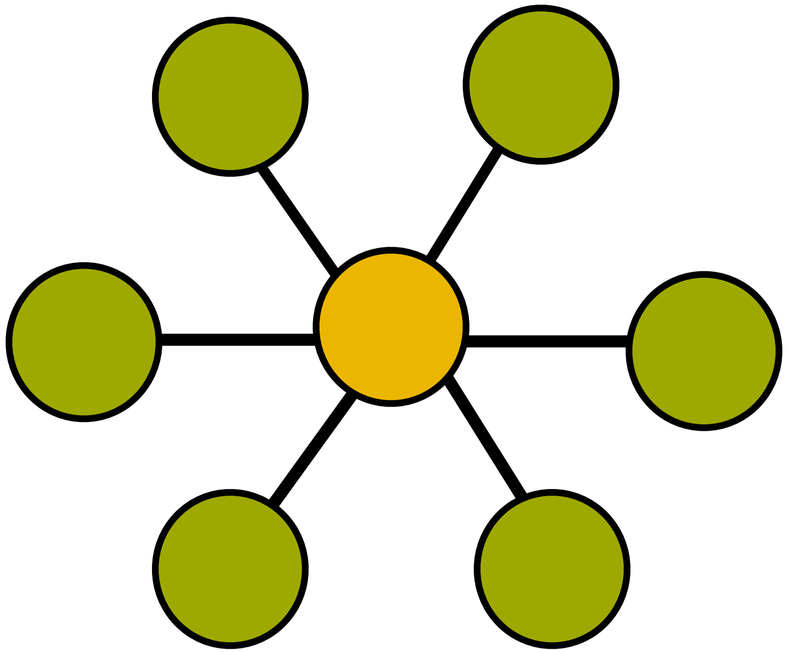
Now we know who they're connected with ("who their alters are")

There may be other nodes/edges, but we don't have any information about them



Egocentric Networks

An **ego-net** is one person's local perspective of a network.



Imagine that there is a whole network which exists, but don't have access to all the information

We survey one respondent

"Who are your friends?"

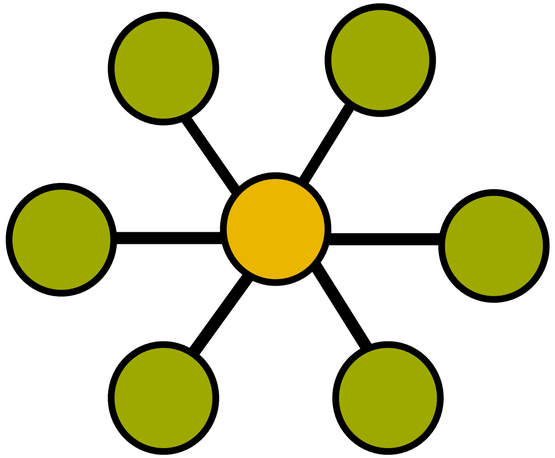
Now we know who they're connected with ("who their alters are")

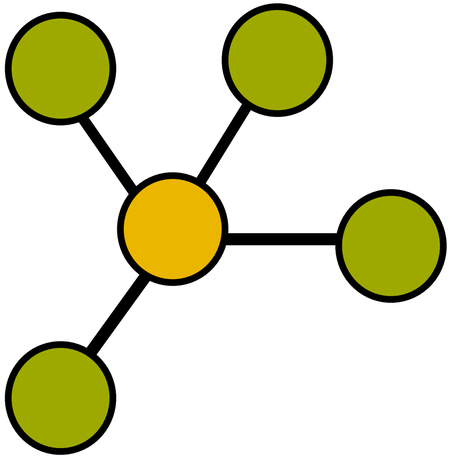
There may be other nodes/edges, but we don't have any information about them

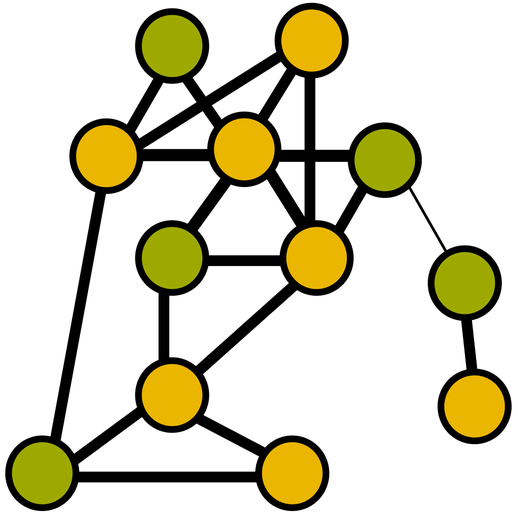
May ask questions like, "which of your friends are friends with each other?" BUT, can be unreliable

Ego-Nets v. Whole Networks

Ego-nets can often be treated as *independent* from each other



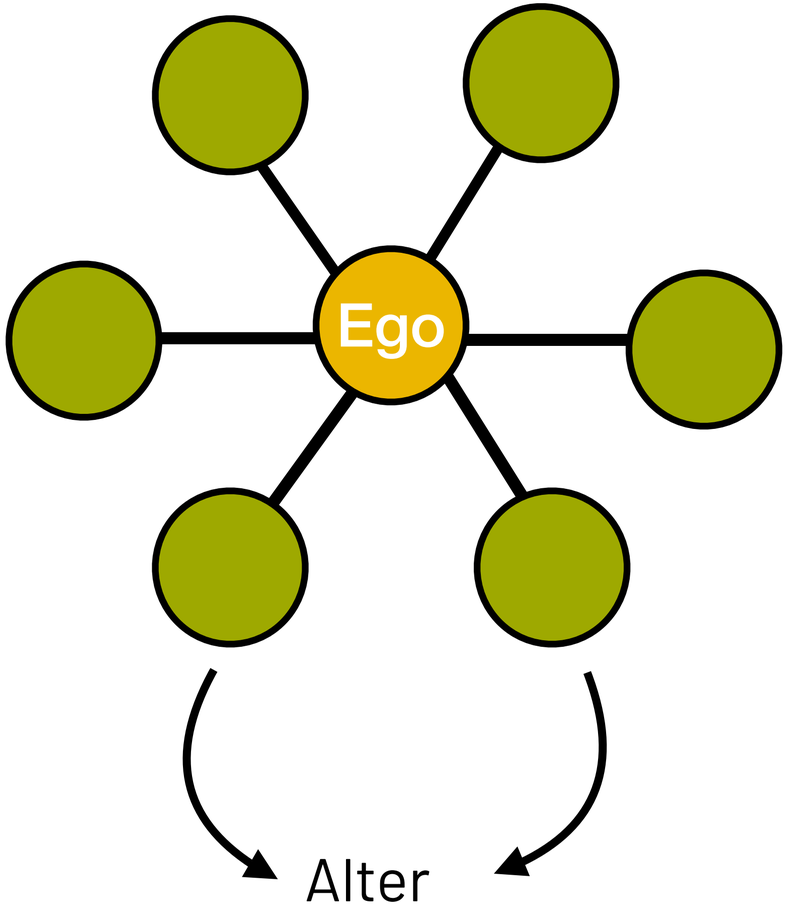




BUT, ego data collection can also be used to try to generate full network data

Example 3: ANES

In the ANES data, it makes the most sense to treat the generated networks as **independent**. It is unlikely that the networks reported by respondents overlap (include the same people)



Nodes:

- ➡ Survey respondents

- ➡ AND people named by survey respondents

2000 | V001699 | Z1. Person R discusses politics with: na

Z1.
Now let's shift our attention to another area.
From time to time, people discuss government, elections and politics with other people. I'd like to ask you about the people with whom you discuss these matters.
These people might or might not be relatives.
Can you think of anyone? R IDENTIFIES NAME 1

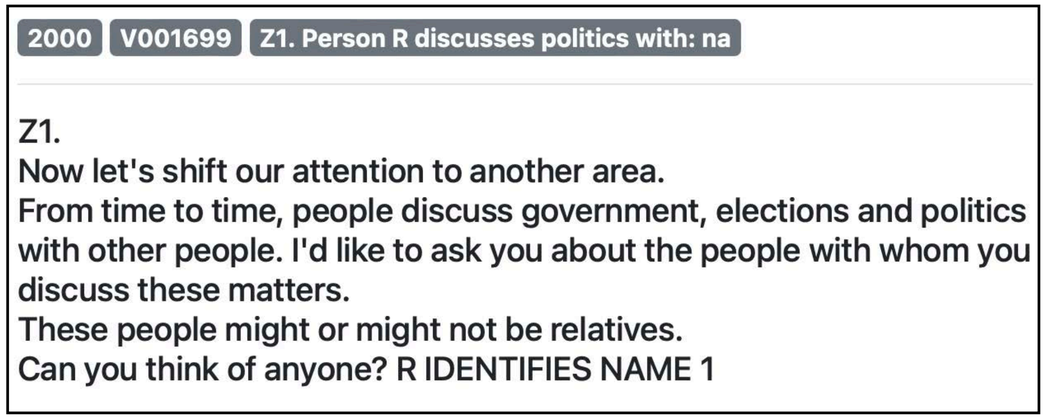
Example 3: ANES

In the ANES data, it makes the most sense to treat the generated networks as **independent**. It is unlikely that the networks reported by respondents overlap (include the same people)

Nodes:

- ➡ Survey respondents

- ➡ AND people named by survey respondents



This has some (good) privacy implications:

- ✦ Ask for / store just alter's *initial* (eg, "S") or *pseudonym*
- ✦ The subject knows who they're talking about, but the alter is non-identifiable to the researcher

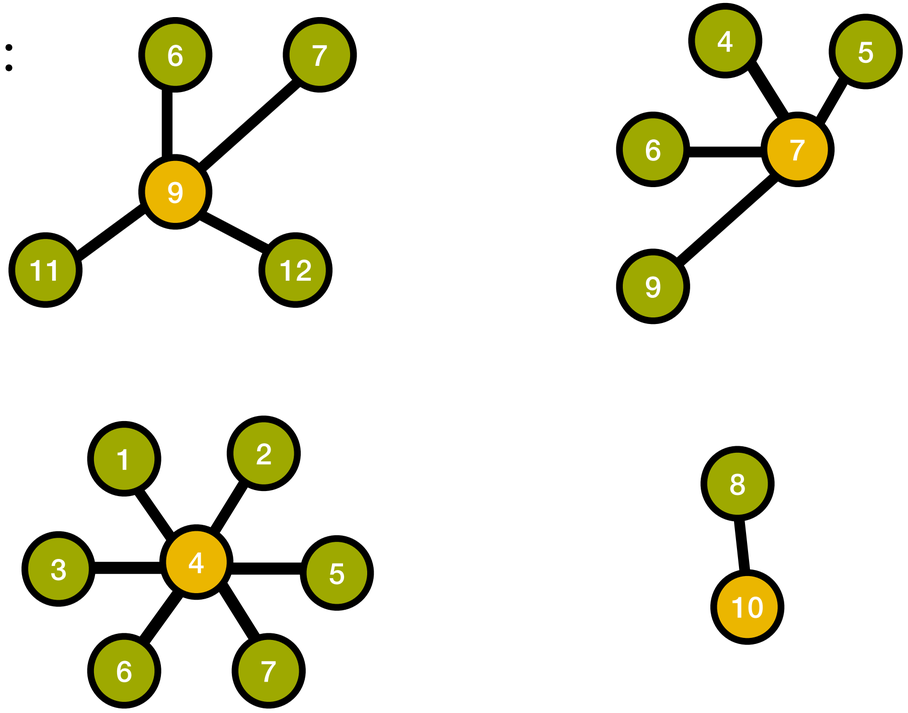
Alters are non-consenting subjects. Be careful with how much data you collect about them!

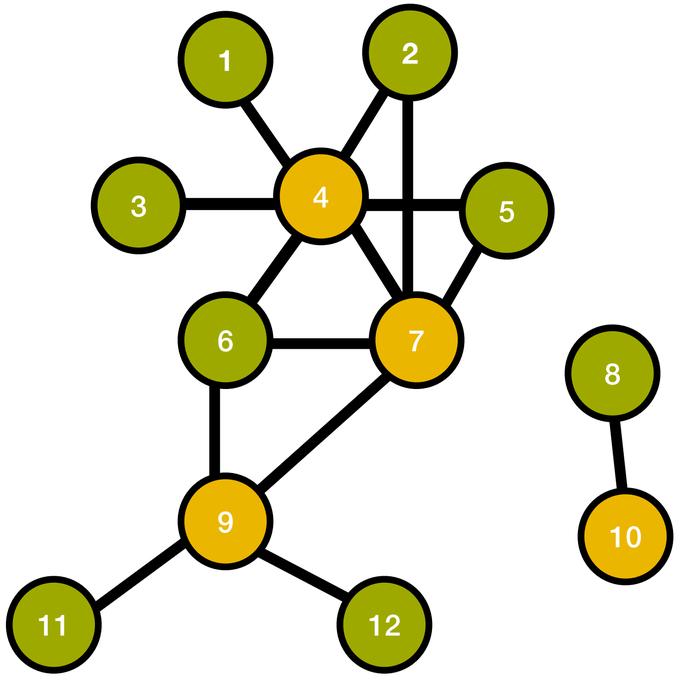
Alternate Example: This class

Imagine I don't have a class roster, but I want to know who is in class

One strategy would be to ask a random sample of people who their friends are, and then ask those friends, etc, **until I get whole network data**

4 ego-nets:





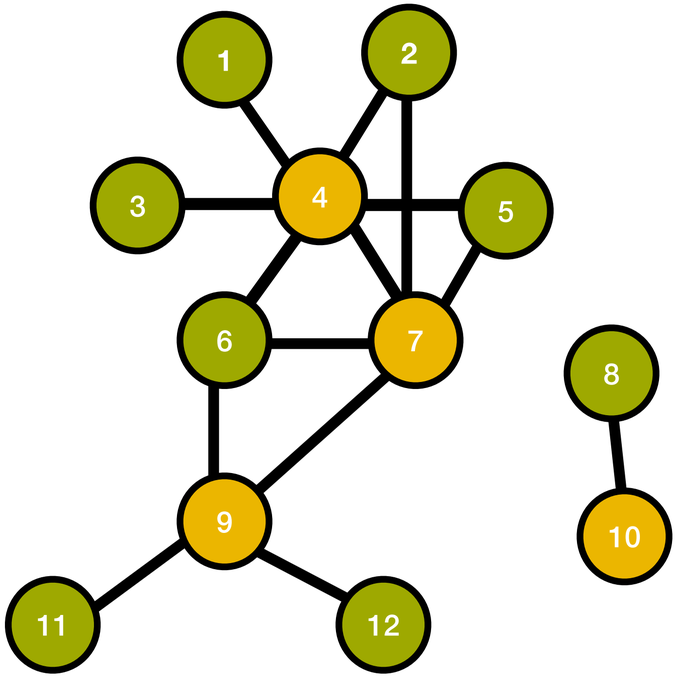
One aggregated not-quite whole network

Alternate Example: This class

Imagine I don't have a class roster, but I want to know who is in class

One strategy would be to ask a random sample of people who their friends are, and then ask those friends, etc, **until I get whole network data**

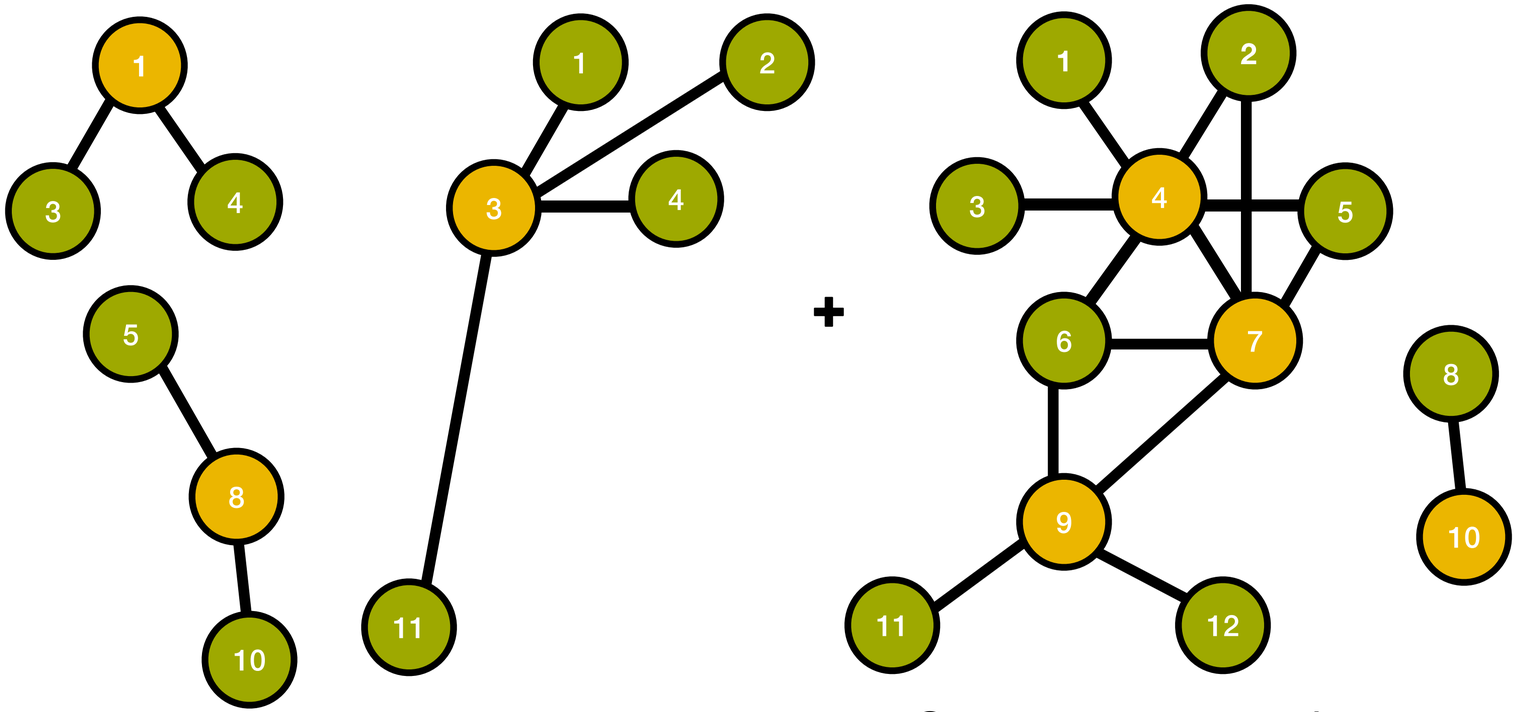
This network may not be complete (may be missing nodes or edges), so now I ask a sample of **alters** who their friends are



One aggregated not-quite whole network

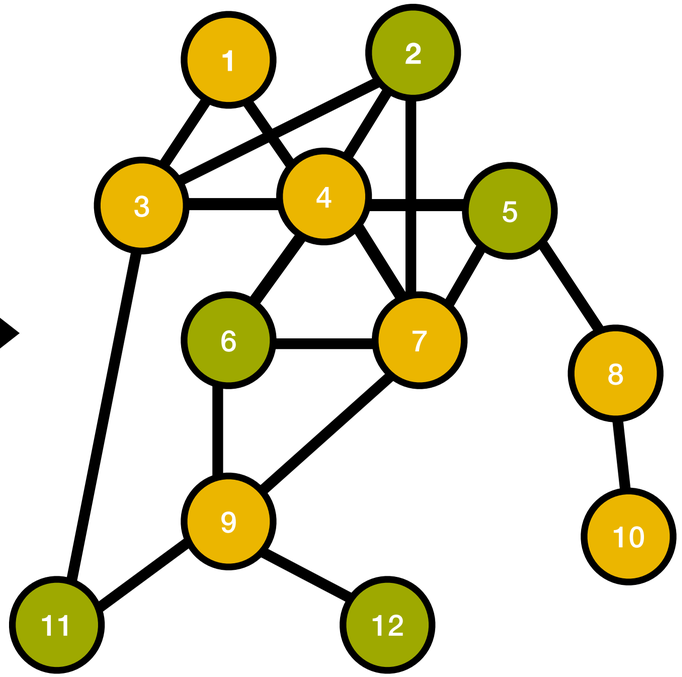
Alternate Example: This class

This network may not be complete (may be missing nodes or edges), so now I ask a random sample of **alters** who their friends are



3 ego-nets

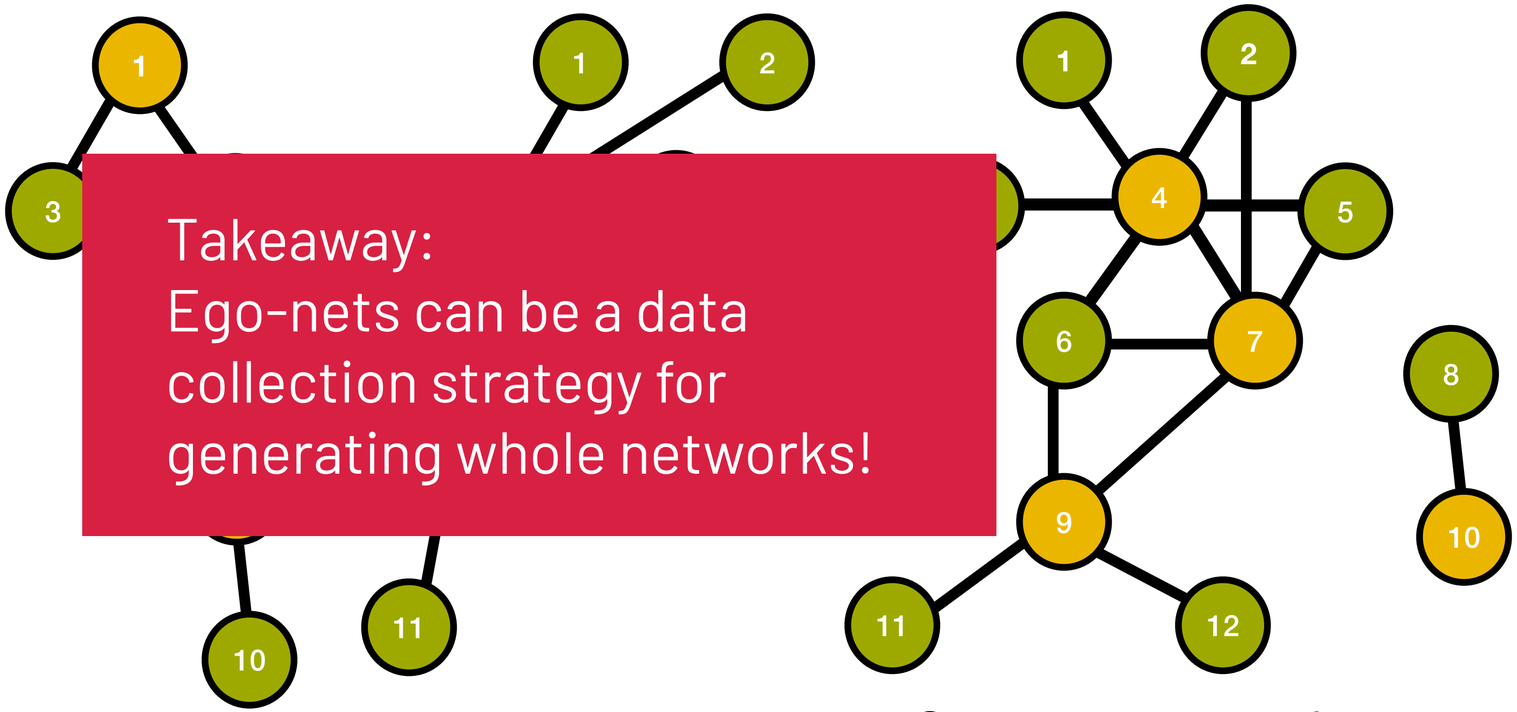
One aggregated not-quite whole network



Closer to whole network

Alternate Example: This class

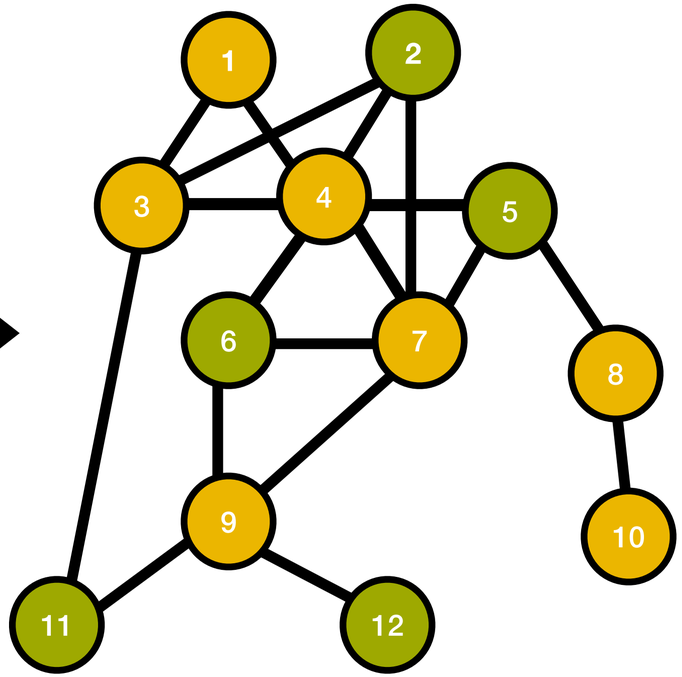
This network may not be complete (may be missing nodes or edges), so now I ask a random sample of **alters** who their friends are



Takeaway:
Ego-nets can be a data collection strategy for generating whole networks!

3 ego-nets

One aggregated not-quite whole network

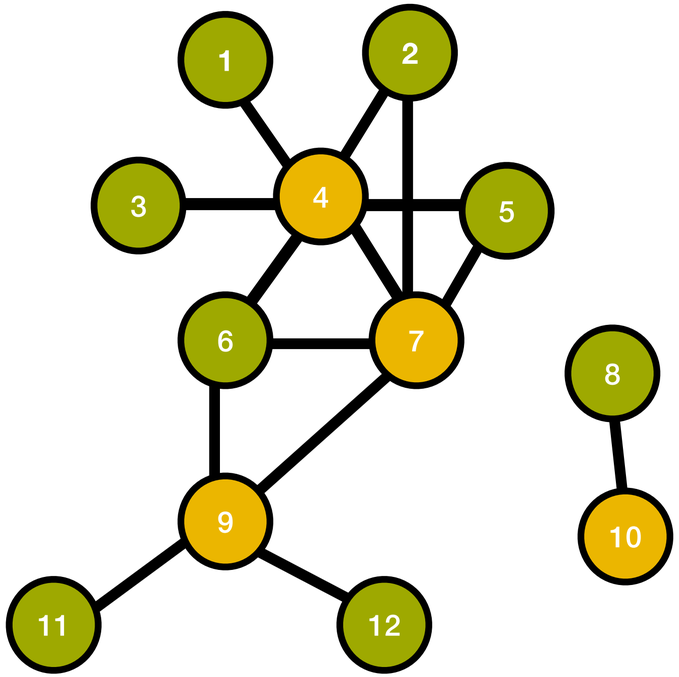


Closer to whole network

Whole Network Strategies

Strategy 1: Snowball sampling

- Iteratively collect data from consenting subjects
- Consenting subjects provide *minimal* information on their alters
- Consenting subjects pass researcher name or survey code on to their contacts
- If alters consent to join study, then you can tie these nodes together



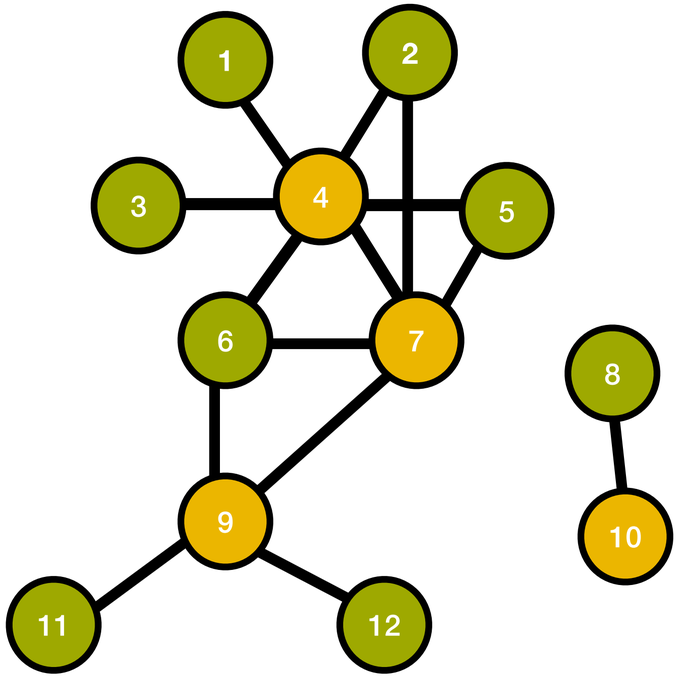
This is very common for vulnerable/ hard to reach populations

Whole Network Strategies

Aggregating ego-nets can be hard because we need to know if two alters are the same person or if one person's alter is also a subject

Strategy 2: Census survey before network survey

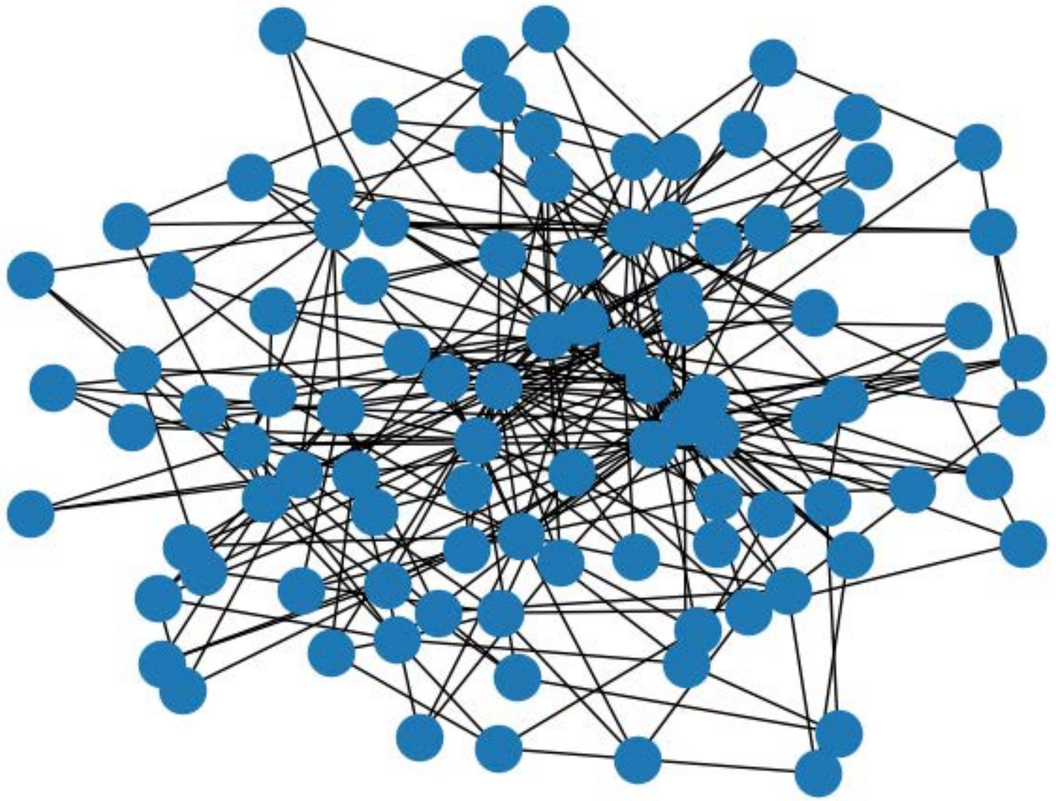
- *Wave 1:* Ask people to consent, collect full census of all nodes — only consenting subjects will appear as nodes in your network
- *Wave 2:* Ask subjects to indicate connections from list of consenting nodes



Regressing Responsibly

Regression?

Network observations are interdependent, which violates the independence assumption of OLS and other regression models



"The assumption of conditional independence in a traditional logit model of conflict implies that the British declaration of war on Germany in 1939 was *entirely* independent of all factors other than, say, regime type similarity and the ratio of relative capabilities between the two states."

Cranmer, 2016

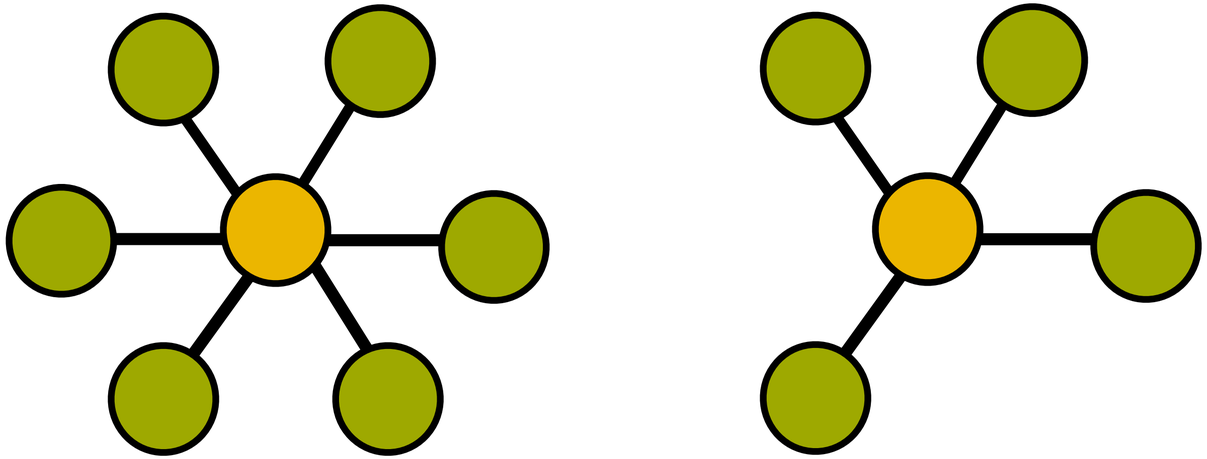
BUT...

Regression?

The ANES ego networks can treated as are independent

...So can we do regression now?

Yes, but always proceed with CAUTION



2000 | V001699 | Z1. Person R discusses politics with: na

Z1.
Now let's shift our attention to another area.
From time to time, people discuss government, elections and politics with other people. I'd like to ask you about the people with whom you discuss these matters.
These people might or might not be relatives.
Can you think of anyone? R IDENTIFIES NAME 1

In the ANES data, it makes the most sense to treat the generated networks as **independent**. It is unlikely that the networks reported by respondents overlap (include the same people)

Regression Framework

$$\mathbf{y} = \beta_0 + \beta_1 X$$

Dependent Variable (DV)
The outcome we are interested in estimating
(A vector)

Intercept:
The estimated value of **y** when x = 0
(A vector)

Coefficients:
The estimated change in **y** for a 1-unit change in x
(A vector)

Independent Variable (IV)
The features we think effect the outcome variable
(A matrix)

Regression Framework

$$\mathbf{y} =$$

Dependent Variable (DV)
The outcome we are interested in estimating
(A vector)

Typically **cannot** use network measures as an **outcome** variable (violates independence)

CAN be okay with **independent** ego-nets:

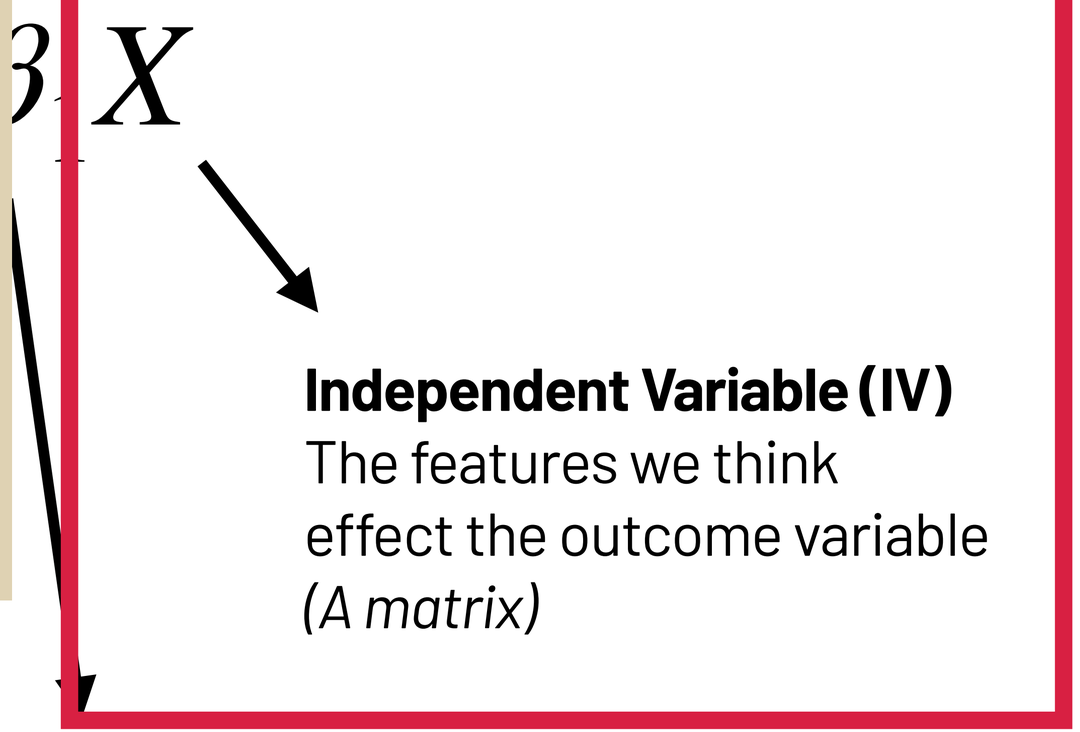
- Do people with XYZ characteristics (IVs) report more contacts, ie, have higher degree (DV)?

Interc
The estimated v
of **y** when x = 0
(A vector)

The estimated change in **y**
for a 1-unit change in x
(A vector)

Regression Framework

More commonly, it **can** be okay use a network measure as an **input** variable



(A vector)

Intercept:
The estimated value of **y** when x = 0
(A vector)

Independent Variable (IV)
The features we think effect the outcome variable
(A matrix)

Coefficients:
The estimated change in **y** for a 1-unit change in x
(A vector)

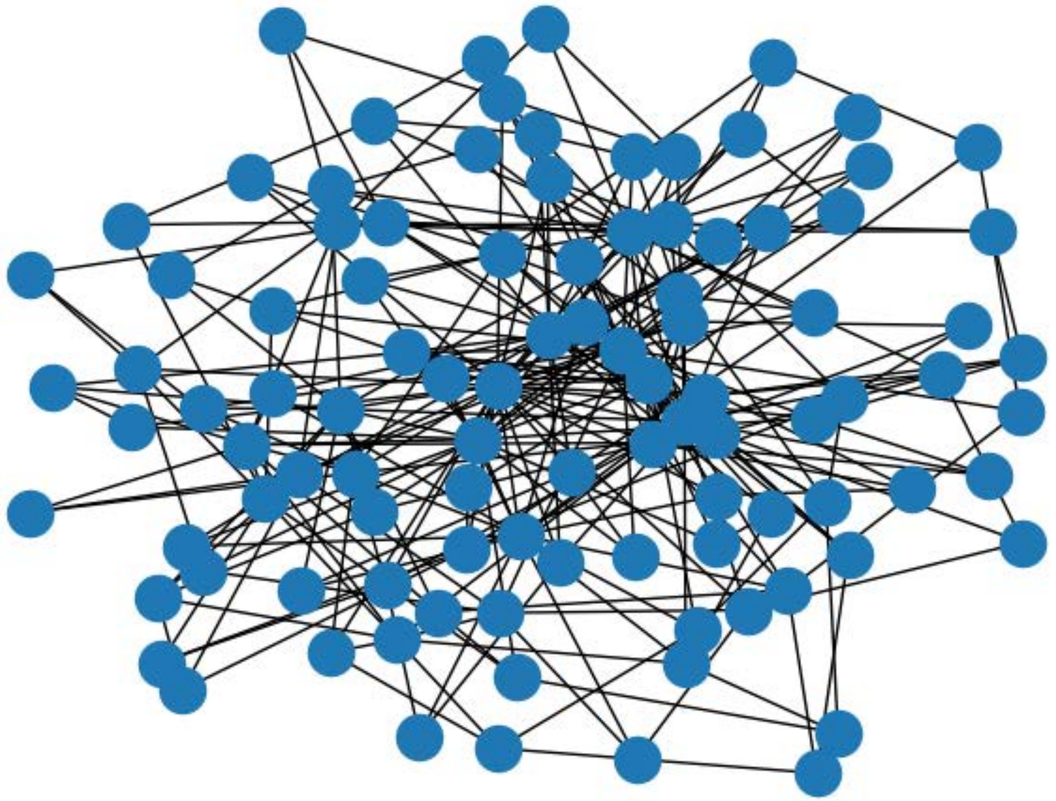
Independent Variables

I have some network

I suspect that a node-level characteristic (for example, degree) is related to some non-network outcome

I can (with caution!) use that node-level measure in a regression

Various forms of node **centrality** are commonly used as independent variables



Regression Example 1

Brokering health policy: Coalitions, parties, and interest group influence

MT Heaney - Journal of Health Politics, Policy and Law, 2006 - read.dukeupress.edu

Assuming a position as broker between disconnected interests is one way for an interest group to influence the making of federal health policy. This study demonstrates how groups use their connections with political parties and lobbying coalitions to augment their brokerage positions and enhance their influence over policy making. Evidence is drawn from statistical analysis of 263 interviews with health policy elites and a qualitative case study of the debate over the Medicare Prescription Drug, Improvement, and Modernization …

☆ Save Cite Cited by 194 Related articles All 10 versions

Estimates effect of an interest group's brokerage (IV, network measure) on their influence (DV) in negotiations around the Affordable Care Act

- Brokerage measured from network constructed through interviews with interest groups
- Influence measured through interviews with congressional staffers

Regression Example 2

Quality over quantity: Amici influence and judicial decision making

JM Box-Steffensmeier, DP Christenson… - … Political Science Review, 2013 - cambridge.org

Interest groups often make their preferences known on cases before the US Supreme Court via amicus curiae briefs. In evaluating the case and related arguments, we posit that judges take into account more than just the number of supporters for the liberal and conservative positions. Specifically, judges' decisions may also reflect the relative power of the groups. We use network position to measure interest group power in US Supreme Court cases from 1946 to 2001. We find that the effect of interest group power is minimal in times of heavily …

☆ Save Cite Cited by 135 Related articles All 13 versions

Estimates effect of an interest group's eigenvector centrality (IV, network measure) on the outcomes of judicial decisions (DV)

- Eigenvector centrality measured from network of co-signers of Amici Briefs
- Judicial Decisions measured through public records

Some notes of caution

Note that both these papers use **only one network measure** among their IVs (they do have other IVs)

- Network measures are often correlated, which can also present a problem for regression

Brokering health policy: Coalitions, parties, and interest group influence
MT Heaney - Journal of Health Politics, Policy and Law, 2006 - read.dukeupress.edu
Assuming a position as broker between disconnected interests is one way for an interest group to influence the making of federal health policy. This study demonstrates how groups use their connections with political parties and lobbying coalitions to augment their brokerage positions and enhance their influence over policy making. Evidence is drawn from statistical analysis of 263 interviews with health policy elites and a qualitative case study of the debate over the Medicare Prescription Drug, Improvement, and Modernization …
☆ Save Cite Cited by 194 Related articles All 10 versions

Note that both these papers construct their DV measurement through an **entirely different method/dataset** from their network approach

Quality over quantity: Amici influence and judicial decision making
JM Box-Steffensmeier, DP Christenson… - … Political Science Review, 2013 - cambridge.org
Interest groups often make their preferences known on cases before the US Supreme Court via amicus curiae briefs. In evaluating the case and related arguments, we posit that judges take into account more than just the number of supporters for the liberal and conservative positions. Specifically, judges' decisions may also reflect the relative power of the groups. We use network position to measure interest group power in US Supreme Court cases from 1946 to 2001. We find that the effect of interest group power is minimal in times of heavily …
☆ Save Cite Cited by 135 Related articles All 13 versions

Network Inference

Model	What you're estimating
Permutation test	Probability of seeing a network-level statistic
ERGM	Probability of seeing an edge between i and j
LSM	Probability of seeing an edge between i and j

The benefit of network inference methods is that they can use network measures as **dependent variables:**

- Probability of seeing a network statistic
- Probability of seeing a network structure
- Probability of seeing a connection between nodes

Regression typically CANNOT do this!

Example 4:
Topical Co-Occurance

What kinds of “things” might we care about?

- People
- Countries
- Organizations
- Dolphins
- Words
- Concepts
- …And more!

In the social sciences, we’re generally interested connections between in **social agents** OR their **artifacts**

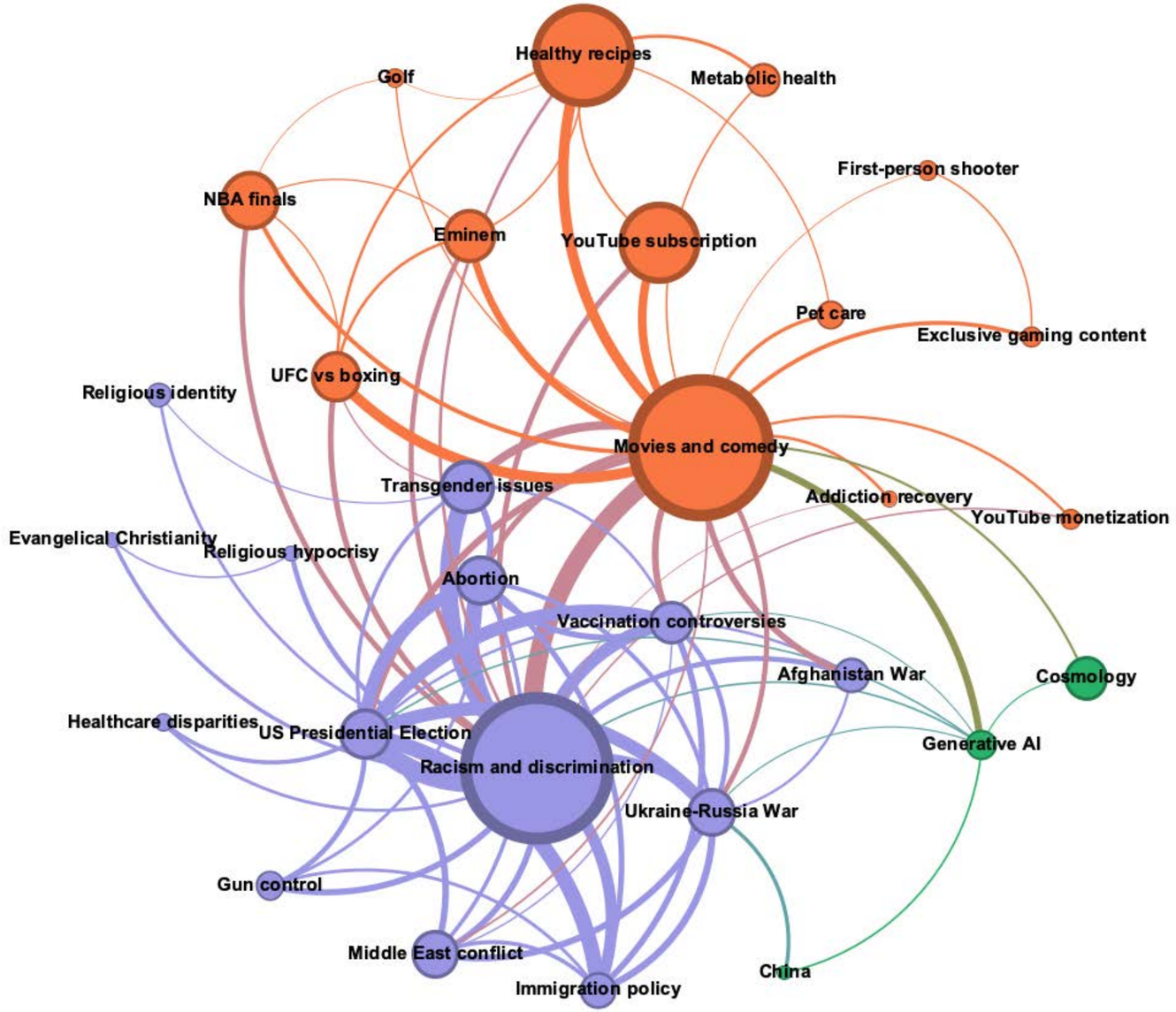
Social created or interpreted “things”

Topical Networks

RQ: What topics do Black conservative YouTubers discuss?

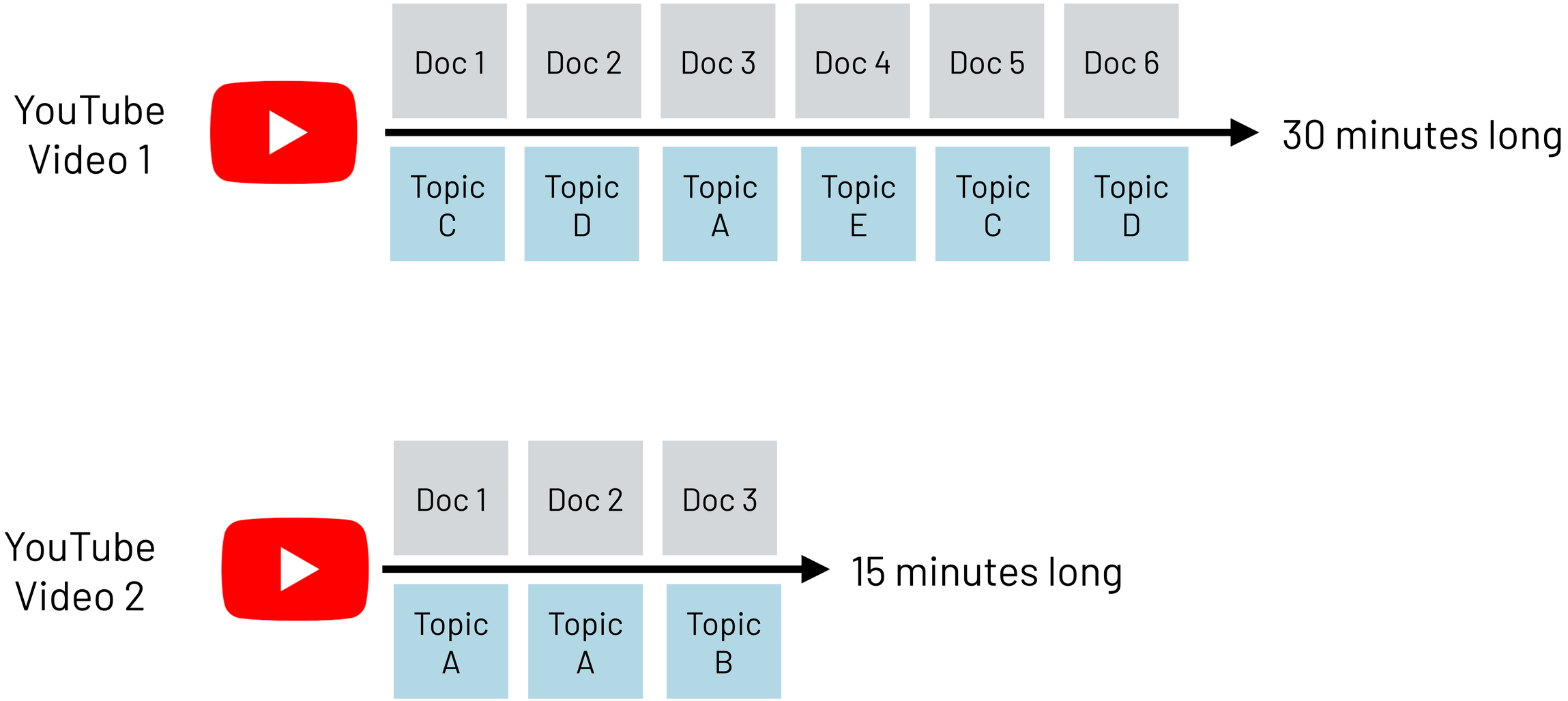
Nodes: "Topics"

Edges: Topics co-occur within the same video (weighted)



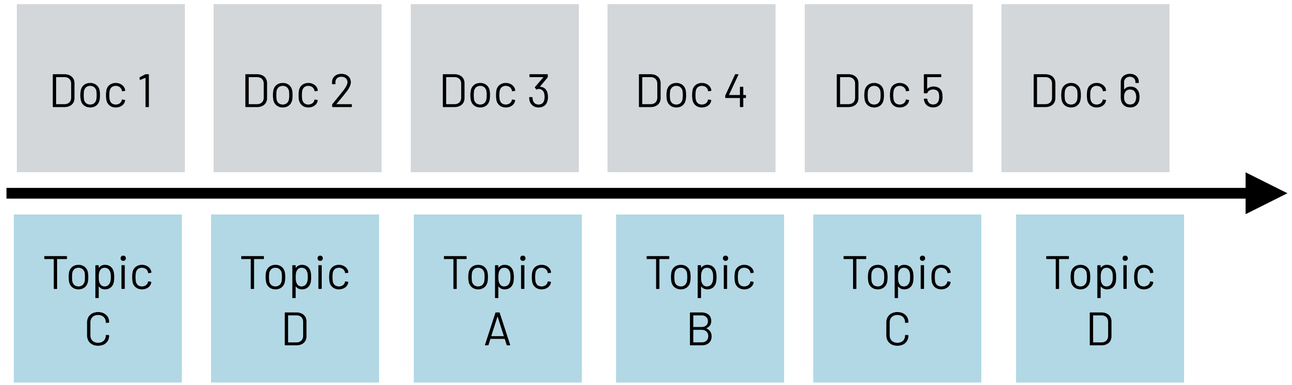
Work in progress, led by Marisa A. Smith, Michigan State University

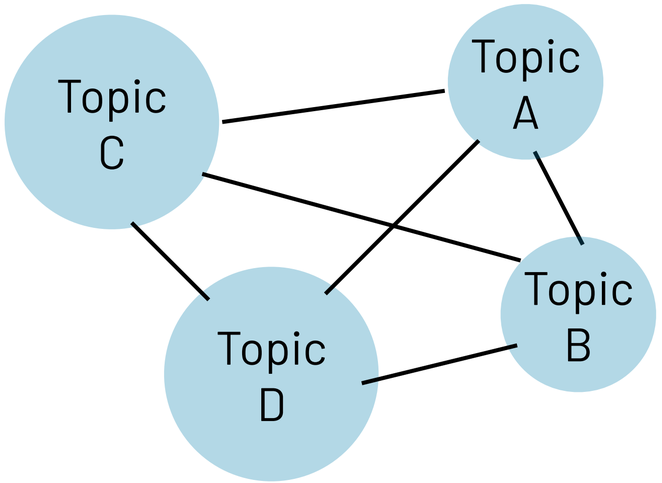
Constructing Topical Networks



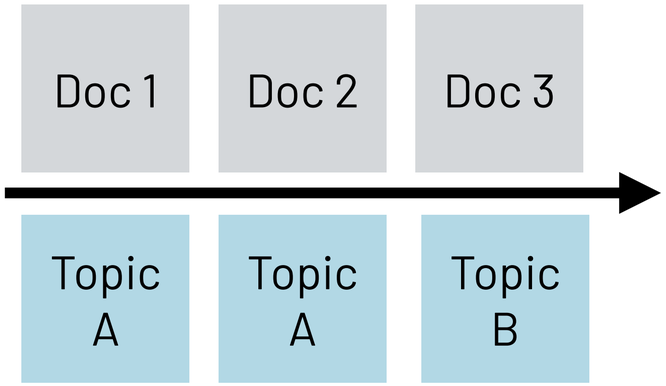
Constructing Topical Networks

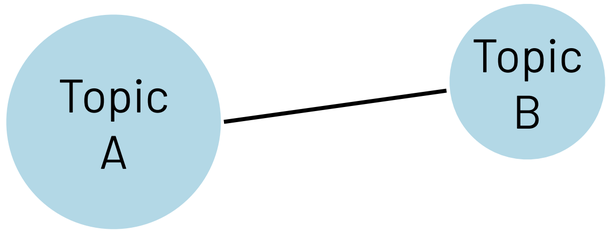
YouTube Video 1





YouTube Video 2



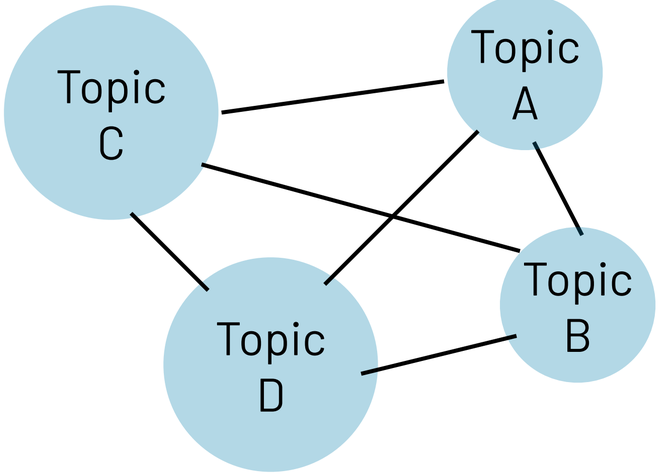


Nodes: Topics. Sized by propotion of occurance

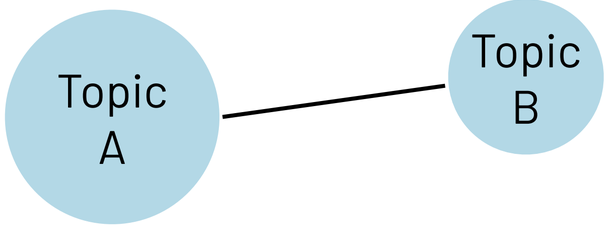
Edges: Weight of 1 for every time topics co-occur within a video

Constructing Topical Networks

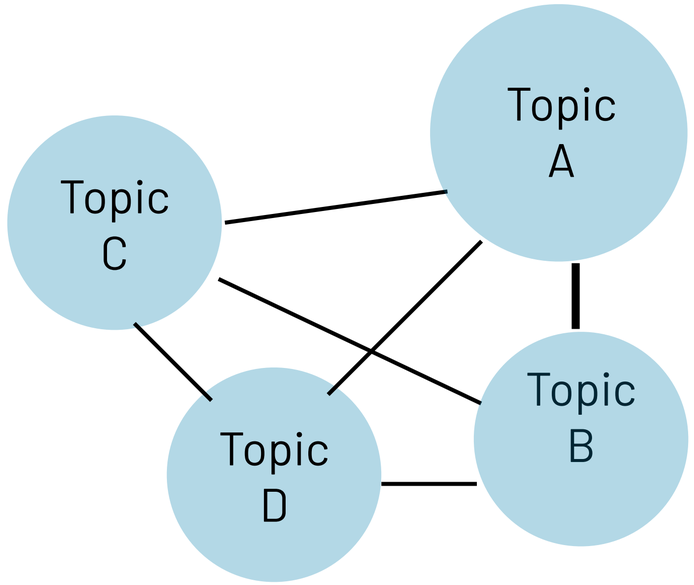
YouTube Video 1



YouTube Video 2



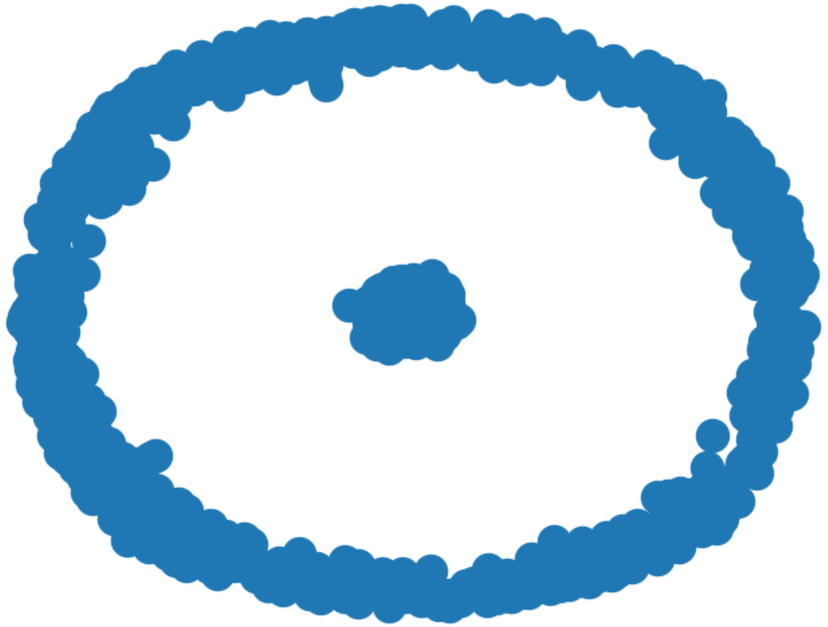
Video 1 + 2



Nodes: Topics. Sized by propotion of occurance

Edges: Weight of 1 for every time topics co-occur within a video

...With real data



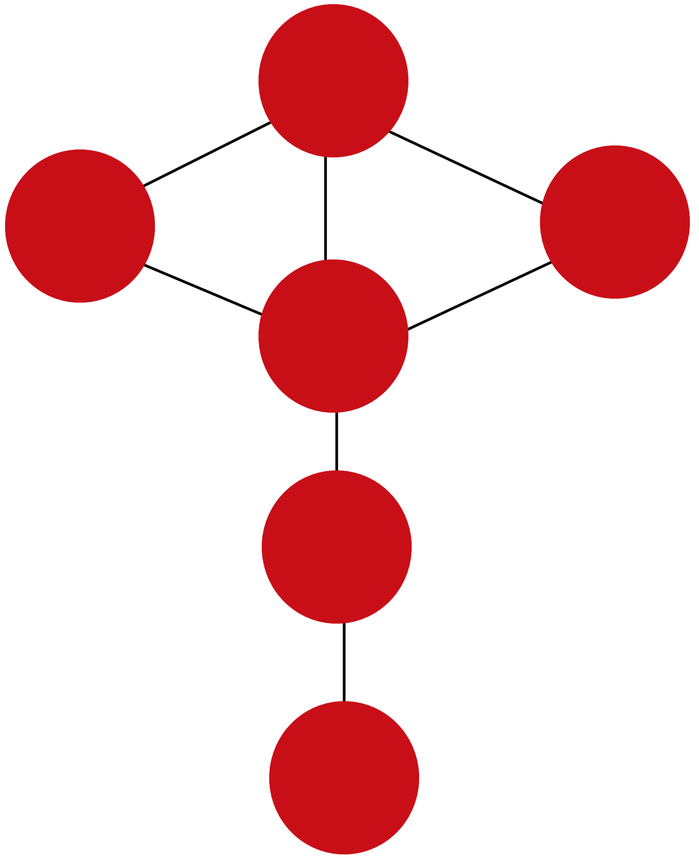
Full topical co-occurance network

Network Pruning & Backboning

Connected Networks

So far, we have mostly looked at **connected networks**

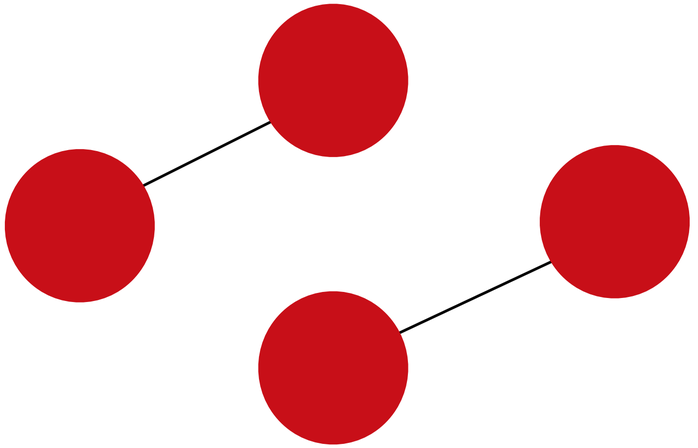
- A network is connected if there is at least one **path** from every node *i* to every node *j*
- Ie, connected means you can get *to* anywhere in the network *from* anywhere else in the network



Disconnected networks

If a network is not connected it is **disconnected**

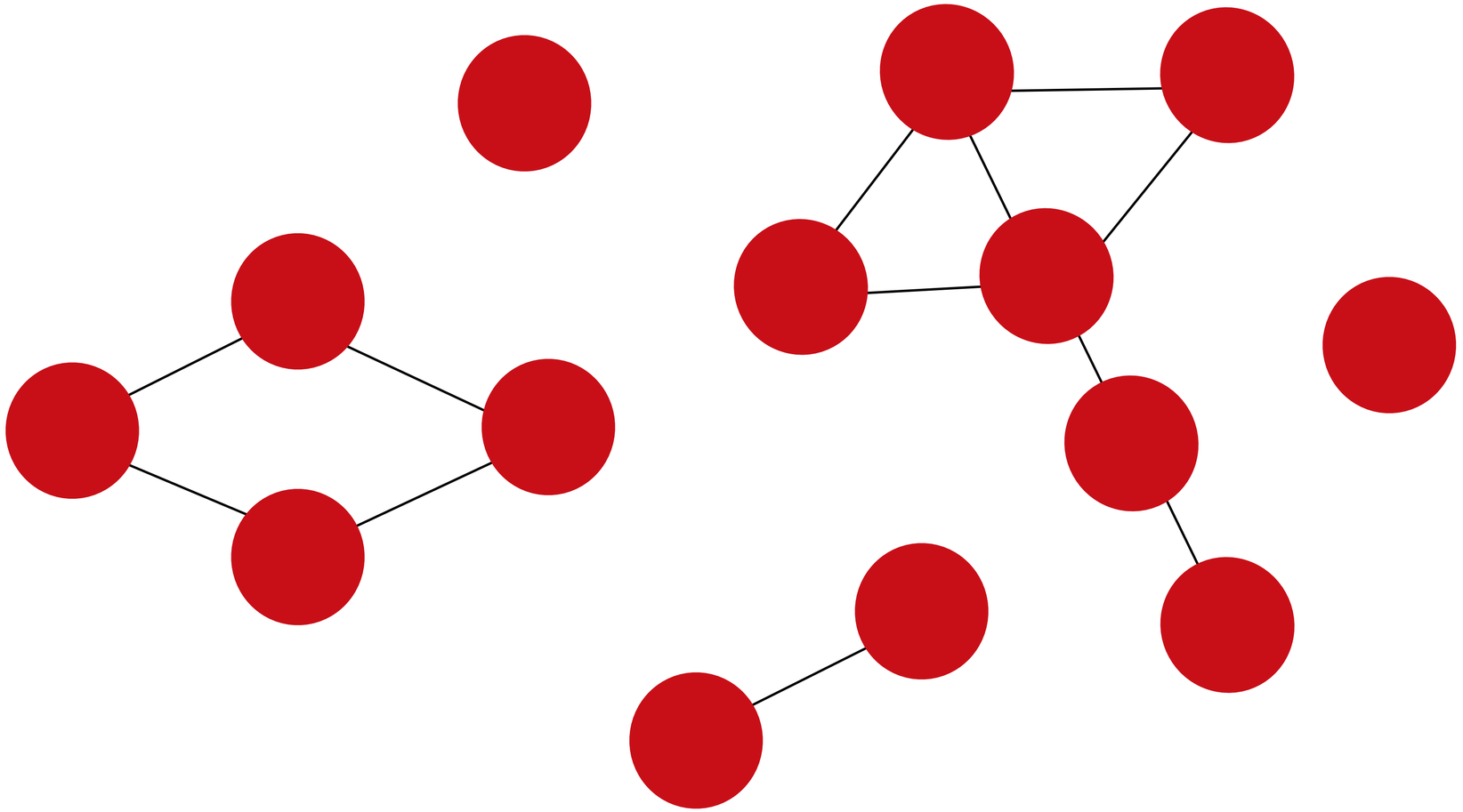
- This means the network has multiple **components**
- *Within* a component you can get from any node to any other node (ie, components are internally connected)
- But there are **no paths *between* components**



Components

Components will often have different **sizes** and **structures**

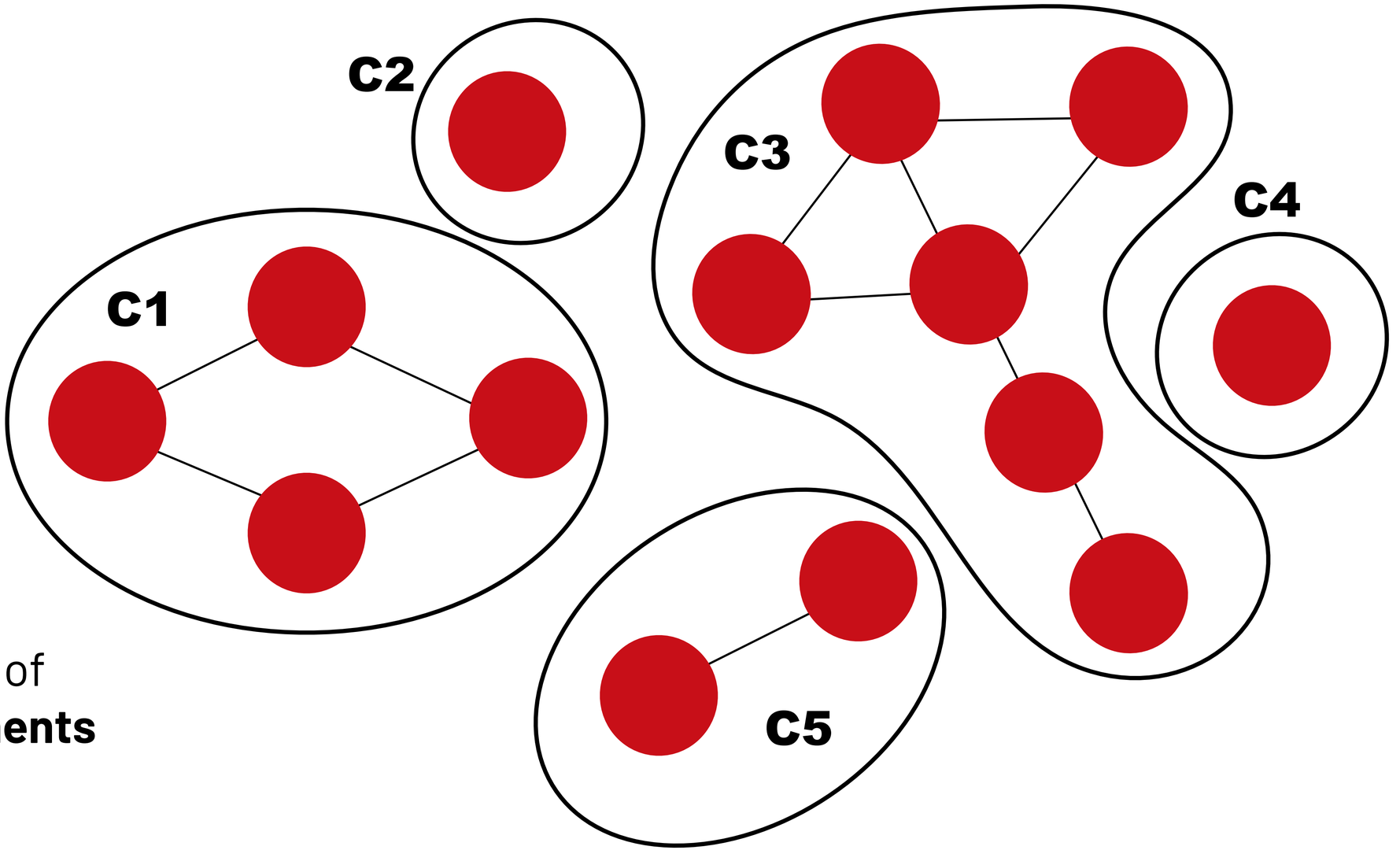
One network with
multiple components



Components

Components will often have different **sizes** and **structures**

One network with
multiple components



This network is made up of
5 disconnected components

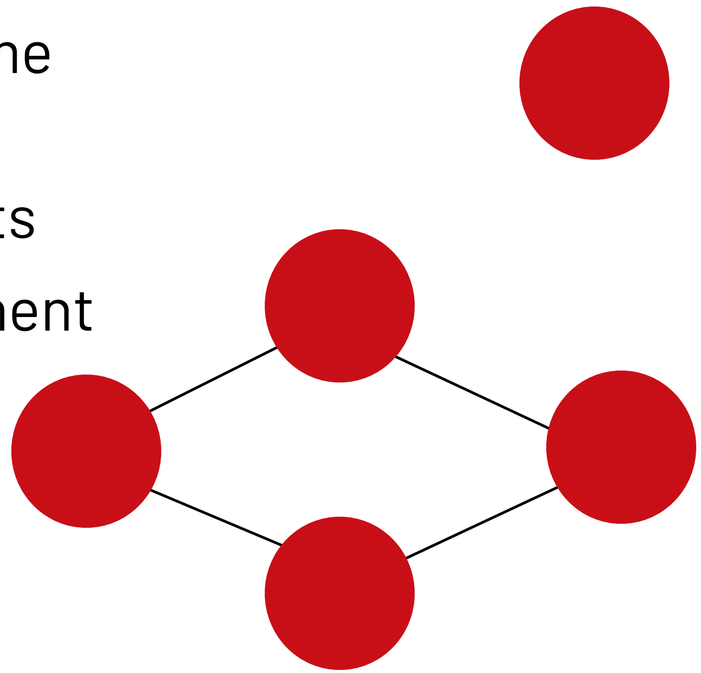
Components

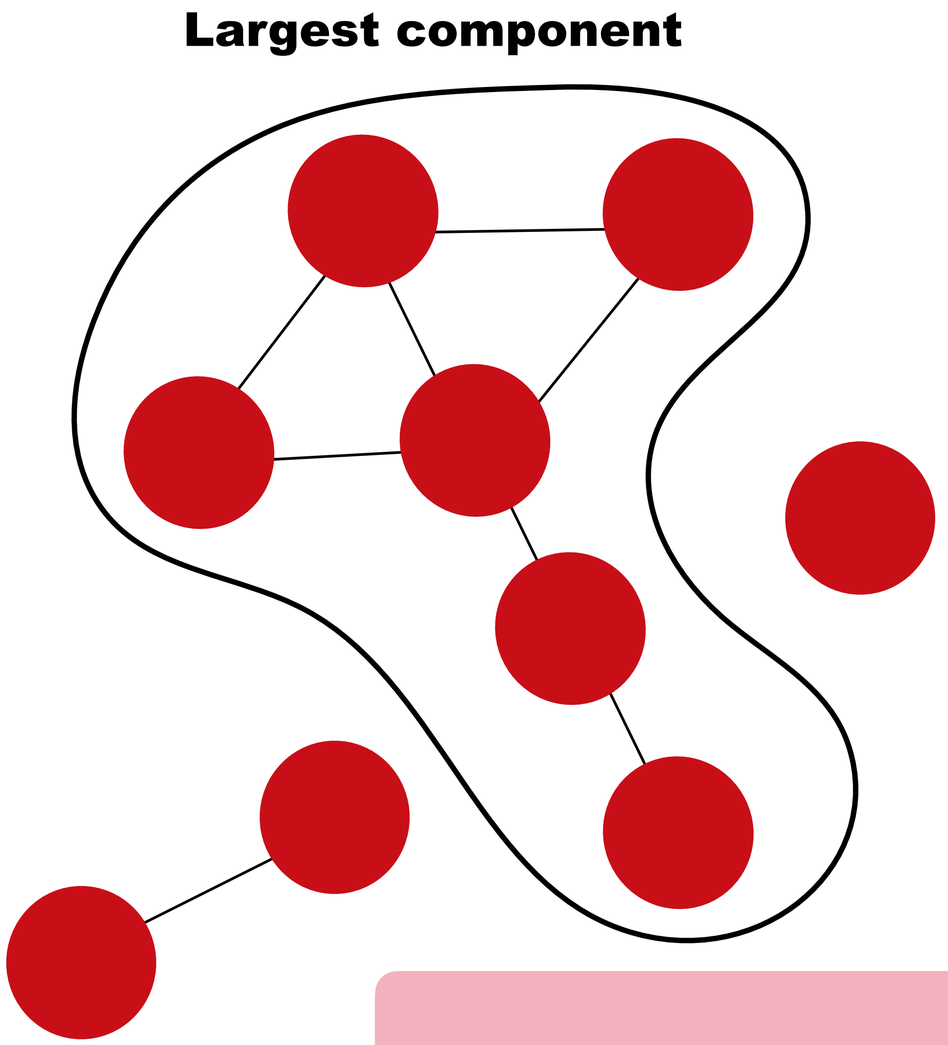
If your network has multiple components

It's common to report some summary statistics:

- Number of components
- Distribution of component size

...And then to do the bulk of your analysis on **just the major component**





One network
with **5 components**

Extracting Components (igraph)

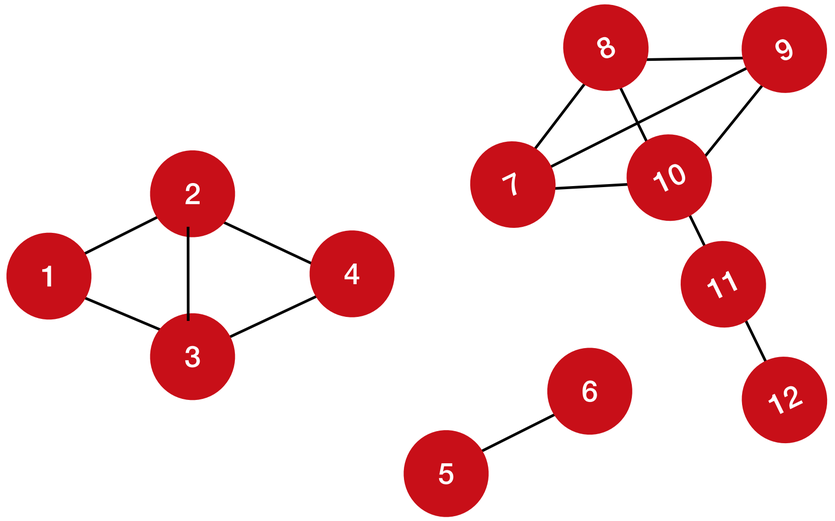
```
library(igraph)
```

Check if our network is connected:

```
is_connected(G)
```

```
> FALSE
```

Network G is **NOT** connected



Graph G

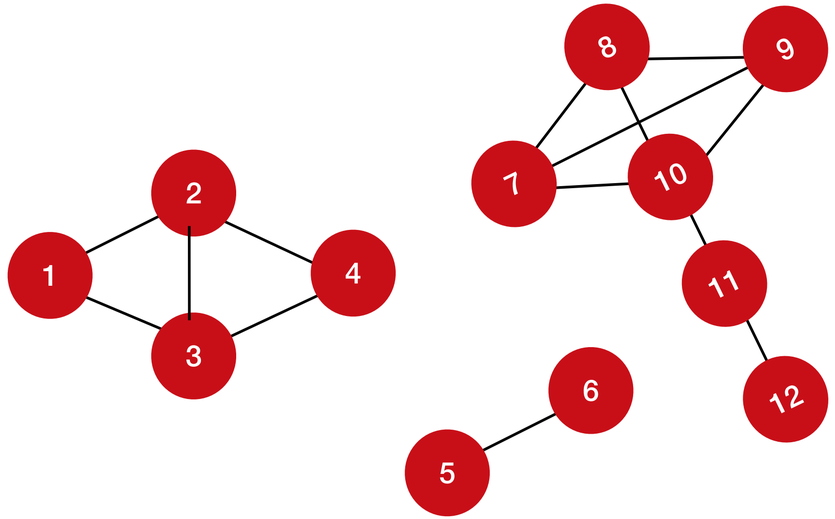
Extracting Components (igraph)

Get summary information about our components

```
comps <- components(G)
```

This will return information about:

- The number of components ($no)
- The size of each component ($csize)
- The component each node is in ($membership)



Graph G

Extracting Components (igraph)

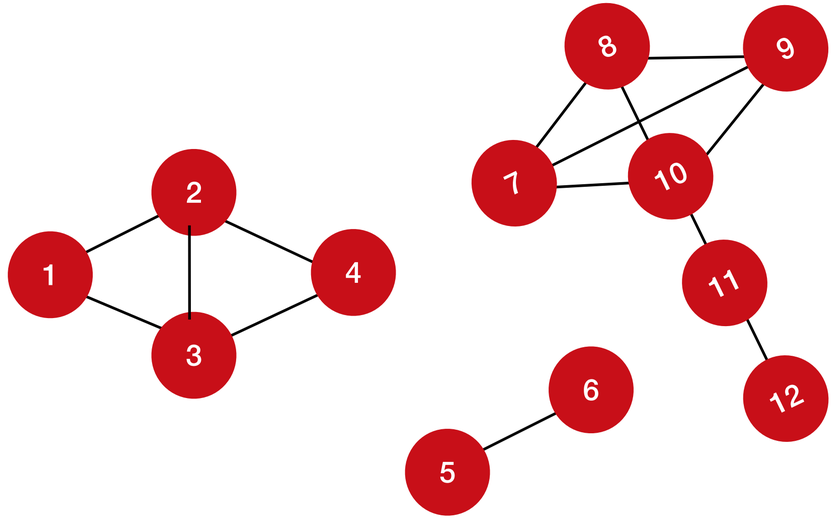
Get summary information about our components

```
comps <- components(G)
```

```
comps$no
```

```
> 3
```

There are 3 components in our graph



Graph G

This will return information about:

- The number of components ($no)
- The size of each component ($csize)
- The component each node is in ($membership)

Extracting Components (igraph)

Get summary information about our components

```
comps <- components(G)
```

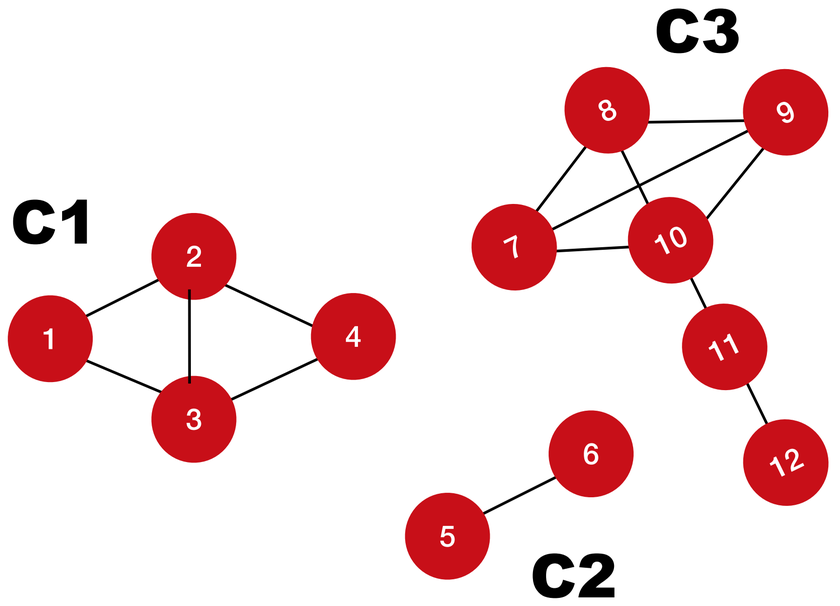
```
comps$csize
```

```
> 4  1  6
```

C1: 4 nodes

C2: 2 nodes

C3: 6 nodes



Graph G

This will return information about:

- The number of components ($no)
- The size of each component ($csize)
- The component each node is in ($membership)

Extracting Components (igraph)

Get summary information about our components

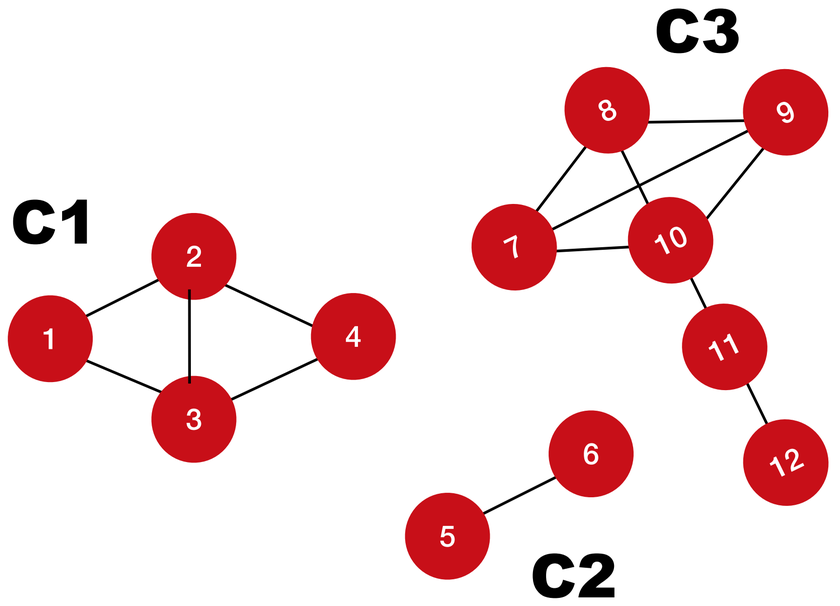
```
comps <- components(G)
```

```
comps$membership
```

```
> 1 1 1 1 2 2 3 3 3 3 3 3
```

The first node is in C1

The last node is in C3



Graph G

This will return information about:

- The number of components ($no)
- The size of each component ($csize)
- The component each node is in ($membership)

Extracting Components (igraph)

Get summary information about our components

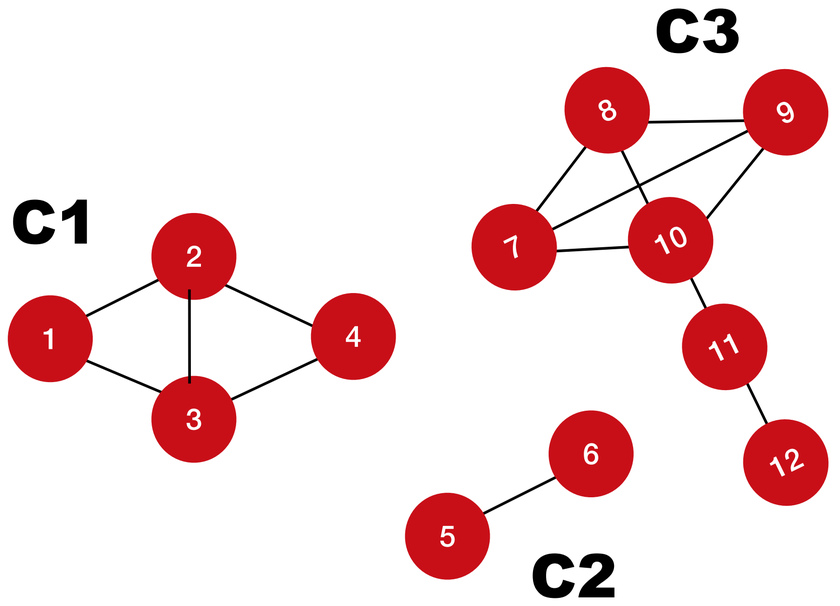
```
comps <- components(G)
```

```
comps$membership
```

```
> 1 1 1 1 2 2 3 3 3 3 3 3
```

Save this as a nodal attribute

```
G <- set_vertex_attr(G,
                     "component",
                     value=comps$membership
)
```



Graph G

Extracting Components (igraph)

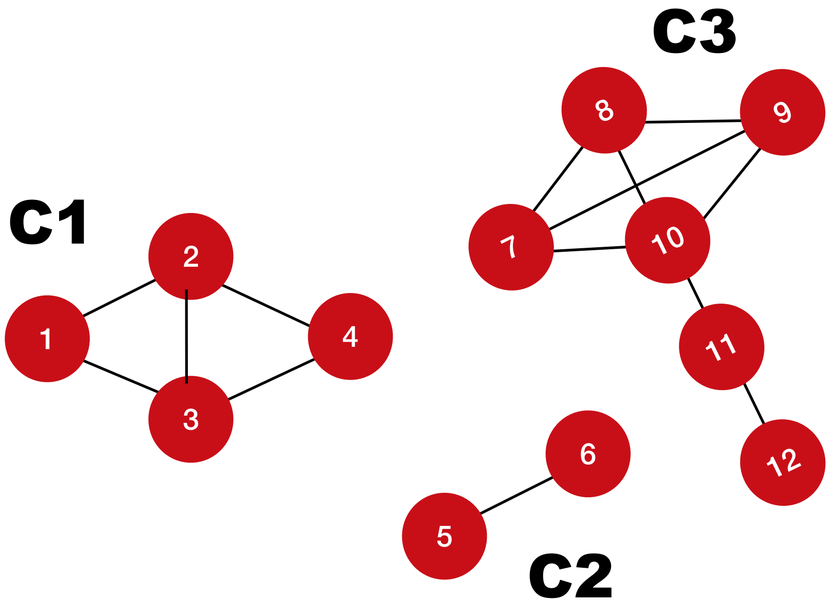
Save component membership as a nodal attribute

```
G <- set_vertex_attr(G,
                     "component",
                     value=comps$membership
)
```

Network object

Name we're giving to attribute

Value of attribute



Graph G

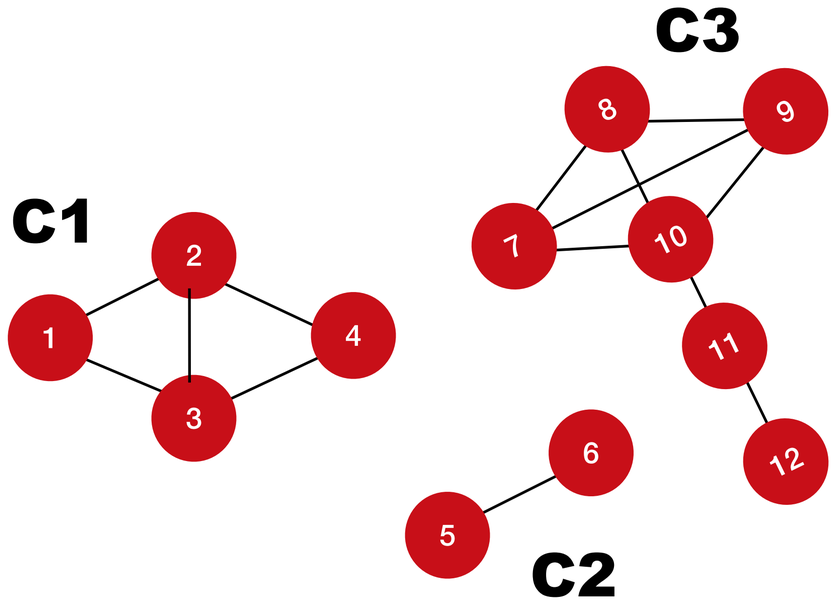
Extracting Components (igraph)

Save component membership as a nodal attribute

```
G <- set_vertex_attr(G,
                     "component",
                     value=comps$membership
)
```

Create a **subgraph** with just the nodes in C3 (membership==3)

```
C3 <- subgraph(G,
               V(G)[V(G)$component==3]
               )
```



Graph G

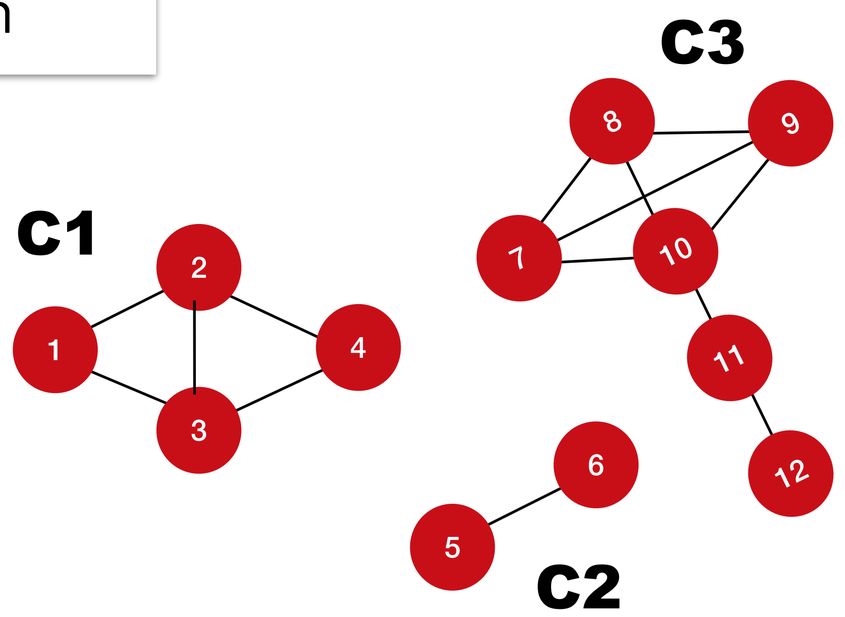
Extracting Components (igraph)

Create a **subgraph** with just the nodes in C3 (membership==3)

```
C3 <- subgraph(G,
                V(G)[V(G)$component==3]
               )
```

Starting graph

Nodes that should be included in the subgraph



Graph G

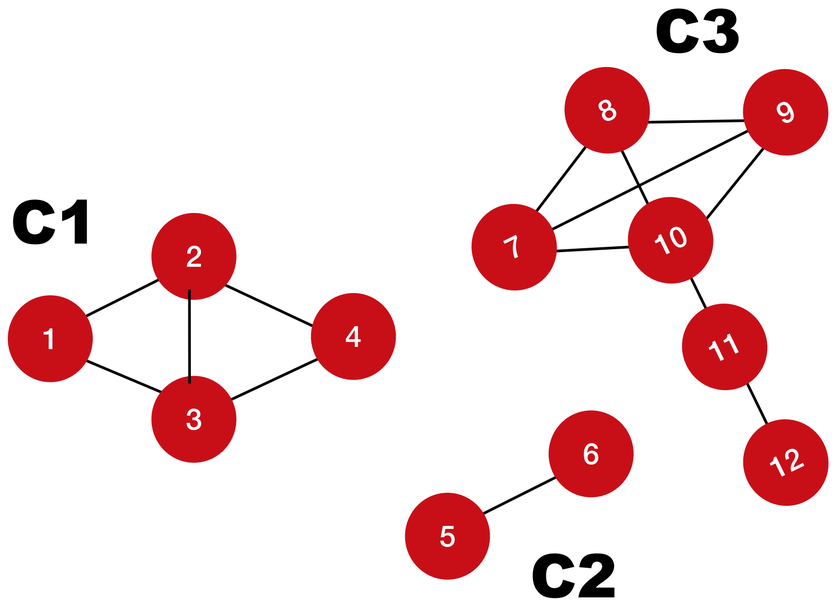
Extracting Components (igraph)

Create a **subgraph** with just the nodes in C3 (membership==3)

```
V(G)$component
```

Vector of node memberships

```
> 1 1 1 1 2 2 3 3 3 3 3 3
```



Graph G

Extracting Components (igraph)

Create a **subgraph** with just the nodes in C3 (membership==3)

```
V(G)$component==3
```

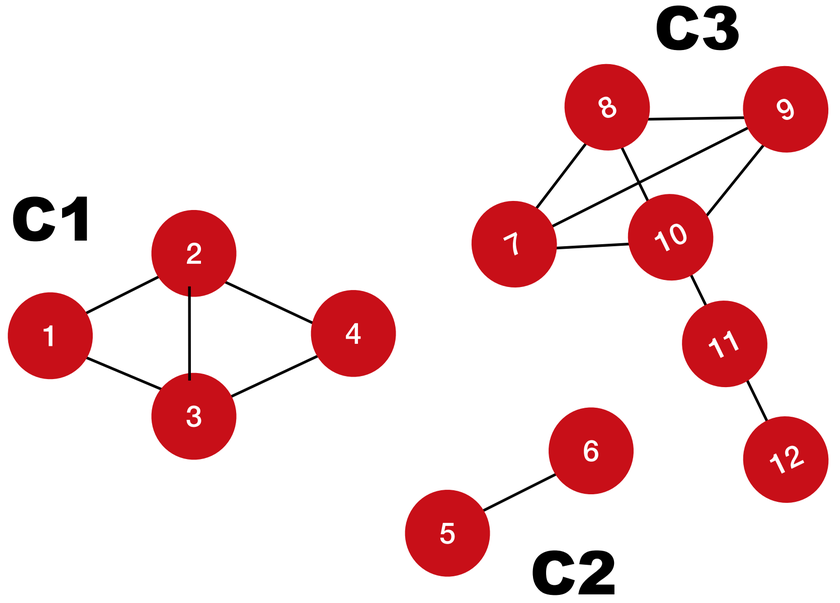
Boolean statement:

Return TRUE if value == 3, else FALSE

Vector of node memberships

```
> 1 1 1 1 2 2 3 3 3 3 3 3
```

```
> F F F F F F T T T T T T
```



Graph G

Extracting Components (igraph)

Create a **subgraph** with just the nodes in C3 (membership==3)

```
V(G)$component==3
```

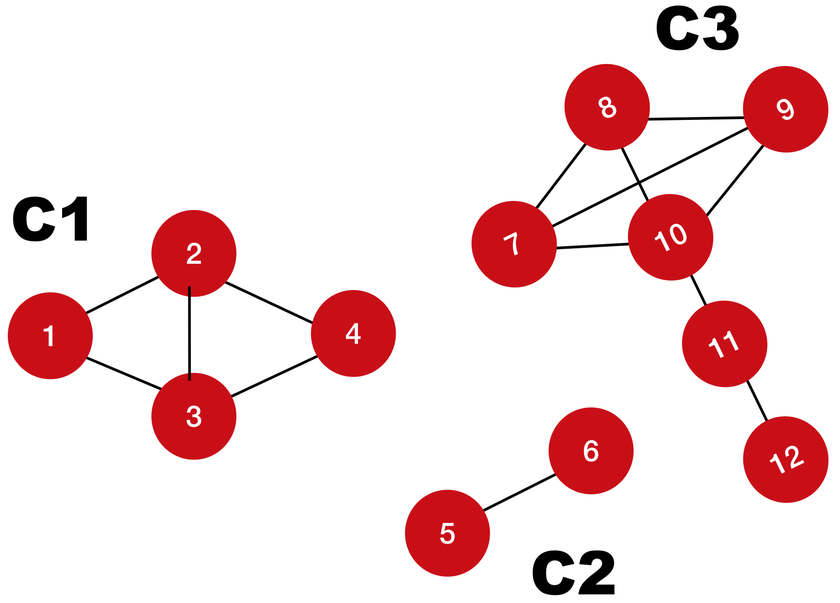
Boolean statement:

Return TRUE if value == 3, else FALSE

Vector of node memberships

```
> 1 1 1 1 2 2 3 3 3 3 3 3
```

```
> F F F F F F T T T T T T
```



Graph G

Extracting Components (igraph)

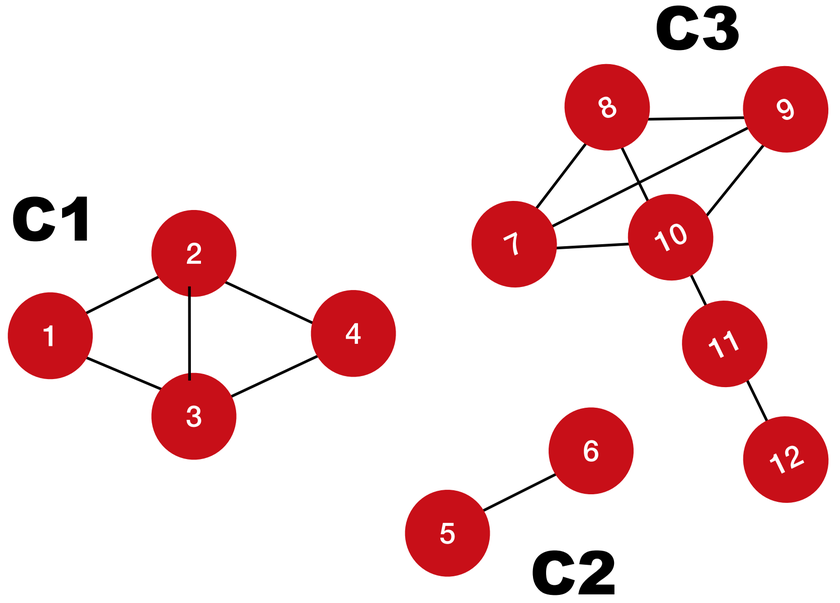
Create a **subgraph** with just the nodes in C3 (membership==3)

```
V(G)[V(G)$component==3]
```

Subset to only nodes where the boolean statement returns TRUE

Vector with TRUE if a node's component == 3, else FALSE

```
> F F F F F F T T T T T T
```



Graph G

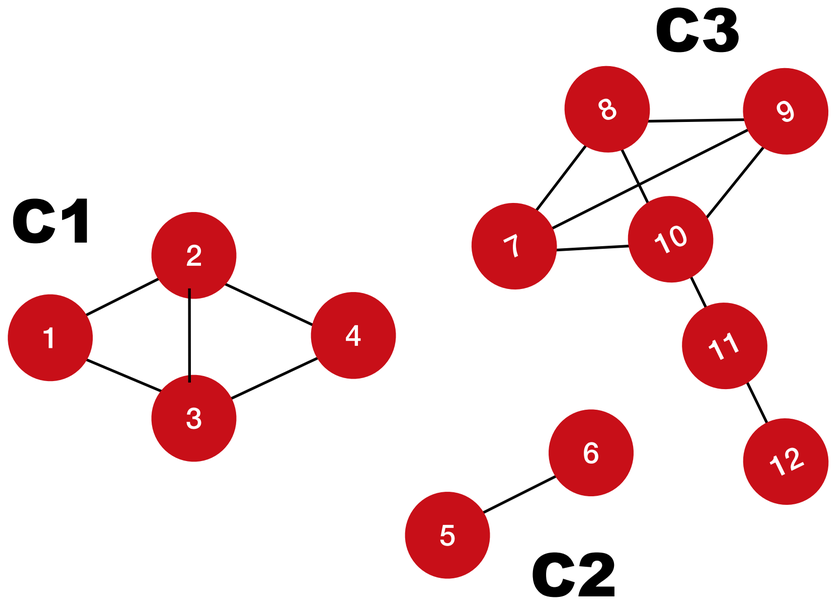
Extracting Components (igraph)

Create a **subgraph** with just the nodes in C3 (membership==3)

```
V(G)[V(G)$component==3]
```

Subset to only nodes where the boolean statement returns TRUE

Node	1	2	3	4	5	6	7	8	9	10	11	12
Value	F	F	F	F	F	F	**T**	**T**	**T**	**T**	**T**	**T**



Graph G

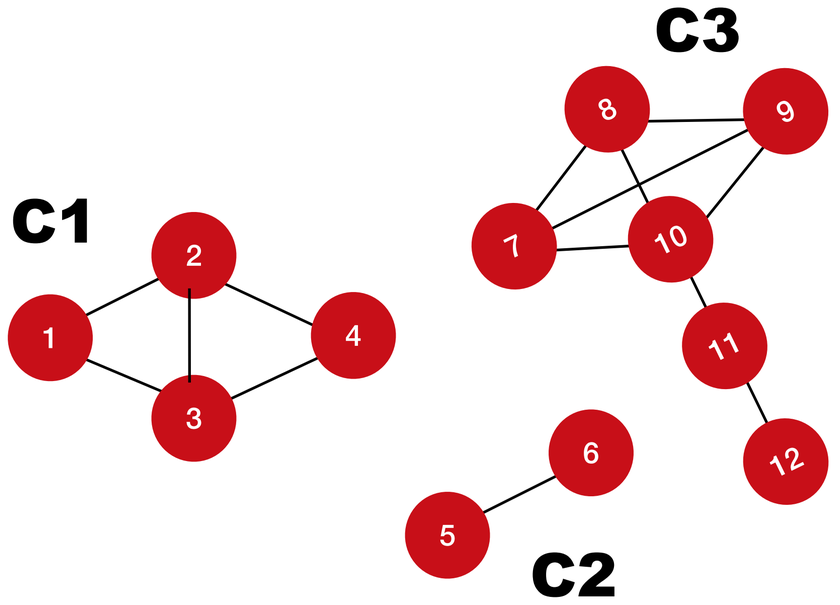
Extracting Components (igraph)

The subgraph() function will return an igraph object over the specified nodes

```
C3 <- subgraph(G,
                V(G)[V(G)$component==3]
               )
```

Subset to only nodes where the boolean statement returns TRUE

Node	1	2	3	4	5	6	7	8	9	10	11	12
Value	F	F	F	F	F	F	**T**	**T**	**T**	**T**	**T**	**T**



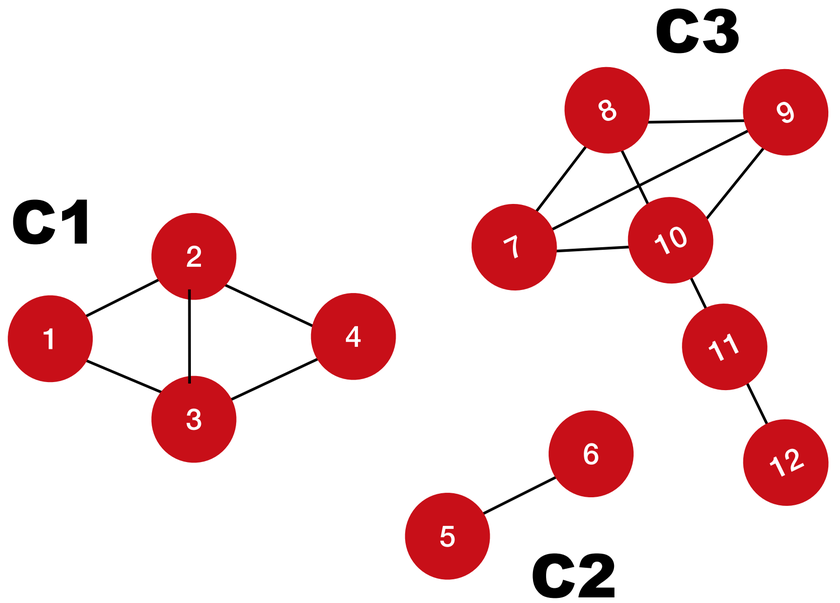
Graph G

Extracting the LARGEST component

What if we're not sure what number is associated with the largest component?

The decompose() function will return a list of all subgraphs as igraph objects

```
components <- decompose(G)
```



Graph G

Extracting the LARGEST component

What if we're not sure what number is associated with the largest component?

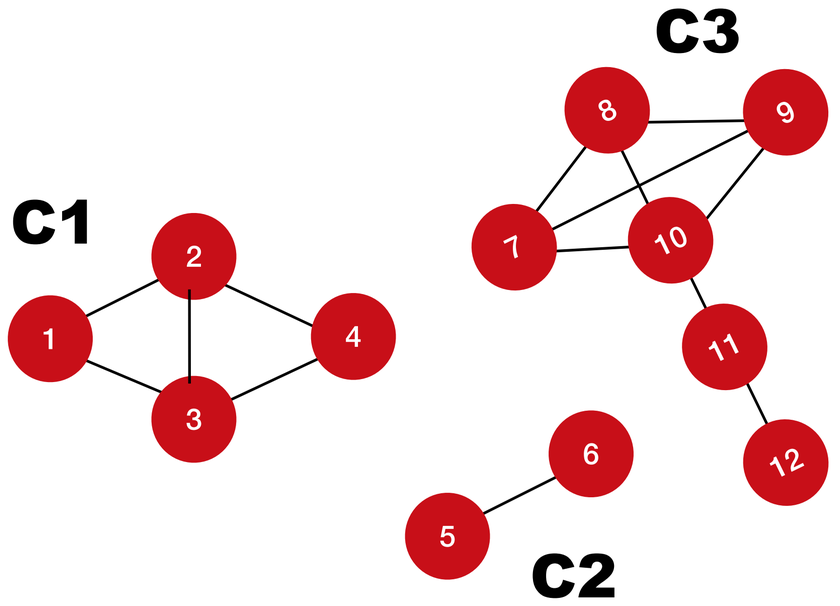
The decompose() function will return a list of all subgraphs as igraph objects

```
components <- decompose(G)
```

Check how many components we get

```
length(components)
```

```
> 3
```



Graph G

Extracting the LARGEST component

What if we're not sure what number is associated with the largest component?

The decompose() function will return a list of all subgraphs as igraph objects

```
components <- decompose(G)
```

Check the number of nodes in each component subgraph

```
gorder(components[[1]])
```

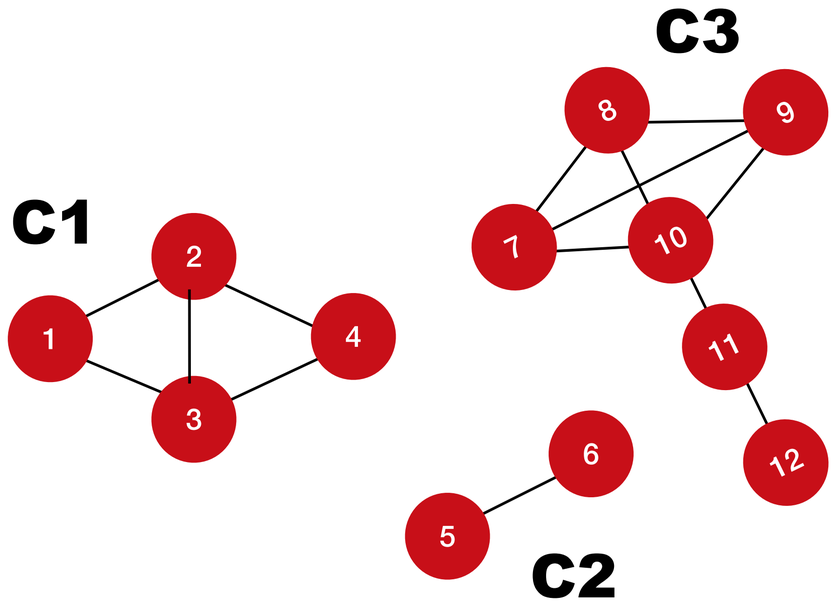
```
> 4
```

```
gorder(components[[2]])
```

```
> 2
```

```
gorder(components[[3]])
```

```
> 6
```



Graph G

Extracting the LARGEST component

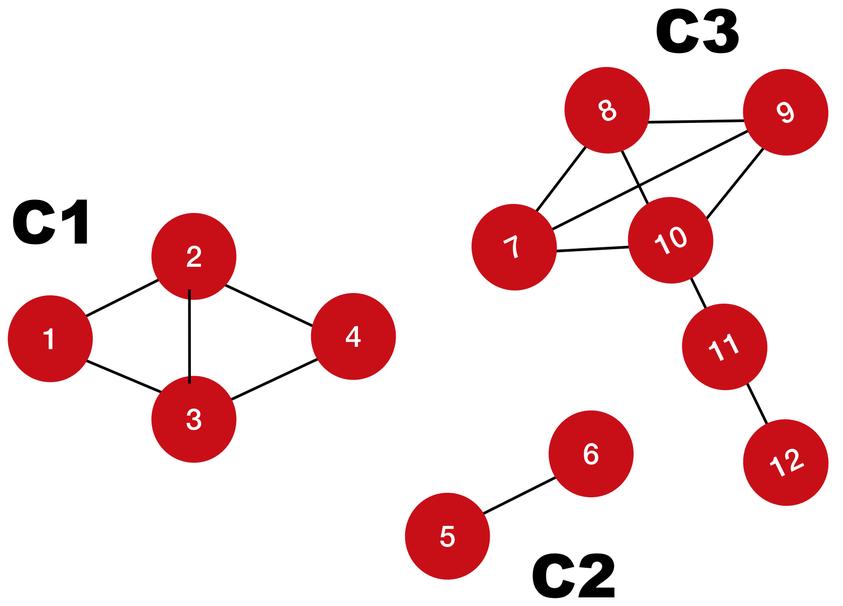
What if we're not sure what number is associated with the largest component?

We can also specify the minimum number of nodes (`min.vertices`) which must be in a subgraph in order for it to be returned

```
comps <- components(G)
```

```
comps$csize
```

```
> 4  1  6
```



Graph G

```
LC <- decompose(G,
                min.vertices=max(comps$csize))[[1]]
```

Only return subgraphs with at least `max(comps$csize)` nodes

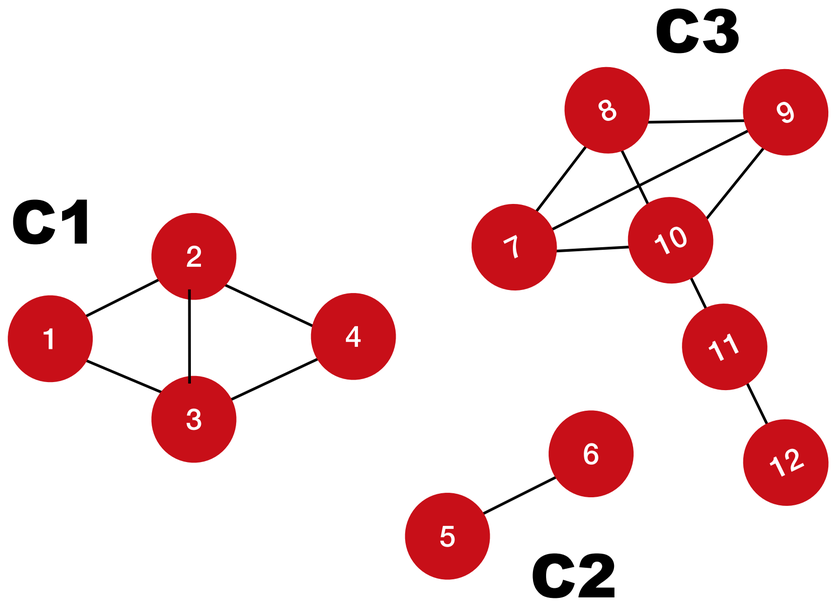
Extracting the LARGEST component

```
LC <- decompose(G,
                min.vertices=max(comps$csize))[[1]]
```

Check how many nodes are in our new graph

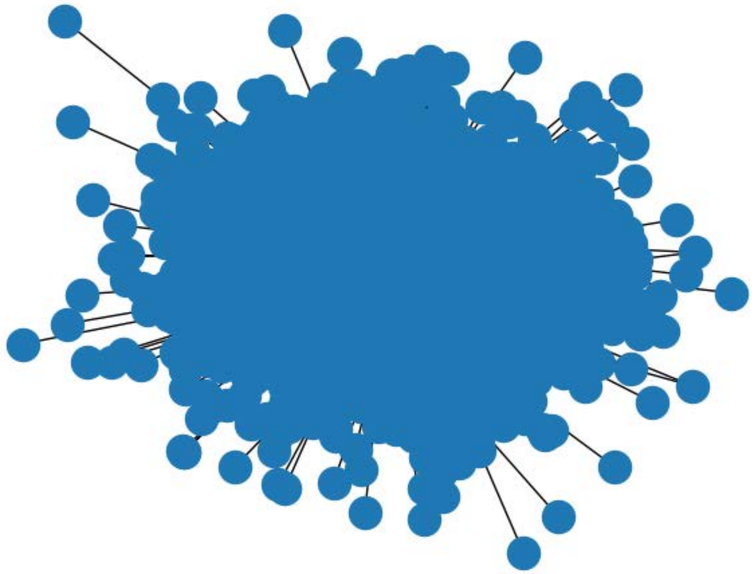
```
gorder(LC)
```

```
> 6
```



Graph G

...With real data



The largest component in my
topical co-occurance network

Two approaches to sparsifying a network

K-core decomposition

Disparity filter network backbone

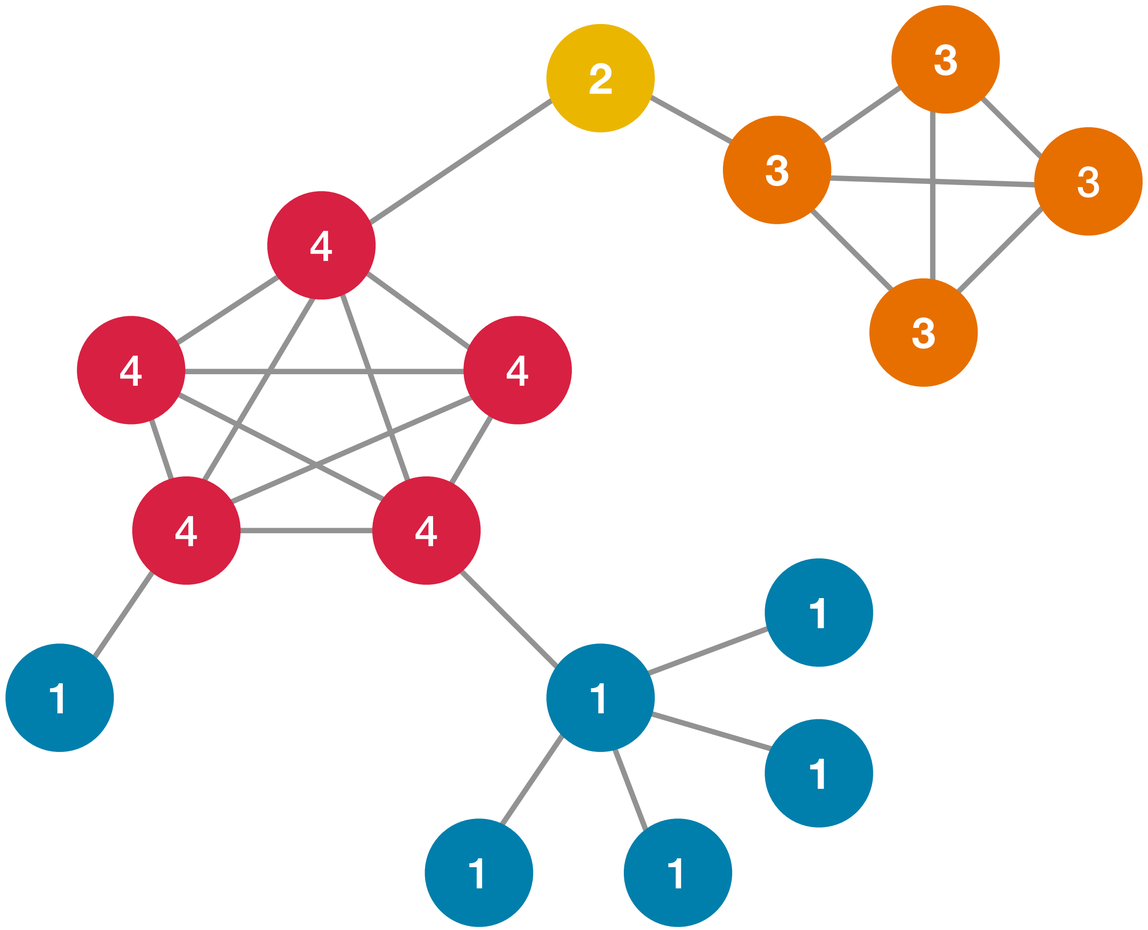
A statistical test for significant edges

Extracting the multiscale **backbone** of complex weighted **networks**
MÁ **Serrano**, M Boguná… - Proceedings of the …, 2009 - National Acad Sciences
… In the following, we will refer to the **network** reduction as the construction of a **network** that … the relevant features of the original **networks**; for instance, the traffic **backbone** of a large-scale …
☆ Save 99 Cite Cited by 790 Related articles All 23 versions

PSA: Generally a bad idea to use a global threshold (ie, edge must have weight > x)

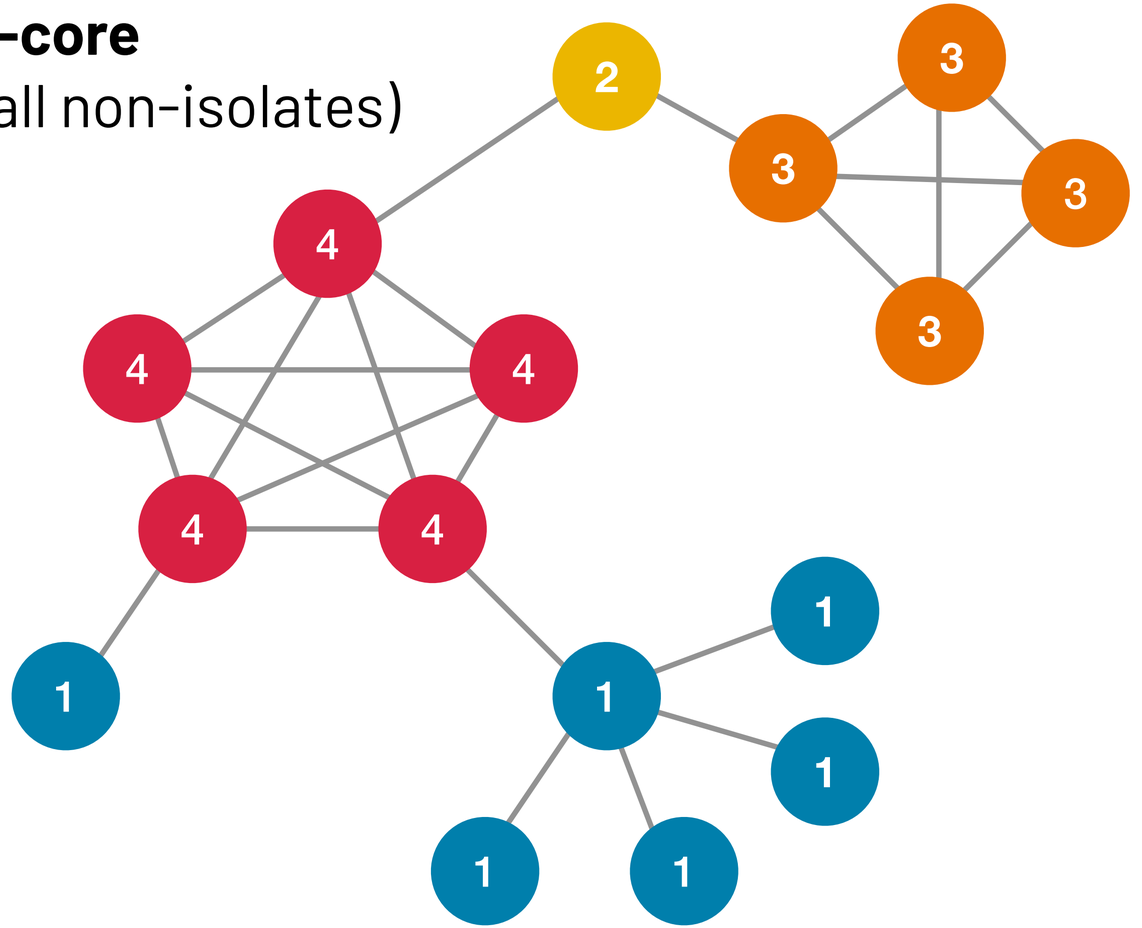
k-core decomposition

Nodes and their k-core



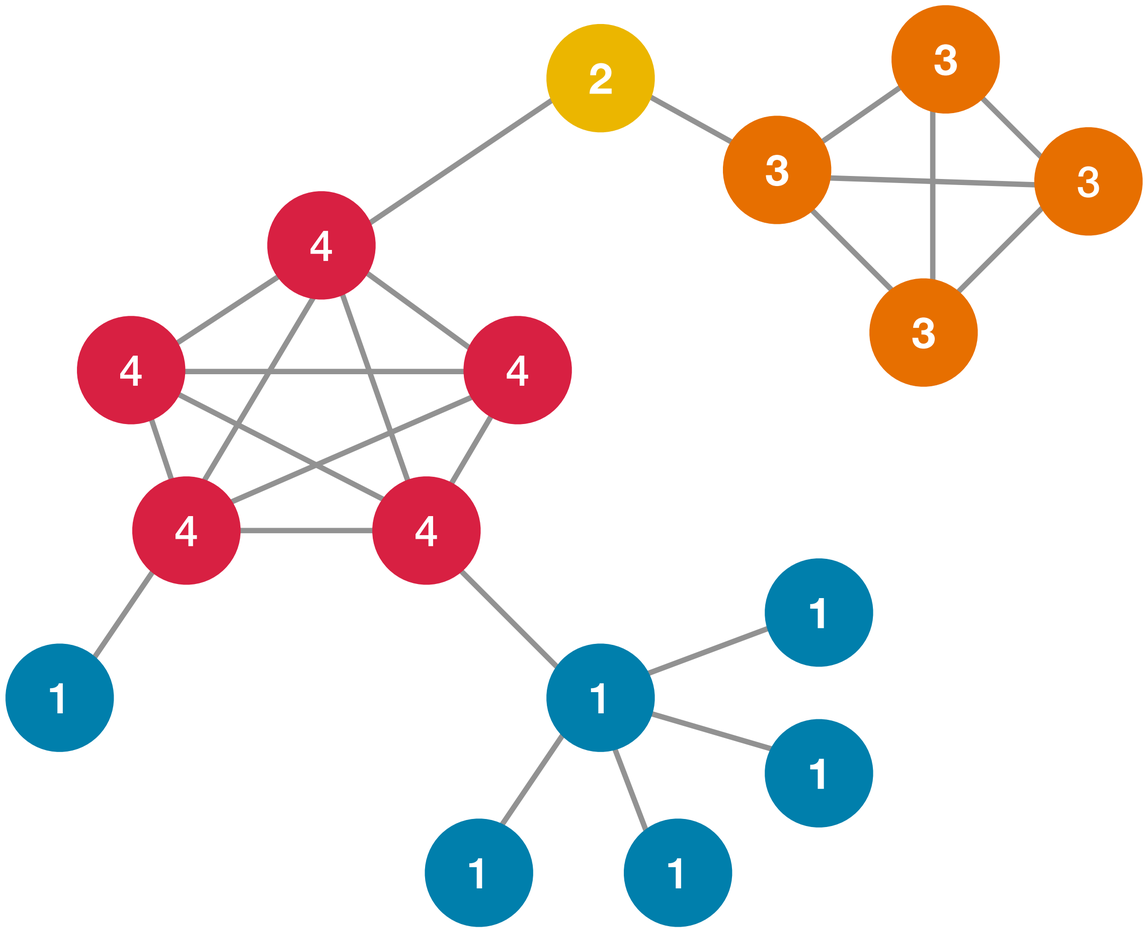
A **k-core** is a **subgraph** where every node has at least coreness k

1-core
(all non-isolates)



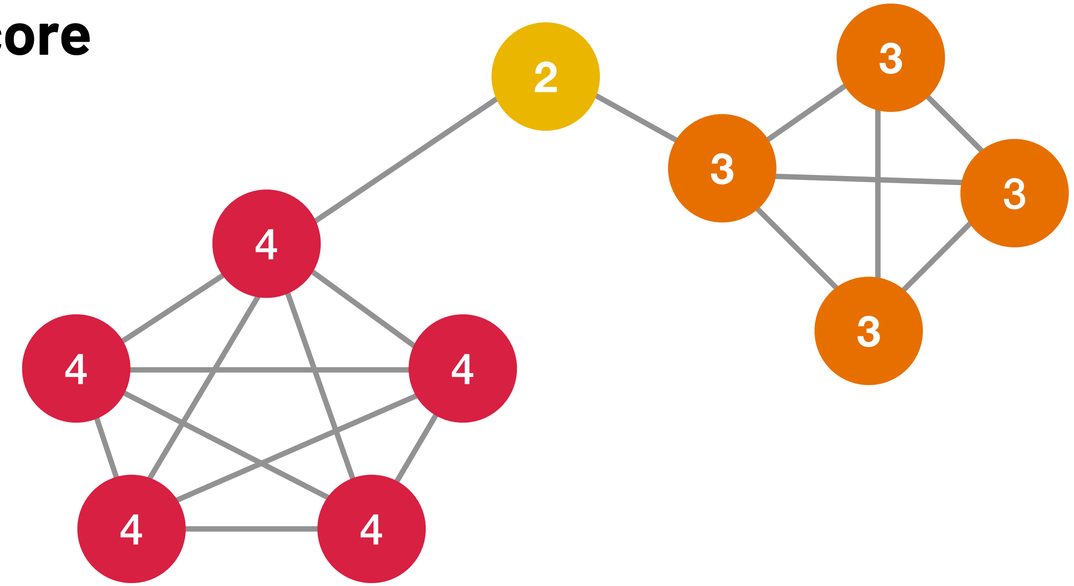
k-core decomposition

Nodes and their k-core



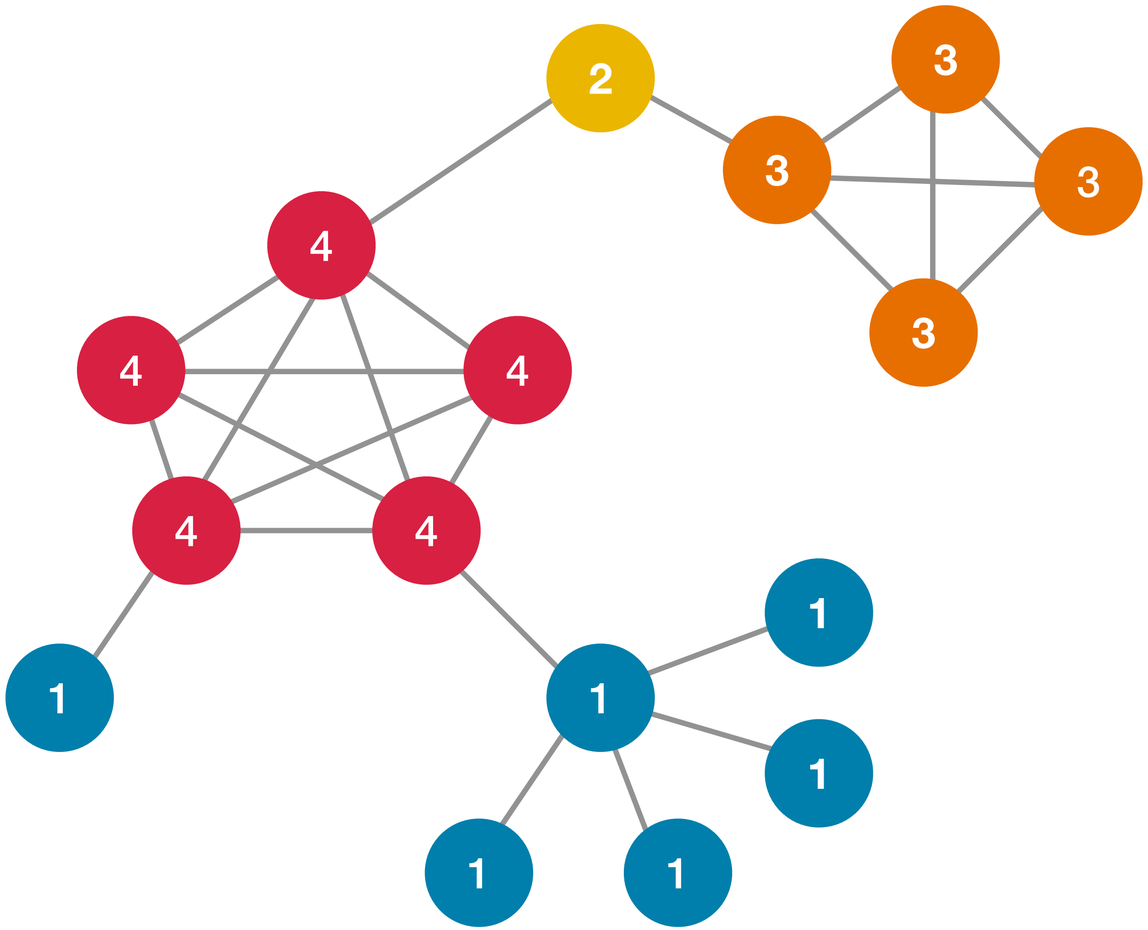
A **k-core** is a **subgraph** where every node has at least coreness k

2-core



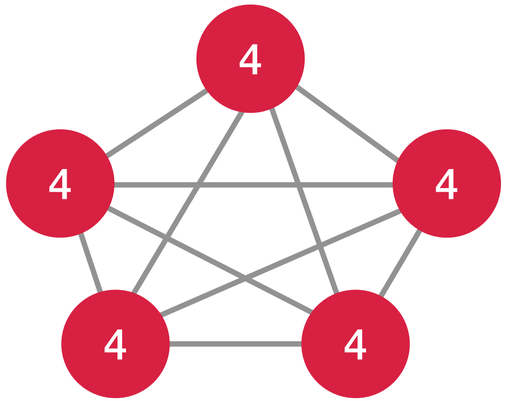
k-core decomposition

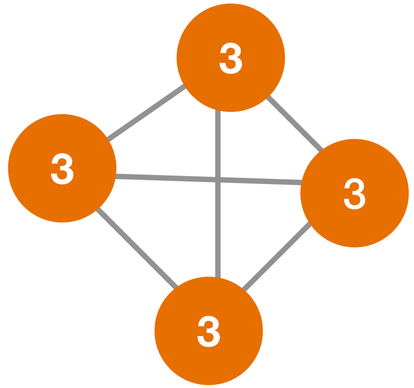
Nodes and their k-core



A **k-core** is a **subgraph** where every node has at least coreness k

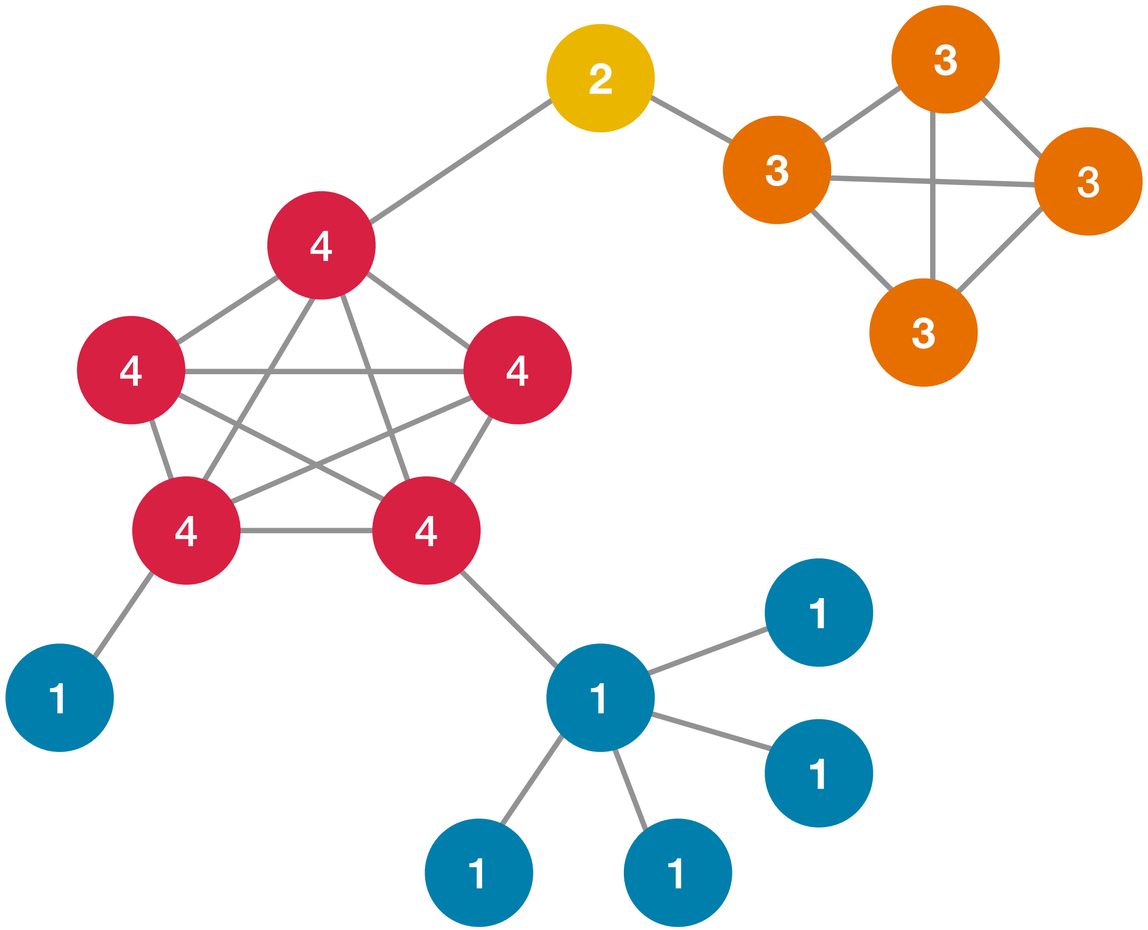
3-core





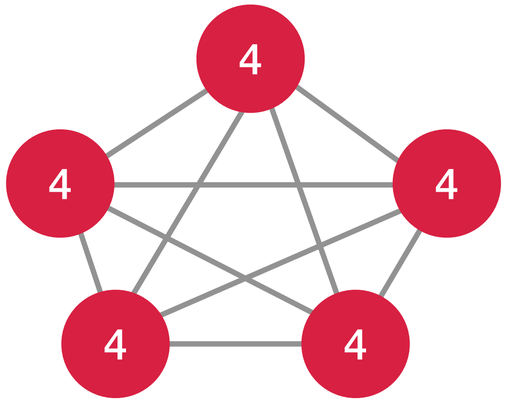
k-core decomposition

Nodes and their k-core



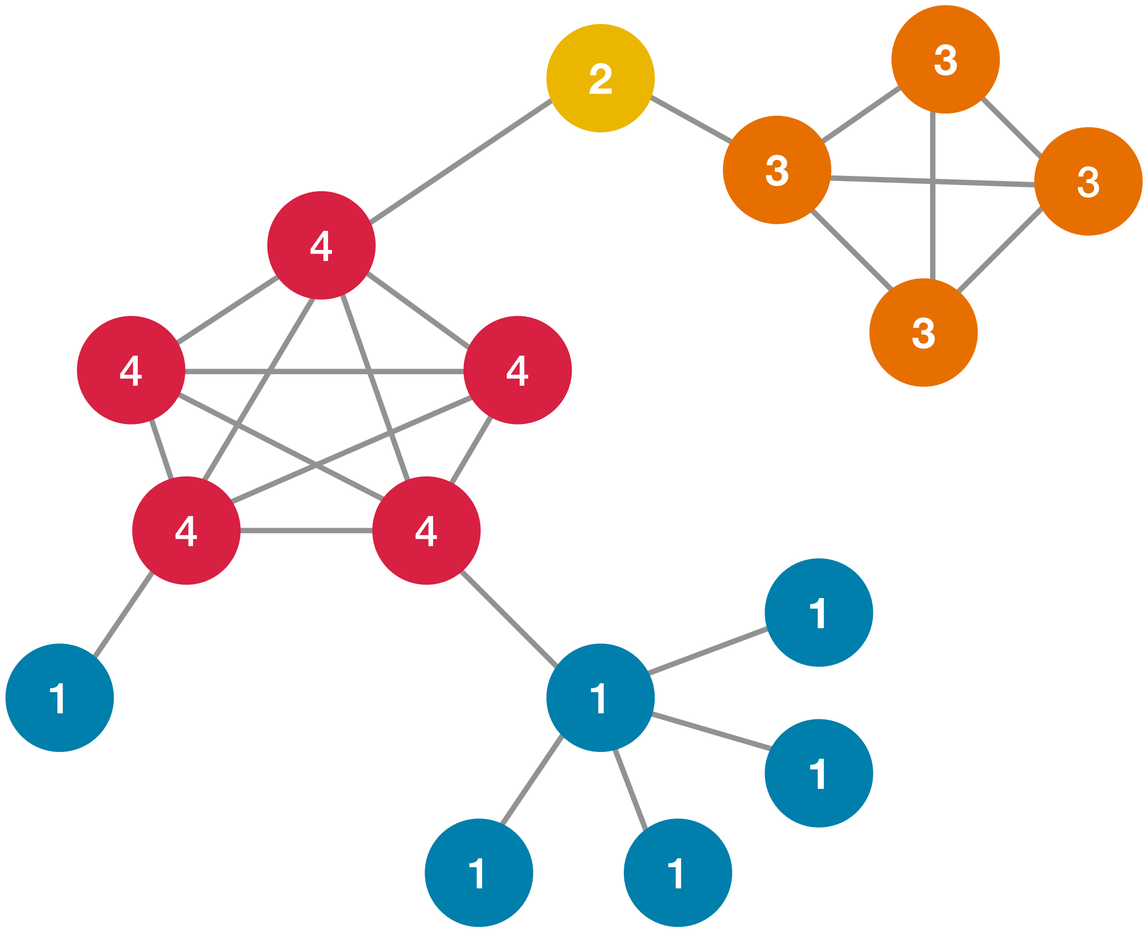
A **k-core** is a **subgraph** where every node has at least coreness k

4-core



k-core decomposition

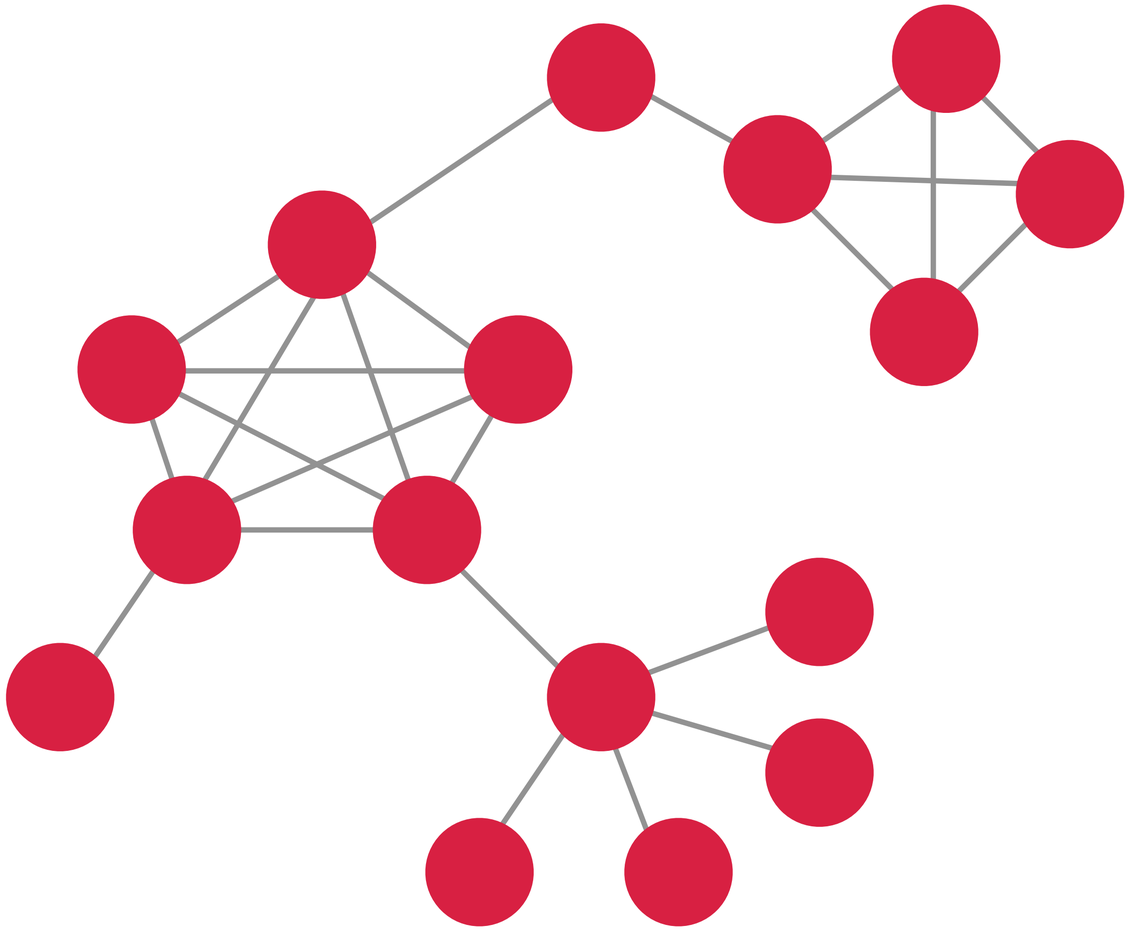
Nodes and their k-core



A **k-core** is a **subgraph** where every node has at least coreness k

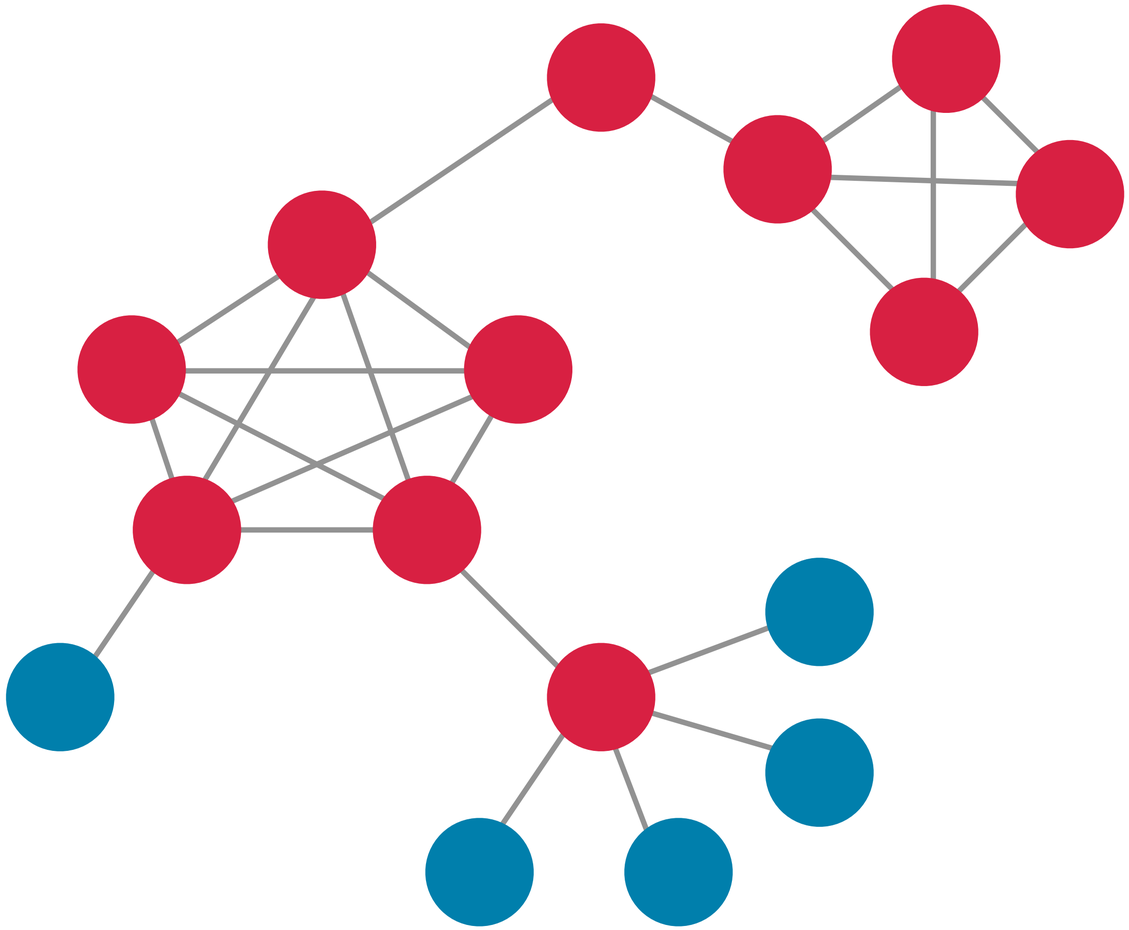
5-core

k-core decomposition



Identify nodes with degree = 1

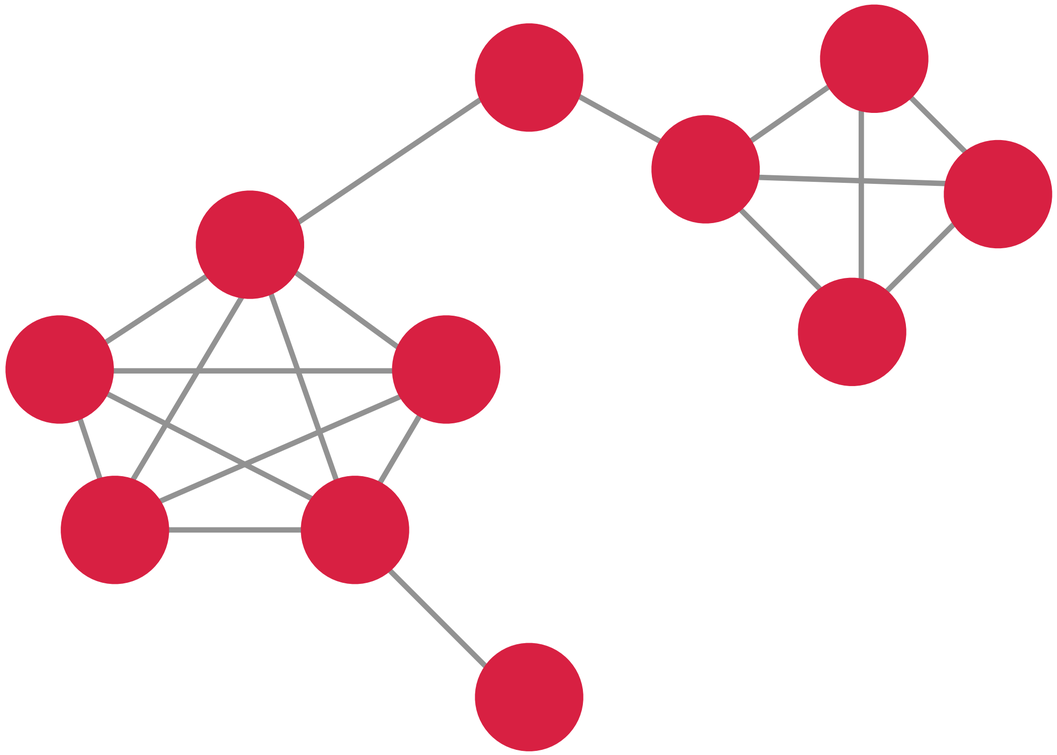
k-core decomposition



Identify nodes with degree = 1

➡ Remove nodes with degree = 1

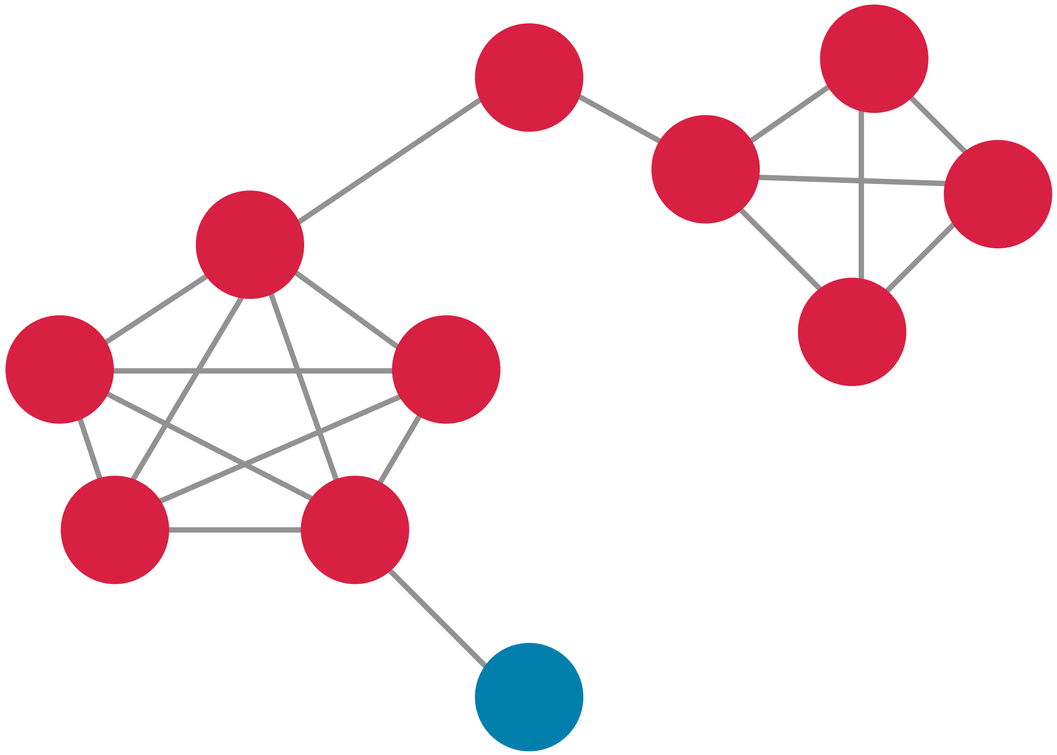
k-core decomposition



Identify nodes with degree = 1

- ➡ Remove nodes with degree = 1
- ➡ Identify and remove and nodes that *now* have degree = 1

k-core decomposition

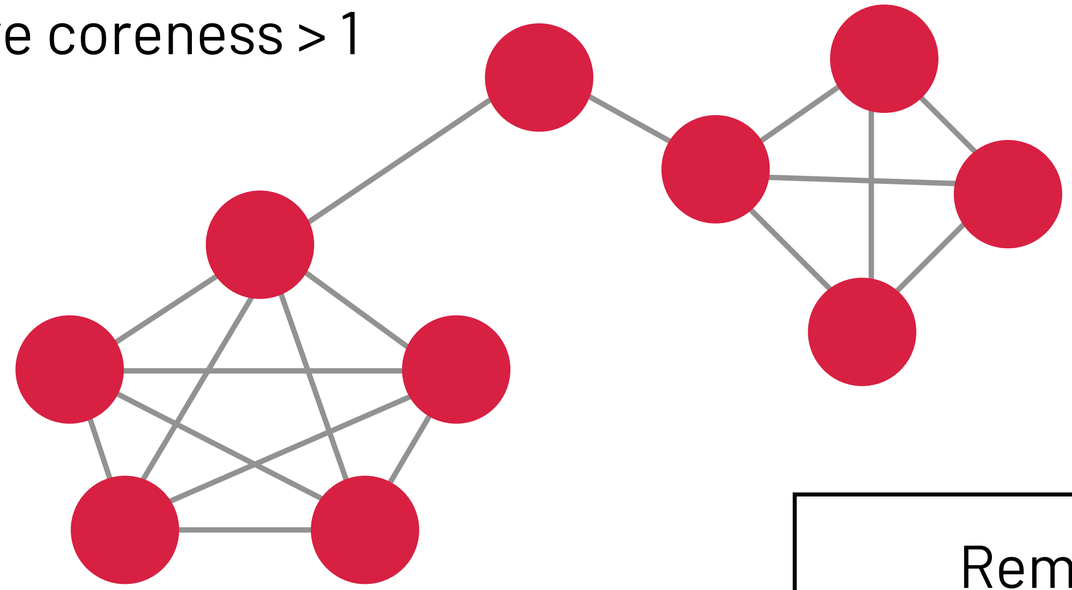


Identify nodes with degree = 1

- ➡ Remove nodes with degree = 1
- ➡ Identify and remove and nodes that *now* have degree = 1

k-core decomposition

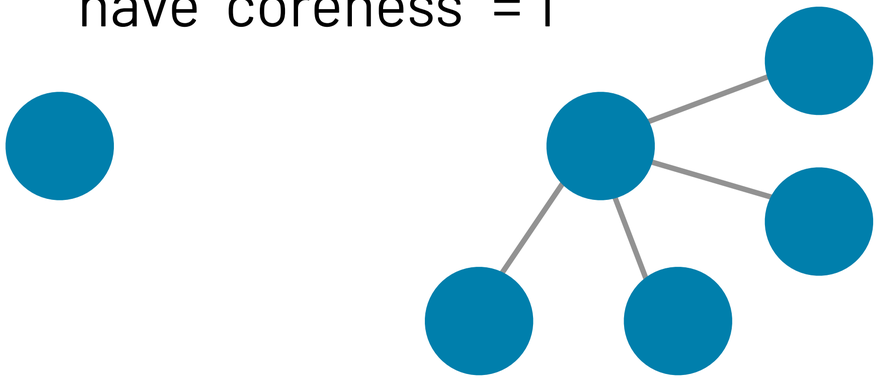
Remaining nodes
have coreness > 1



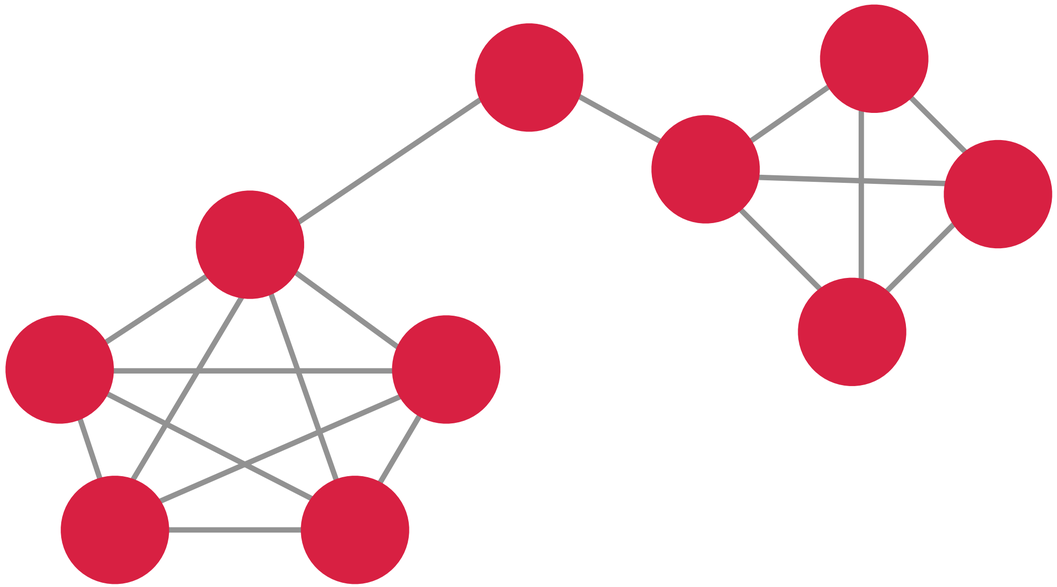
Identify nodes with degree = 1

- ➡ Remove nodes with degree = 1
- ➡ Identify and remove and nodes that *now* have degree = 1
- ➡ Do this iteratively until there are no nodes with degree = 1

Removed nodes
have “coreness” = 1

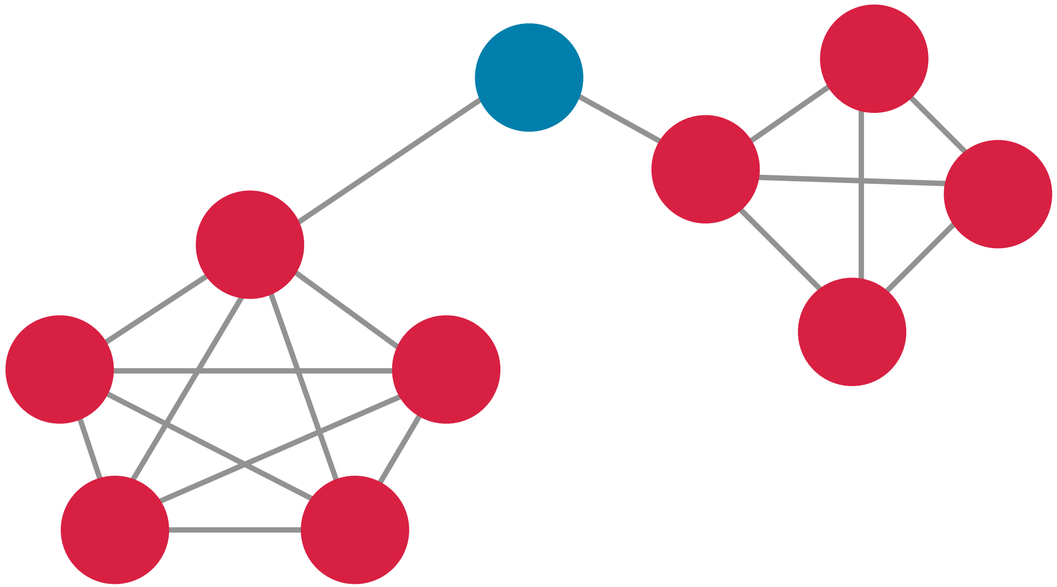


k-core decomposition



In our remaining network, identify nodes with degree = **2**

k-core decomposition

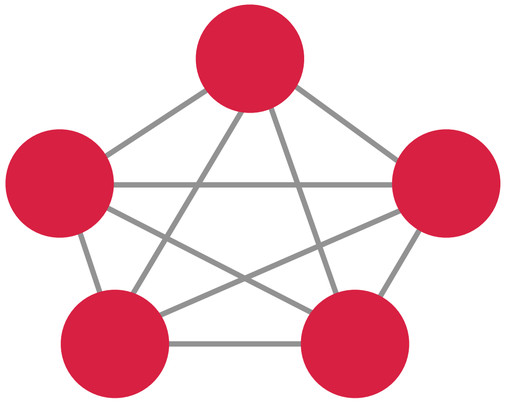


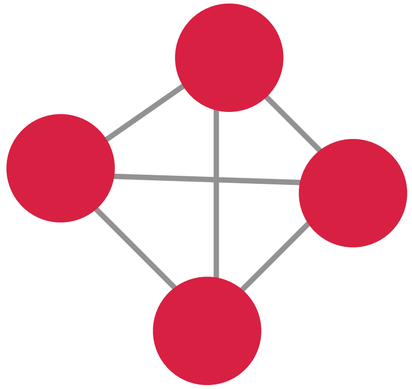
In our remaining network,
identify nodes with degree = **2**

➡ Remove nodes with degree = 2

k-core decomposition

Remaining nodes
have coreness > 2



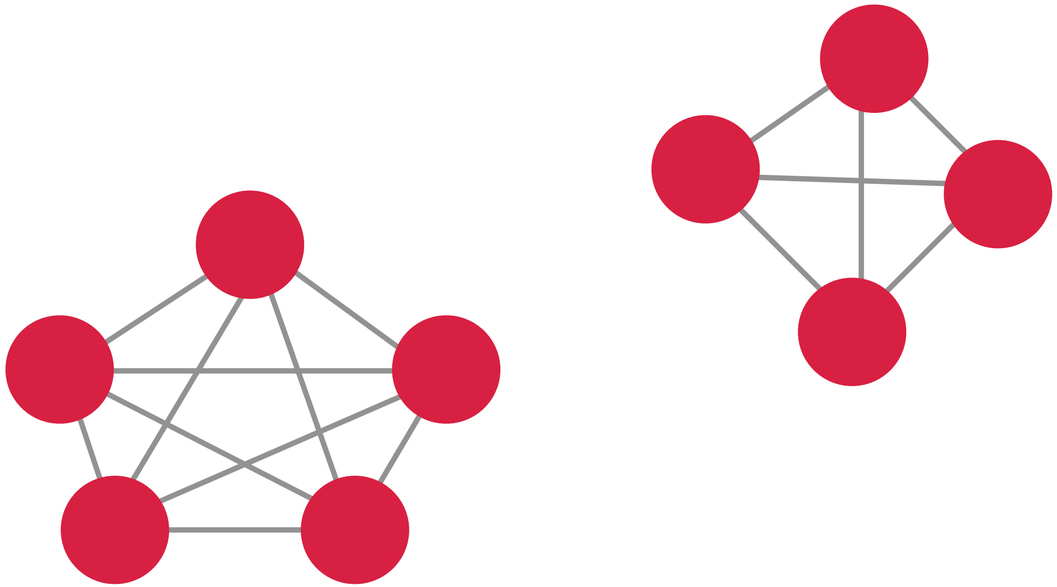


In our remaining network,
identify nodes with degree = **2**

- ➡ Remove nodes with degree = 2
- ➡ Do this iteratively until there are no nodes with degree = 2

Nodes removed in this
step have coreness = 2

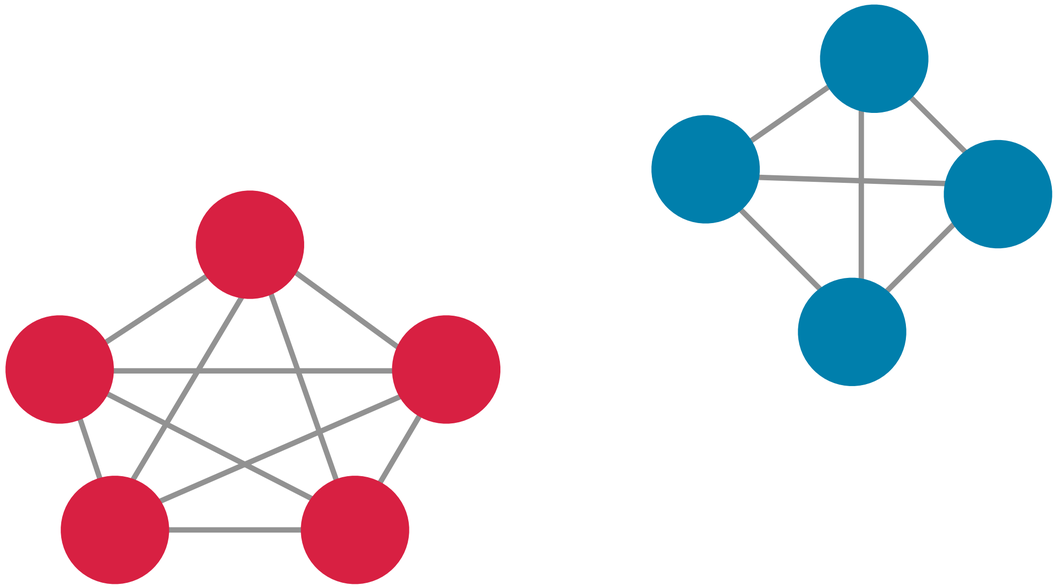
k-core decomposition



In our remaining network, identify nodes with degree = **3**

- ➡ Remove nodes with degree = **3**
- ➡ Do this iteratively until there are no nodes with degree = **3**

k-core decomposition

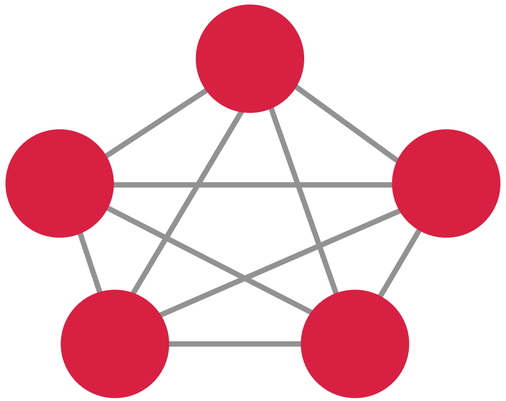


In our remaining network, identify nodes with degree = **3**

- ➡ Remove nodes with degree = **3**
- ➡ Do this iteratively until there are no nodes with degree = **3**

k-core decomposition

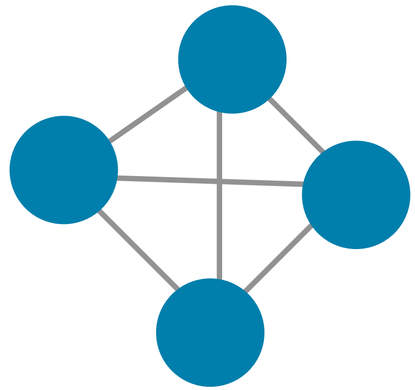
Remaining nodes
have coreness > 4



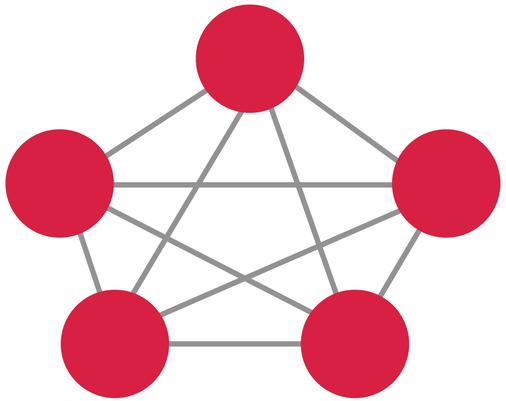
In our remaining network,
identify nodes with degree = **3**

- ➡ Remove nodes with degree = **3**
- ➡ Do this iteratively until there are no nodes with degree = **3**

Nodes removed in this
step have coreness = 3



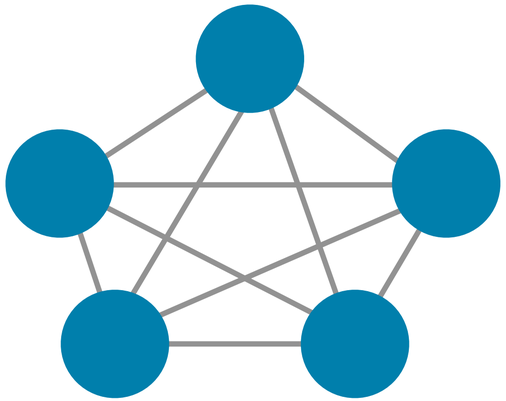
k-core decomposition



In our remaining network, identify nodes with degree = **4**

- ➡ Remove nodes with degree = **4**
- ➡ Do this iteratively until there are no nodes with degree = **4**

k-core decomposition



In our remaining network, identify nodes with degree = **4**

- ➡ Remove nodes with degree = **4**
- ➡ Do this iteratively until there are no nodes with degree = **4**

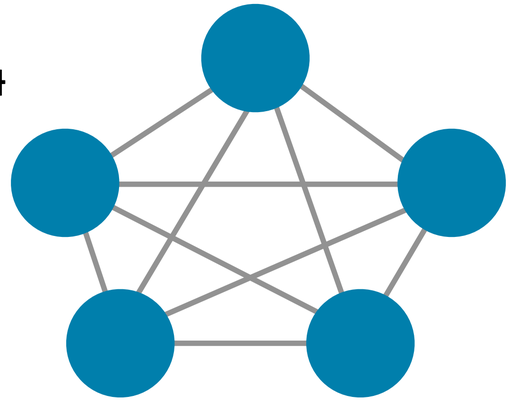
k-core decomposition

Once we have no more nodes, we are done!

In our remaining network, identify nodes with degree = **4**

- ➡ Remove nodes with degree = **4**
- ➡ Do this iteratively until there are no nodes with degree = **4**

Nodes removed in this step have coreness = 4



Finding k-cores (igraph)

```
library(igraph)
```

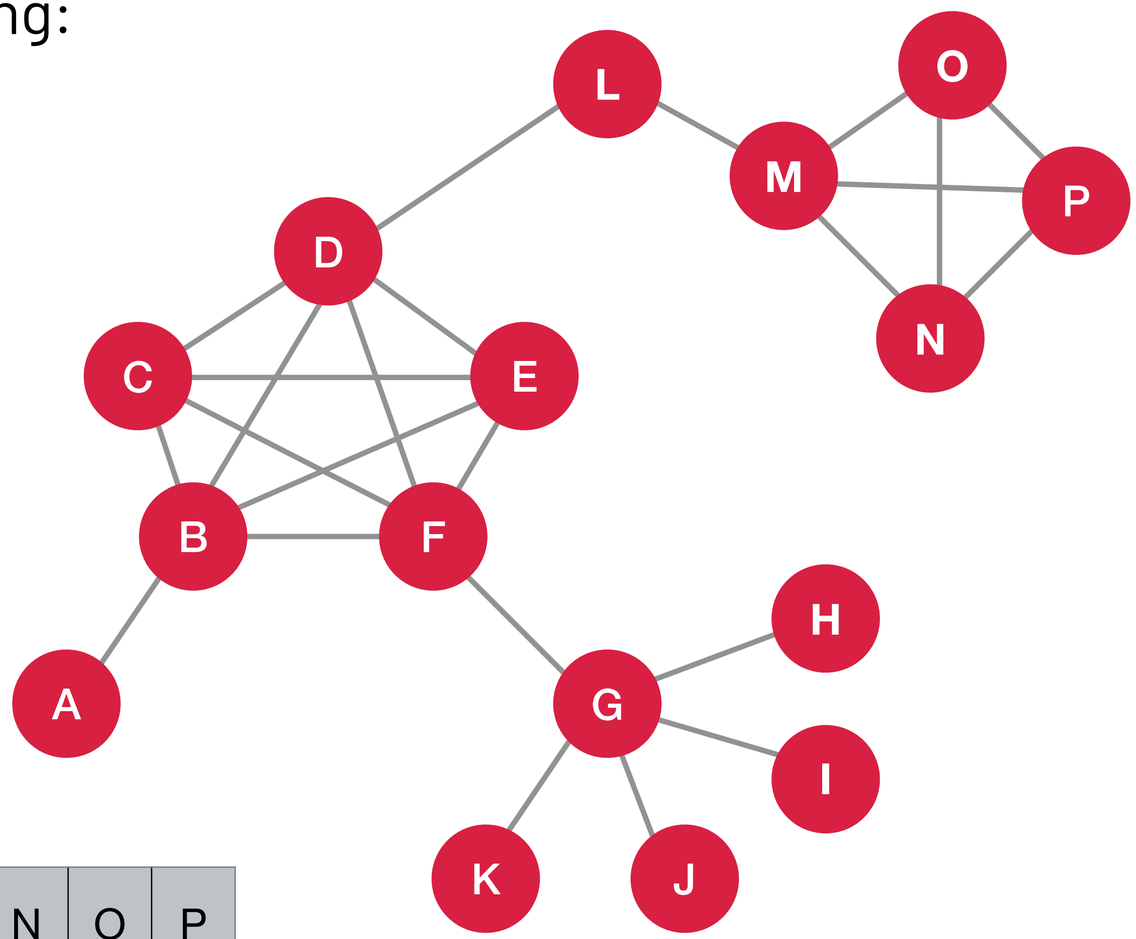
Calculate the coreness of every node in G using:

```
coreness(G)
```

```
A B C D E F G H I J K L M N O P
1 4 4 4 4 4 1 1 1 1 1 2 3 3 3 3
```

Gives us the coreness of every node

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core	1	4	4	4	4	4	1	1	1	1	1	2	3	3	3	3



Network G

Finding k-cores (igraph)

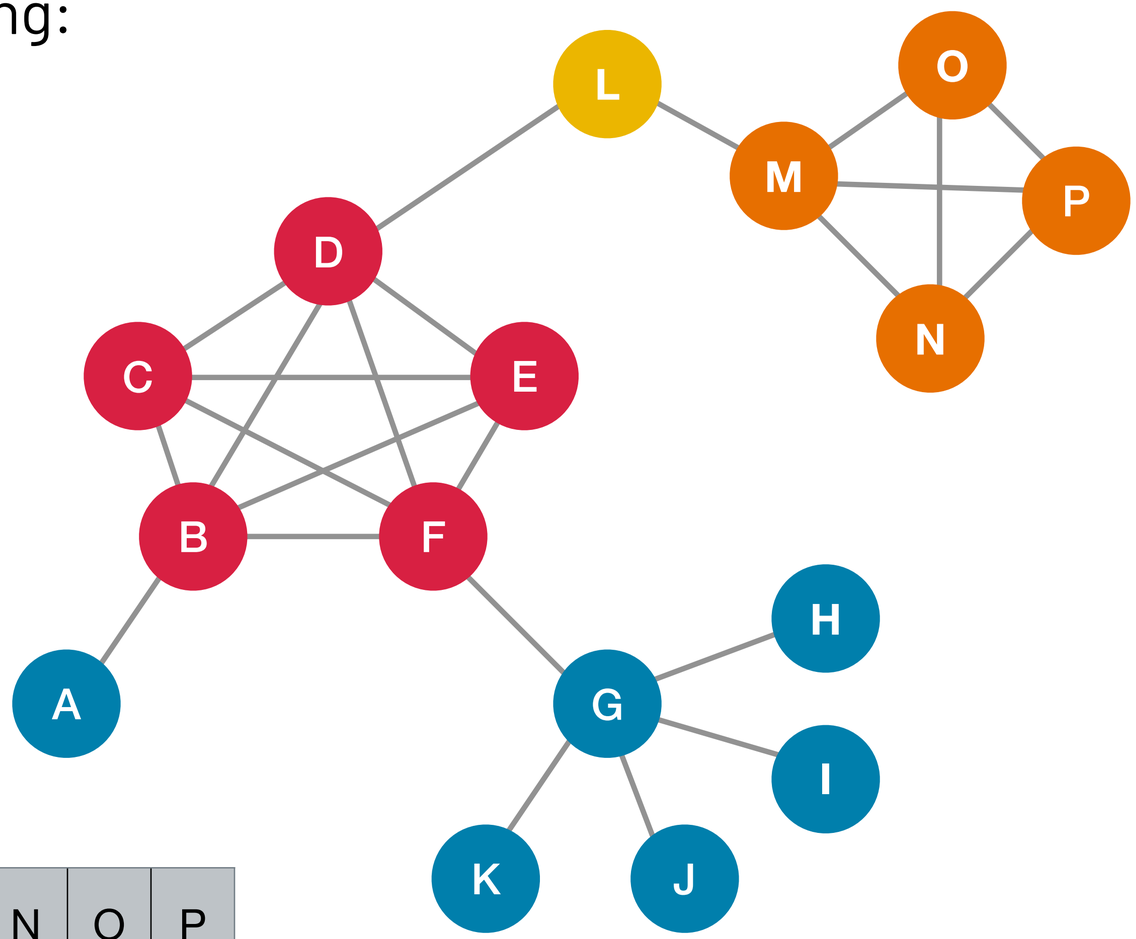
Calculate the coreness of every node in G using:

```
coreness(G)
```

```
A B C D E F G H I J K L M N O P
1 4 4 4 4 4 1 1 1 1 1 2 3 3 3 3
```

Gives us the coreness of every node

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core	1	4	4	4	4	4	1	1	1	1	1	2	3	3	3	3



Network G

Finding k-cores (igraph)

Calculate the coreness of every node in G using:

```
coreness(G)
```

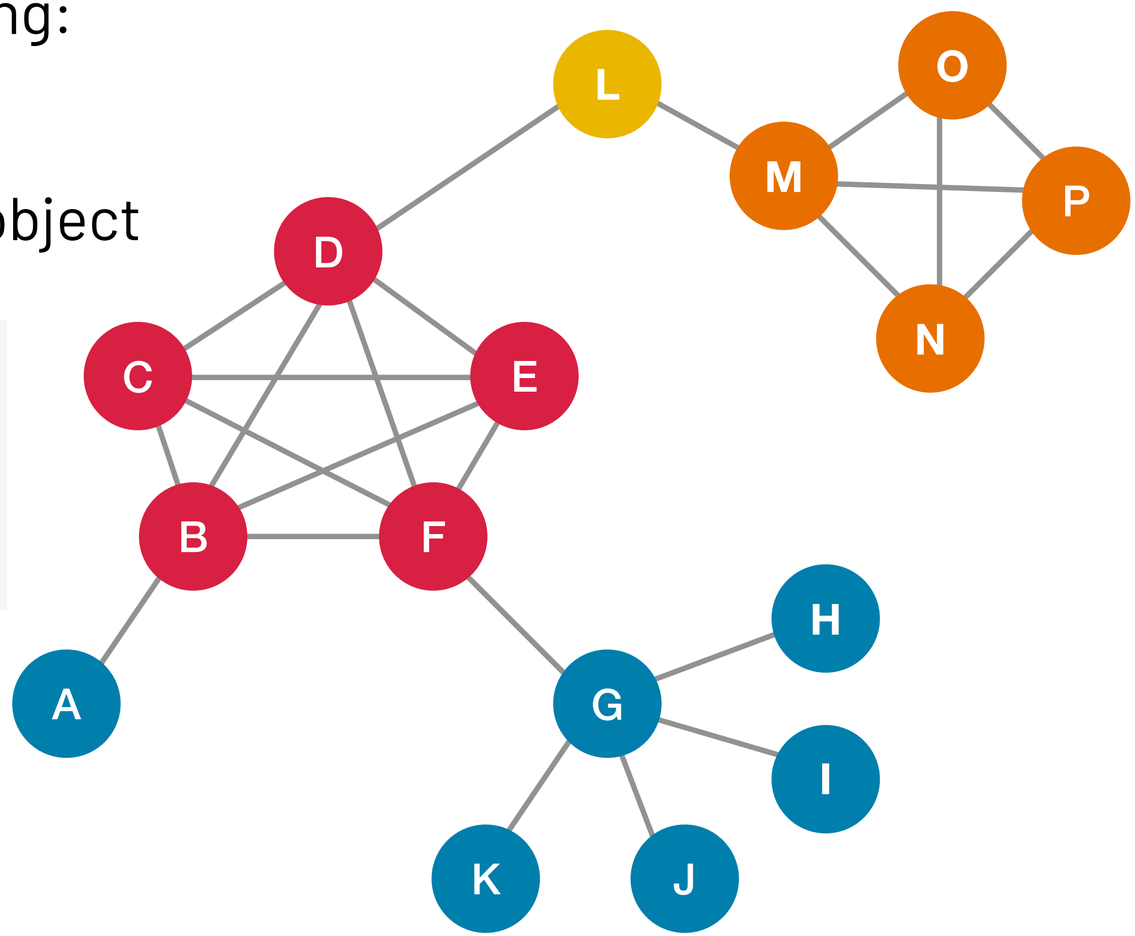
Save coreness a node attribute:

Network object

```
G <- set_vertex_attr(G,
                     "core",
                     value=coreness(G)
)
```

Name we're giving to attribute

Value of attribute



Network G

Plotting k-cores (igraph)

Save coreness a node attribute:

```
G <- set_vertex_attr(G, "core",
                     value=coreness(G))
```

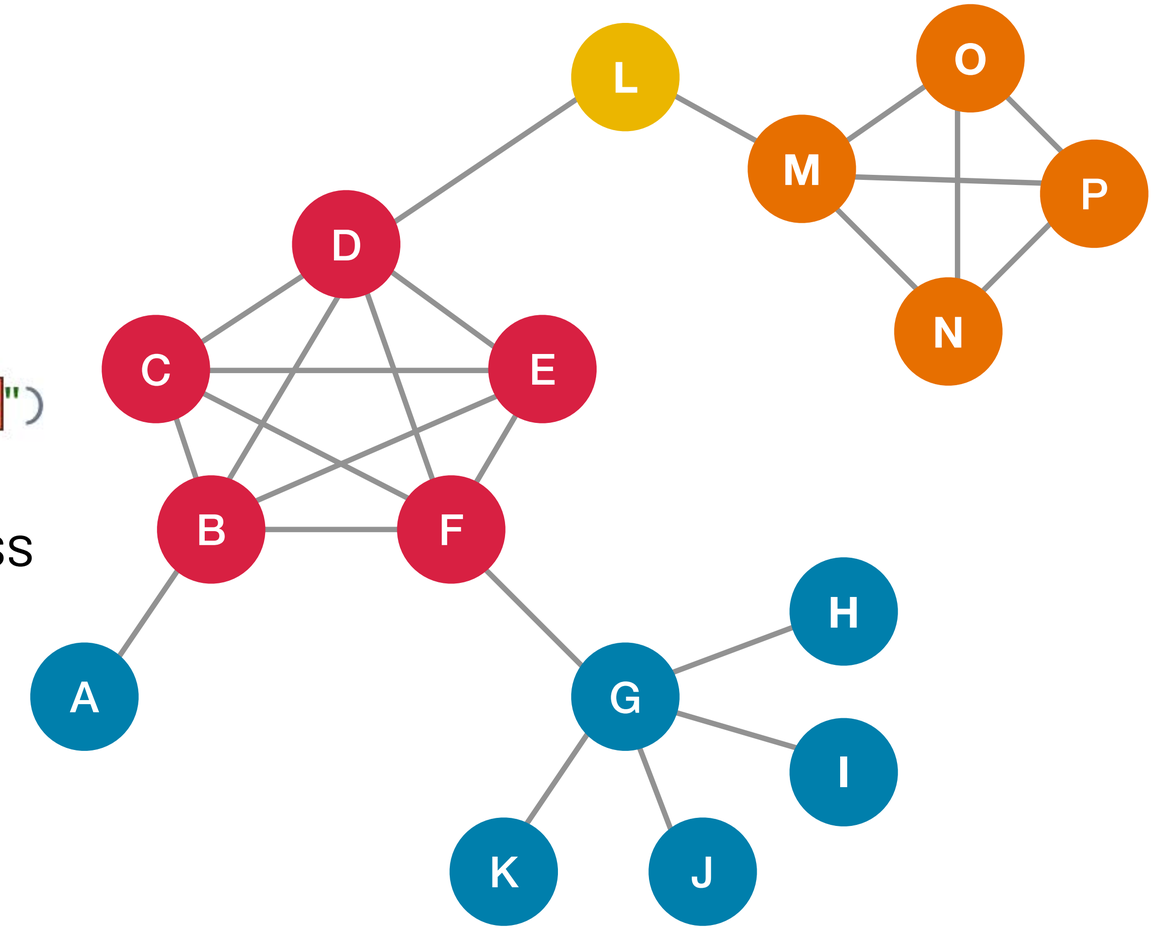
Create list of colors to use:

```
color_list <- c("steelblue","gold","orange", "tomato")
```

Save color as node attribute based on coreness

```
V(G)$color <- color_list[V(G)$core]
```

Find value of color_list at index V(G)$core



Network G

Plotting k-cores (igraph)

Save coreness a node attribute:

```
G <- set_vertex_attr(G, "core",
                     value=coreness(G))
```

Create list of colors to use:

```
color_list <- c("steelblue","gold","orange", "tomato")
```

Save color as node attribute based on coreness

```
V(G)$color <- color_list[V(G)$core]
```

Find value of color_list at index V(G)$core

color_list

index	1	2	3	4
color	“steelblue”	“gold”	“orange”	“tomato”

V(G)$core

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core	1	4	4	4	4	4	1	1	1	1	1	2	3	3	3	3

color_list[V(G)$core]

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core																

Plotting k-cores (igraph)

Save coreness a node attribute:

```
G <- set_vertex_attr(G, "core",
                     value=coreness(G))
```

Create list of colors to use:

```
color_list <- c("steelblue","gold","orange", "tomato")
```

Save color as node attribute based on coreness

```
V(G)$color <- color_list[V(G)$core]
```

Save as node attribute named “color”

color_list

index	1	2	3	4
color	“steelblue”	“gold”	“orange”	“tomato”

V(G)$core

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core	1	4	4	4	4	4	1	1	1	1	1	2	3	3	3	3

color_list[V(G)$core]

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core																

Plotting k-cores (igraph)

Save color as node attribute based on coreness

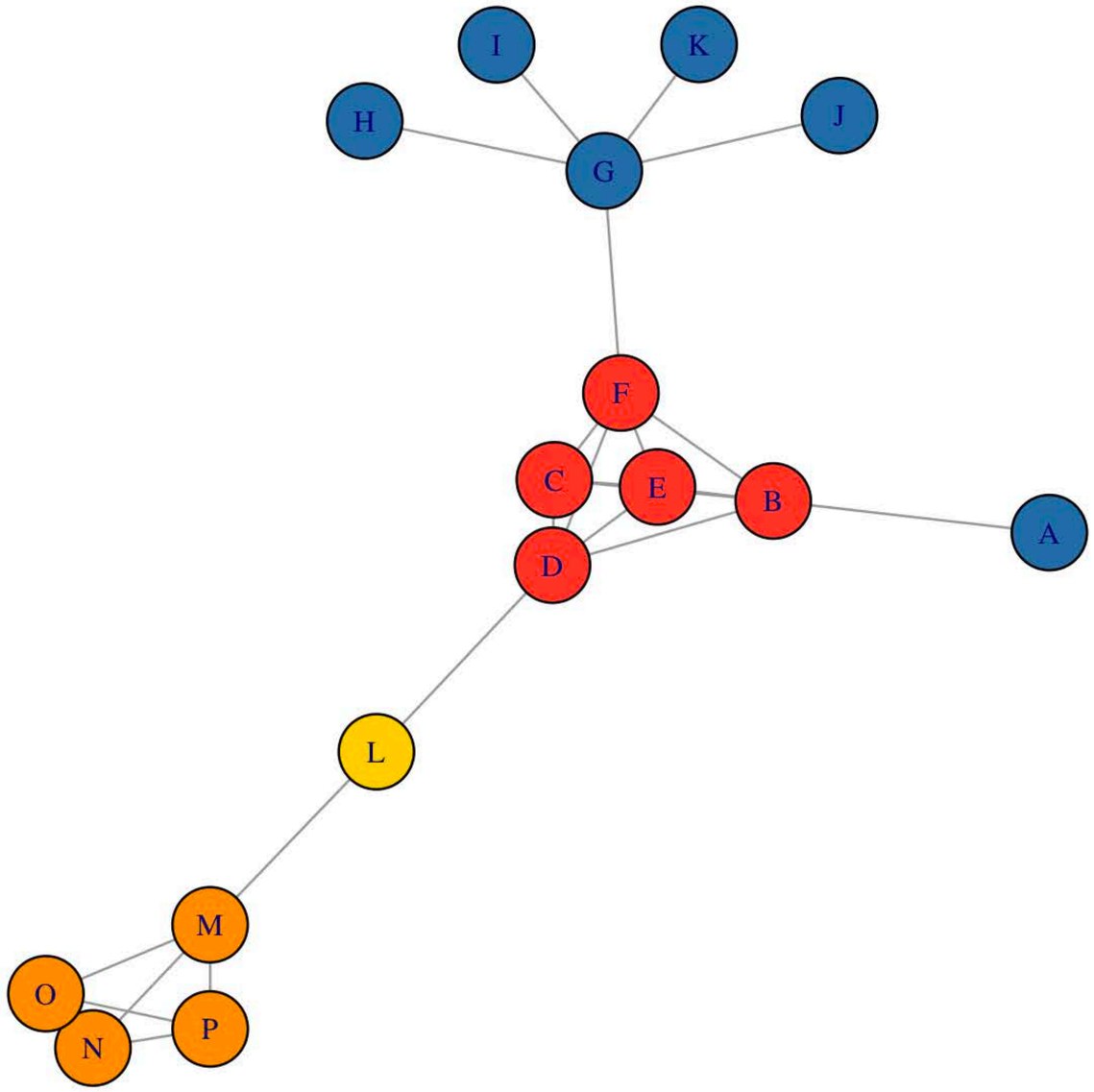
```
V(G)$color <- color_list[V(G)$core]
```

Node	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Core																

```
plot(G)
```

Since we have saved color as a node attribute, this is equivalent to

```
plot(G,
     vertex.color=V(G)$color)
```



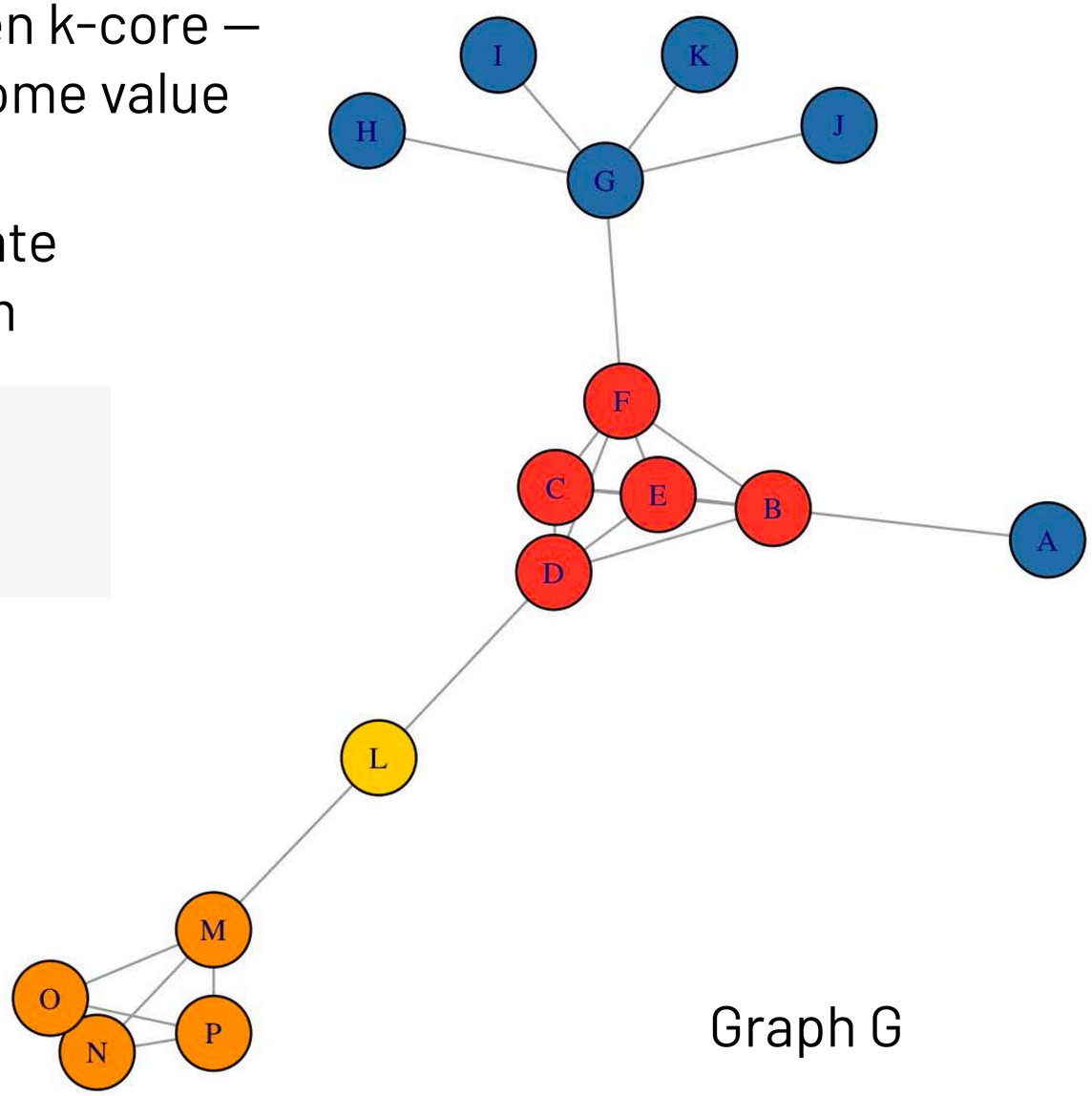
Extracting k-cores

We can also create a **subgraph** with just a given k-core — ie, only includes nodes with coreness above some value

Network to create a subgraph from

```
Gk <- subgraph(G,
                V(G)[V(G)$core>=2]
       )
```

Nodes to include in subgraph



Graph G

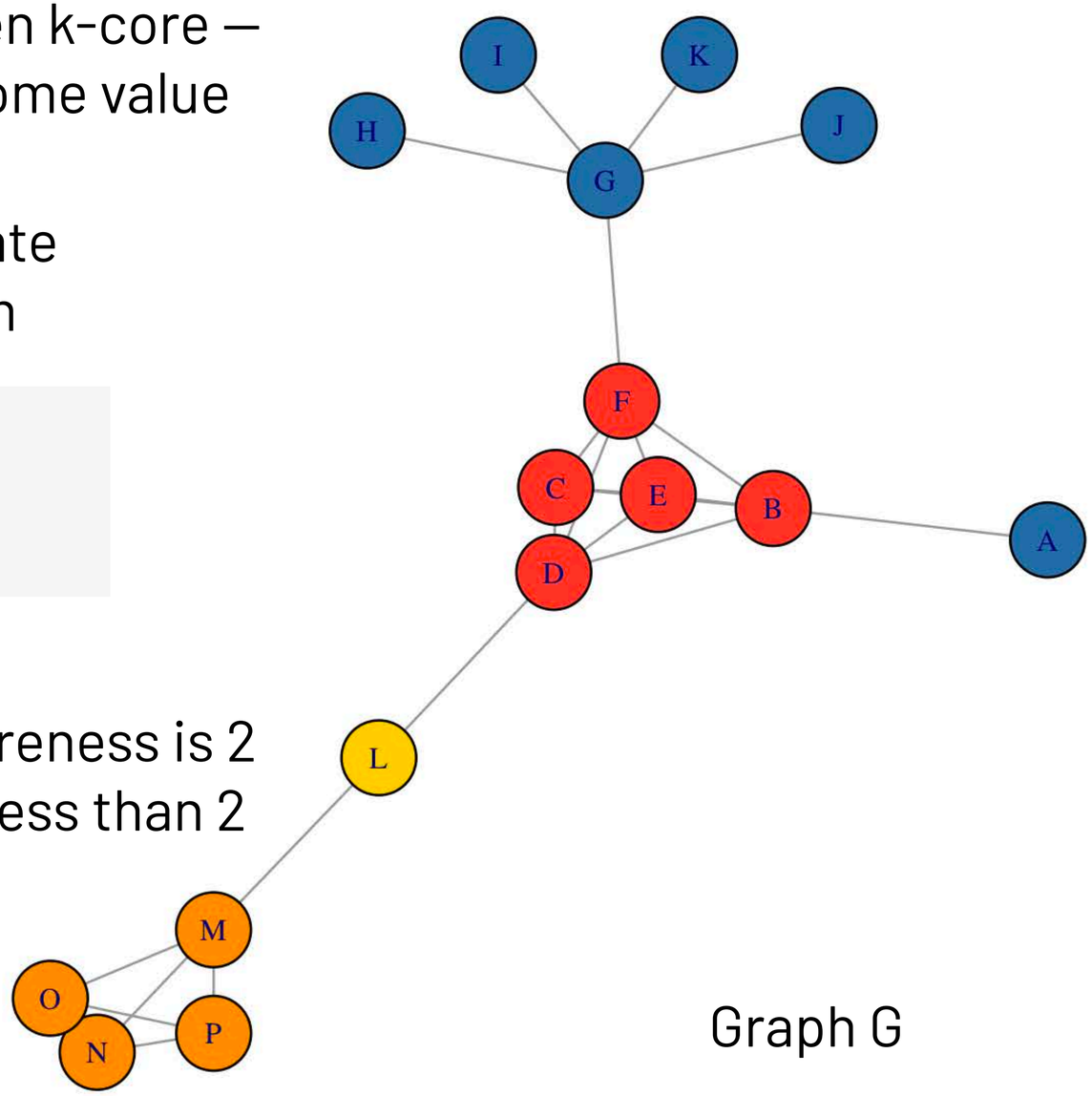
Extracting k-cores

We can also create a **subgraph** with just a given k-core — ie, only includes nodes with coreness above some value

Network to create a subgraph from

```
Gk <- subgraph(G,
                [V(G)$core>=2]
       )
```

Boolean statement: return TRUE if node's coreness is 2 or more. Return FALSE if node's coreness is less than 2



Graph G

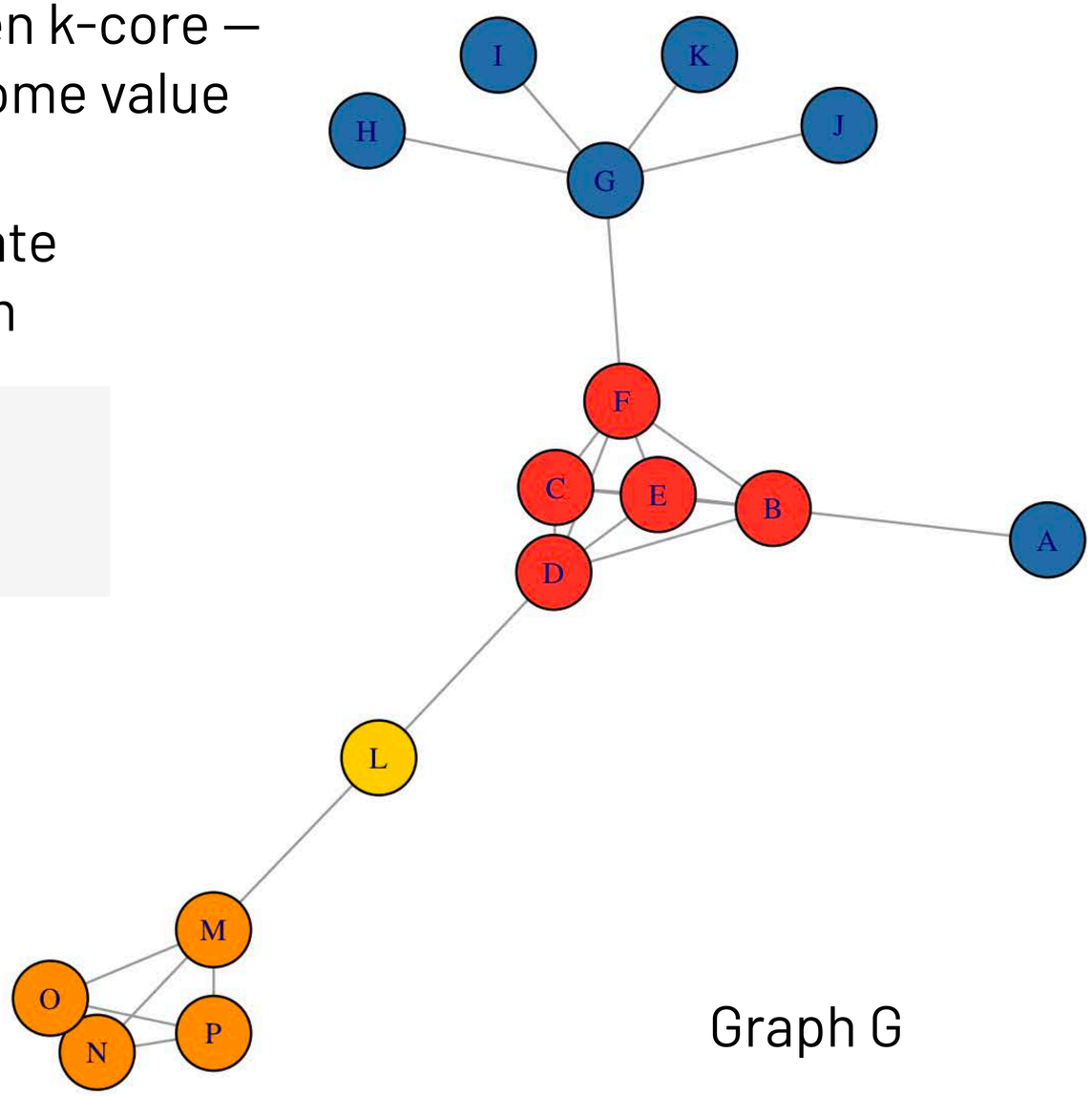
Extracting k-cores

We can also create a **subgraph** with just a given k-core — ie, only includes nodes with coreness above some value

Network to create a subgraph from

```
Gk <- subgraph(G,
                V(G)[                    ]
       )
```

Subset to just nodes where boolean statement is TRUE



Graph G

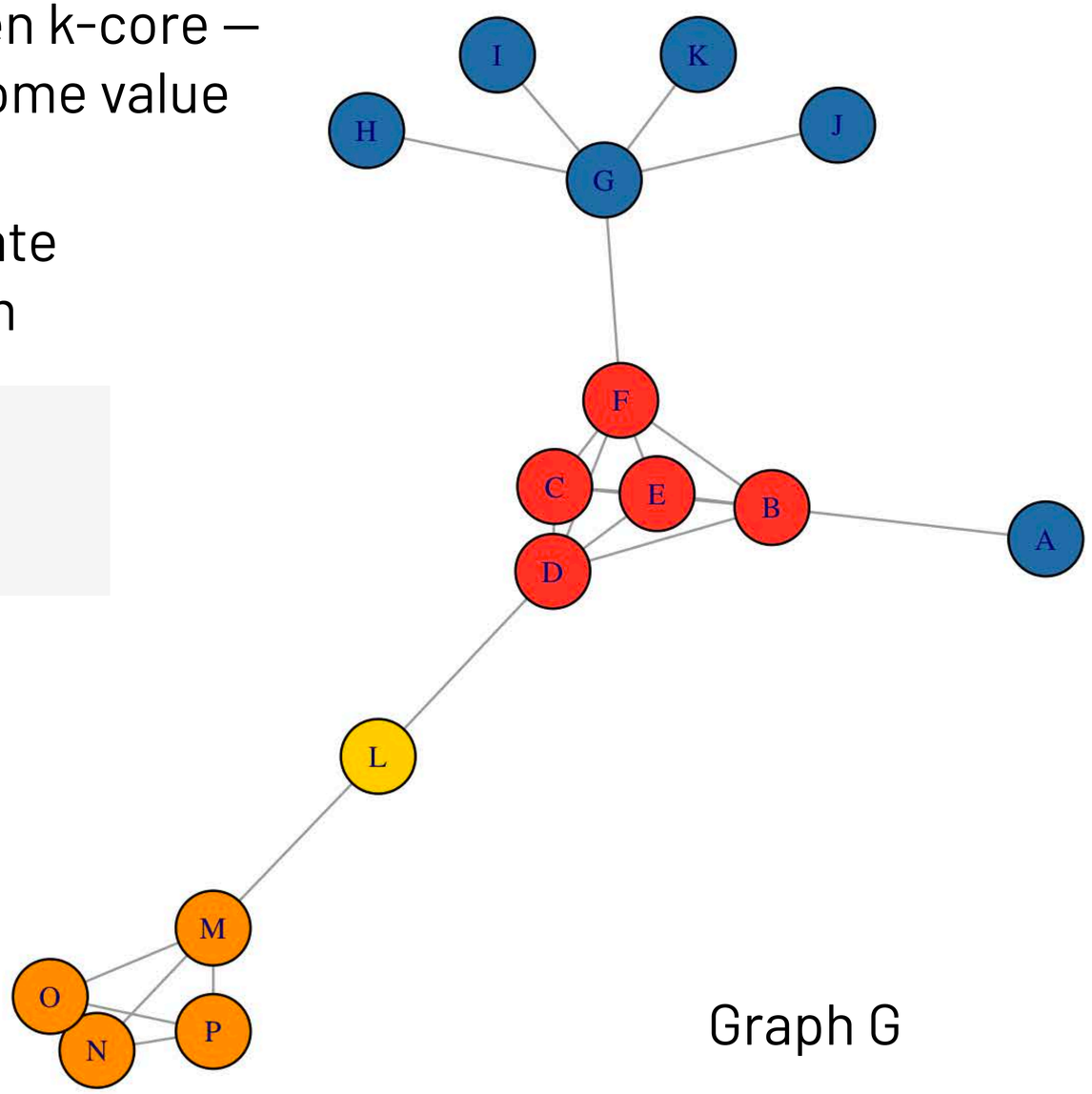
Extracting k-cores

We can also create a **subgraph** with just a given k-core — ie, only includes nodes with coreness above some value

Network to create a subgraph from

```
Gk <- subgraph(G,
                V(G)[V(G)$core>=2]
       )
```

Only include nodes whose coreness is at least 2

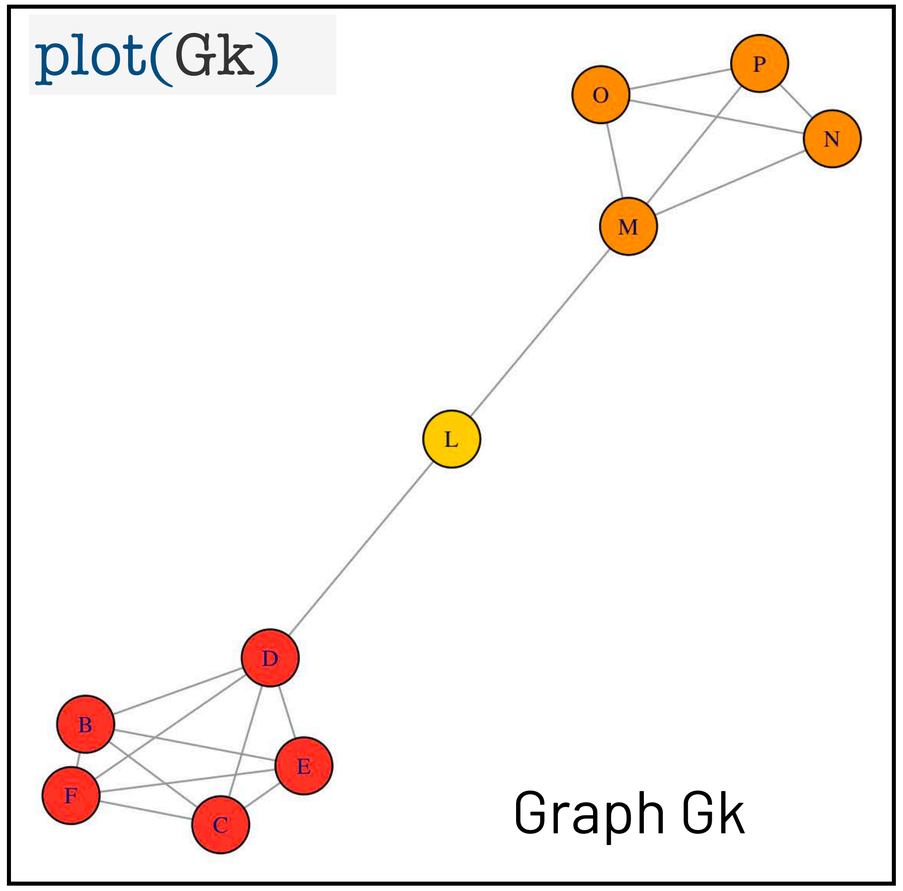


Graph G

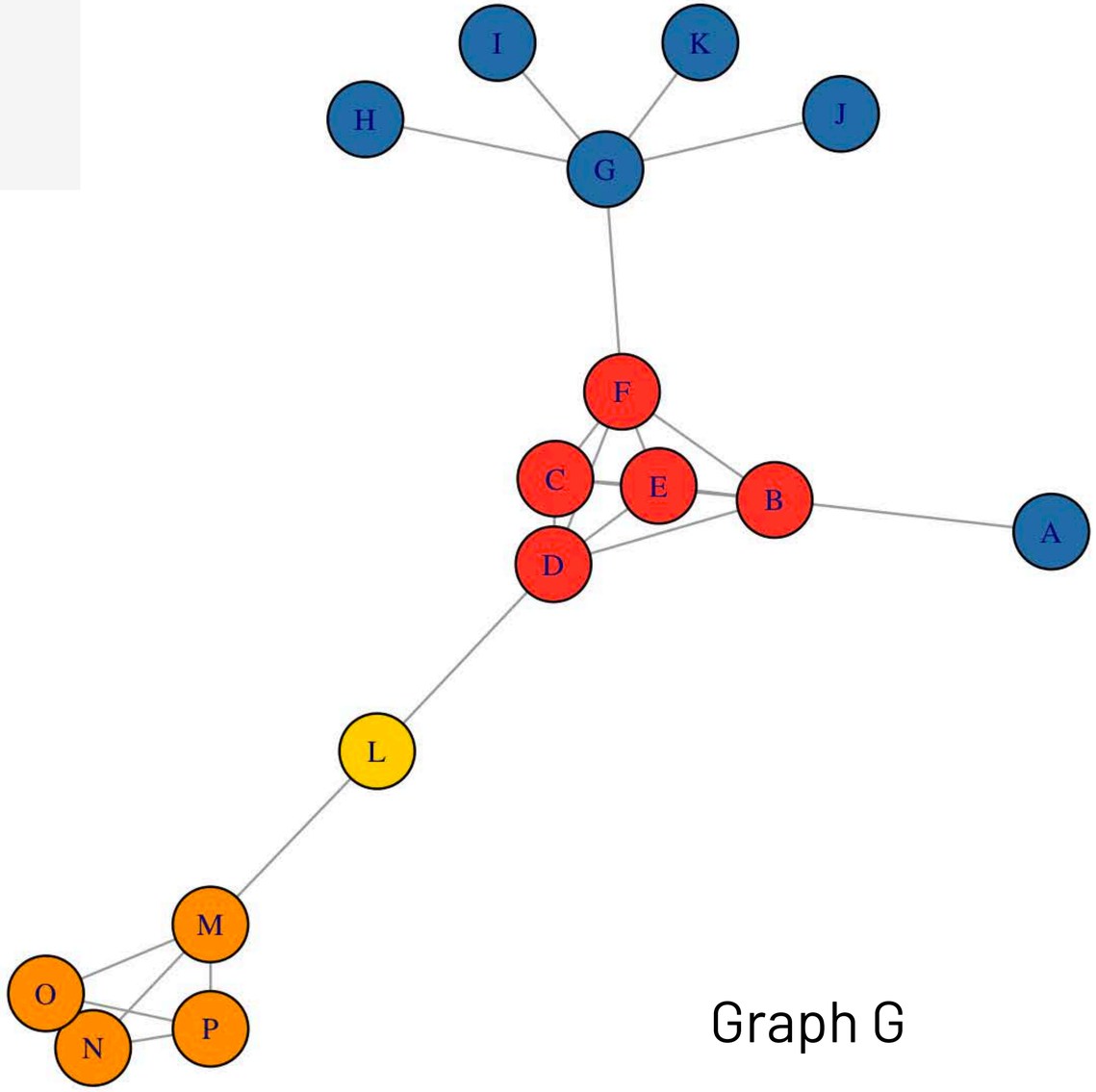
Extracting k-cores

```
Gk <- subgraph(G,
                    V(G)[V(G)$core>=2]
       )
```

```
plot(Gk)
```



Graph Gk



Graph G

Final Thoughts

Always start by asking:

What are your nodes and what are your edges?

There are no perfect choices, but what you choose should be **justifiable** based on the **research question**

PLEASE KEEP IN TOUCH!

Dr. Sarah Shugars
sarah.shugars@rutgers.edu
they/them/theirs