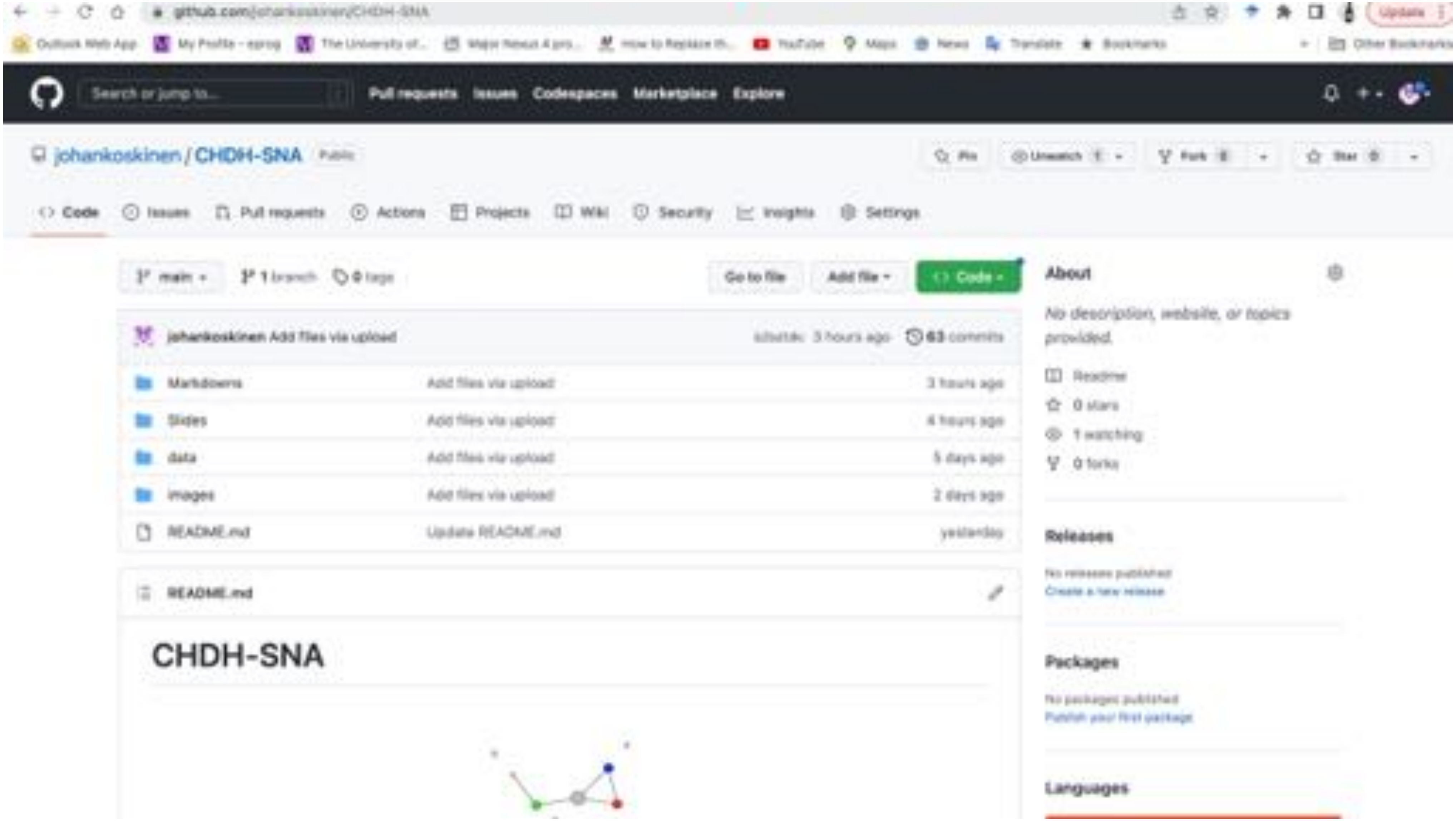
Introduction to statistical network analysis

CHDH SNA 1

22 February 2023



Search or jump to...
Pull requests
Issues
Codespaces
Marketplace
Explore
johankoskinen / CHDH-SNA
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
main
Go to file
Add file
Code
johankoskinen Add files via upload
Markdowns
Slides
data
images
README.md
Add files via upload
Add files via upload
Add files via upload
Add files via upload
Update README.md
yesterday
README.md
CHDH-SNA
About
No description, website, or topics provided.
Readme
0 stars
1 watching
0 forks
Releases
Packages
Languages

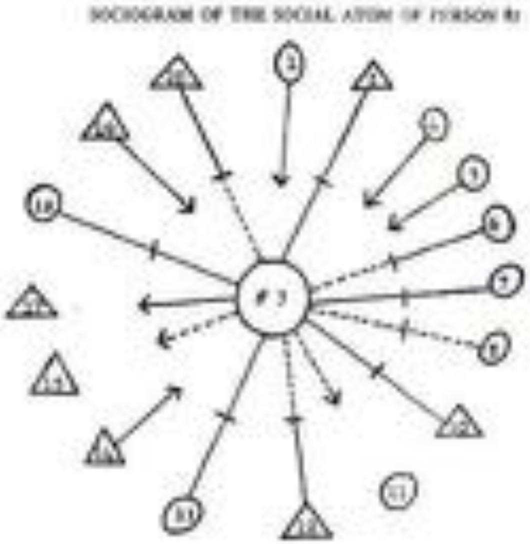
Network data and graph metrics

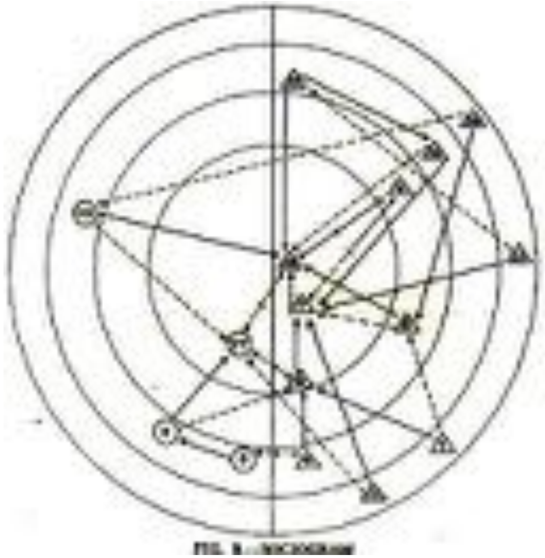
Network data and graph metrics

- Adjacency matrix, basic manipulation
- Network non-parametric approaches
- Ego-nets
- Network regression

https://raw.githubusercontent.com/johankoskinen/CHDH-SNA/main/Markdowns/CHDH-SNA-1.Rmd")

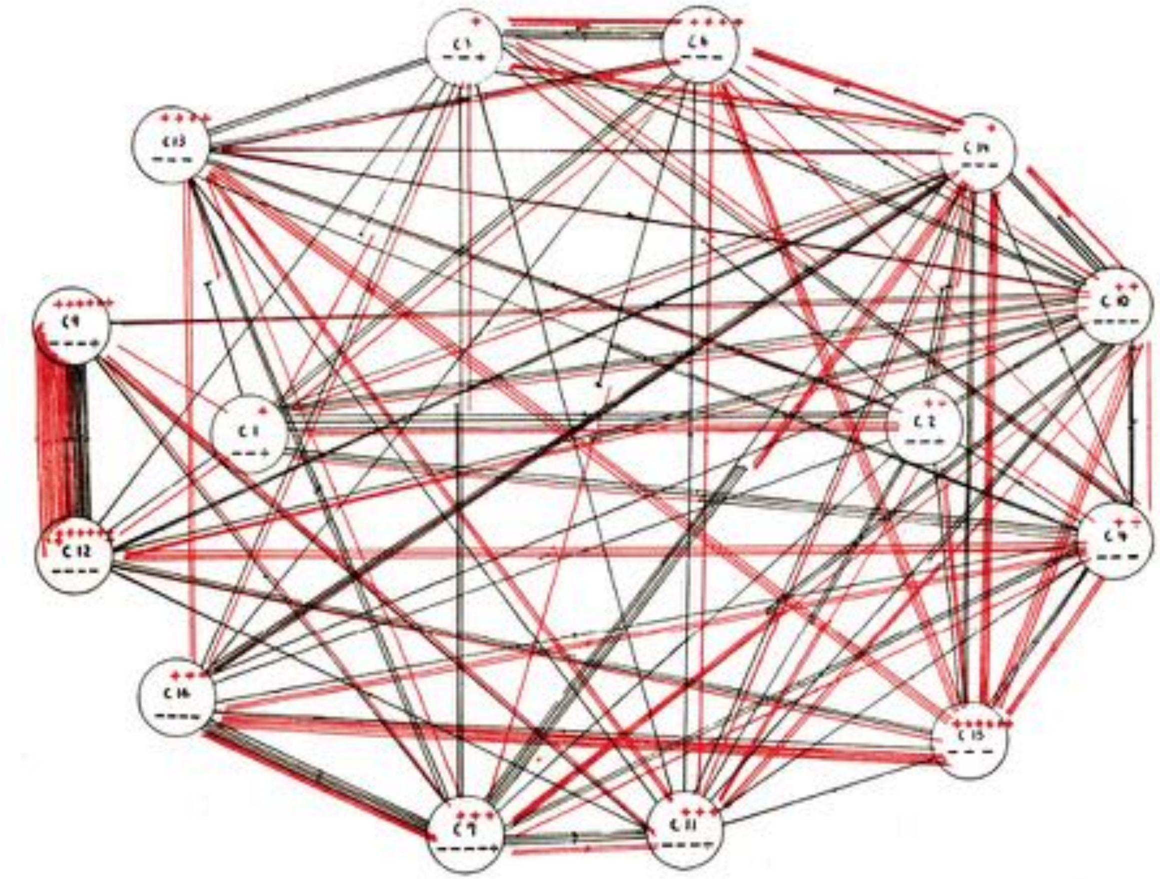
Moreno (1934) – Sociometrics





Jacob L Moreno

Aim: to create a social equivalent of Psychometrics and map out the forces that affect the individual



Psychological Geography Map IV. A Reduction Sociogram (Moreno, 1934)

– Attraction
– Rejection
– indifference

… and put these maps together to create a narrative

Module: Network Models of Social Processes

- Network data and graph metrics (Lecture + tutorial)
- Network structure and network mechanisms (Lecture + tutorial)
- Network contagion(Lecture + tutorial)
- Empirical analysis of network structure (Lecture)

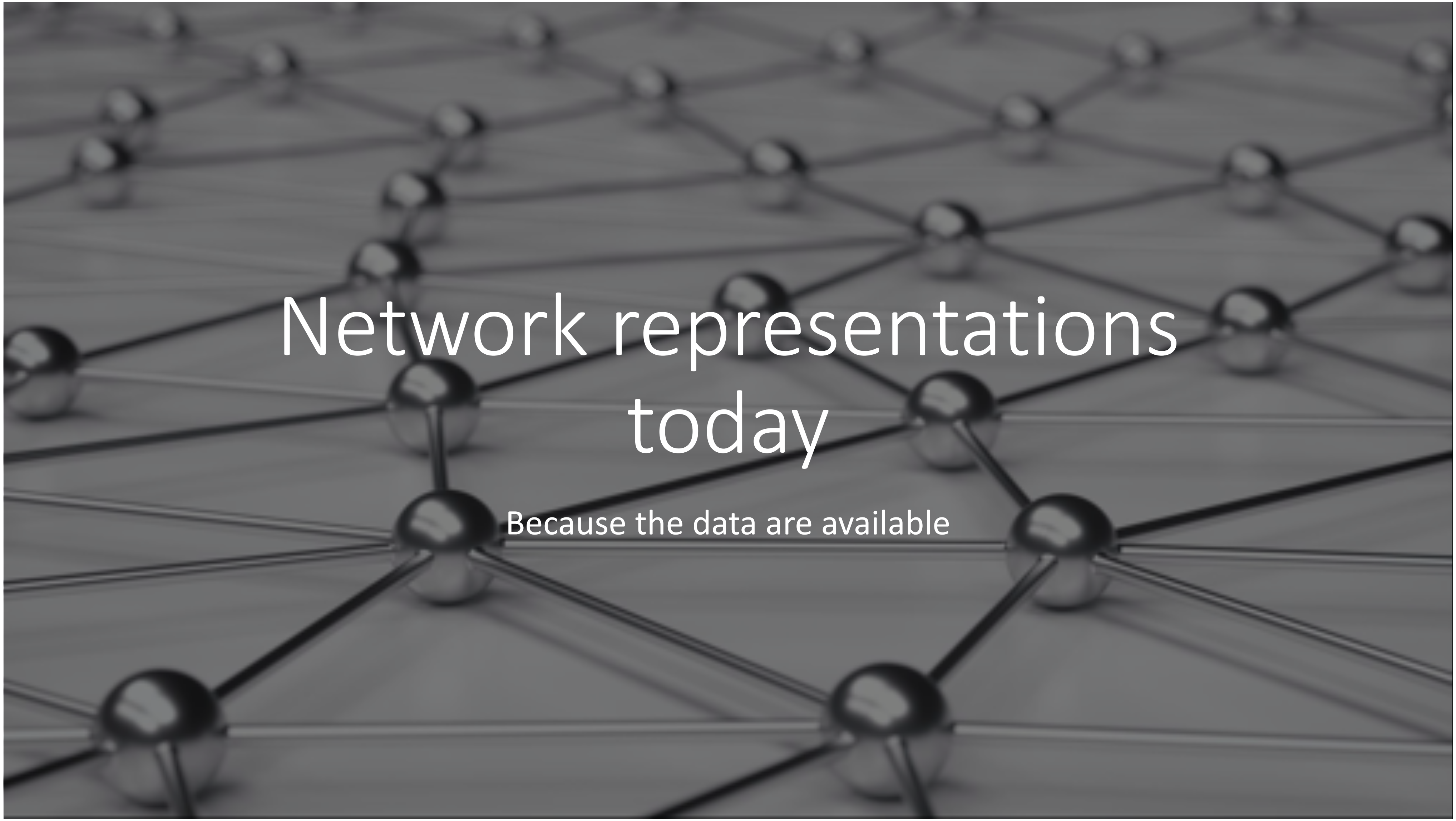
Last week: revision tutorial for entire subject

Network data and graph metrics: Agenda

- Introduction to networks
- History (Freeman, 2011; Borgatti et al. 2009)
- Why are networks important? (Brandes et al., 2013)
- Different types of networks
- Network notation, definitions, and concepts
    - Representation: Graph, sets, edge list, adjacency matrix
    - Degree: density, degree, degree distribution
    - Reach: path, geodesic, distance, diameter
    - Clustering: clique, triads, closure
- Global and local

History and background

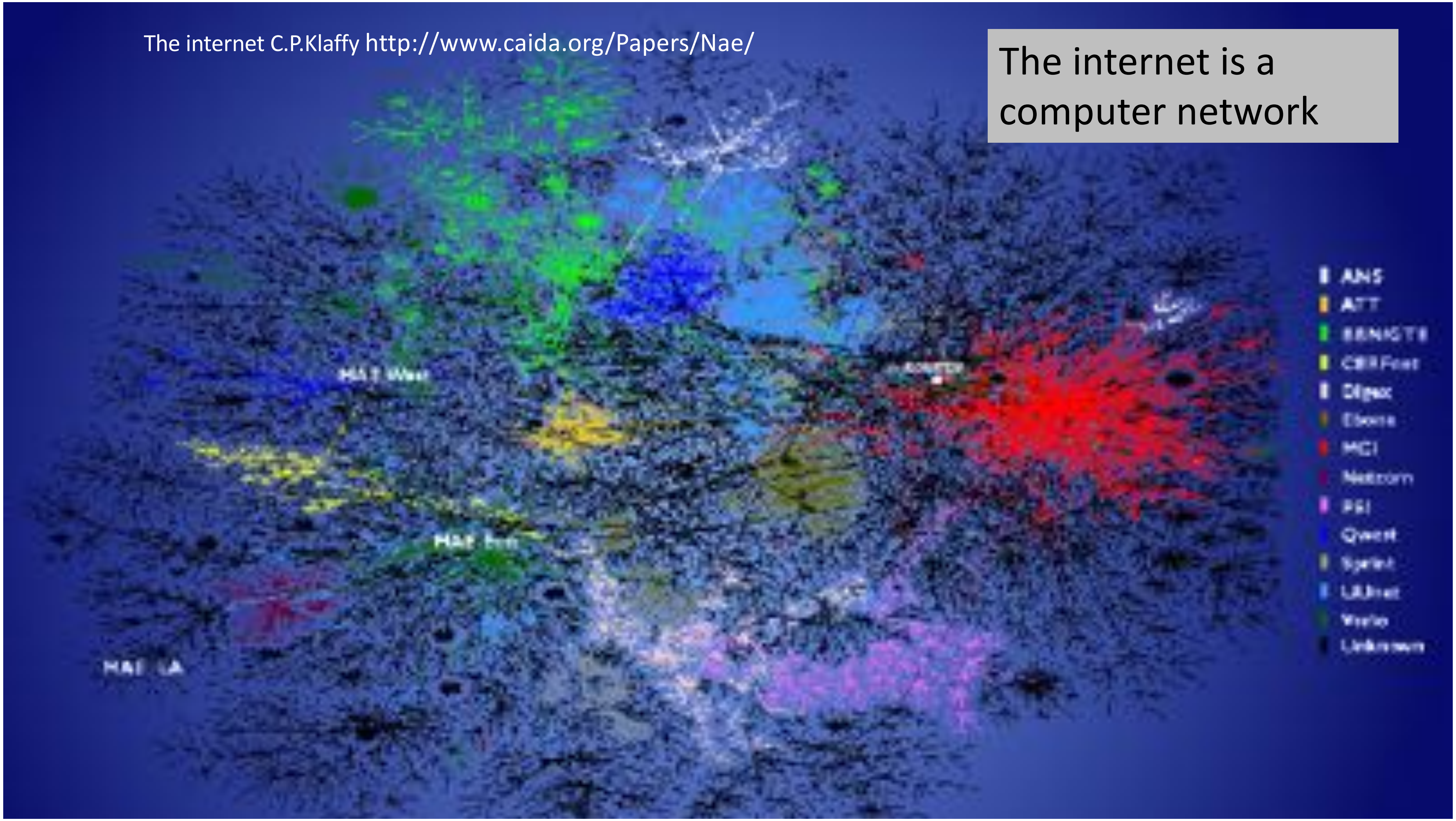
- Freeman, L. C. (2011). The development of social network analysis–with an emphasis on recent events. *The SAGE handbook of social network analysis*, *21*(3), 26-39.
- Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, *1*(687), 159-167.
- Brandes, U., Robins, G., McCranie, A., & Wasserman, S. (2013). What is network science?. *Network science*, *1*(1), 1-15.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science*, *323*(5916), 892-895.


Network representations today
Because the data are available

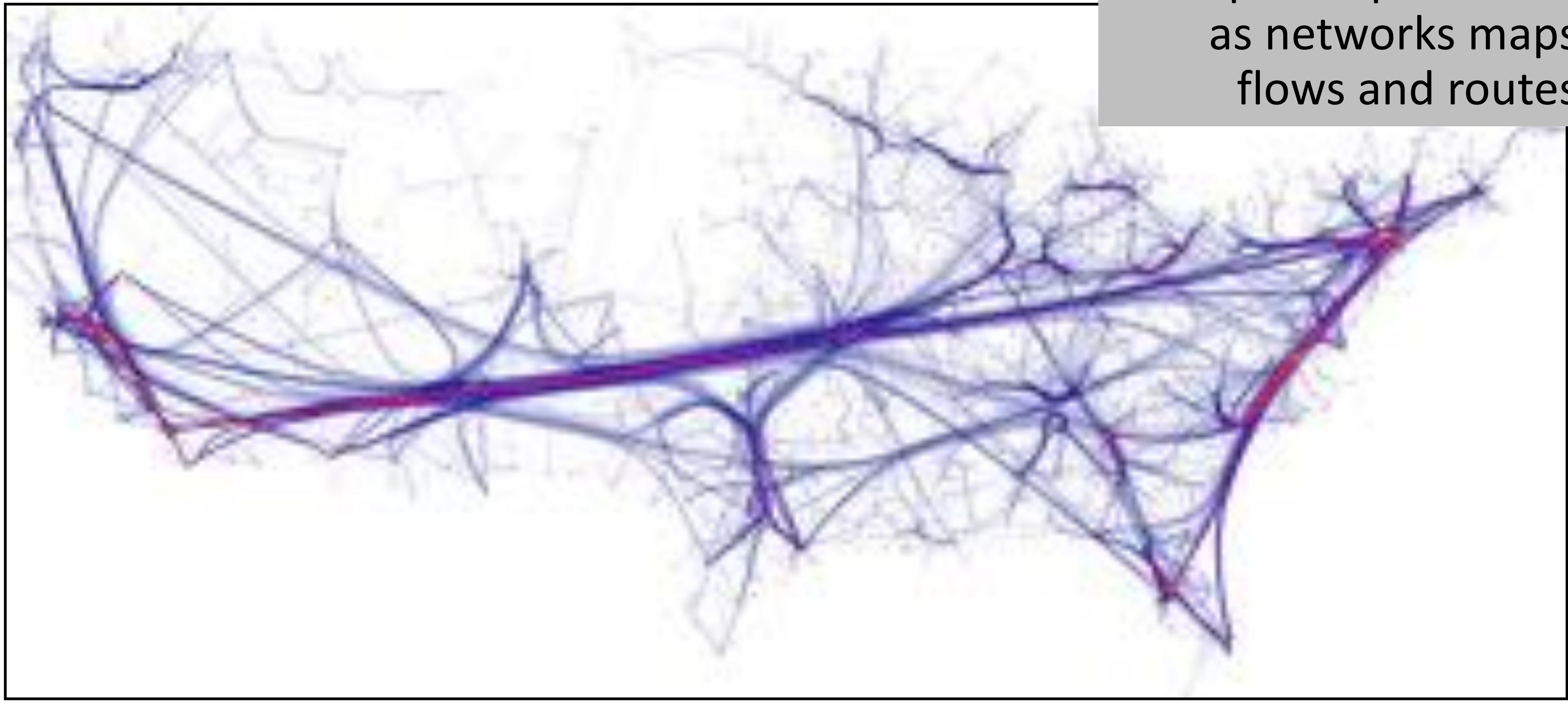
Social media is relationally organised

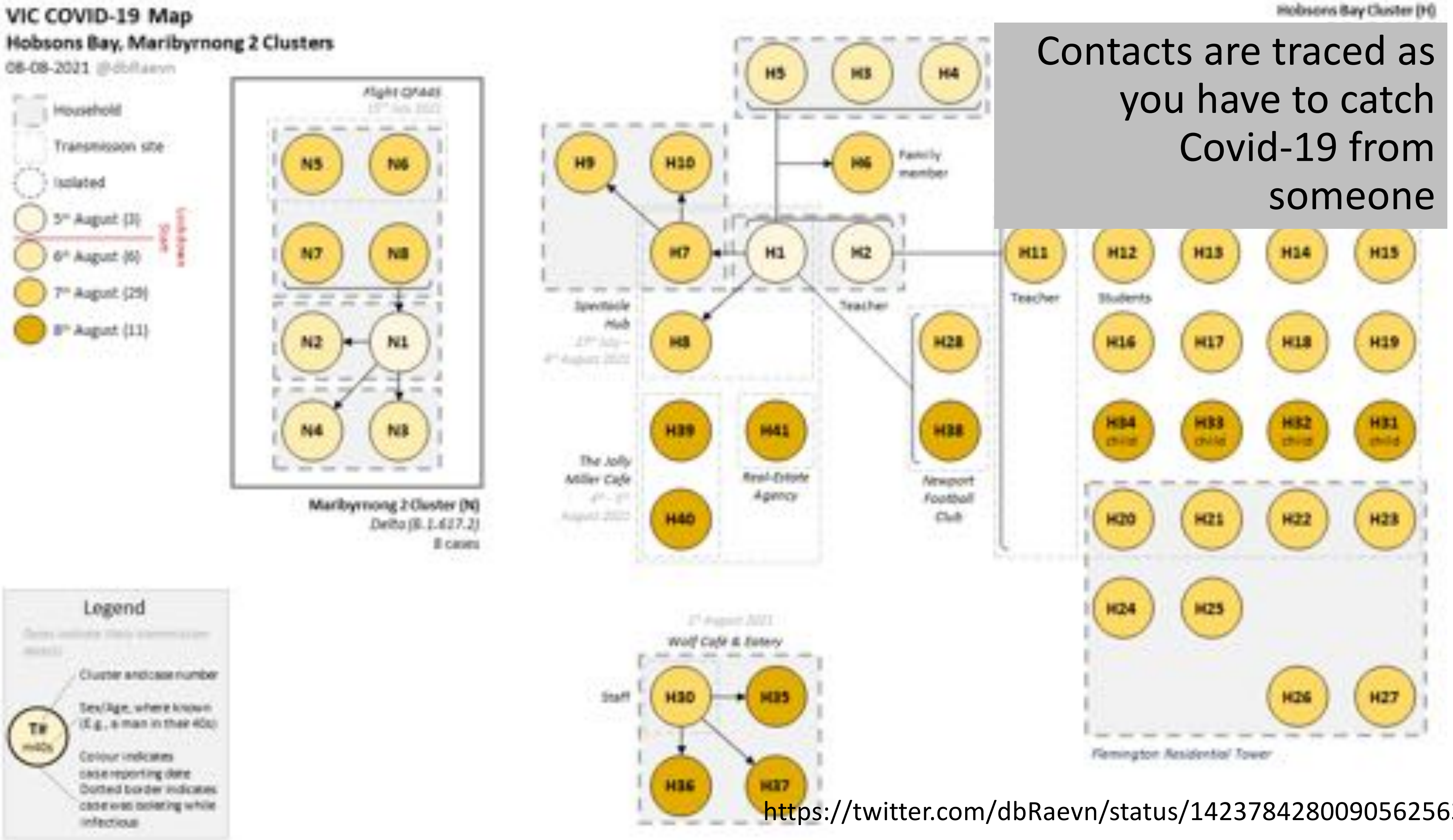
The internet C.P.Klaffy http://www.caida.org/Papers/Nae/

The internet is a computer network



Transport represented
as networks maps
flows and routes

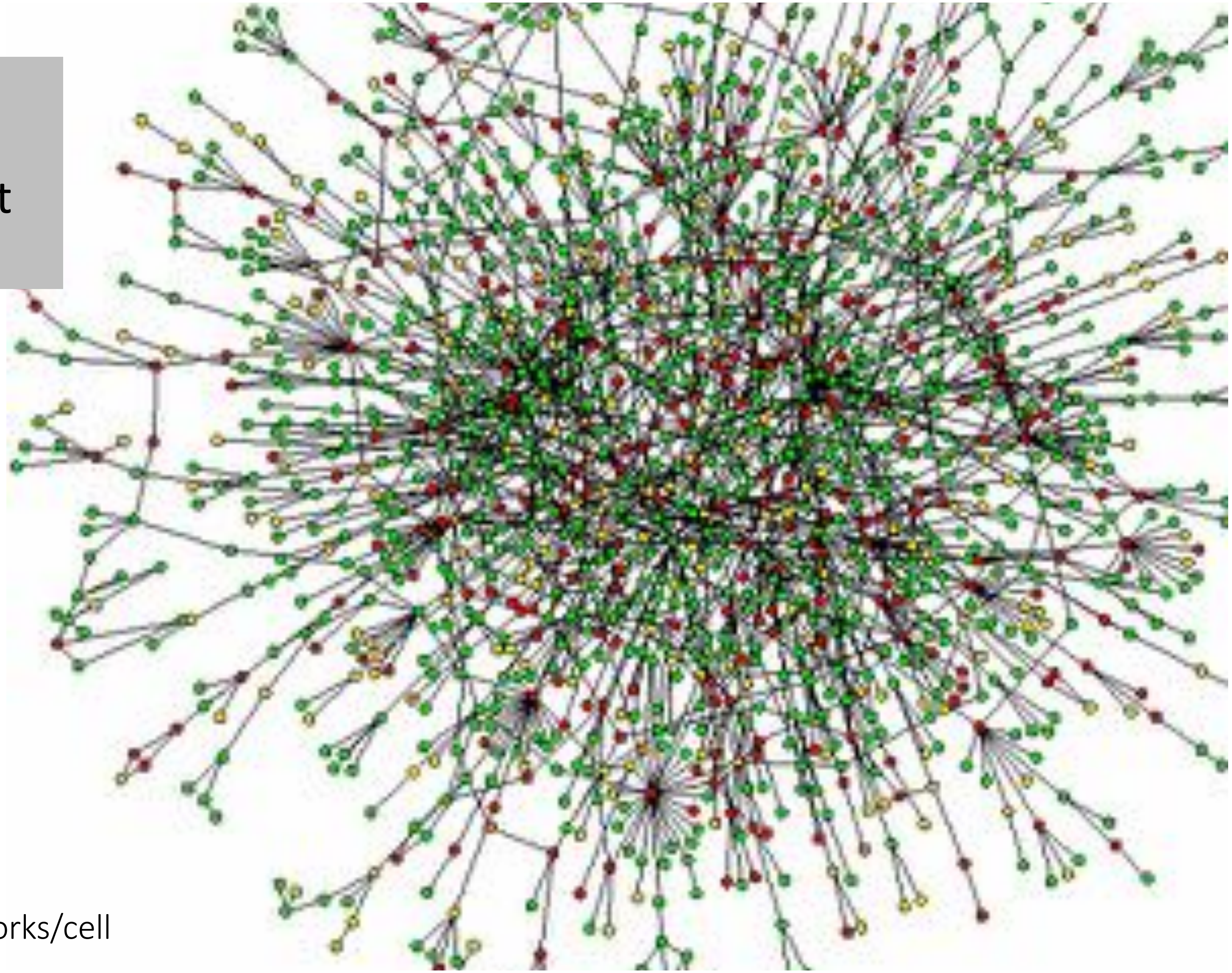




Contacts are traced as you have to catch Covid-19 from someone

https://twitter.com/dbRaevn/status/1423784280090562563

Associations can be mapped as networks for things that are not networks

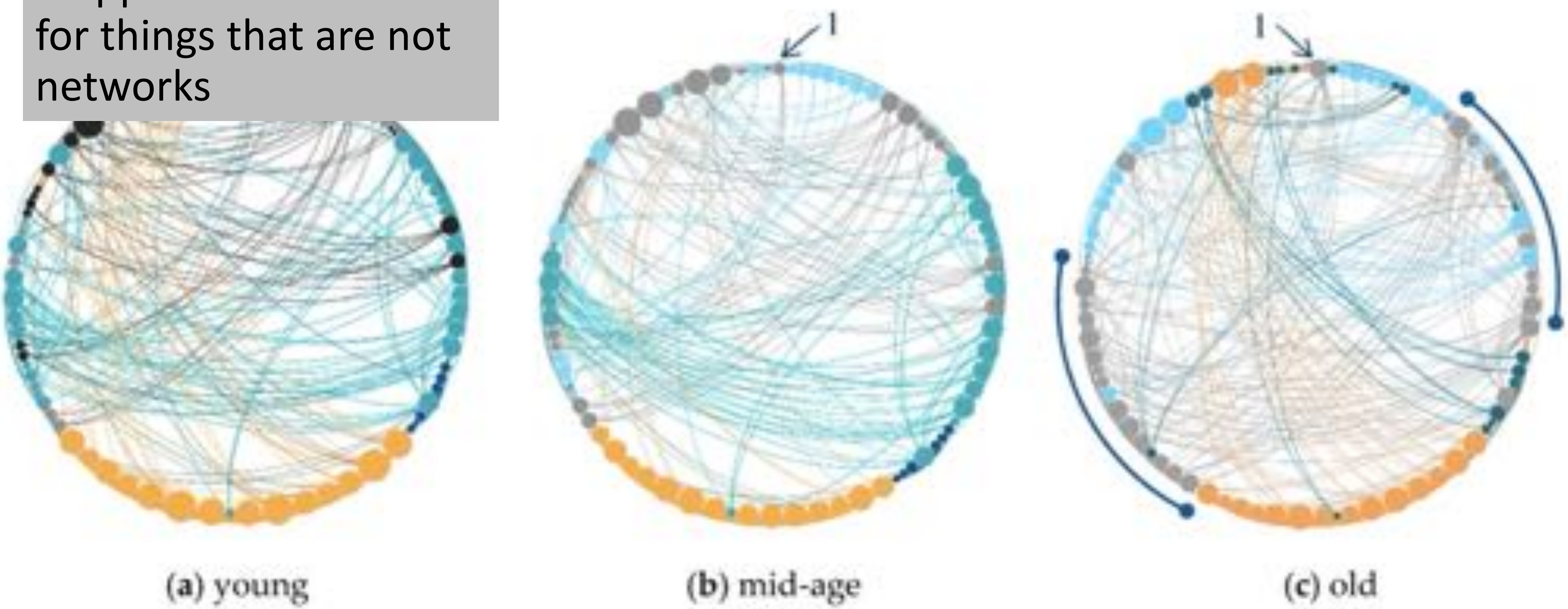


Protein-protein interactions

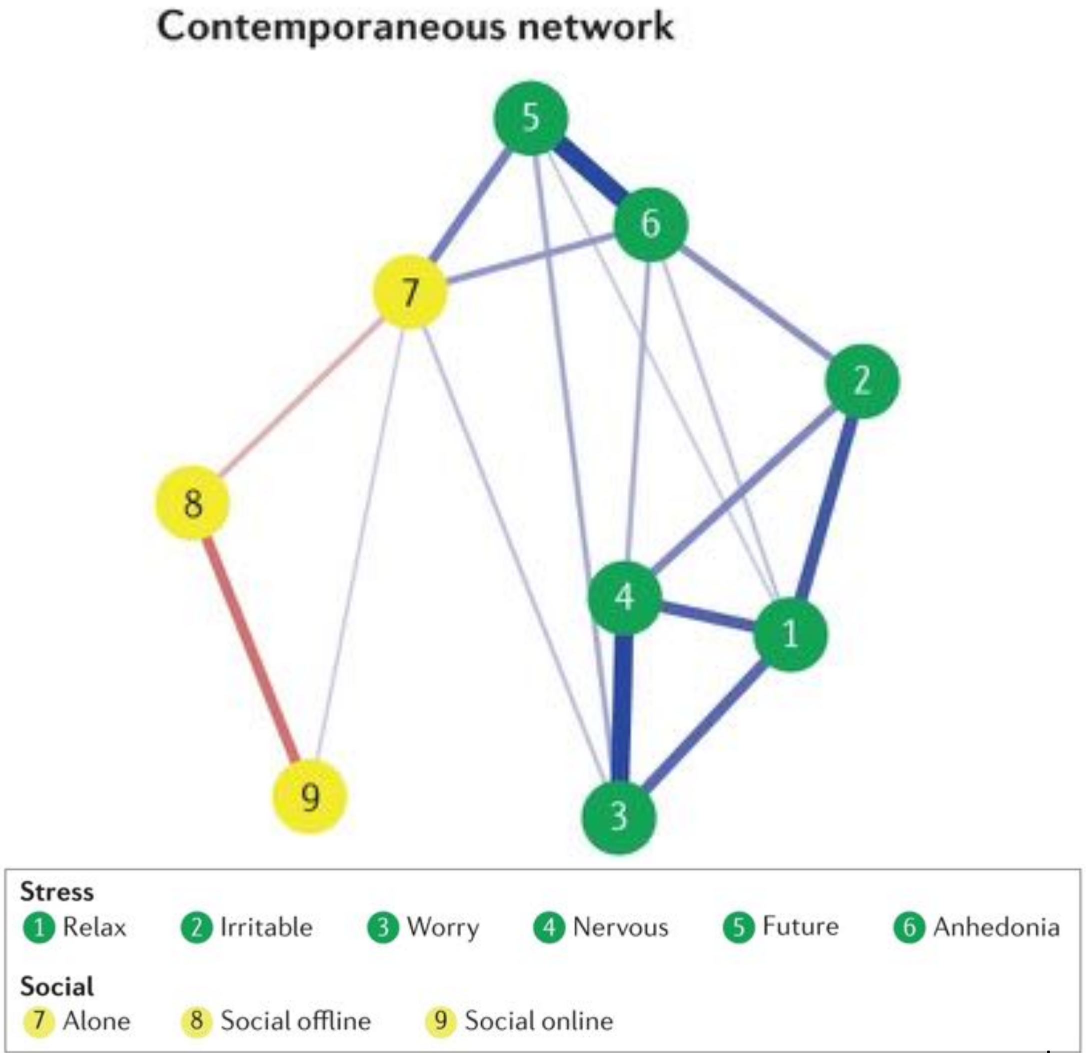
Hawoong Jeong

http://www.nd.edu/%7Enetworks/cell

Associations can be mapped as networks for things that are not networks



Brain network



https://www.nature.com/articles/s43586-021-00055-w/figures/6

Associations can be mapped as networks for things that are not networks

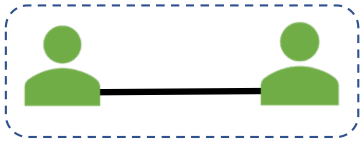
(Social) Networks

Developing graph-theoretic representation of social interaction

Simmel (1908; 1922): Interactions, groups, and society

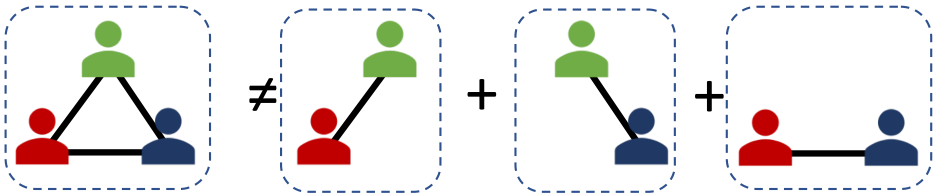
“Society exists where a number of individuals enter into interaction.”

The Web of Group Affiliations (1922):



Dyad

Triad



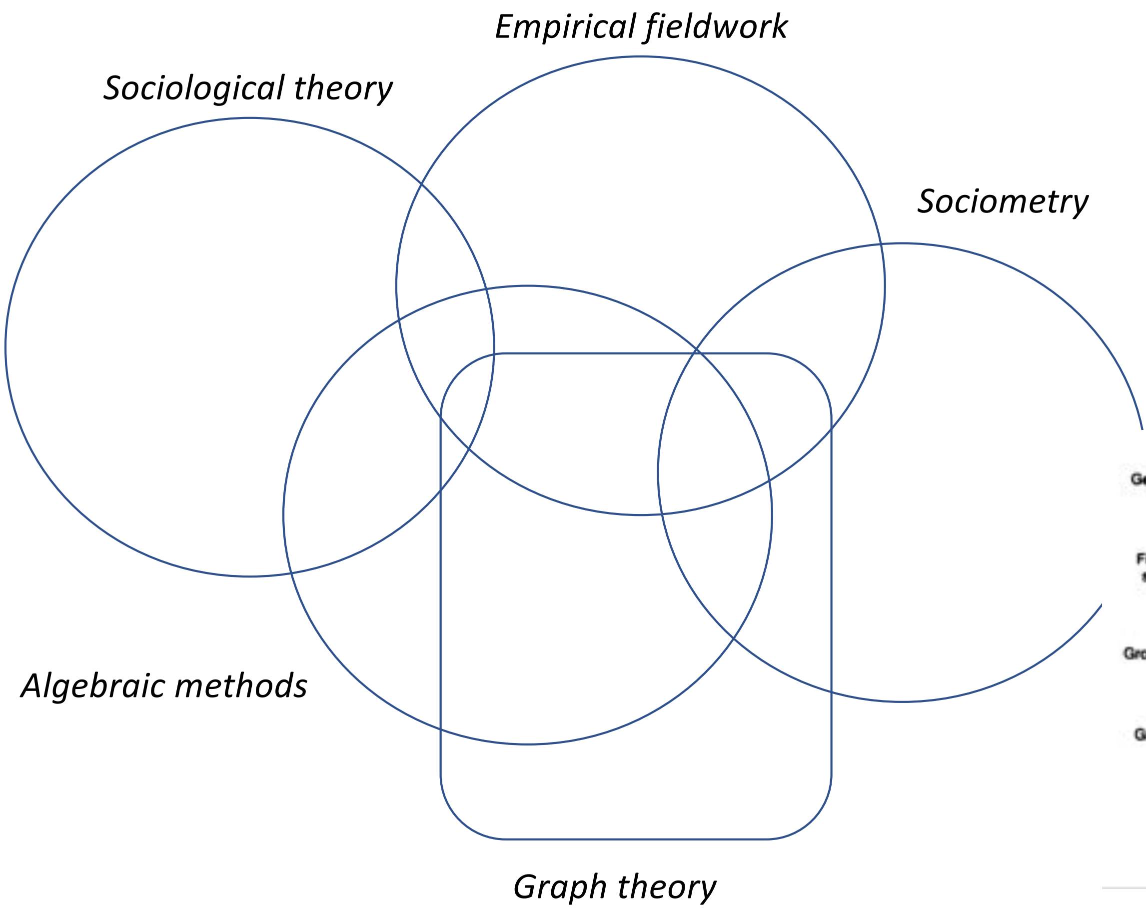
Antropology

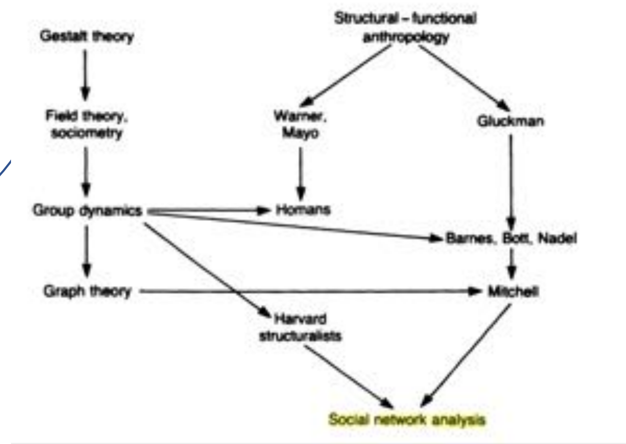
“the whole of social life…[could be seen as]… a set of points some of which are joined by lines… [to form a]… total network… [of relations]”

John Barnes

Multidisciplinary field developing

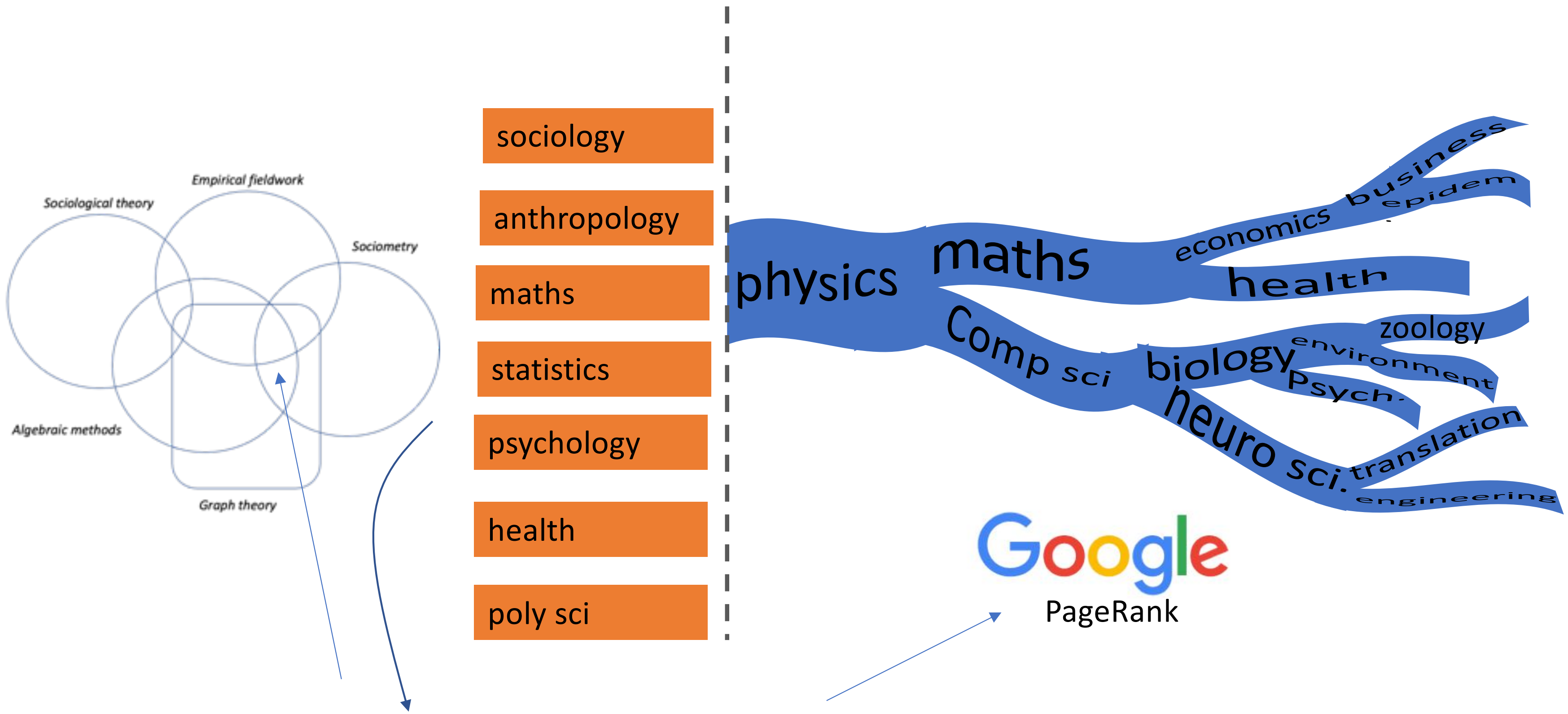
- Network representation
- Data collection approaches
- Theory
- Mathematics
- Statistics
- etc





Freeman ‘The Development of Social Network Analysis with an Emphasis on Recent Events’

1998: Network Science



Katz (1953) “A new status index derived from **sociometric** analysis," Psychometrika

Why are networks important

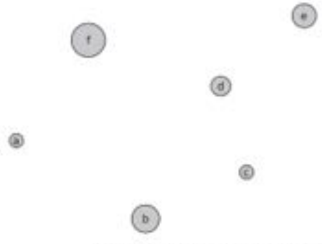
Network explanation v non-network explanation

Brandes et al. (2013)

Explanations that treat people as in a vacuum (a)

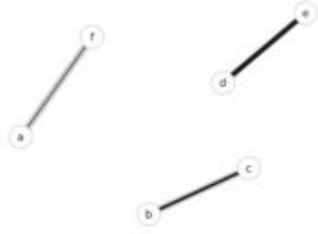
What does (b) tell us that (a) doesn’t tell us?

What does (c) tell us that (b) doesn’t tell us?



A	x
a	1
b	4
c	1
d	2
e	3
f	6

(a) standard table: variables in columns indexed with unrelated entities



D	x
(a,f)	1
(d,e)	5
(b,c)	3

(b) dyadic: variables in columns indexed with unrelated pairs of entities

D	x
(a,b)	0
(a,c)	1
(a,d)	0
(a,e)	1
(a,f)	0
(b,c)	1
(b,d)	2
(c,d)	0
⋮	

$x(D)$	a	b	c	d	e	f
a	·	0	1	0	1	0
b	0	·	1	2	·	·
c	1	1	·	0	2	0
d	0	2	0	·	1	4
e	1	·	2	1	·	2
f	0	·	0	4	2	·

(c) network: variables in columns indexed with incident pairs of entities, or in matrices

Fig. 2. Data formats distinguished by the structure of the domain.

Relational perspectives and explanations

Dr D eats (predominantly) vegetarian food...

Vegetarian
partner

Relational perspectives and explanations (alt.)

Relational perspectives and explanations

Someone close to you is unhappy...

... will you remain unaffected?

Relational perspectives and explanations

Equal opportunities based on our individual qualities …

…

What are network data

Brandes et al. (2013)

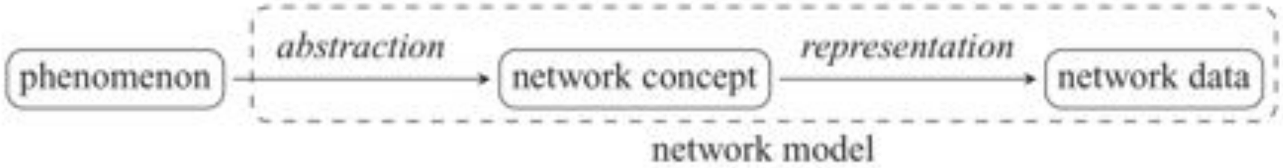
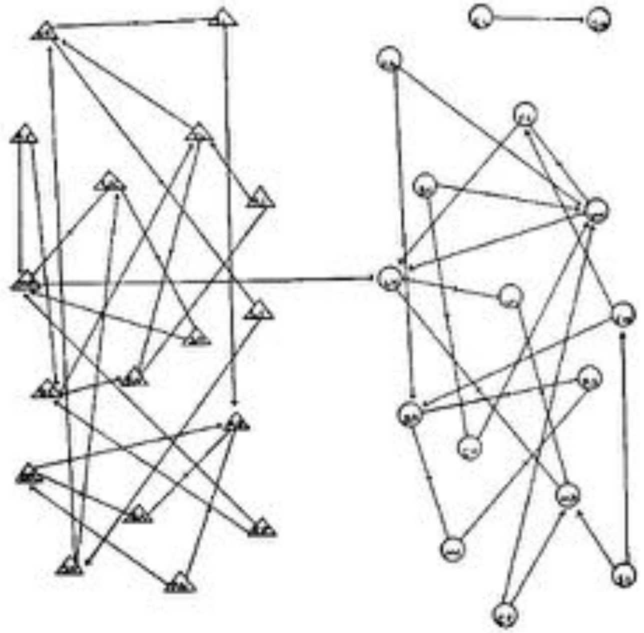


Fig. 1. The elements of network models.

1. A specification of how the phenomenon (in general, i.e., more generally than this particular instantiation) is abstracted to a network.
2. A specification of how this conceptual network is represented in data (e.g., measured or observed).

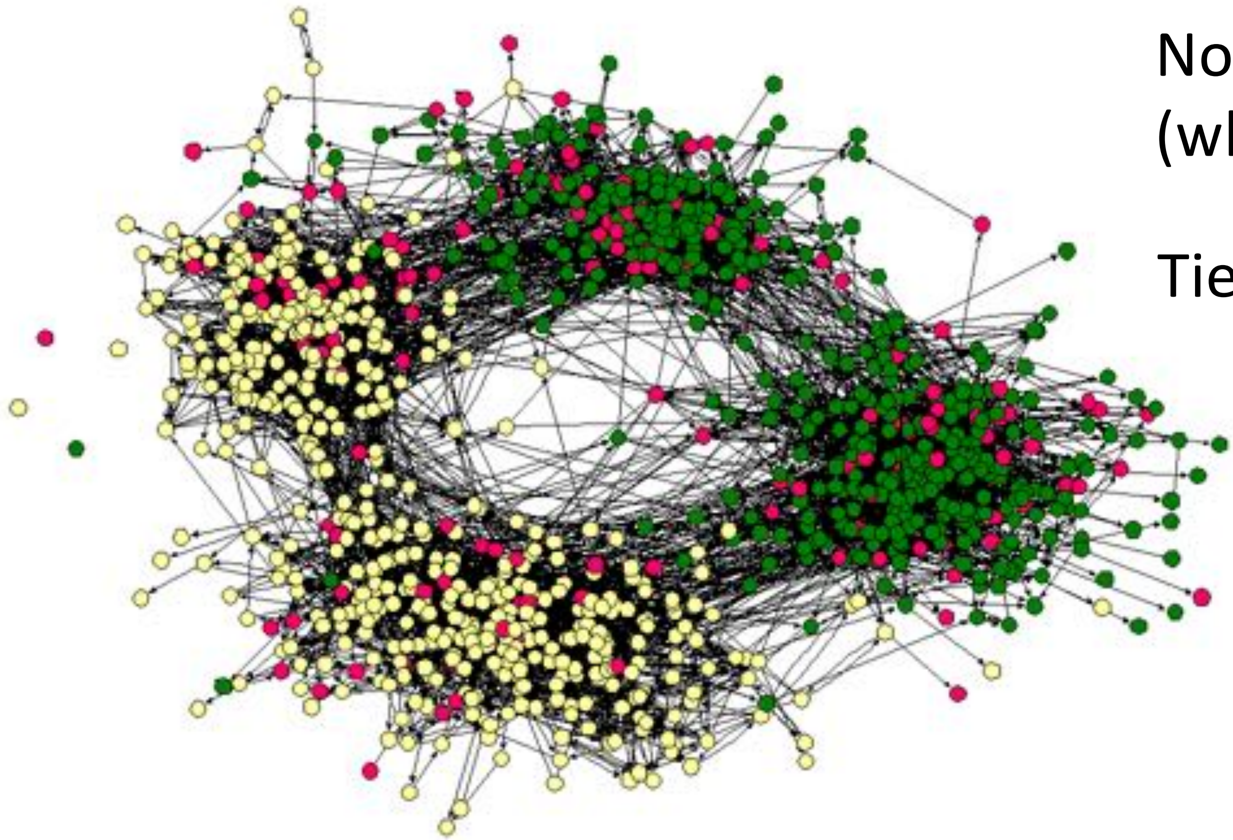
Moreno (1934) fourth-grade class



Nodes: fourth-grade student (boy/girl)

Ties: friendship (whom they chose to sit next to while studying or playing)

Moody (2001)

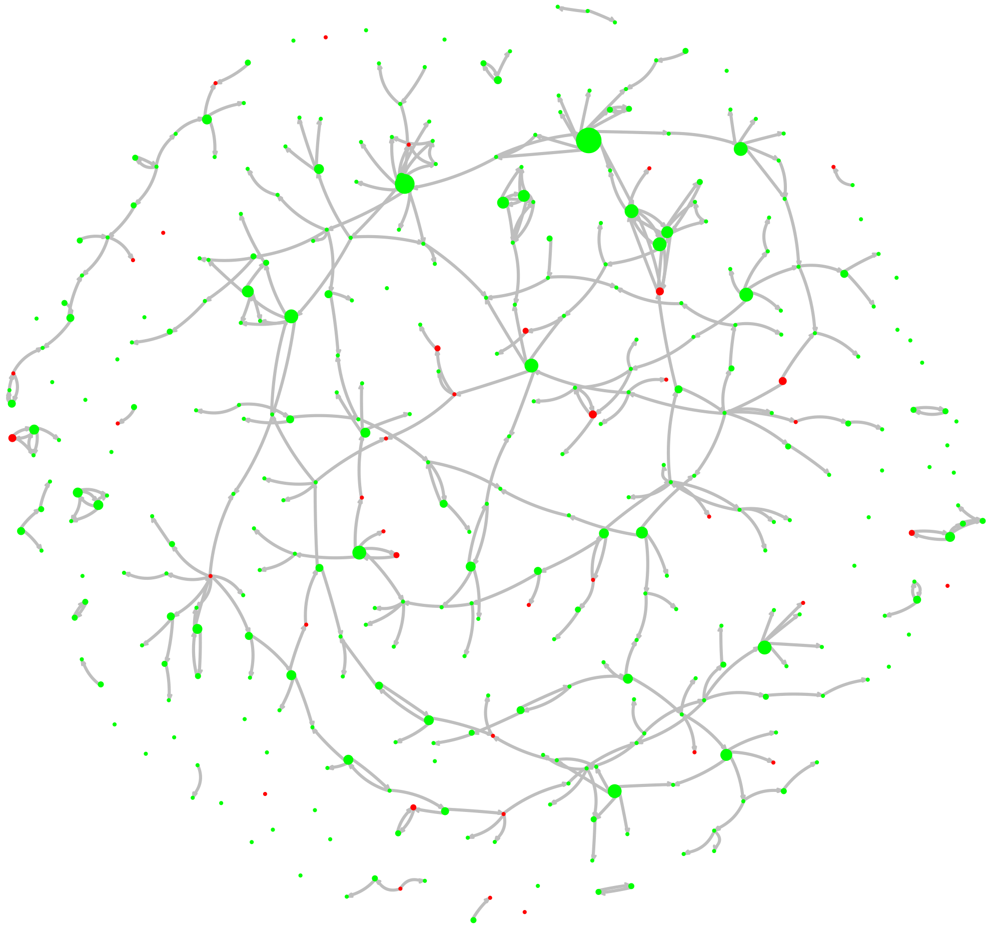


Nodes: high-school students (white/black/Hispanic)

Ties: friendship

Romantic/ sexual relationships at a US high school (Bearman, Moody & Stovel, 2004)

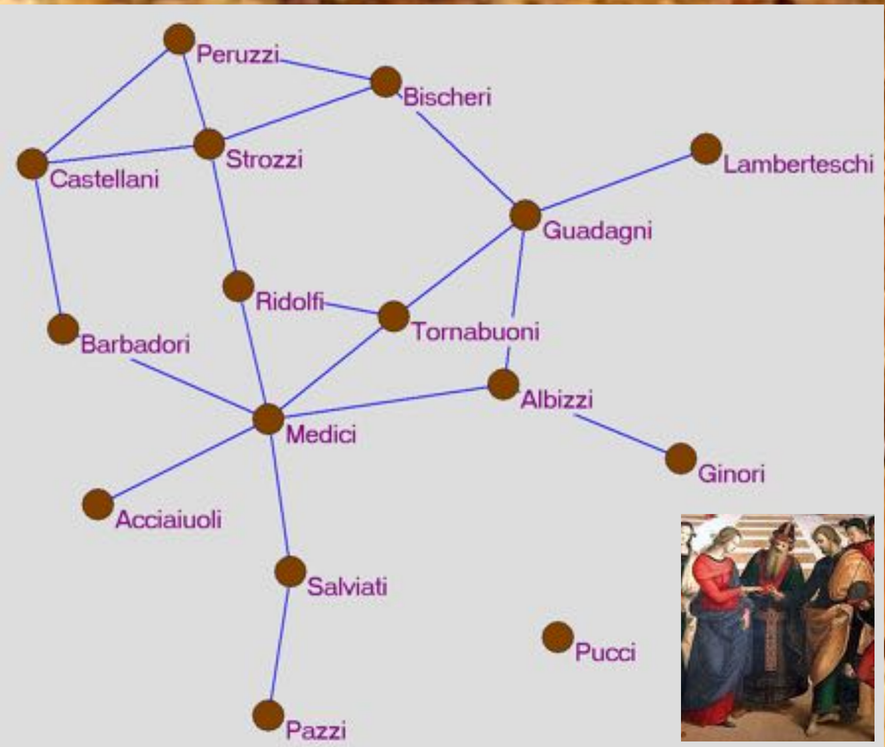
Bushfire-affected community Victoria



Nodes: person (high/low depression score)

Ties: social support

Padgett and Ansell (1993)

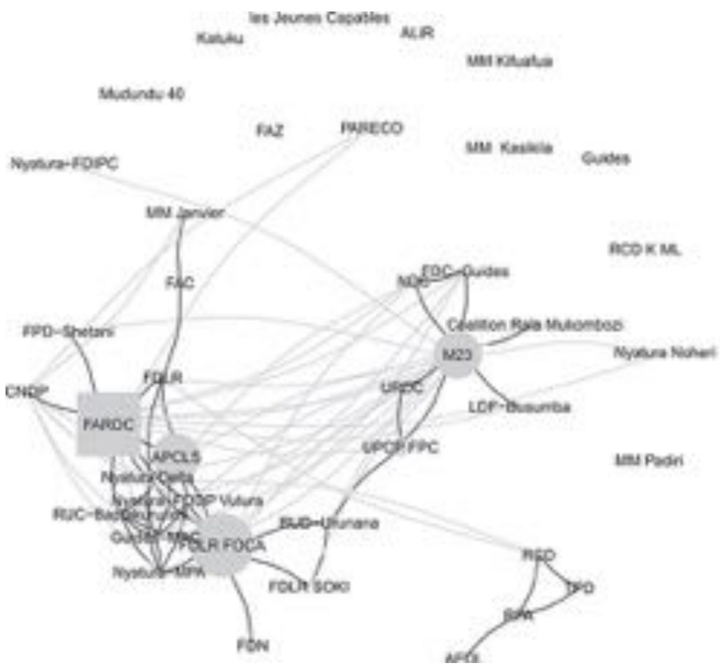


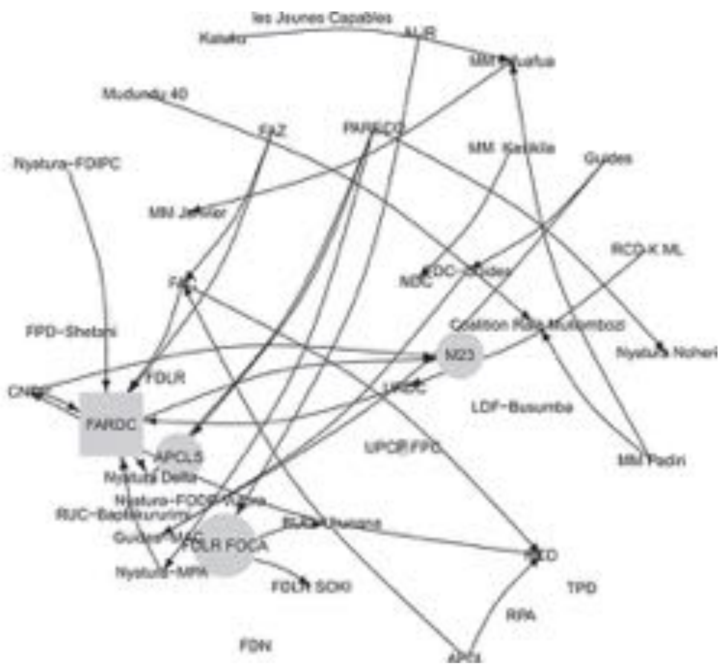
Nodes: families

Ties: marriage

Stys et al. (2020)

Nodes: rebel group

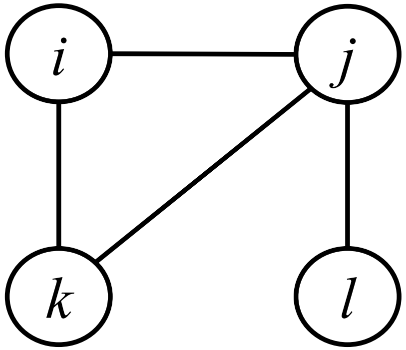




Ties: Alliance (black) enmity (grey) Ties: progeneities

Graph theoretic representation

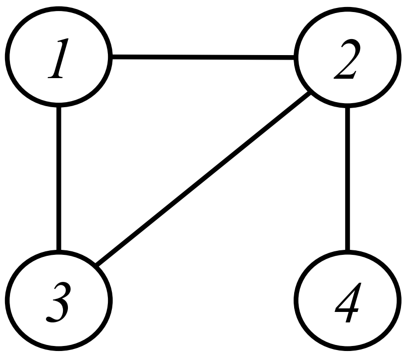
A Graph $G(V,E)$, is a collection of



Vertices/Nodes: $V=\{i,j,k,l\}$

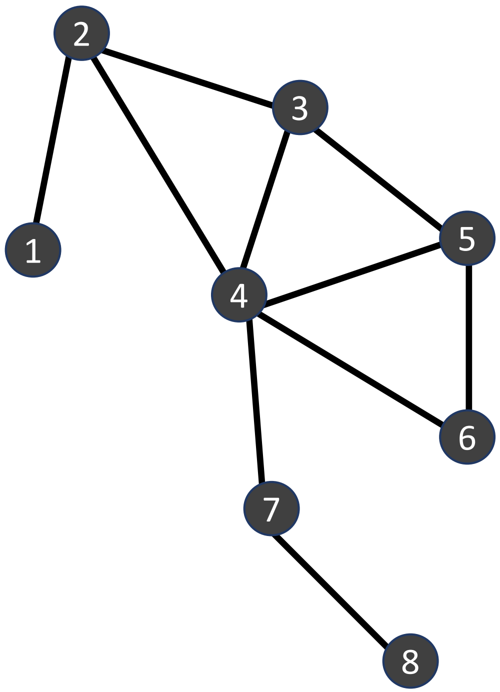
Edges/ties: $E=\{\{i,j\},\{i,k\},\{k,j\},\{l,j\}\}$

Edgelist



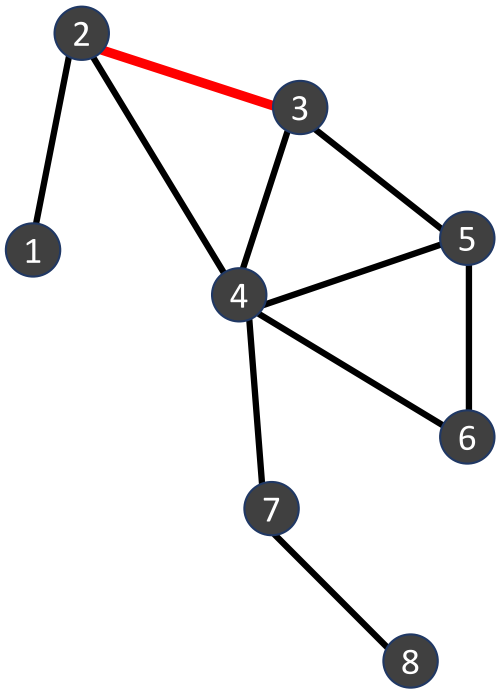
Sender	Receiver	Strength
1	2	1
1	3	1
2	3	1
2	4	1
2	1	1
2	3	1
2	4	1
3	1	1
3	2	1
4	2	1

Network graph – Adjacency matrix



	1	2	3	4	5	6	7	8
1	-	1						
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

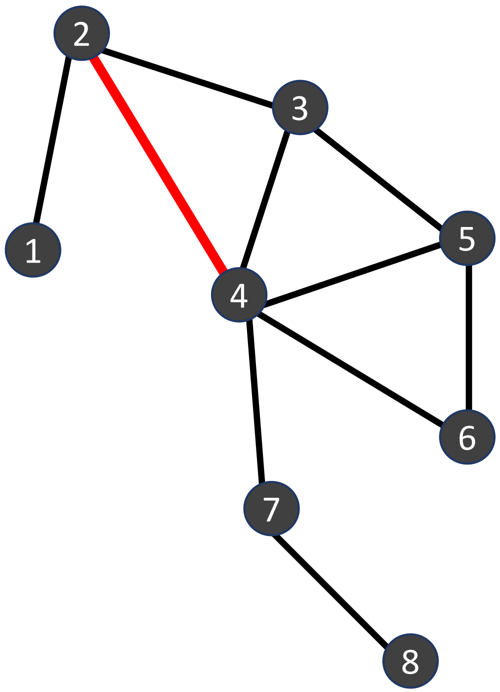
Network graph – Adjacency matrix



Tie between 2 and 3

	1	2	3	4	5	6	7	8
1	-	1						
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

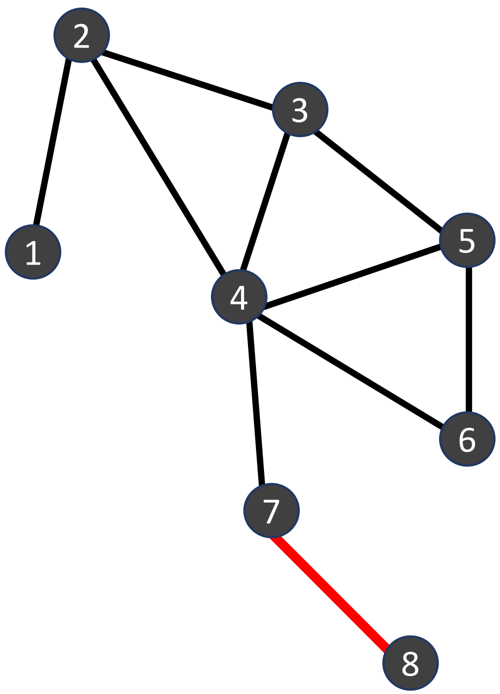
Network graph – Adjacency matrix



Tie between 2 and 4

	1	2	3	4	5	6	7	8
1	-	1						
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

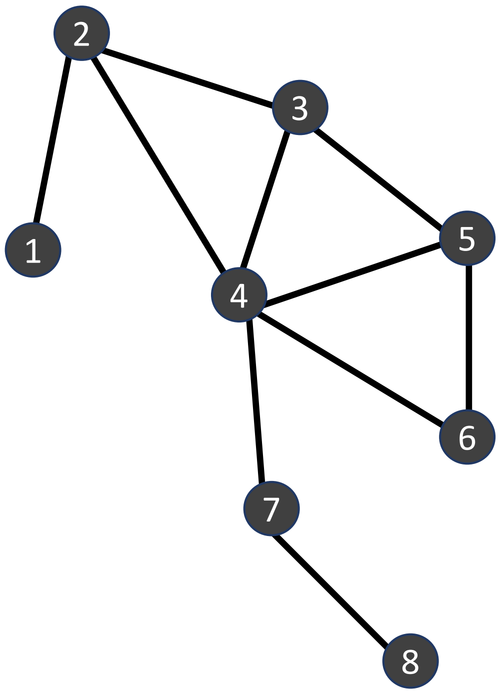
Network graph – Adjacency matrix



Tie between 7 and 8

	1	2	3	4	5	6	7	8
1	-	1						
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

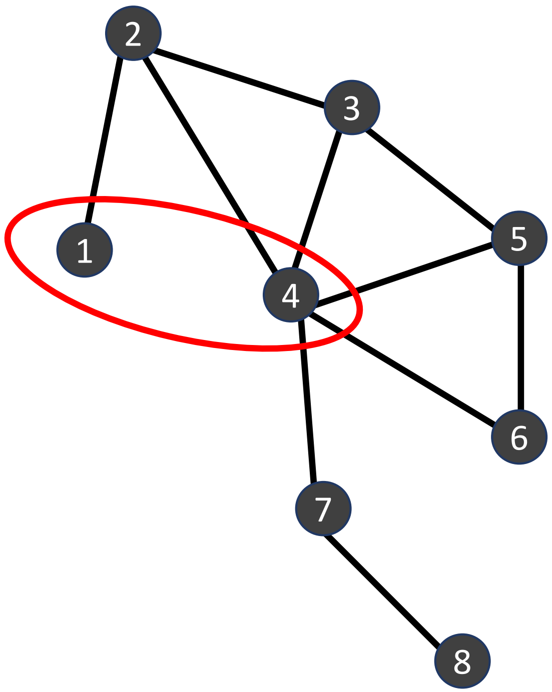
Network graph – Adjacency matrix



NO self-ties

	1	2	3	4	5	6	7	8
1	-	1						
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

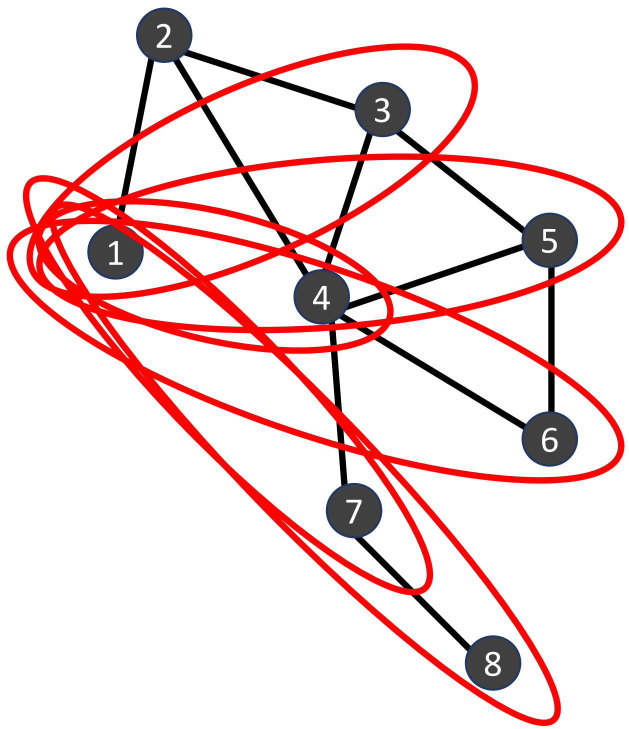
Network graph – Adjacency matrix



NO Tie between 1 and 4

	1	2	3	4	5	6	7	8
1	-	1		0				
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

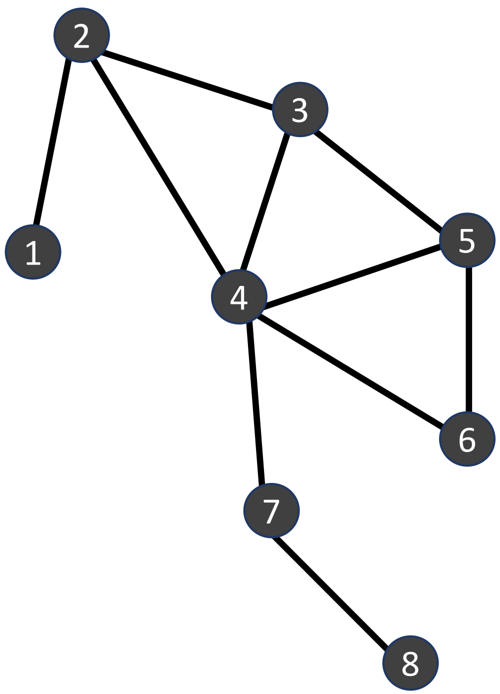
Network graph – Adjacency matrix



NO Tie between 1 and 3,4,5,6,6,8

	1	2	3	4	5	6	7	8
1	-	1	0	0	0	0	0	0
2		-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

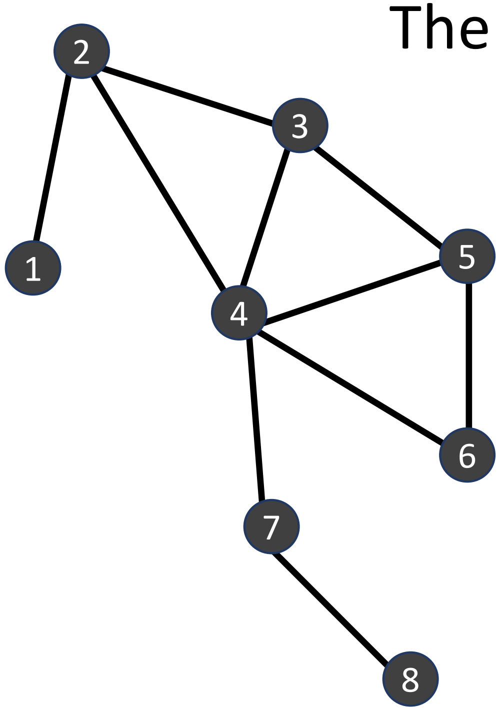
Network graph – Adjacency matrix



Set all NON-Ties to 0

	1	2	3	4	5	6	7	8
1	-	1	0	0	0	0	0	0
2		-	1	1	0	0	0	0
3			-	1	1	0	0	0
4				-	1	1	1	0
5					-	1	0	0
6						-	0	0
7							-	1
8								-

Network graph – Adjacency matrix

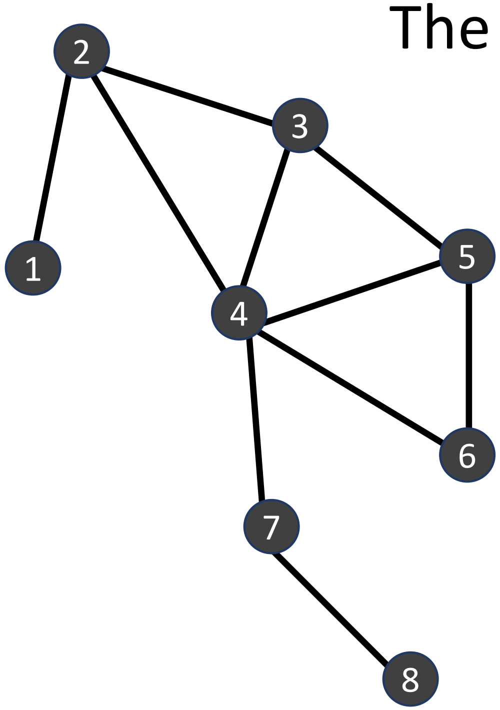


The network is undirected = the matrix is symmetric

	1	2	3	4	5	6	7	8
1	-	1						
2	1	-	1	1				
3			-	1	1			
4				-	1	1	1	
5					-	1		
6						-		
7							-	1
8								-

1 -> 2 = 2 -> 1

Network graph – Adjacency matrix

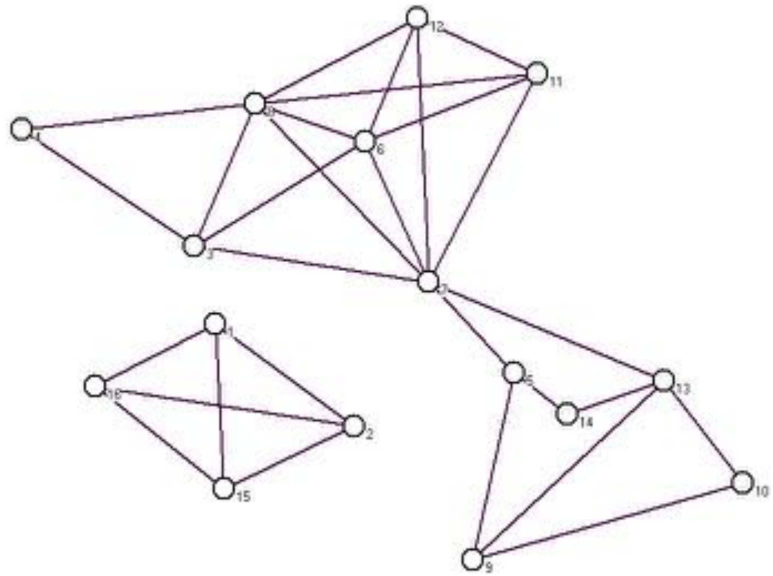


The network is undirected = the matrix is symmetric

	1	2	3	4	5	6	7	8
1	-	1						
2	1	-	1	1				
3		1	-	1	1			
4		1	1	-	1	1	1	
5			1	1	-	1		
6				1	1	-		
7				1			-	1
8							1	-

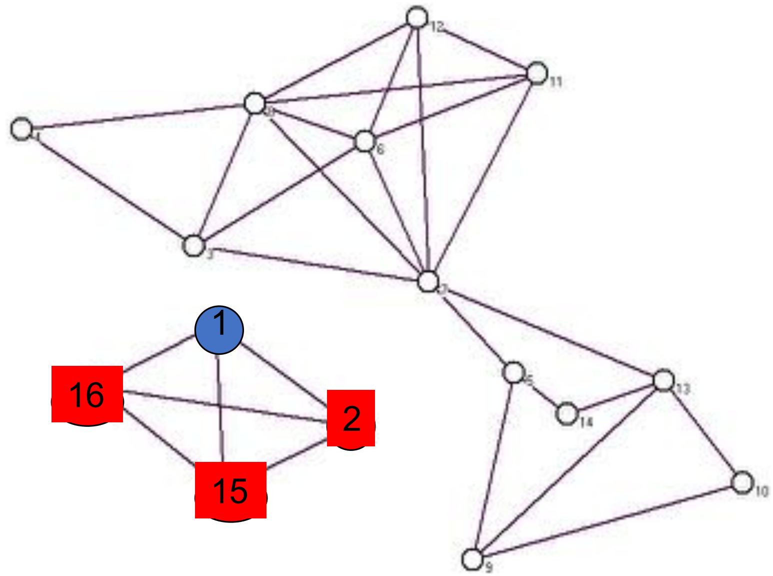
Read Highland tribes

0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1
1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
0	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0
0	0	1	0	1	1	0	1	0	0	1	1	1	0	0	0
0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	0
0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0
0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0



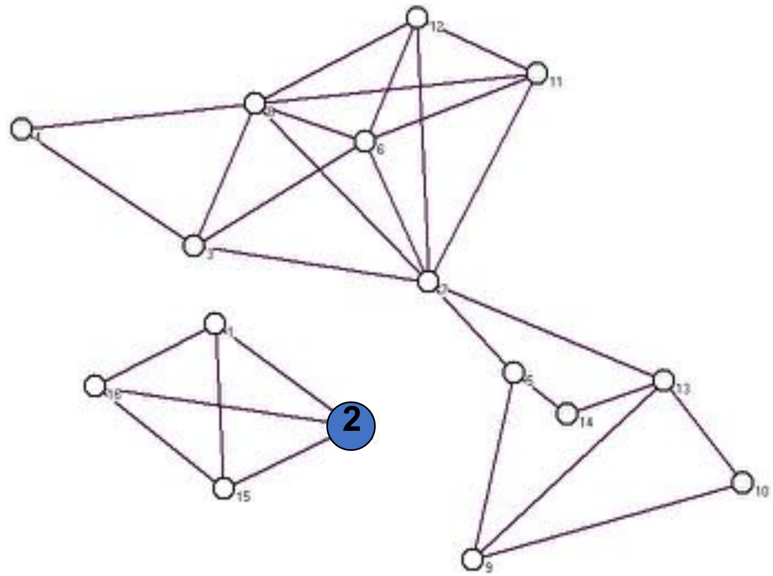
Read Highland tribes

0 1 0 0 0 0 0 0 0 0 0 0 0 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0
0 0 1 0 0 0 1 1 0 0 1 1 0 0 0 0
0 0 1 0 1 1 0 1 0 0 1 1 1 0 0 0
0 0 1 1 0 1 1 0 0 0 1 1 0 0 0 0
0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0
0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 0
0 0 0 0 0 1 1 1 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 1 1 0 0 0 1 0 0
0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0



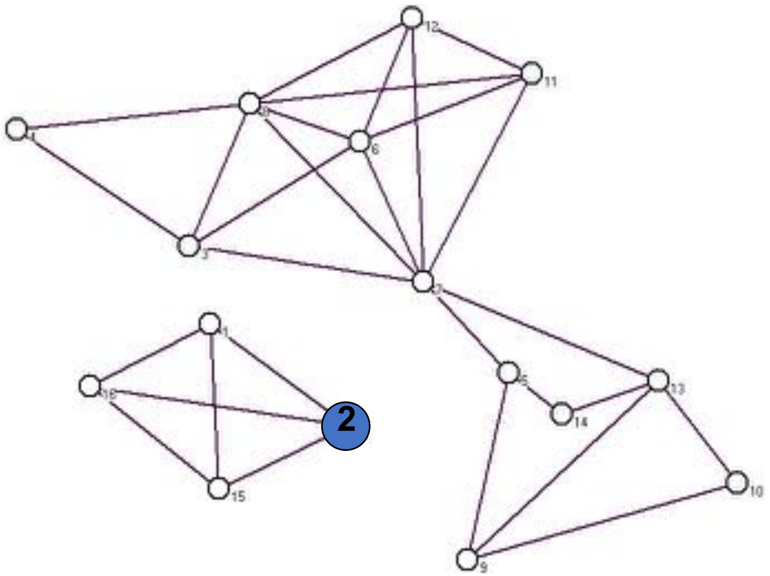
Read Highland tribes

0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0
0 0 1 0 0 0 1 1 0 0 1 1 0 0 0 0
0 0 1 0 1 1 0 1 0 0 1 1 1 0 0 0
0 0 1 1 0 1 1 0 0 0 1 1 0 0 0 0
0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0
0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 0
0 0 0 0 0 1 1 1 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 1 1 0 0 0 1 0 0
0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0



Read Highland tribes

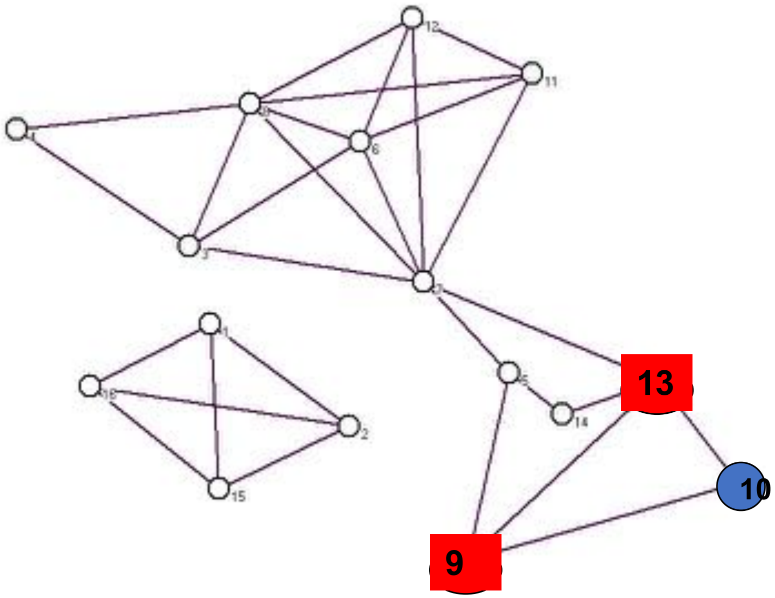
0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0
0 0 1 0 0 0 1 1 0 0 1 1 0 0 0 0
0 0 1 0 1 1 0 1 0 0 1 1 1 0 0 0
0 0 1 1 0 1 1 0 0 0 1 1 0 0 0 0
0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0
0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 0
0 0 0 0 0 1 1 1 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 1 1 0 0 0 1 0 0
0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0



Symmetric for a non-directed network

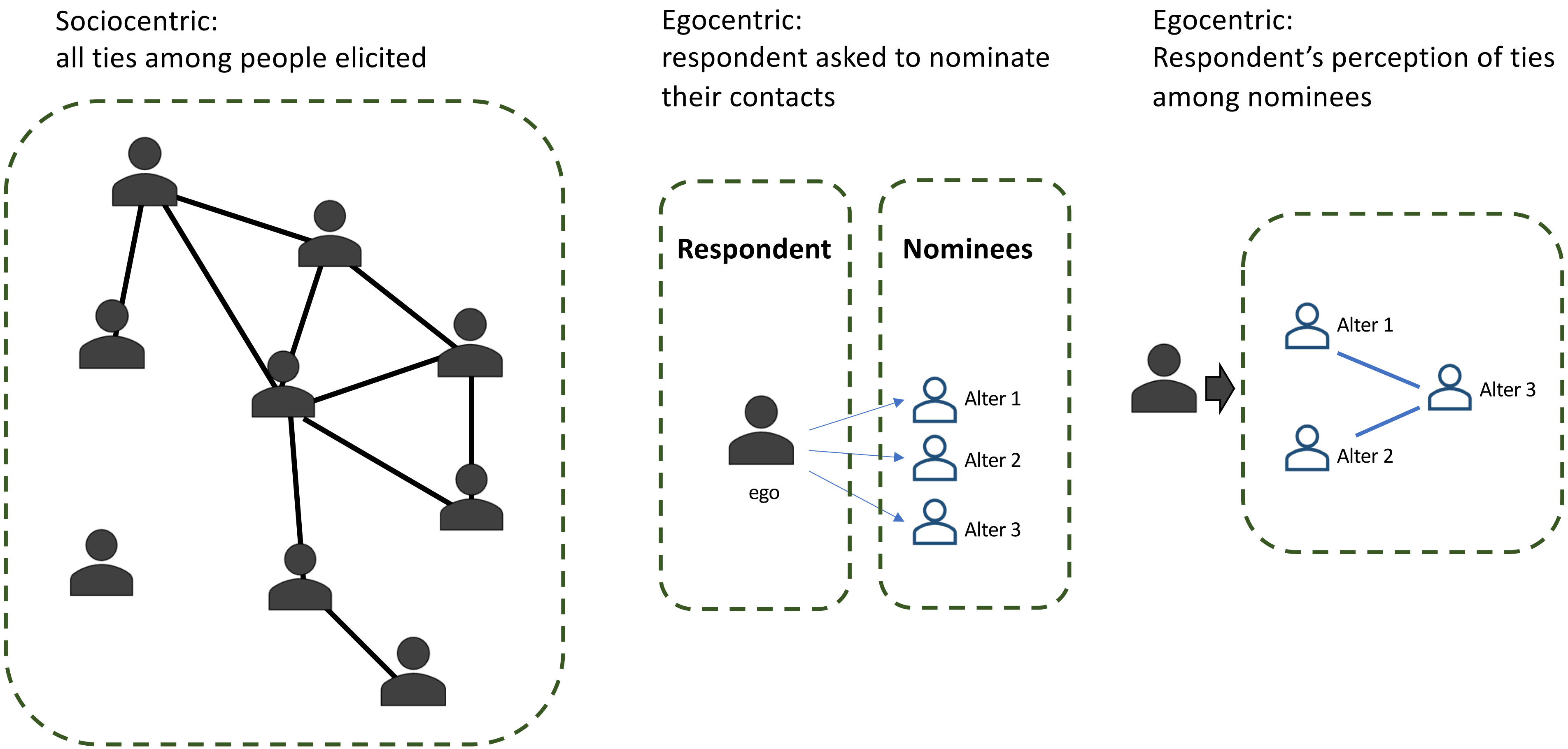
Read Highland tribes

0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1
0 0 0 1 0 1 1 1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0
0 0 1 0 0 0 1 1 0 0 1 1 0 0 0 0
0 0 1 0 1 1 0 1 0 0 1 1 1 0 0 0
0 0 1 1 0 1 1 0 0 0 1 1 0 0 0 0
0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0
0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 0
0 0 0 0 0 1 1 1 0 0 1 0 0 0 0 0
0 0 0 0 0 0 1 0 1 1 0 0 0 1 0 0
0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1
1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0



Where do network ties come from – sociometric surveys

- Ethnographic
  - Kapferer (1972)
- Archival
  - Padgett and Ansell (Marriage and business records)
  - Bright, Koskinen, Malm (court records)
- Name generator
- Resource generator
- Position generator
- Roster method
- Free recall
- Participant aided sociograms



Sociocentric:
all ties among people elicited
Egocentric:
respondent asked to nominate their contacts
Respondent
Nominees
ego
Alter 1
Alter 2
Alter 3
Egocentric:
Respondent's perception of ties among nominees
Alter 1
Alter 3
Alter 2

Lazerga’s (2001) lawfirm partners

"Here is the list of all the members of your Firm."

- Strong coworkers network:
  - "Because most firms like yours are also organized very informally, it is difficult to get a clear idea of how the members really work together. Think back over the past year, consider all the lawyers in your Firm. Would you go through this list and check the names of those with whom you have worked with. [By "worked with" I mean that you have spent time together on at least one case, that you have been assigned to the same case, that they read or used your work product or that you have read or used their work product; this includes professional work done within the Firm like Bar association work, administration, etc.]
- "Basic advice network:
  - "Think back over the past year, consider all the lawyers in your Firm. To whom did you go for basic professional advice? For instance, you want to make sure that you are handling a case right, making a proper decision, and you want to consult someone whose professional opinions are in general of great value to you. By advice I do not mean simply technical advice.
- "'Friendship' network:
  - "Would you go through this list, and check the names of those you socialize with outside work. You know their family, they know yours, for instance. I do not mean all the people you are simply on a friendly level with, or people you happen to meet at Firm functions."

van de Bunt (1999) students

Rate each person on a scale on the six point scale

Label	*Description of the response categories*
1. Best friendship	Persons whom you would call your 'real' friends
2. Friendship	Persons with whom you have a good relationship, but whom you do not (yet) consider a 'real' friend
3. Friendly relationship	Persons with whom you regularly have pleasant contact during classes. The contact could grow into a friendship
4. Neutral relationship	Persons with whom you have not much in common. In case of an accidental meeting the contact is good. The chance of it growing into a friendship is not large
0. Unknown person	Persons whom you do not know
5. Troubled relationship	Persons with whom you can't get on very well, and with whom you definitely do not want to start a relationship. There is a certain risk of getting into a conflict

Sociometric free recall

ID Number________

Who are your five BEST FRIENDS in this class?

Write their names on the lines below starting with your best friend in this class. After you write their name, look at the list of names on the roster that has been provided. Match the name to the number and write the number in the boxes. If you cannot think of five people in this class, then leave the extra lines blank.

For example, your best friend's name may be John Angeles. Then you would write his name and then look up his number, which is 1 2 3 and then write that in the boxes. It is written in as an example below.

	FIRST NAME	LAST NAME	ROSTER NUMBER
	John	*Angeles*	*1 2 3*
1			
2			
3			
4			
5			

General Social Survey – name generator

From time to time, most people discuss important matters with other people. Looking back over the last six months - who are the people with whom you discussed matters important to you? Just tell me their first names or initials. If LESS THAN 5 NAMES MENTIONED, PROBE, Anyone else? ONLY RECORD FIRST 5 NAMES.

LIST ALL NAMES IN ORDER ACROSS THE TOP OF THE MATRIX (SEE 2 PAGES AHEAD). THEN WRITE NAMES 2-5 DOWN THE SIDE OF THE MATRIX.

A. INTERVIEWER CHECK: HOW MANY NAMES WERE MENTIONED?

Name interpreters

- Present respondent with name
  - ➢ Do you feel very close to this person
  - ➢ Do you socialise regularly with this person outside of working hours
  - ➢ Are you required by the organisation to report to this person on important tasks
- Order
  - By name, or
  - By interpreter item

General Social Survey – name interpreter

From time to time, most people discuss important matters with other people. Looking back over the last six months - who are the people with whom you discussed matters important to you? Just tell me their first names or initials. If LESS THAN 5 NAMES MENTIONED, PROBE, Anyone else? ONLY RECORD FIRST 5 NAMES.
LIST ALL NAMES IN ORDER ACROSS THE TOP OF THE MATRIX (SEE 2 PAGES AHEAD). THEN WRITE NAMES 2-5 DOWN THE SIDE OF THE MATRIX.
A. INTERVIEWER CHECK: HOW MANY NAMES WERE MENTIONED?

Here is a list of some of the ways in which people are connected to each other. Some people can be connected to you in more than one way. For example, a man could be your brother and he may belong to your church and be your lawyer. When I read you a name, please tell me all of the ways that person is connected to you. How is (NAME) connected to you? PROBE: What other ways? (The options were presented on a card: Spouse, Parent, Sibling, Child, Other family, C—worker, Member of group, Neighbour, Friend, Advisor, Other.)

Position generator

Of your relatives, friends and social associates, is there anyone who has the jobs listed below? What is your relationship to them? What is his/her ethnicity if not the same as yours? Does he or she give you help or advice?

Occupation	Do you know people who have this job? Please answer all that applies.	What is his or her relationship to you? (Show card)	Is he or she of the same ethnicity as you? If not, what is his or her ethnicity? (Show card)	If you need help or advice in setting up or running your business, will you turn to him/her for help?	Do you sometimes talk with him or her about your business plans/worries?	How long have you known each other?	If you need a large sum of money, will you turn to him or her for help?
1. Solicitor 2. Bank/building society manager 3. Accountant 4. Business person 5. Insurance manager 6. Gov business advisor 7. Sales manager 8. University lecturer 9. Real estate agent 10. Hotelier 11. Restaurant owner 12. Someone running a take-away 13. Pharmacist 14. Taxi driver 15. Retailer (shop or news agent)							

Resource generator

"Do you know[1] anyone who…"

1 can repair a car, bike, etc.
2 owns a car
3 is handy repairing household equipment
4 can speak and write a foreign language
5 can work with a personal computer

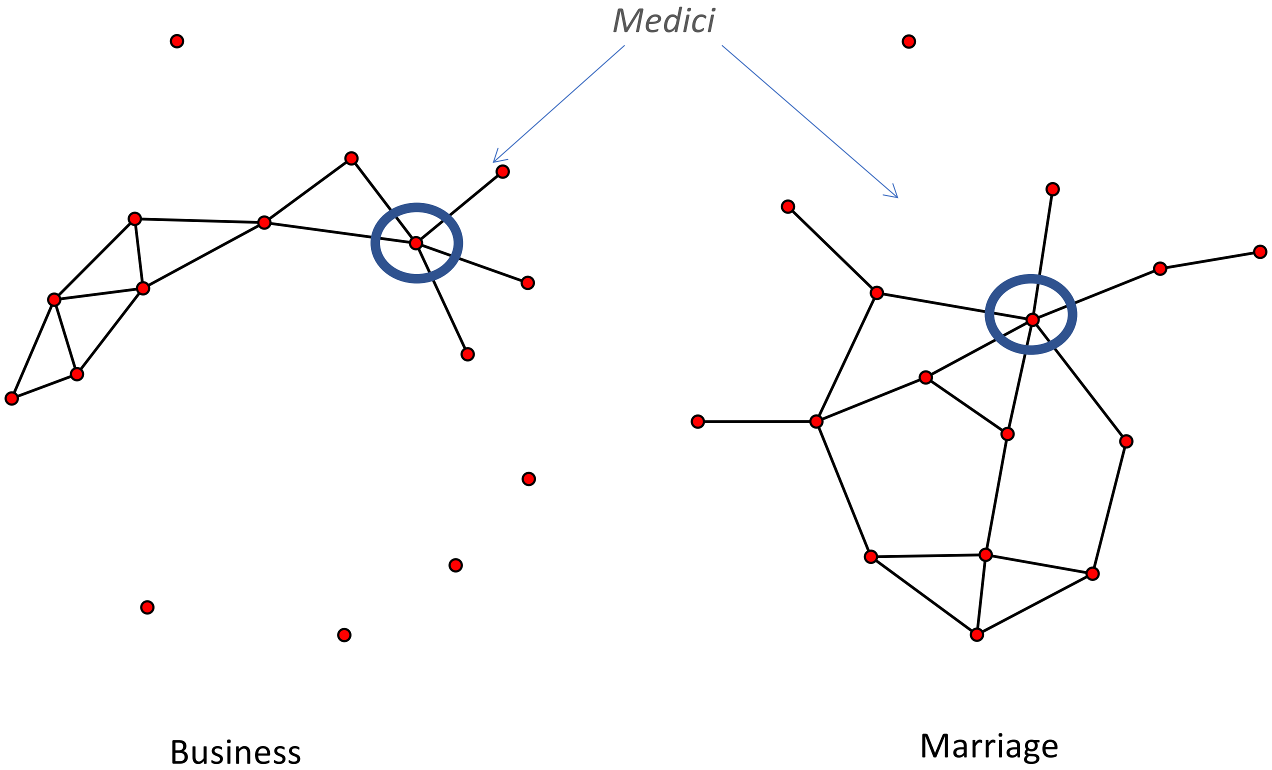
Table B: The SSND Resource Generator and responses: percentage of sample who mentioned at least one alter per resource item in any relationship, and strongest relationship when known (Survey on the Social Networks of the Dutch (SSND), 1999-2000; $N=1{,}004$).

		% 'yes'	*if yes, access through*			
	"Do you know[1] anyone who…"		acq.	friend	family member	scale[2]
1	can repair a car, bike, etc.	83	16	18	66	
2	owns a car	87	0	3	97	g
3	is handy repairing household equipment	72	12	17	71	
4	can speak and write a foreign language	87	4	11	84	g
5	can work with a personal computer	90	2	9	89	g
6	can play an instrument	79	10	16	74	
7	has knowledge of literature	70	9	23	67	p
8	has senior high school (VWO) education	87	6	14	81	p
9	has higher vocational (HBO) education	94	6	13	82	p
10	reads a professional journal	78	7	13	81	g
11	is active in a political party	34	34	26	39	s
12	owns shares for at least Dfl.10,000[3]	54	11	21	67	
13	works at the town hall	42	44	23	34	
14	earns more than Dfl.5,000 monthly	76	10	19	71	p
15	own a holiday home abroad	41	34	26	41	p
16	is sometimes in the opportunity to hire people	65	21	23	57	s
17	knows a lot about governmental regulations	69	23	25	52	
18	has good contacts with a newspaper, radio- or TV station	32	36	24	41	p
19	knows about soccer	80	7	16	77	
20	has knowledge about financial matters (taxes, subsidies)	81	15	22	64	s
21	can find a holiday job for a family member	61	29	23	47	
22	can give advice concerning a conflict at work	73	22	32	46	s
23	can help when moving house (packing, lifting)	95	4	17	79	s
24	can help with small jobs around the house (carpenting, painting)	91	9	20	70	
25	can do your shopping when you (and your household members) are ill	96	11	24	64	
26	can give medical advice when you are dissatisfied with your doctor	56	20	31	48	
27	can borrow you a large sum of money (Dfl.10,000)	60	3	13	84	
28	can provide a place to stay for a week if you have to leave your house temporarily	95	2	15	83	
29	can give advice concerning a conflict with family members	83	3	33	64	s
30	can discuss which political party you are going to vote for	65	5	27	68	
31	can give advice on matters of law (problems with landlord, boss, or municipality)	64	24	32	44	
32	can give a good reference when you are applying for a job	65	37	37	26	s
33	can babysit for your children	57	12	17	71	

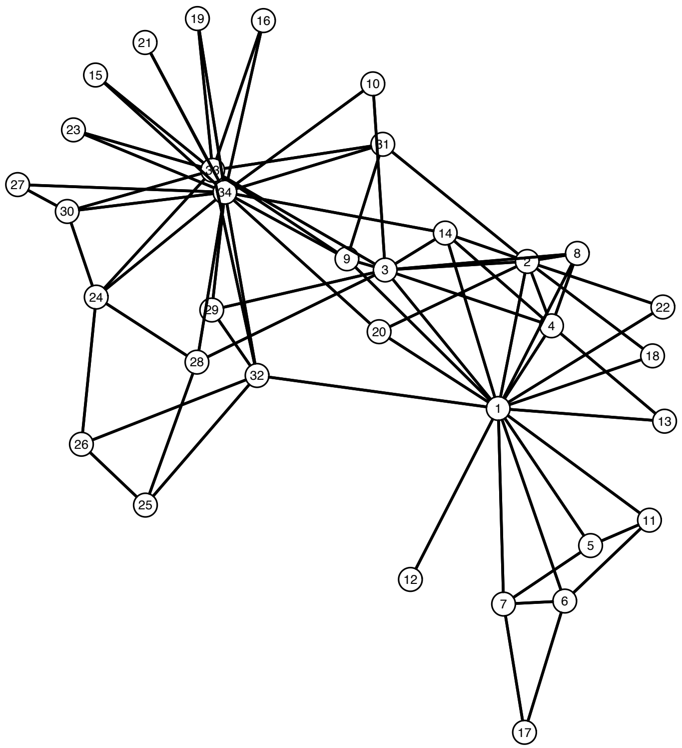
Degree

Density, degree, degree distribution

Padgett and Ansell (1993) Florentine families

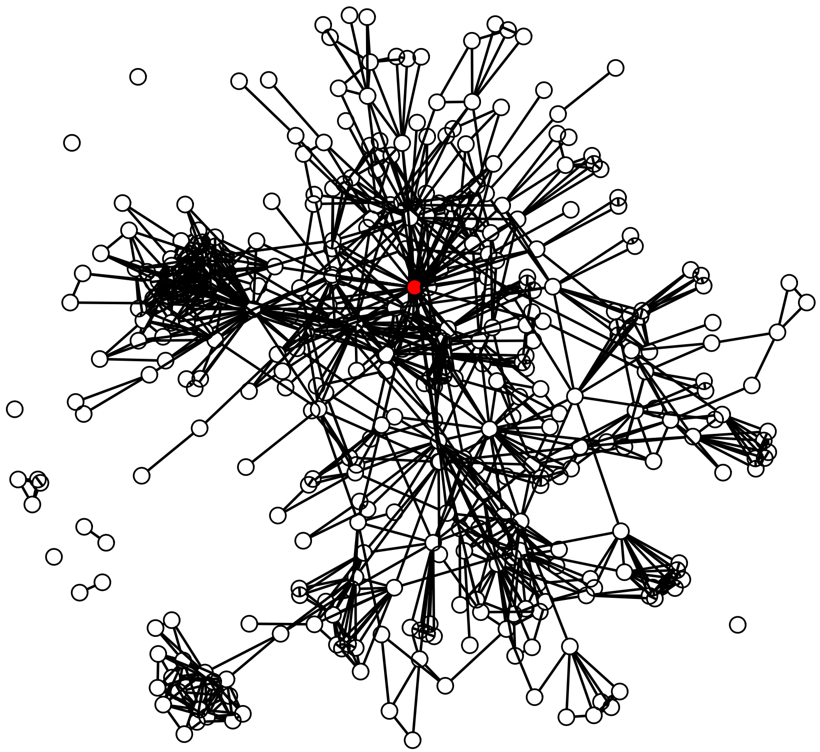


Zachary’s (1977) Karate club



… a story about two rivalling leaders

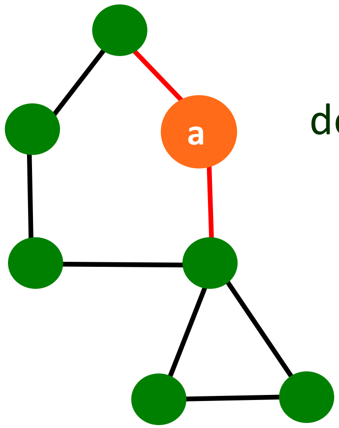
Sageman's (2011) Al Queda dataset



... Osama bin Laden

Ties defined as in Koskinen, Robins, Wang, and Pattison. 2013. Social Networks

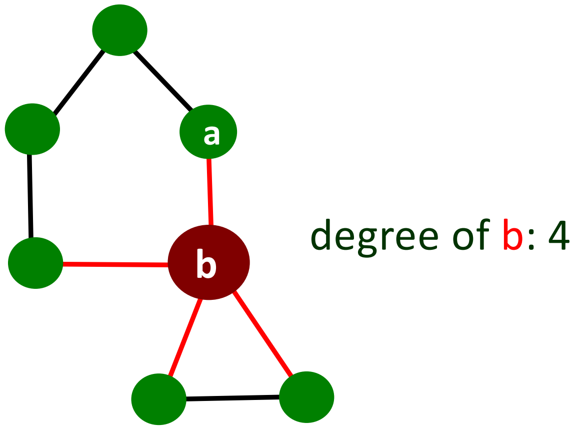
Degree of a node



degree of a: 2

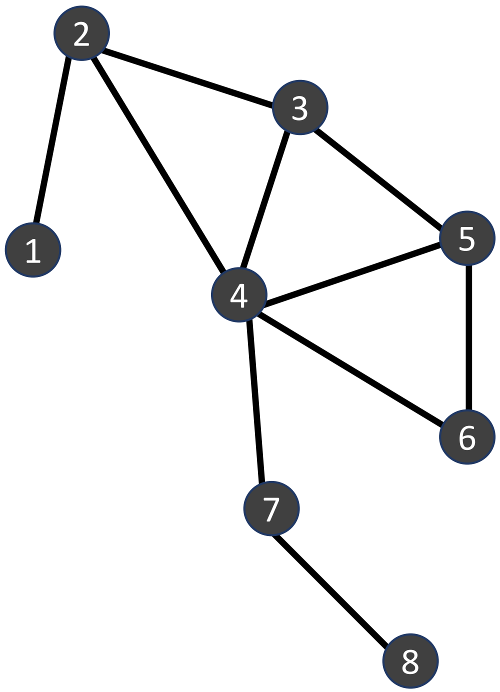
The degree of a node is the number of edges incident to it

Degree of a node



The degree of a node is the number of edges incident to it

Each cell is a **variable** $X_{\text{row,column}}$ that can be 1 or 0

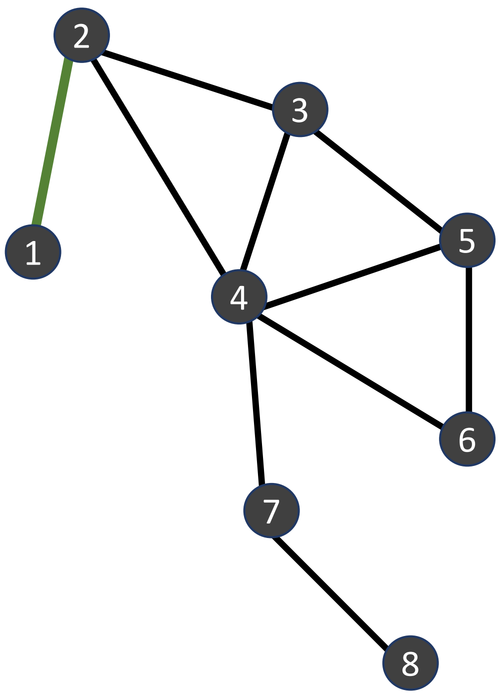


$X_{1,2} = 1$

$X_{1,3} = 0$

	1	2	3	4	5	6	7	8
1	-	1	0	0	0	0	0	0
2	1	-	1	1	0	0	0	0
3	0	1	-	1	1	0	0	0
4	0	1	1	-	1	1	1	0
5	0	0	1	1	-	1	0	0
6	0	0	0	1	1	-	0	0
7	0	0	0	1	0	0	-	1
8	0	0	0	0	0	0	1	-

Degree of a node: number of ties

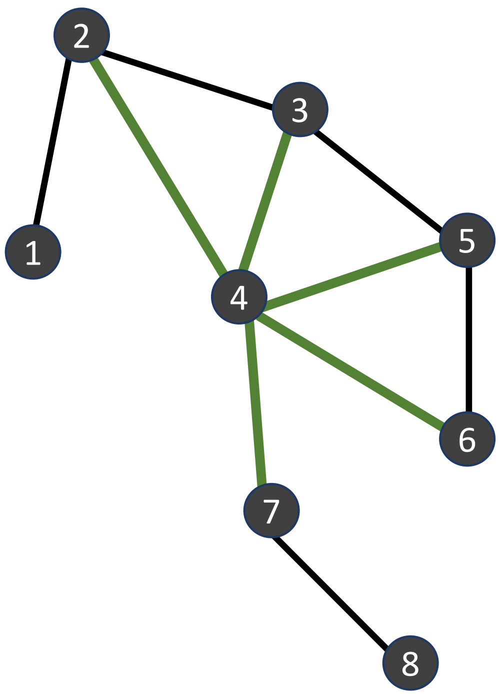


$$X_{1,2} + X_{1,3} + X_{1,4} + X_{1,5} + X_{1,6} + X_{1,7} + X_{1,8} = 1$$

	1	2	3	4	5	6	7	8
1	-	1	0	0	0	0	0	0
2	1	-	1	1	0	0	0	0
3	0	1	-	1	1	0	0	0
4	0	1	1	-	1	1	1	0
5	0	0	1	1	-	1	0	0
6	0	0	0	1	1	-	0	0
7	0	0	0	1	0	0	-	1
8	0	0	0	0	0	0	1	-

= 1

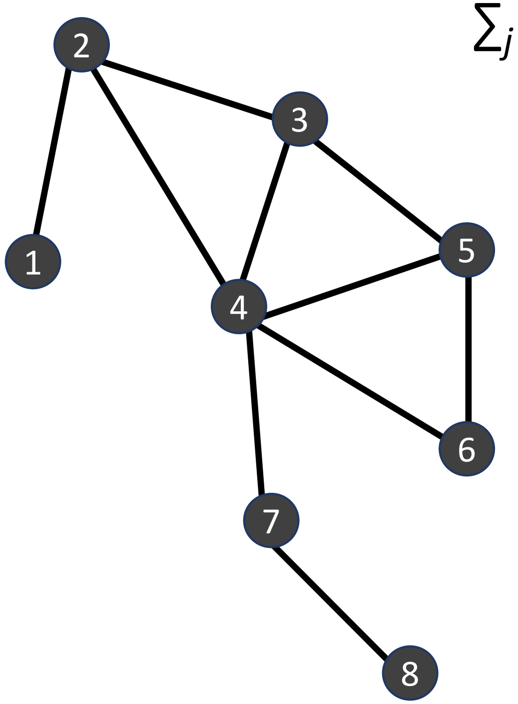
Degree of a node: number of ties



$$X_{4,1} + X_{4,2} + X_{4,3} + X_{4,5} + X_{4,6} + X_{4,7} + X_{4,8} = 5$$

	1	2	3	4	5	6	7	8	
1	-	1	0	0	0	0	0	0	= 1
2	1	-	1	1	0	0	0	0	
3	0	1	-	1	1	0	0	0	
4	0	1	1	-	1	1	1	0	= 5
5	0	0	1	1	-	1	0	0	
6	0	0	0	1	1	-	0	0	
7	0	0	0	1	0	0	-	1	
8	0	0	0	0	0	0	1	-	

Degree of a node: number of ties



$$\sum_j X_{ij} = X_{i,1} + X_{i,2} + X_{i,3} + X_{i,4} + X_{i,5} + X_{i,6} + X_{i,7} + X_{i,8} = d_i$$

	1	2	3	4	5	6	7	8	
1	-	1	0	0	0	0	0	0	= 1
2	1	-	1	1	0	0	0	0	= 3
3	0	1	-	1	1	0	0	0	= 3
4	0	1	1	-	1	1	1	0	= 5
5	0	0	1	1	-	1	0	0	= 3
6	0	0	0	1	1	-	0	0	= 2
7	0	0	0	1	0	0	-	1	= 2
8	0	0	0	0	0	0	1	-	= 1

Degree of a node – Florentine families

Executable file | 16 lines (16 sloc) | 5

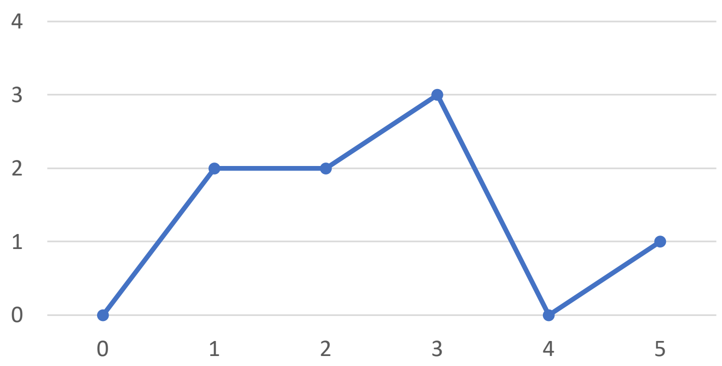
```
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 0 0 1 0 1 0 0 0 0 0
0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0
0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0
0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0
0 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0
0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 1
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 1 1 1 0 0 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
```

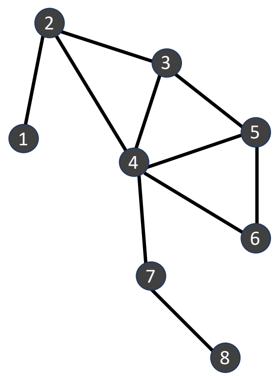
```
> rowSums(padgettbus)
 [1] 0 0 4 3 3 2 2 4 5 1 4 0 0 1 0 1

> degree(padgettbus,gmode='graph')
 [1] 0 0 4 3 3 2 2 4 5 1 4 0 0 1 0 1
```

Degree and degree centrality

- Degree centrality is the degree of a node (possibly normalised)
- Why is centrality important?
    - *Network Process: what network processes make nodes central?*
    - *Network Position: does having more support/collaboration ties protect you/give you access to information?*
    - *Network Properties: is spread more likely if there are high-degree nodes (super spreaders)*

Degree distribution: frequencies of degrees

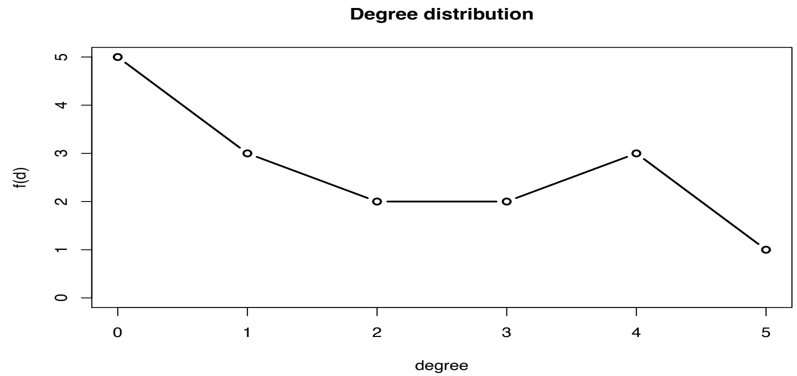




Degree	0	1	2	3	4	5
#nodes	0	2	2	3	0	1

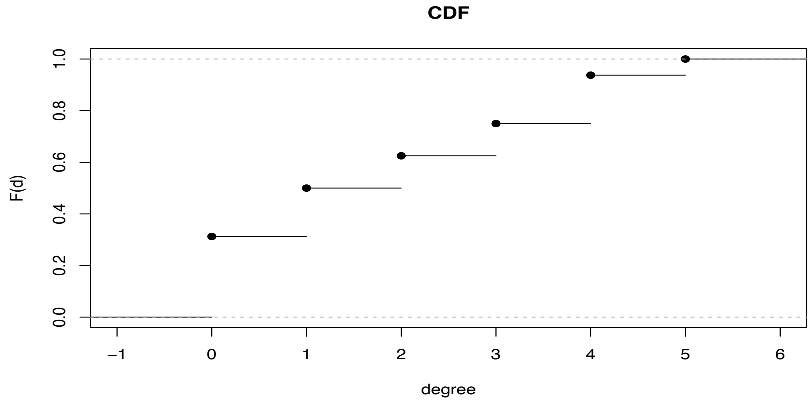
	1	2	3	4	5	6	7	8	
1	-	1							= 1
2	1	-	1	1					= 3
3		1	-	1	1				= 3
4		1	1	-	1	1	1		= 5
5			1	1	-	1			= 3
6				1	1	-			= 2
7				1			-	1	= 2
8							1	-	= 1

```r
plot( table( rowSums( padgettbus ) ),
      type='b',
      xlab = 'degree',
      ylab = 'f(d)',
      main='Degree distribution')
```



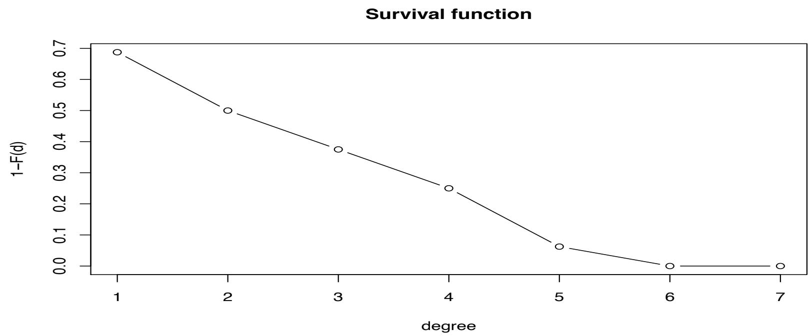
Pr($d_i = k$)

```r
Fn <- ecdf( rowSums( padgettbus ) )
plot(Fn,  main='CDF',
      xlab='degree',
      ylab='F(d)')
```

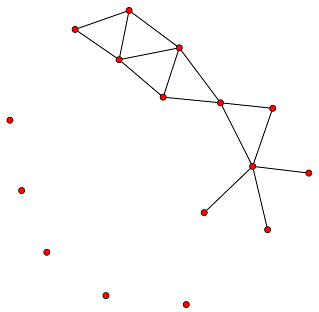


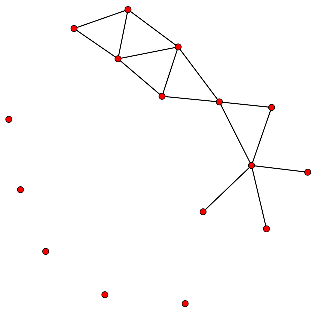
Pr($d_i \leq k$)

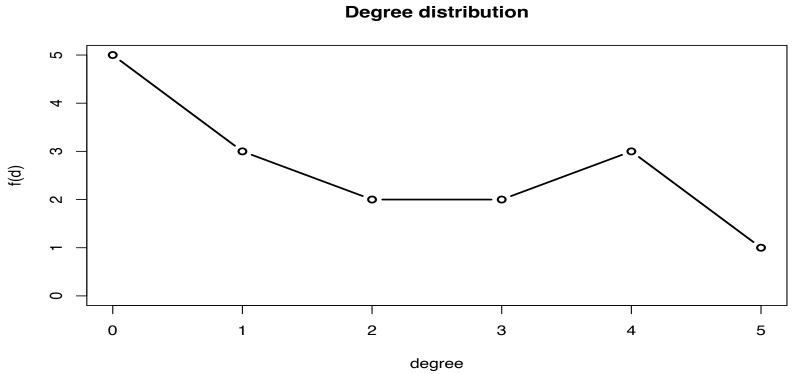
```r
plot(1-Fn(0:6),
      main ='Survival function' ,
      type='b',
      xlab = 'degree',
      ylab = '1-F(d)')
```



Pr($d_i > k$)



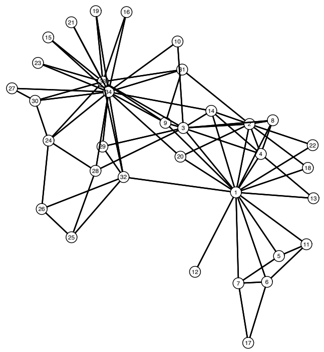


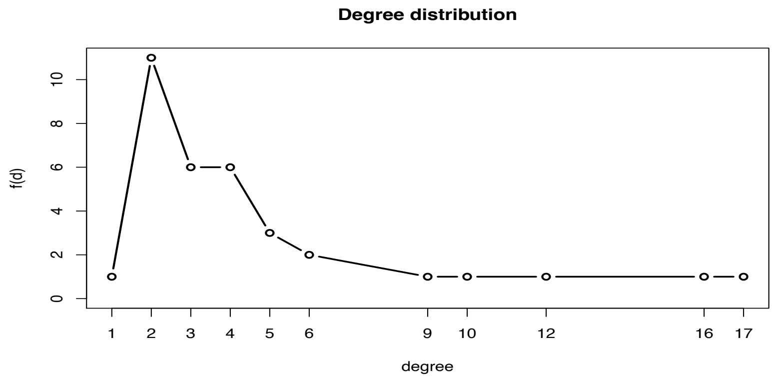


Average degree: $\sum d_i / n$

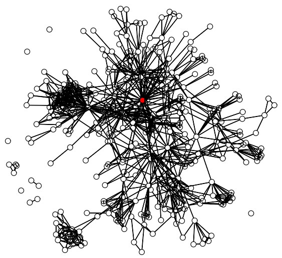
```
> sum( padgettbus )/16
[1] 1.875
```

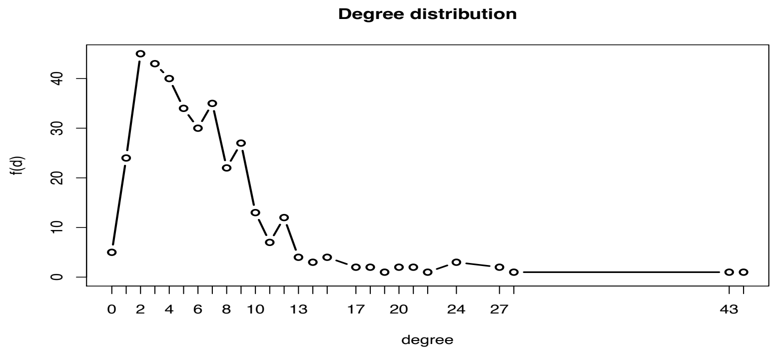
1.88





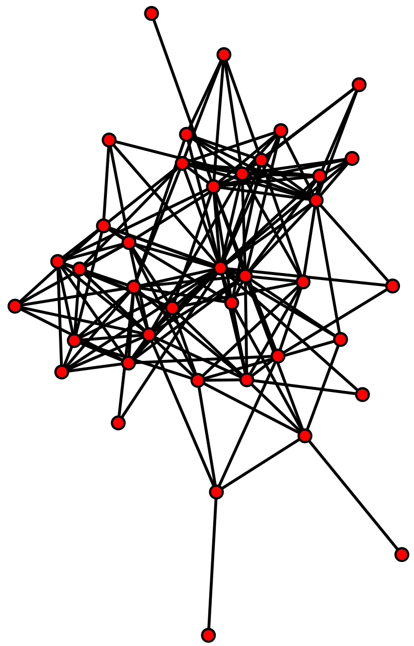
4.59

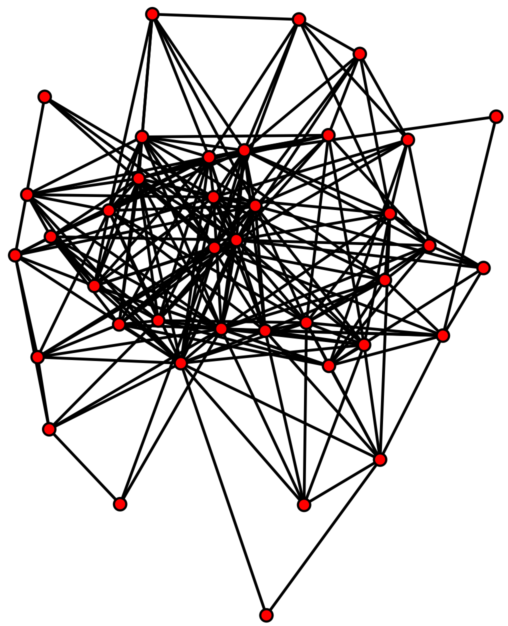




6.44

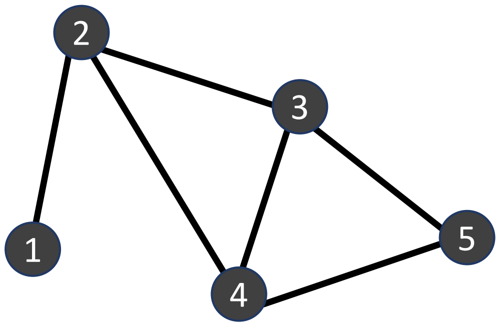
Density: Kapferer's (1972) tailors time 1 and 2





Which network is more 'dense'?

Density: proportion of ties out of all possible



	1	2	3	4	5
1	-	1	0	0	0
2	1	-	1	1	0
3	0	1	-	1	1
4	0	1	1	-	1
5	0	0	1	1	-

	1	2	3	4	5
1	-	X_{12}	X_{13}	X_{14}	X_{15}
2	X_{21}	-	X_{23}	X_{24}	X_{25}
3	X_{31}	X_{32}	-	X_{34}	X_{35}
4	X_{41}	X_{42}	X_{43}	-	X_{45}
5	X_{51}	X_{52}	X_{53}	X_{54}	-

n

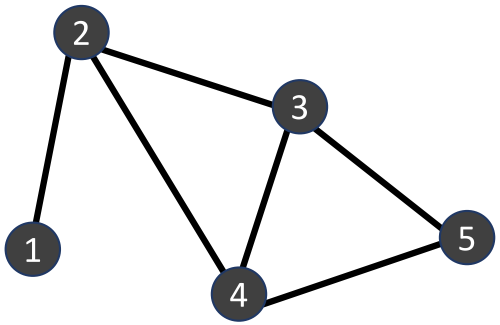
n

$$\frac{\sum_{i<j} X_{ij}}{n(n-1)/2} = \frac{\sum_{i,j} X_{ij}}{n(n-1)} = \frac{2\times 6}{5\times 4} = \frac{3}{5}$$

n^2 cells – n diagonal cells: $n(n-1)$

$X_{ij} = X_{ji}$ so double counting

Average degree



	1	2	3	4	5
1	-	1	0	0	0
2	1	-	1	1	0
3	0	1	-	1	1
4	0	1	1	-	1
5	0	0	1	1	-

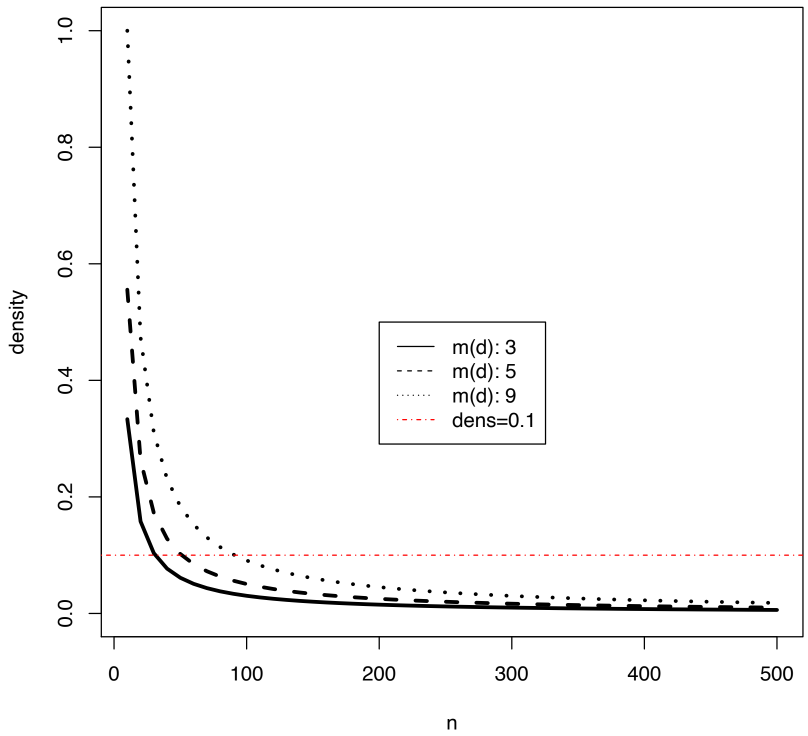
n

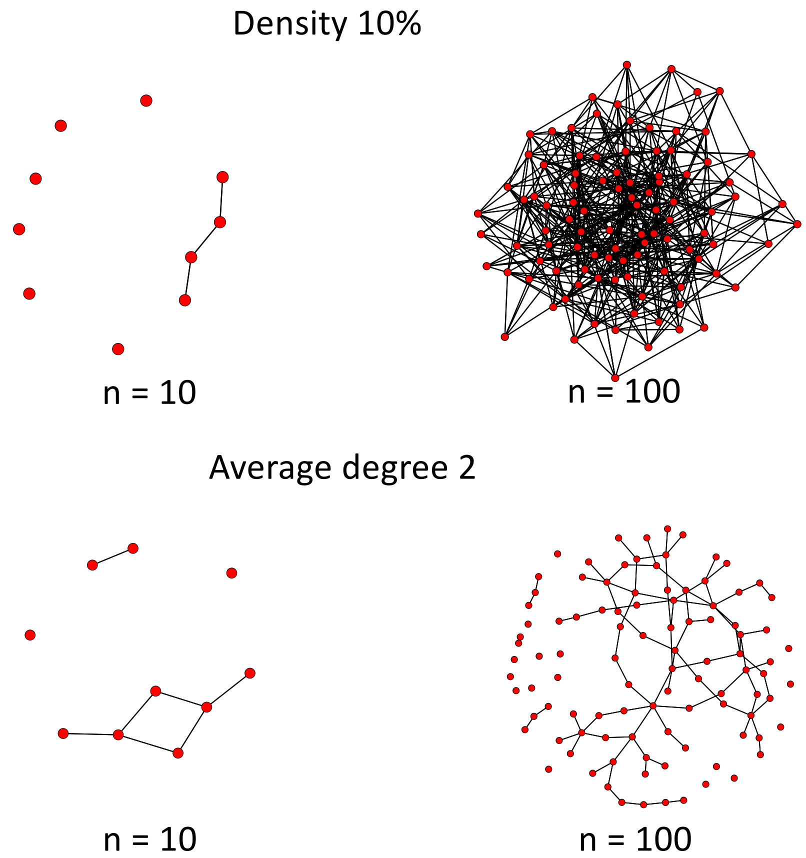
	1	2	3	4	5
1	-	X_{12}	X_{13}	X_{14}	X_{15}
2	X_{21}	-	X_{23}	X_{24}	X_{25}
3	X_{31}	X_{32}	-	X_{34}	X_{35}
4	X_{41}	X_{42}	X_{43}	-	X_{45}
5	X_{51}	X_{52}	X_{53}	X_{54}	-

$d_1 = \sum_j X_{1j}$

$$\frac{\sum_i \sum_j X_{ij}}{n} = \frac{\sum_i d_i}{n} = \frac{2\times 6}{5} = 2.4$$

How does density scale

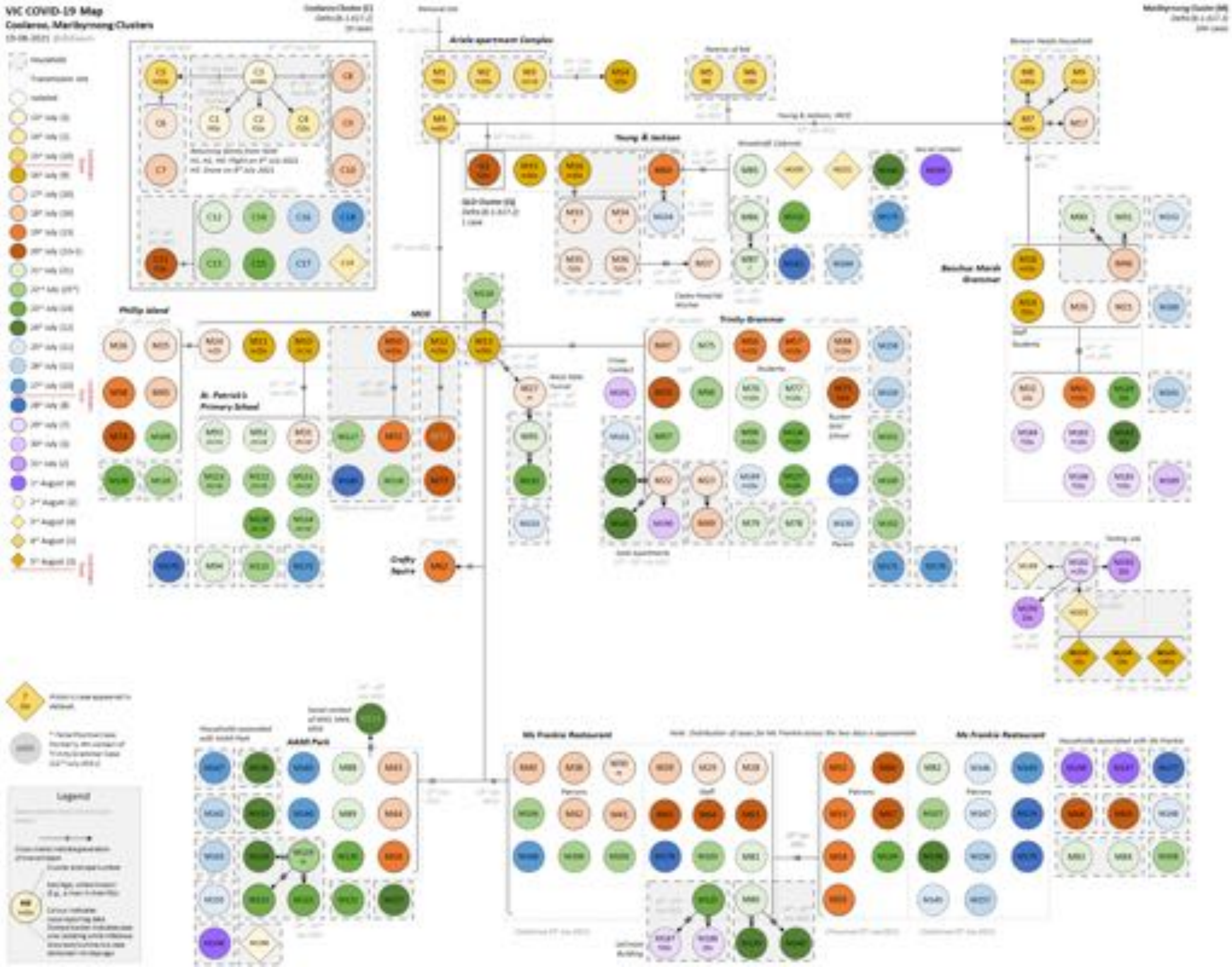




Reach

path, geodesic, distance, diameter

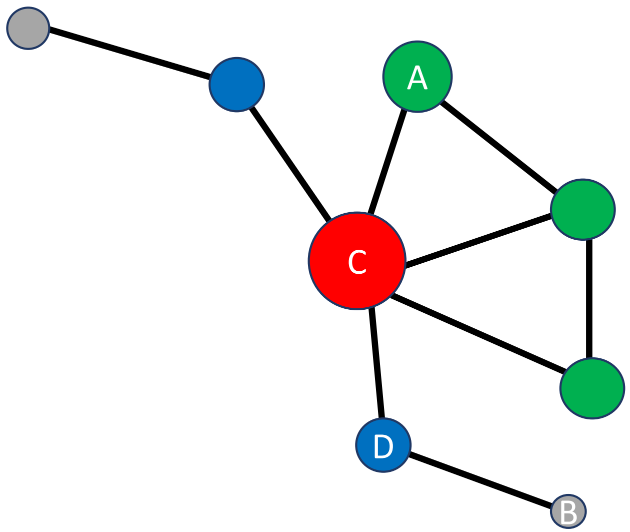
https://www.theguardian.com/world/datablog/ng-interactive/2021/sep/03/how-contagious-delta-variant-covid-19-r0-r-factor-value-number-explainer-see-how-coronavirus-spread-infectious-flatten-the-curve



VIC COVID-19 Map
Coolaroo, Maribyrnong Clusters
Ariele apartment Complex
Young & Jackson
Phillip Island
St. Patrick's Primary School
Trinity Grammar
Crafty Squire
AAMI Park
Legend

Path

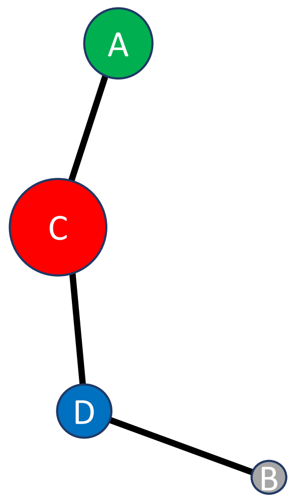
- A **path** is a sequence of ties connecting two nodes



$e_1, e_2, \ldots, e_k$

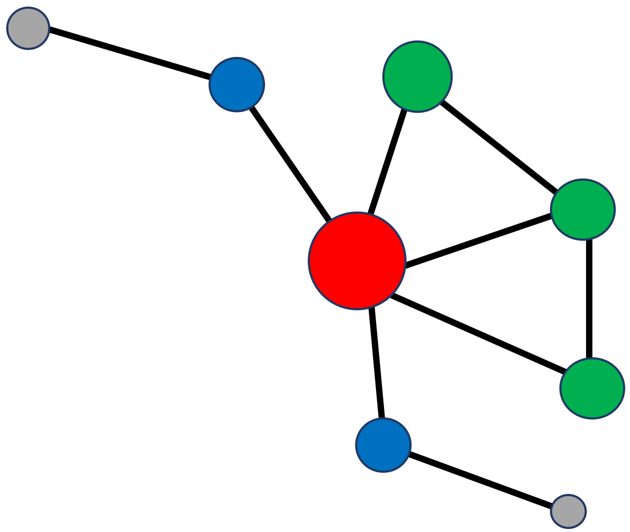
$e_1 = \{A,C\}$, $e_2 = \{C,D\}$, $e_3 = \{D,B\}$

Node sequence: *A,C,D,B*

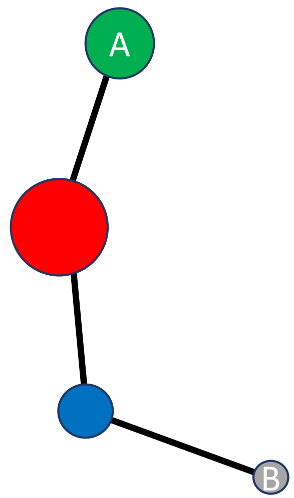


Path

- A **path** is a sequence of ties connecting two nodes



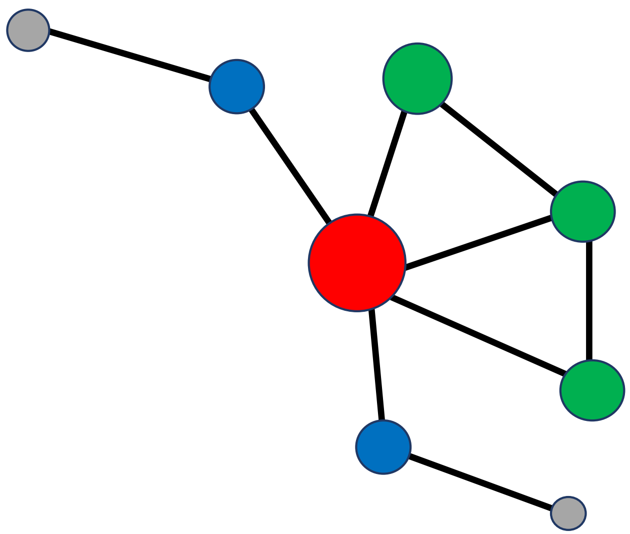
A path is of length 3 between A and B



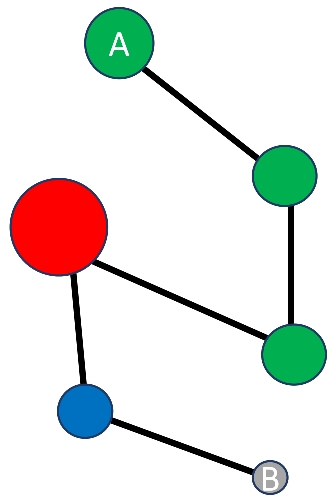
$$|\{A,C\},\{C,D\},\{D,B\}|=3$$

Path

- A **path** is a sequence of ties connecting two nodes

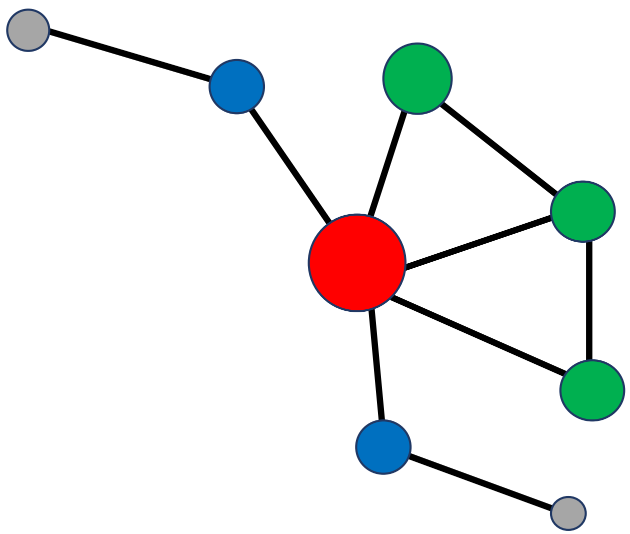


A path is of length 5 between A and B

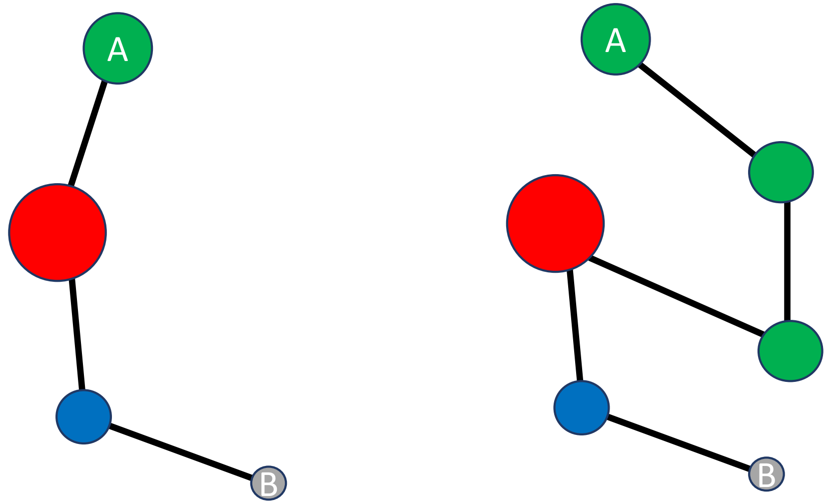


Geodesic

- A **geodesic** is the shortest path between two nodes

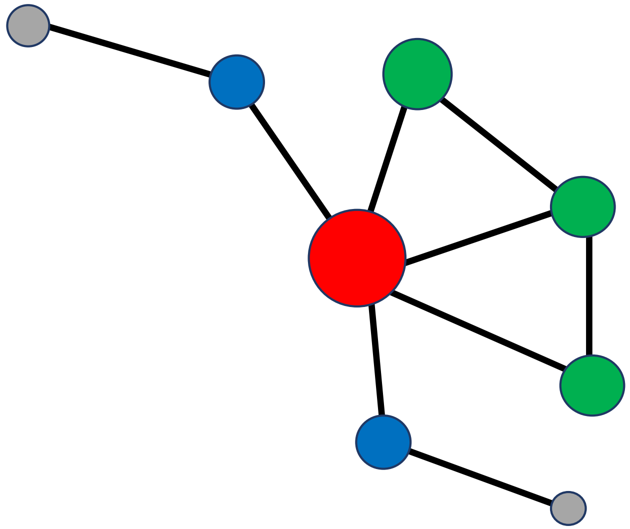


Which one is the geodesic?



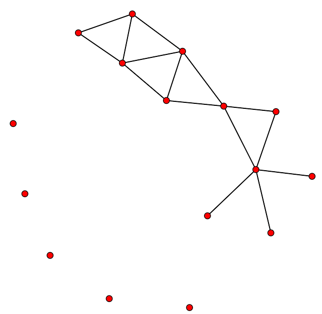
Geodesic distance

- The **geodesic distance** is the length of the geodesic between two nodes



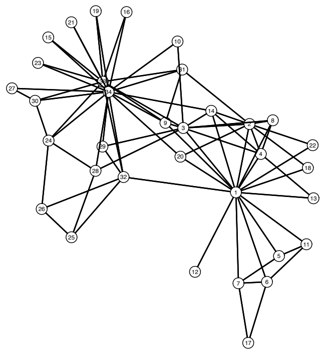
- The **diameter** of a graph is the length of the longest geodesic

Geodesic distance

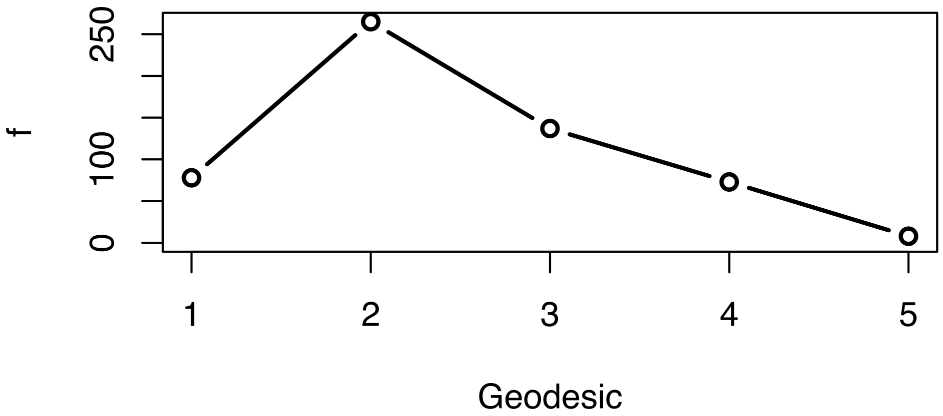


```
> gd<-geodist( padgettbus )
> gd$gdist
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16]
 [1,]    0  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf   Inf   Inf   Inf   Inf   Inf   Inf   Inf
 [2,]  Inf    0  Inf  Inf  Inf  Inf  Inf  Inf  Inf   Inf   Inf   Inf   Inf   Inf   Inf   Inf
 [3,]  Inf  Inf    0    2    1    1    3    2    1     2     1   Inf   Inf     2   Inf     2
 [4,]  Inf  Inf    2    0    2    3    1    1    3     4     1   Inf   Inf     4   Inf     4
 [5,]  Inf  Inf    1    2    0    2    2    1    2     3     1   Inf   Inf     3   Inf     3
 [6,]  Inf  Inf    1    3    2    0    4    3    1     2     2   Inf   Inf     2   Inf     2
 [7,]  Inf  Inf    3    1    2    4    0    1    4     5     2   Inf   Inf     5   Inf     5
 [8,]  Inf  Inf    2    1    1    3    1    0    3     4     1   Inf   Inf     4   Inf     4
 [9,]  Inf  Inf    1    3    2    1    4    3    0     1     2   Inf   Inf     1   Inf     1
[10,]  Inf  Inf    2    4    3    2    5    4    1     0     3   Inf   Inf     2   Inf     2
[11,]  Inf  Inf    1    1    1    2    2    1    2     3     0   Inf   Inf     3   Inf     3
[12,]  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf   Inf   Inf     0   Inf   Inf   Inf   Inf
[13,]  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf   Inf   Inf   Inf     0   Inf   Inf   Inf
[14,]  Inf  Inf    2    4    3    2    5    4    1     2     3   Inf   Inf     0   Inf     2
[15,]  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf  Inf   Inf   Inf   Inf   Inf   Inf     0   Inf
[16,]  Inf  Inf    2    4    3    2    5    4    1     2     3   Inf   Inf     2   Inf     0
```

Geodesic distance



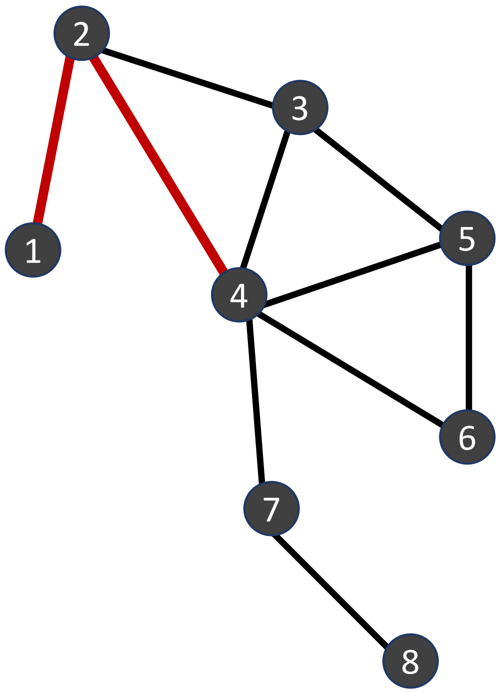
```
gd<-geodist(ZacharyBinary )
plot( table( gd$gdist[upper.tri( gd$gdist) ] ),
      type='b',
      xlab= 'Geodesic',
      ylab='f')
```



Linear algebra

How do we calculate the path lengths using the adjacency matrix?

Paths: calculation



$X_{1,2}X_{4,2} = 1$

$\sum_k X_{i,k} X_{j,k} = |\{k \in V: \{i,k\},\{j,k\} \in E\}|$

$XX^T = (\sum_k X_{i,k} X_{j,k})_{i,j}$

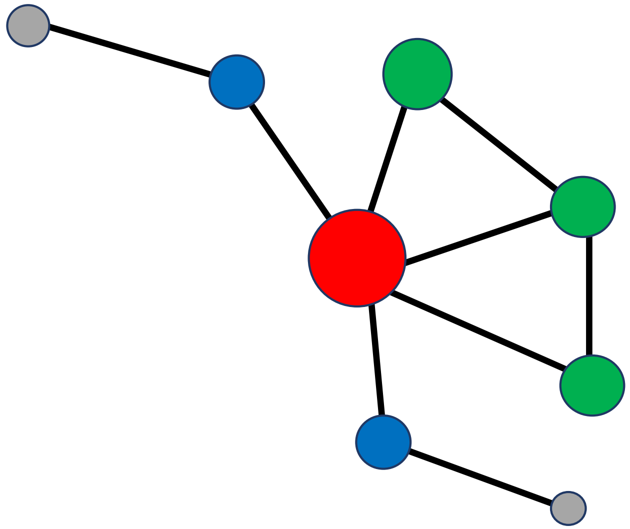
	1	2	3	4	5	6	7	8
1	-	1	0	0	0	0	0	0
2	1	-	1	1	0	0	0	0
3	0	1	-	1	1	0	0	0
4	0	1	1	-	1	1	1	0
5	0	0	1	1	-	1	0	0
6	0	0	0	1	1	-	0	0
7	0	0	0	1	0	0	-	1
8	0	0	0	0	0	0	1	-

	1	2	3	4	5	6	7	8
1	-	1	0	0	0	0	0	0
2	1	-	1	1	0	0	0	0
3	0	1	-	1	1	0	0	0
4	0	1	1	-	1	1	1	0
5	0	0	1	1	-	1	0	0
6	0	0	0	1	1	-	0	0
7	0	0	0	1	0	0	-	1
8	0	0	0	0	0	0	1	-

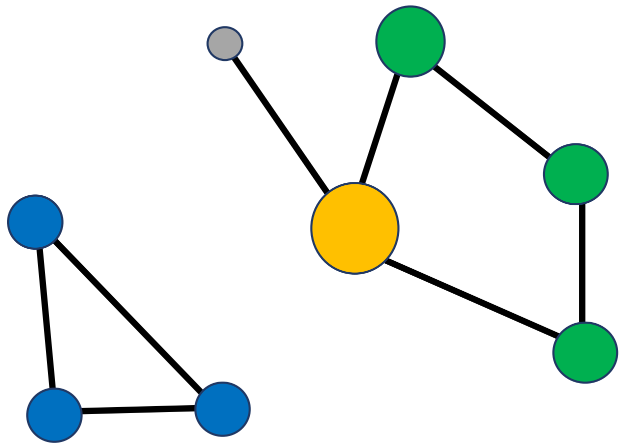
Connectedness

A graph is **connected** if there is a path between any two nodes

A graph is connected if every node is reachable (= there is a path) from any other node



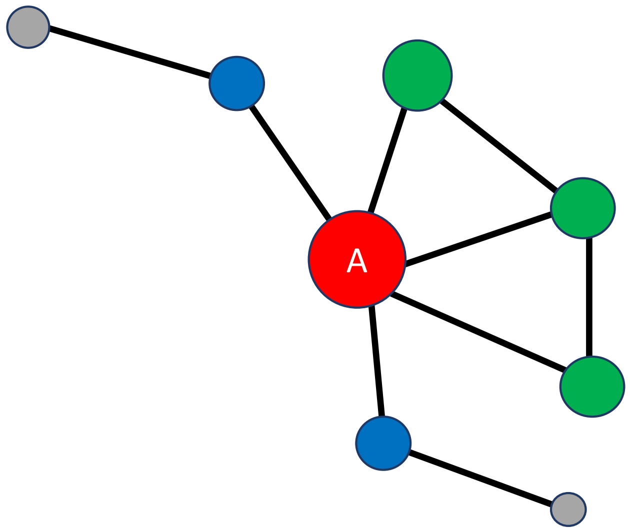
connected



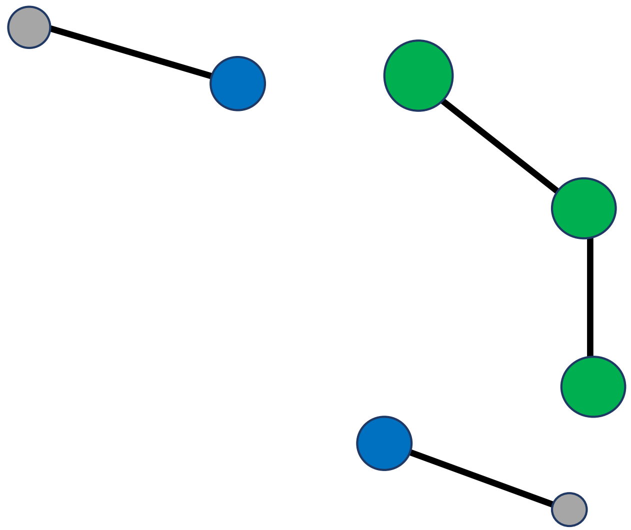
Not connected

Cutpoint

- A **cutpoint** is a node 'connects' the network



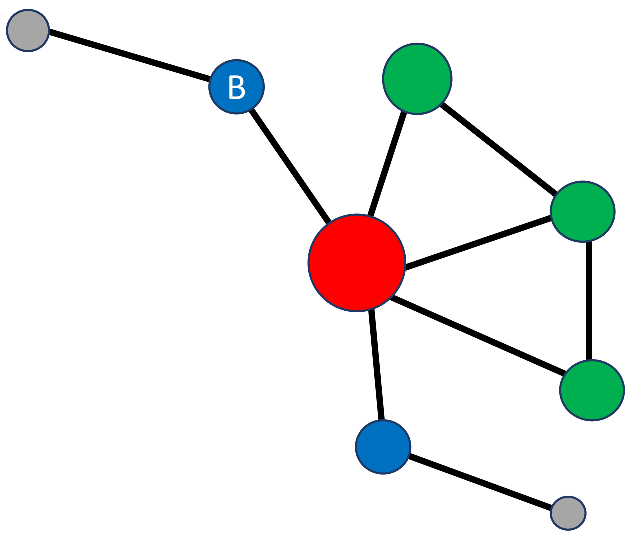
Remove A

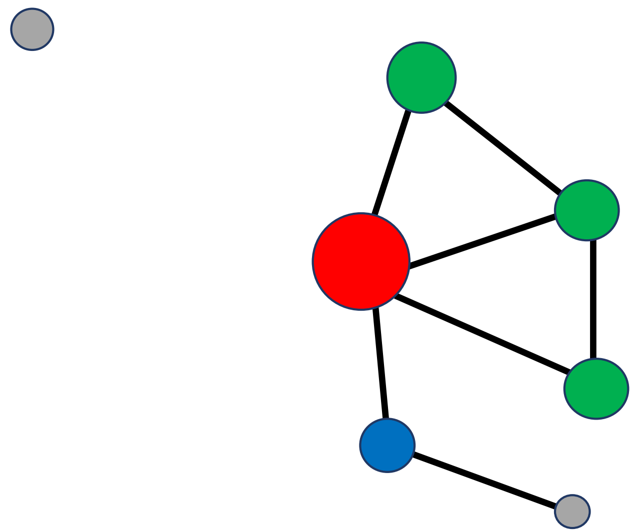


Cutpoint

- A **cutpoint** is a node 'connects' the network

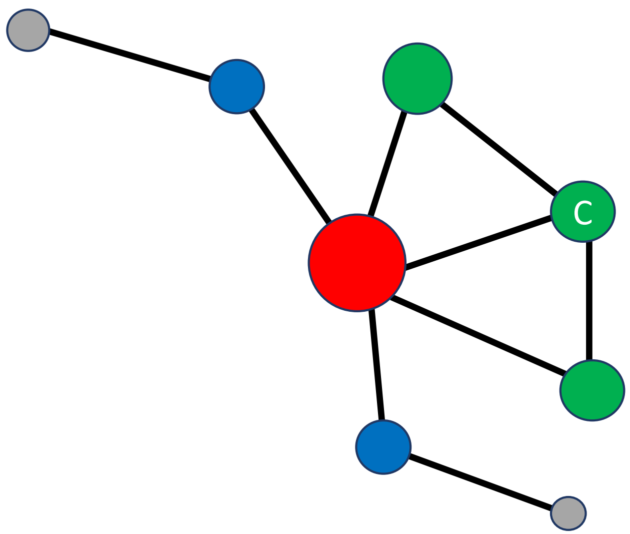
Remove B



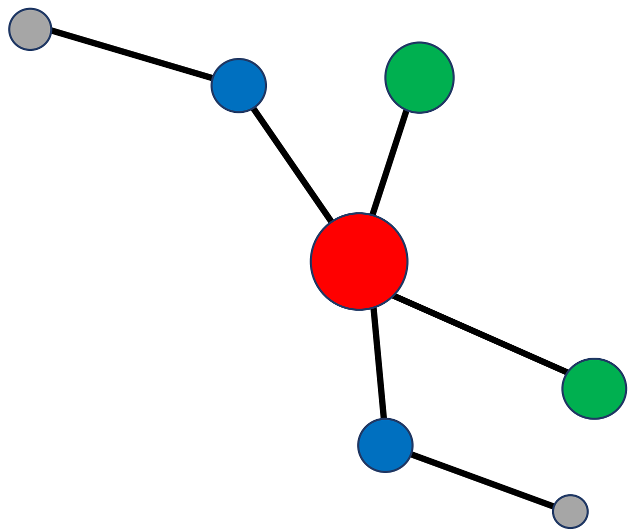


Cutpoint

- A **cutpoint** is a node ‘connects’ the network

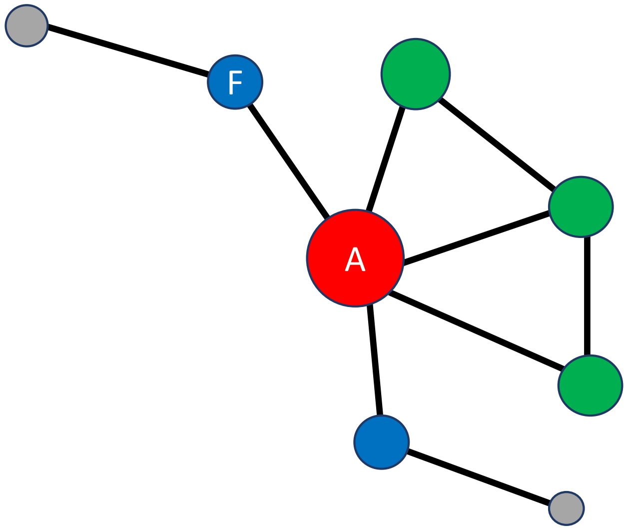


Remove C

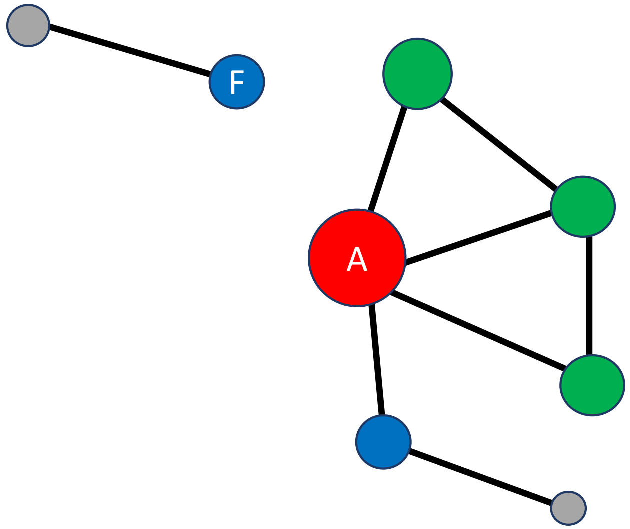


Bridge

- A **bridge** is an edge that ‘connects’ the network

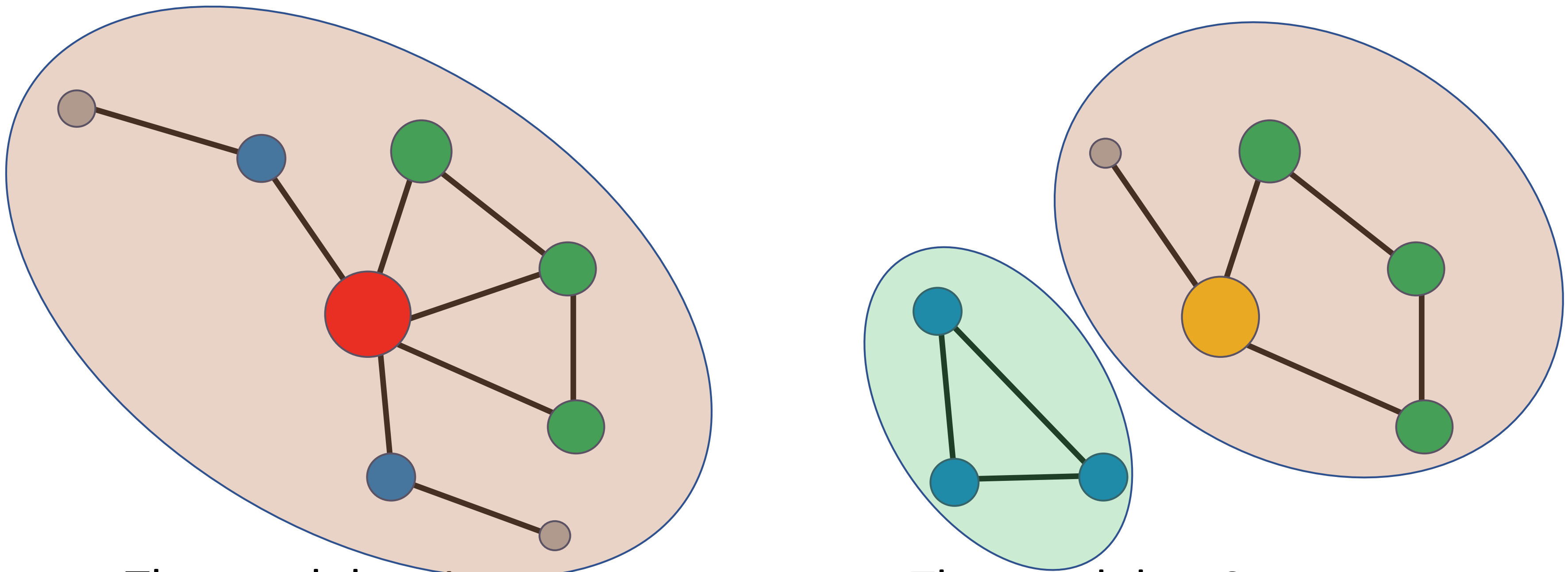


Remove {A,F}



Components

A **component** is a subgraph that is maximally connected



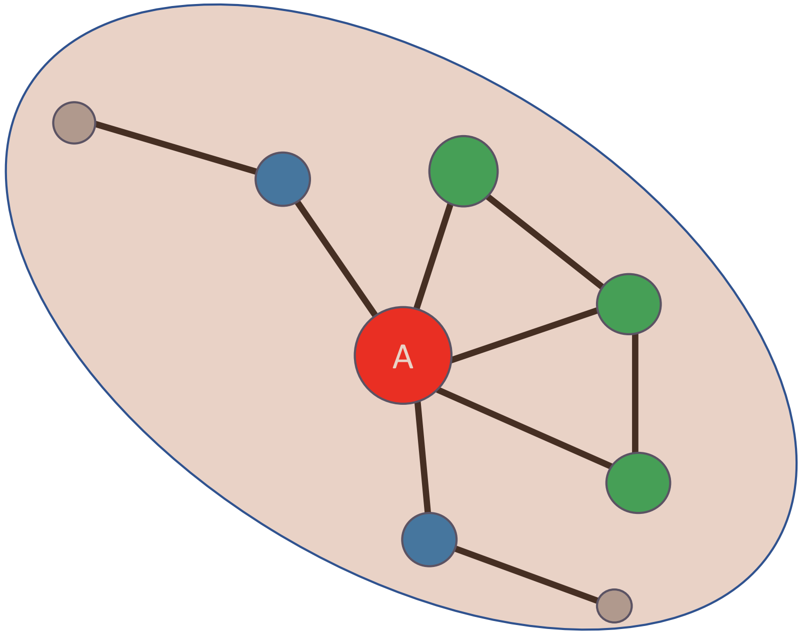
The graph has 1 component

The graph has 2 components

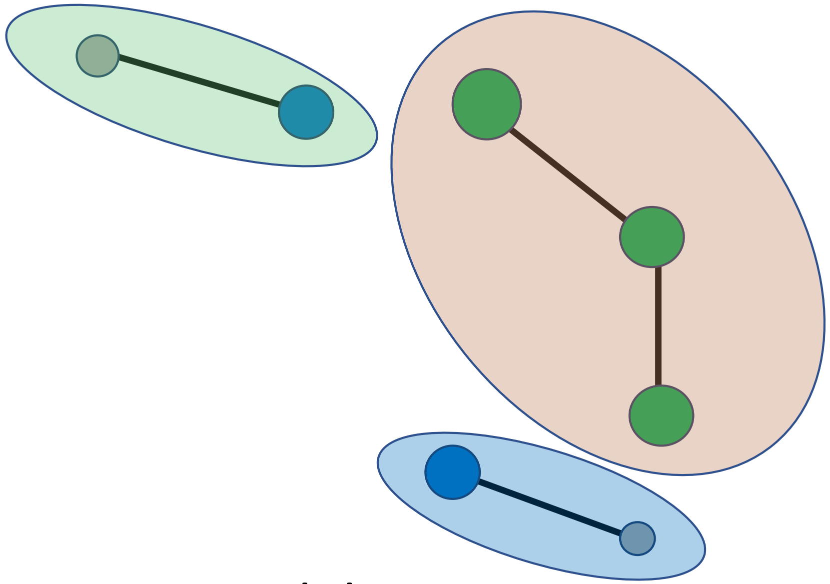
Cutpoint – path, connected, components

- A cutpoint is a node…

… so that if you remove it increases the number of components



graph has 1 component



graph has 3 components

Paths: usage

- How connected is a network?
  - Is it connected?
  - What is the average path-length?
  - What is the longest distance?
- How connected is a person – e.g. betweenness centrality

Reach

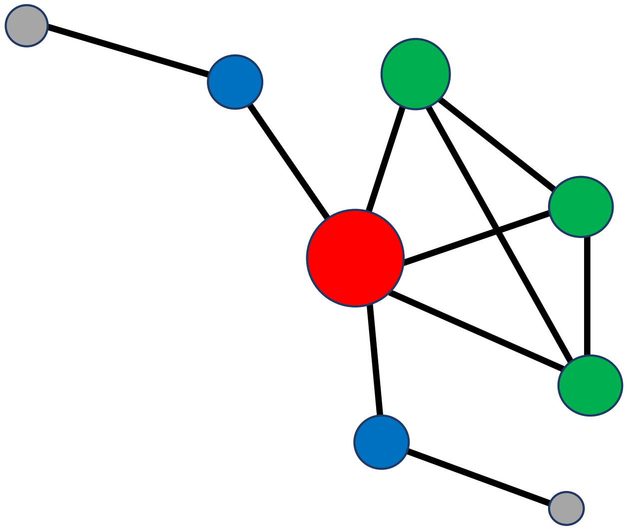
How many handshakes away from Joe Biden are you?

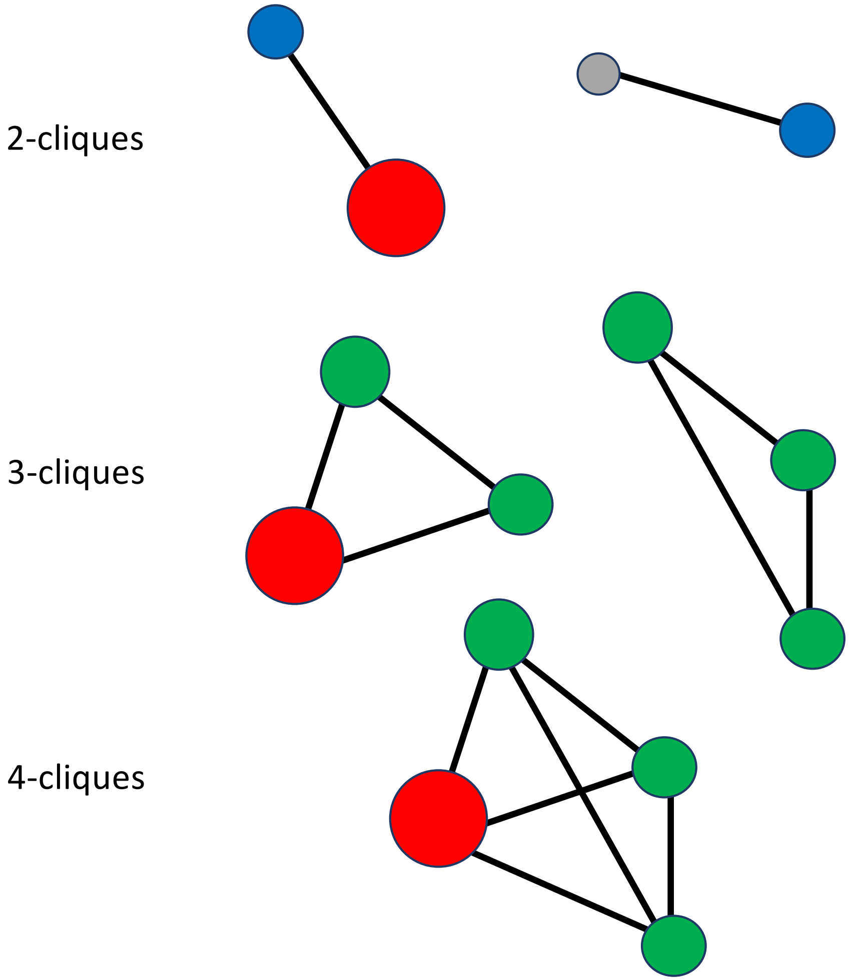
Clustering

clique, triads, closure

Cliques

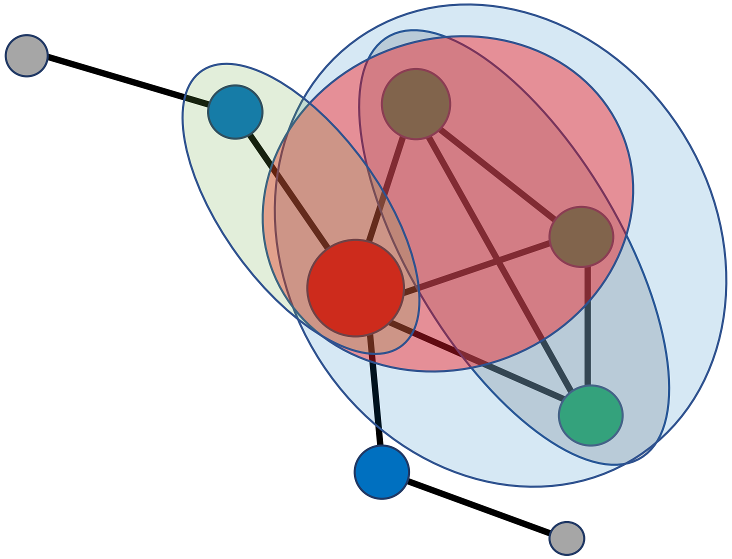
- A k-**clique** is a subset of k nodes that are all connected





Cliques

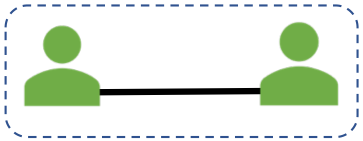
- A k-**clique** is a subset of k nodes that are all connected



- Cliques may be overlapping
- A node can belong to several cliques
- Every subgraph of a clique is a clique
  - everyone in a clique is ‘equal’

Triads

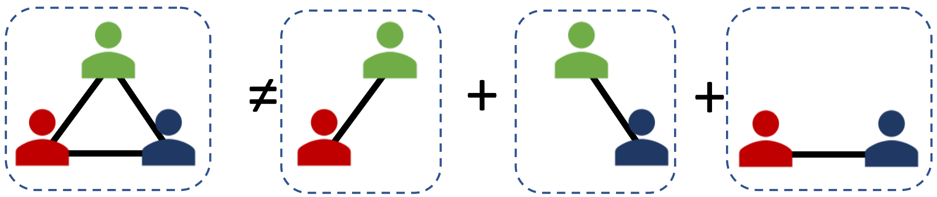
The Web of Group Affiliations (Simmel, 1922):



Dyad

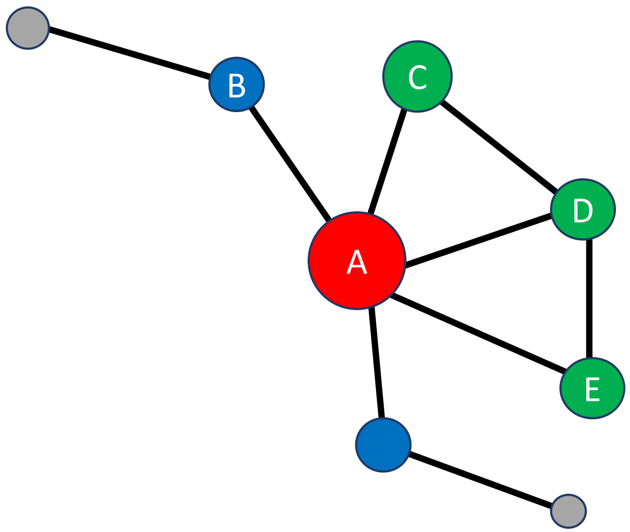
$$X_{i,k}$$

Triad

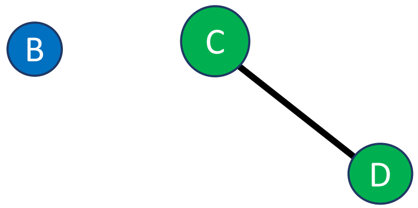


$$X_{i,k}\ ,X_{j,k}\ ,X_{i,j}$$

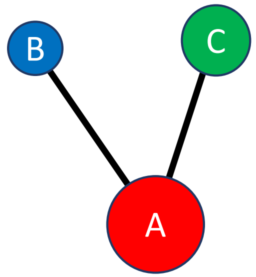
Triads



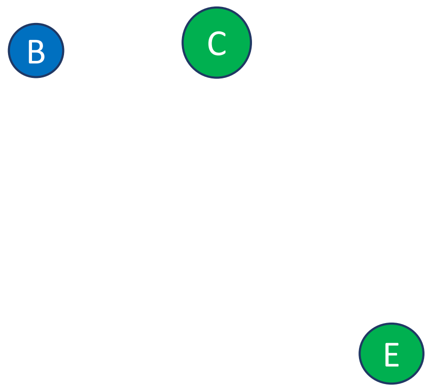
$$X_{BC} + X_{BD} + X_{CD} = 1$$



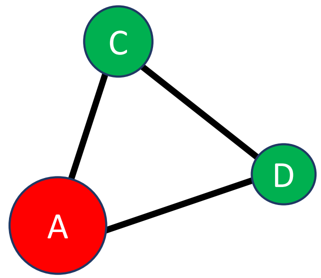
$$X_{BA} + X_{BC} + X_{AC} = 2$$



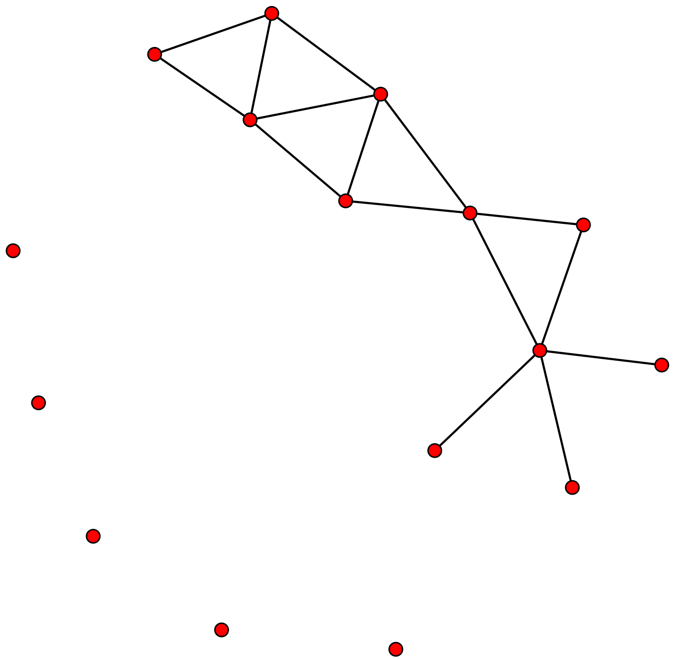
$$X_{BC} + X_{BE} + X_{CE} = 0$$



$$X_{AC} + X_{AB} + X_{CD} = 3$$



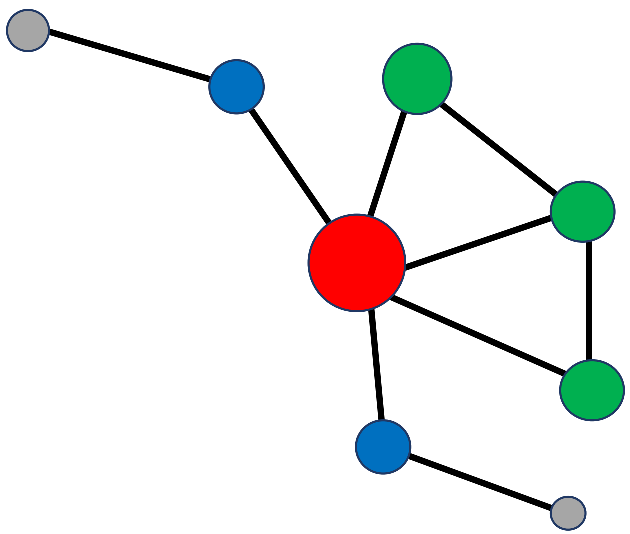
Triad census



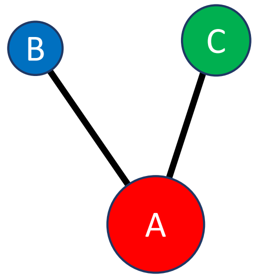
```
> triad.census(padgettbus, mode ='graph')
        0   1  2 3
[1,] 381 153 21 5
```

Triad closure

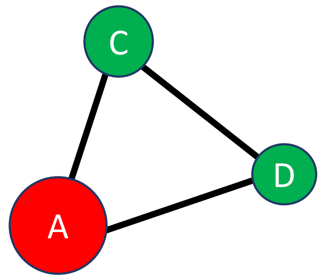
- A **closed triad** is a triangle



Open triad



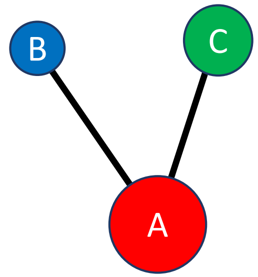
Closed triad



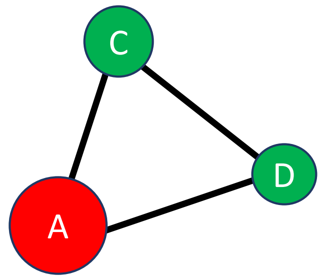
Clustering coefficient

$$\frac{3\sum_{i<j<k} X_{ij}X_{ik}X_{jk}}{3\sum_{i<j<k} X_{ij}X_{ik}X_{jk}+\sum_i \sum_{j<k}X_{ij}X_{ik}\,(1\text{-}X_{jk})}$$

Open triad



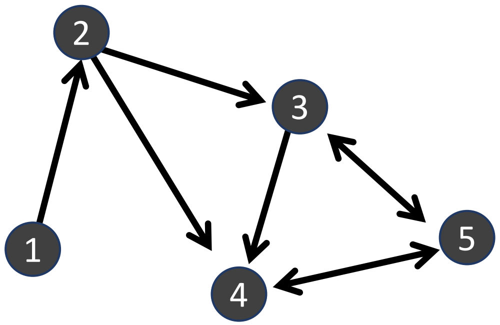
Closed triad



If there are $n(n-1)/2$ dyads, how many triads are there?

Extensions to directed networks

Directed networks: density



	1	2	3	4	5
1	-	1	0	0	0
2	0	-	1	1	0
3	0	0	-	1	1
4	0	0	0	-	1
5	0	0	1	1	-

	1	2	3	4	5
1	-	X_{12}	X_{13}	X_{14}	X_{15}
2	X_{21}	-	X_{23}	X_{24}	X_{25}
3	X_{31}	X_{32}	-	X_{34}	X_{35}
4	X_{41}	X_{42}	X_{43}	-	X_{45}
5	X_{51}	X_{52}	X_{53}	X_{54}	-

n

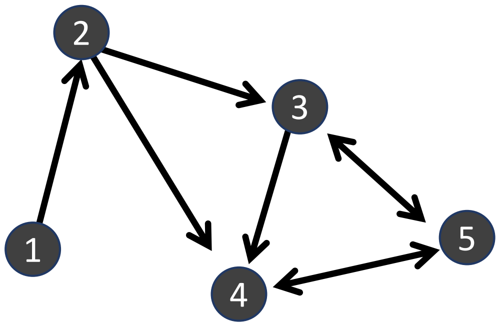
n

$$\frac{\sum_{i,j} X_{ij}}{n(n\text{-}1)} = \frac{8}{5\times 4} = \frac{2}{5}$$

n^2 cells – n diagonal cells: $n(n$-1)

X_{ij} *not necessarily equal to* X_{ji}

Directed networks: dyad census



	1	2	3	4	5
1	-	1	0	0	0
2	0	-	1	1	0
3	0	0	-	1	1
4	0	0	0	-	1
5	0	0	1	1	-

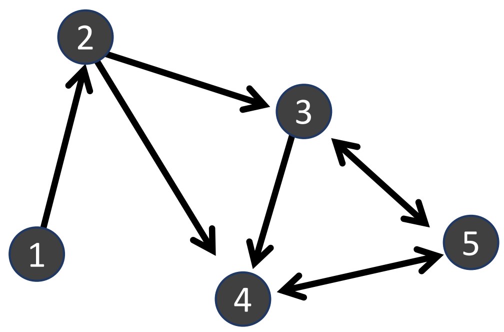
$$n\left\{\underbrace{\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & - & X_{12} & X_{13} & X_{14} & X_{15} \\ 2 & X_{21} & - & X_{23} & X_{24} & X_{25} \\ 3 & X_{31} & X_{32} & - & X_{34} & X_{35} \\ 4 & X_{41} & X_{42} & X_{43} & - & X_{45} \\ 5 & X_{51} & X_{52} & X_{53} & X_{54} & - \end{array}}_{n}\right.$$

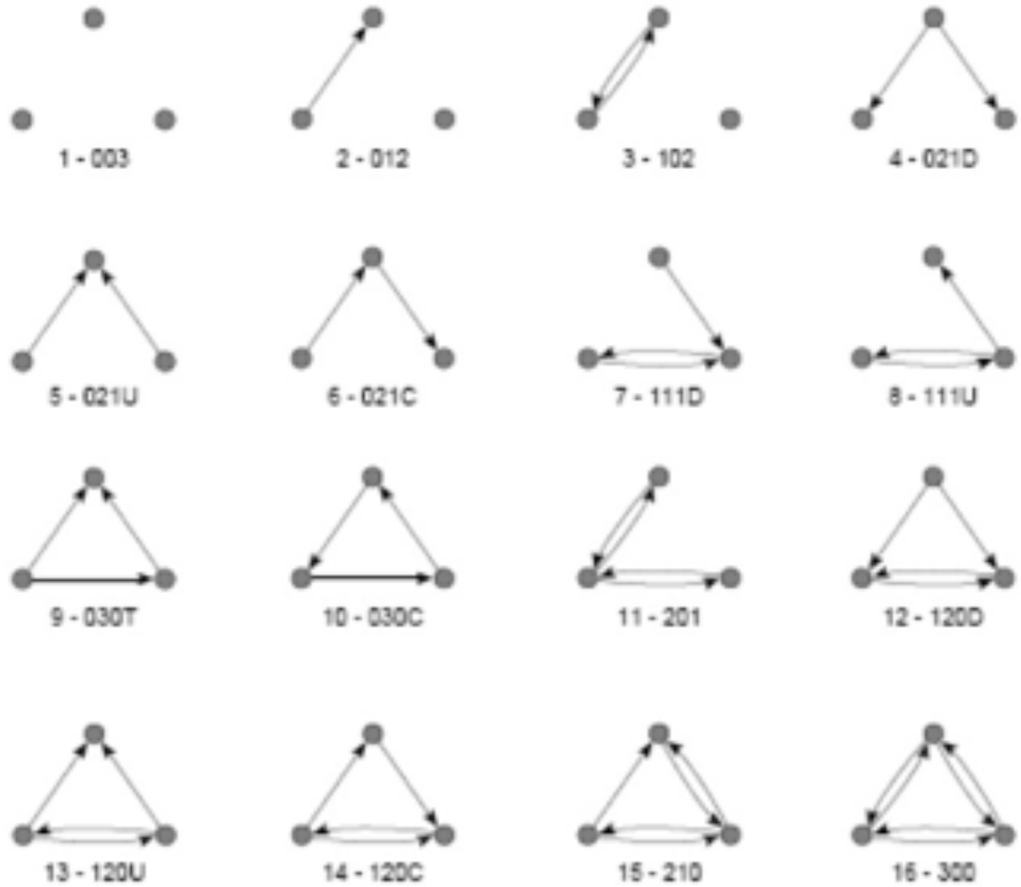
Number of mutual dyads: $\sum_{i>j} X_{ij}X_{ji}$

assymetric dyads: $\sum_{i>j} (X_{ij}\,(1-X_{ij}) + (1- X_{ij})X_{ji})$

null dyads: $\sum_{i>j} (1- X_{ij})(1-X_{ji})$

Directed networks: triad census





Global and local properties

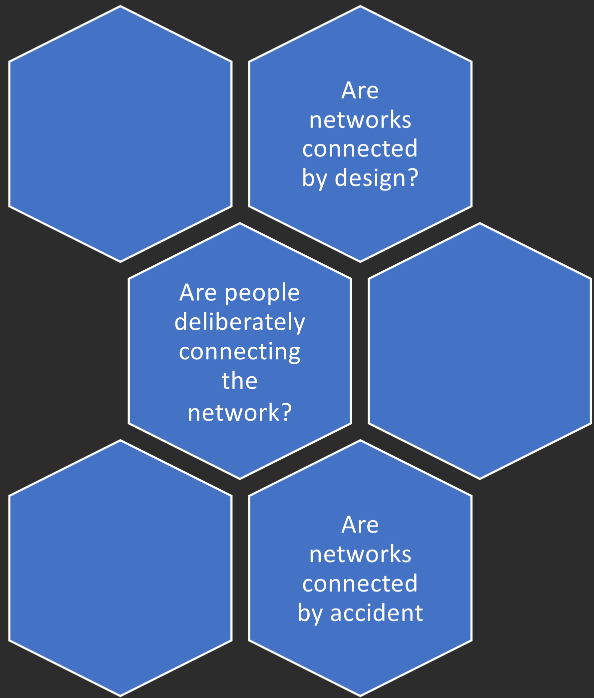
Network questions

Who is most powerful in an organisation?

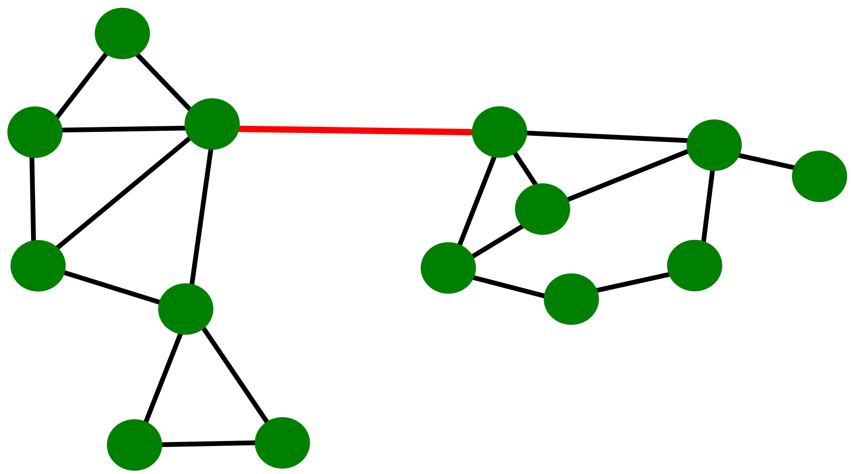
Is this organisation effective?

How many handshakes away from Joe Biden are you?

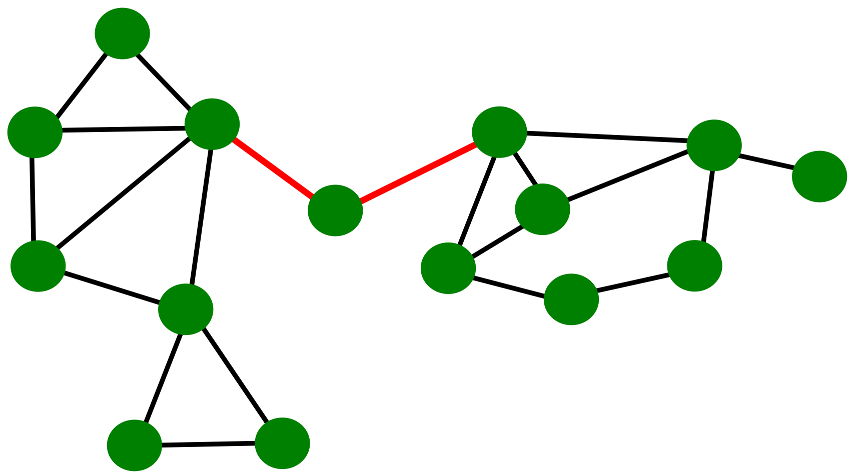
Are networks connected



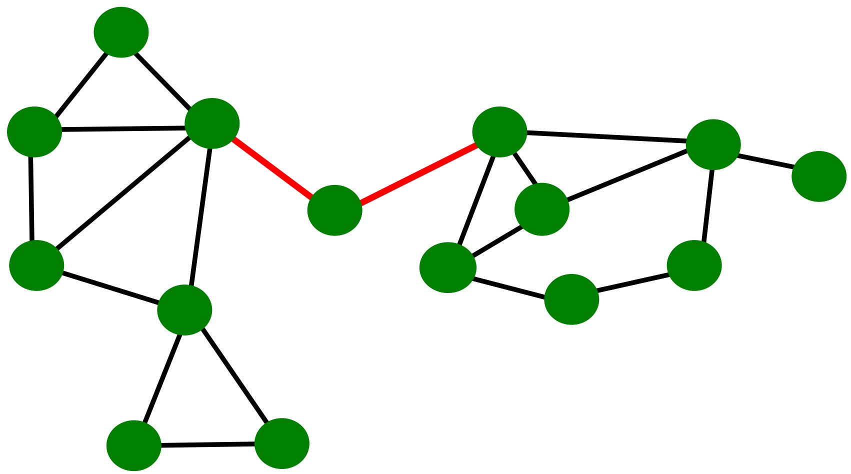
Bridge



Cut-point



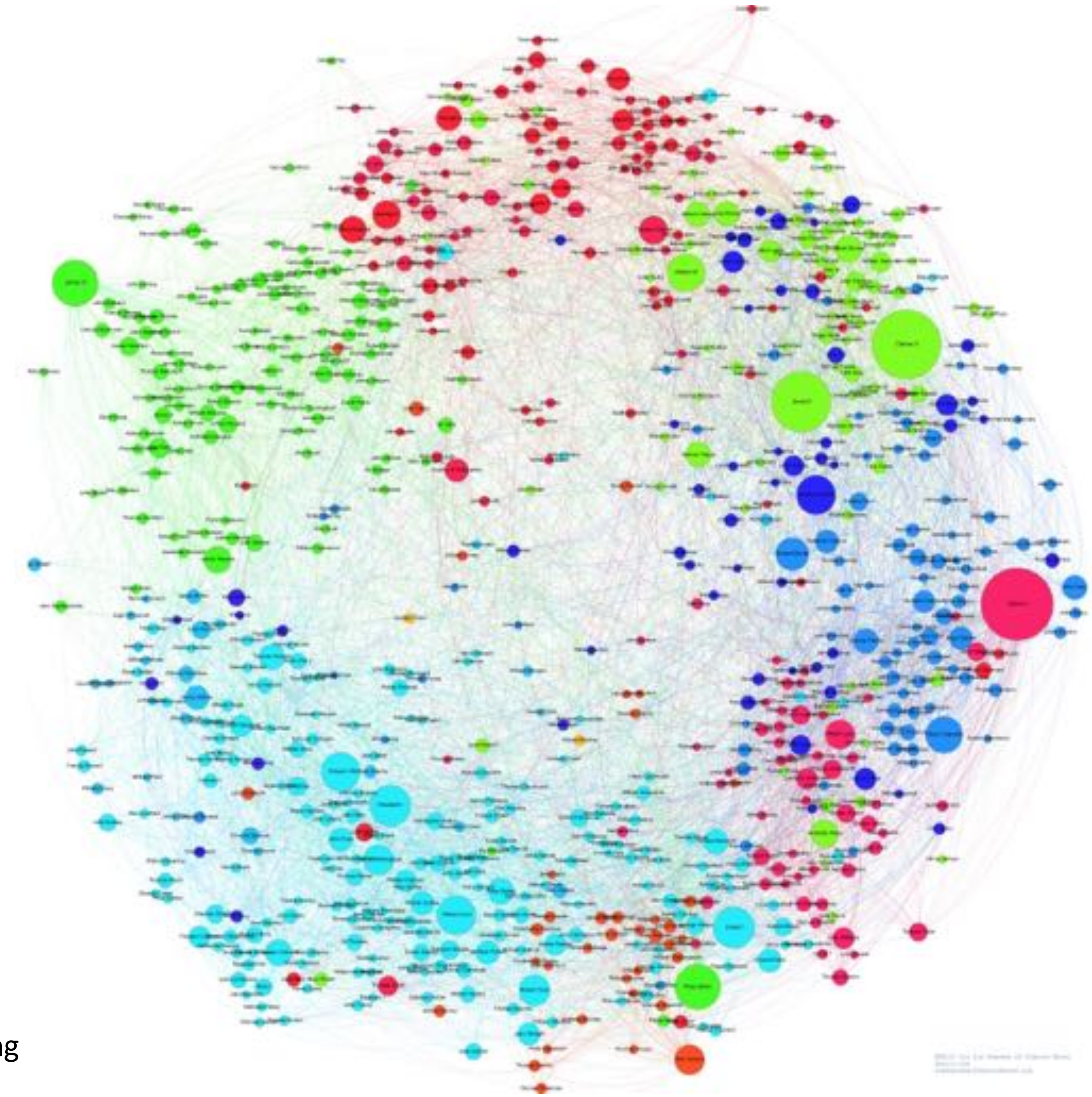
- Graph connected
- Smaller average path-lengths (diameter)



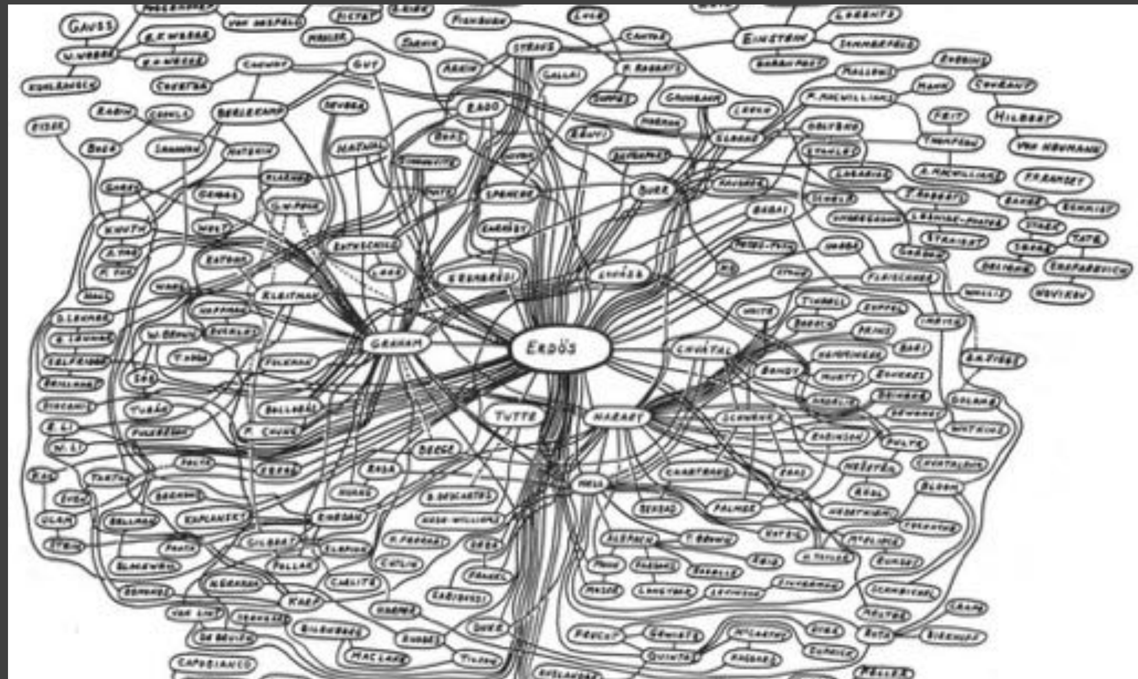
- Does connectivity of the graph (global) depend on brokers?
  - Who is a broker?
    - ppl seek to be brokers?
    - Ppl become brokers through local processes
- Is connectivity (global) not because of brokers but
  - an accident?
  - a result of how ppl form ties?

Kevin Bacon

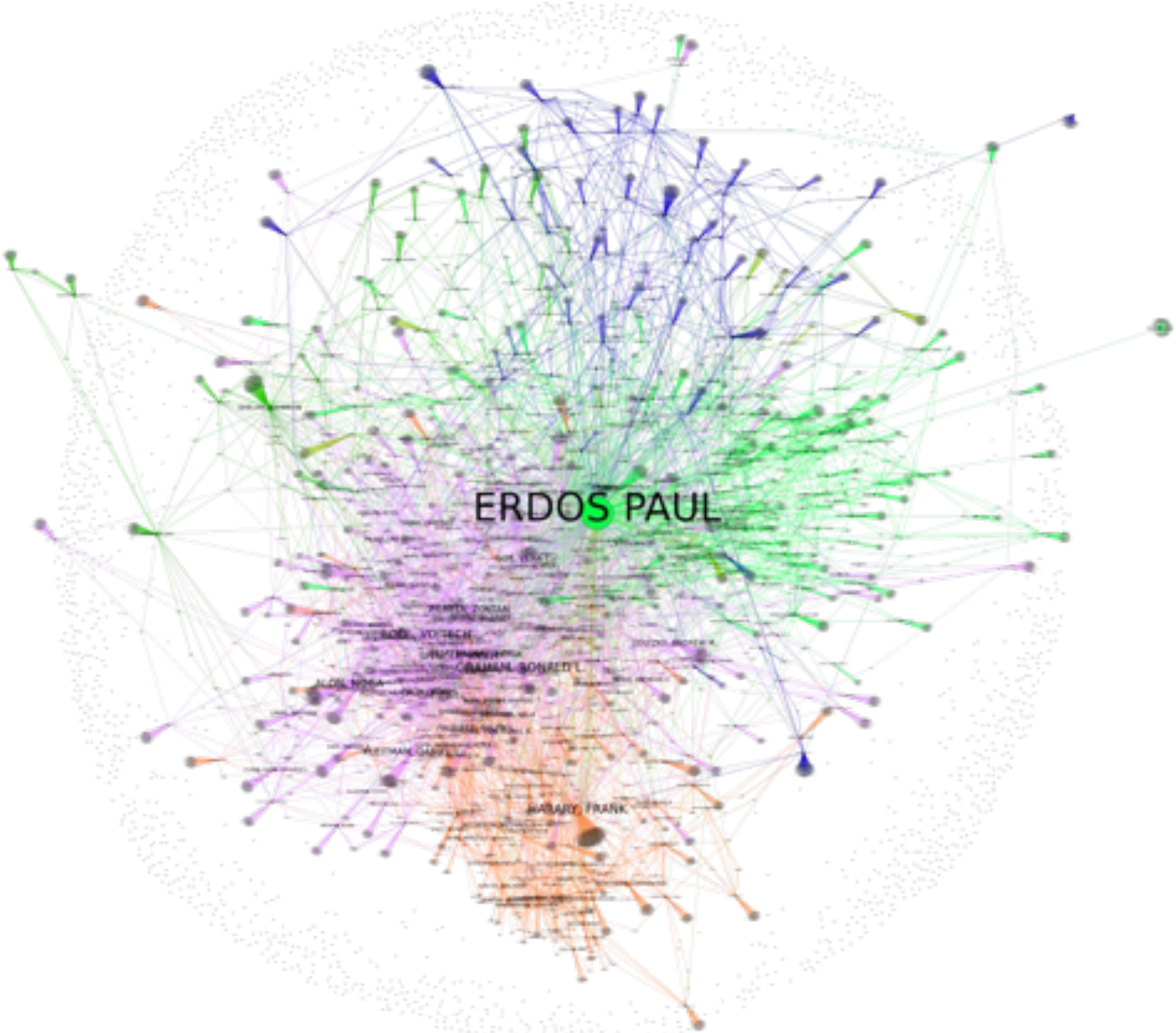
"Kevin Bacon is the Center of the Universe" – 70+ films



http://www.andrew.cmu.edu/user/lawrencw/SDFB3Filter.png



Paul Erdös - ”My mind is open” – 500+ co-authors



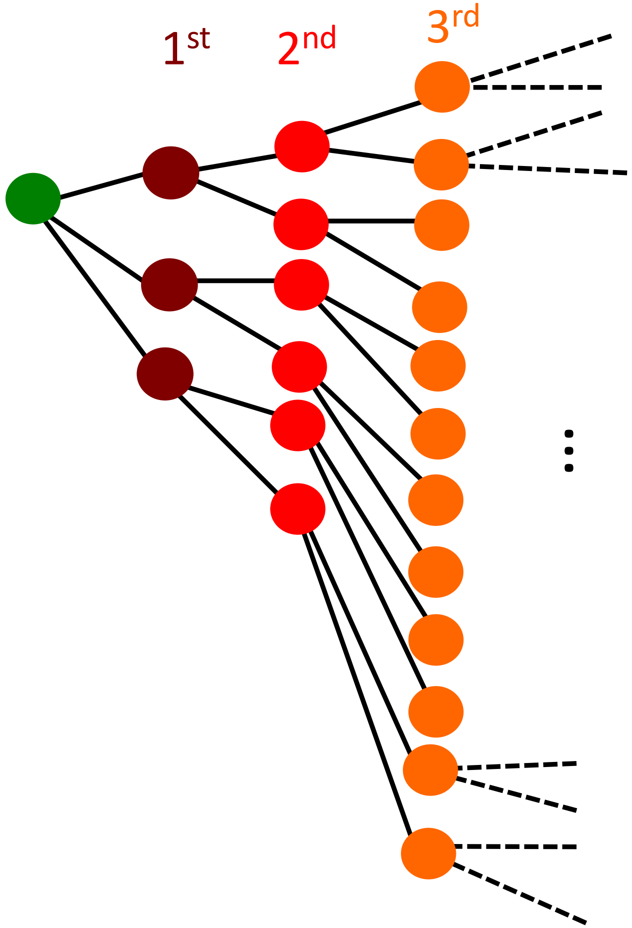
ERDOS PAUL
HARARY, FRANK

Stanley Milgram

Milgram, Stanley (1967). "The Small World Problem". Psychology Today 2: 60–67

So how many steps from Biden are you?

Assume ‘on average 3 friends’



6th

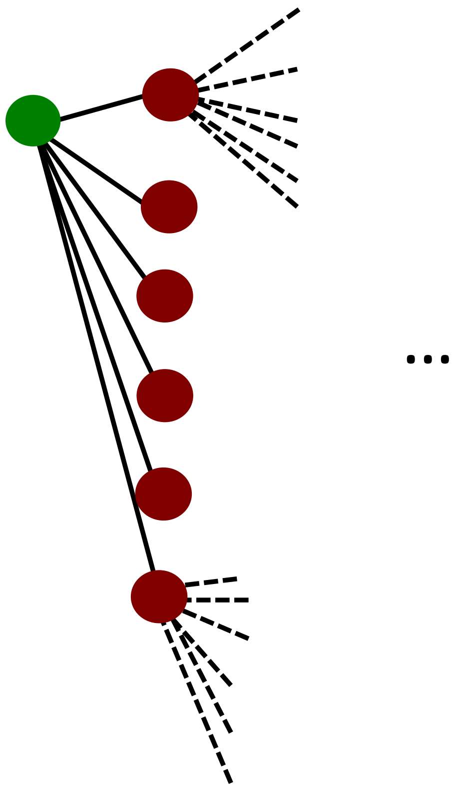
...

How many people at 6th degrees?

$$3 \times 2 \times 2 \times \ldots = ?$$

Assume ‘on average7 friends’

1st 2nd 3rd 6th



How many people at 6th degrees?

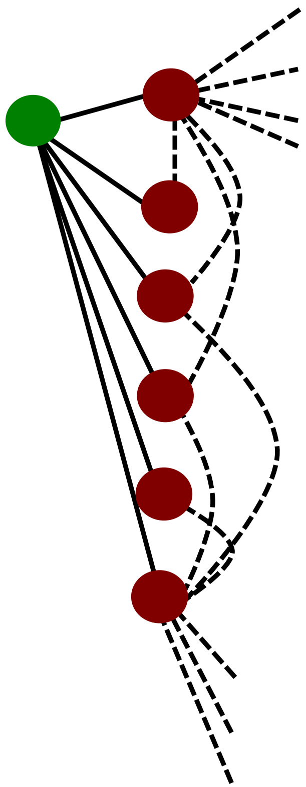
7.9 Billion (2021)

The current world population is **7.9 billion** as of September 2021 [1] according to the most recent United Nations estimates elaborated by Worldometer.

The term "World Population" refers to the human population (the total number of humans currently living) of the world.

$$6^{12} = 2.18\text{bn}$$
$$6^{13} = 13\text{bn}$$

$$7 \times 6 \times 6 \times \ldots \quad = ?$$



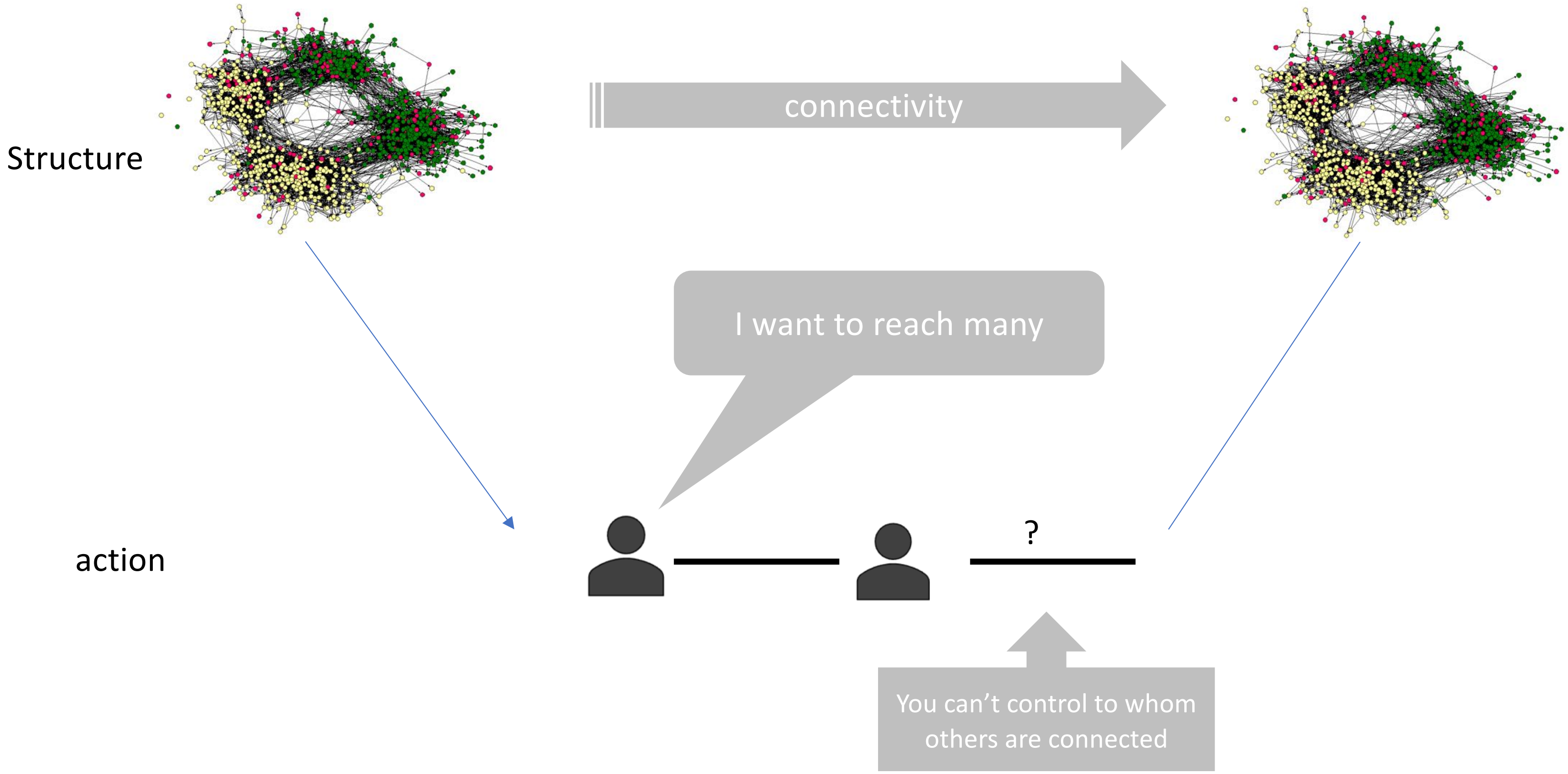
BUT: closure seems empirical regularity in real networks

...

How many people at
n^{th} degrees?

$7 \times (6\text{-}3) \times (6\text{-}3) \times \ldots \qquad = ?$

Micro-macro


Structure
connectivity
I want to reach many
action
?
You can't control to whom others are connected

More graph theory?

- Walks, trails
- K-clans, k-plexes
- Automorphisms, iso-morphism
- Structural equivalence, Automorphic equivalence, regular equivalence
- Centrality measures…

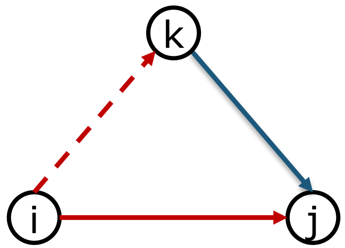
Network non-parametric methods

Ross: *What, you are not mad?*

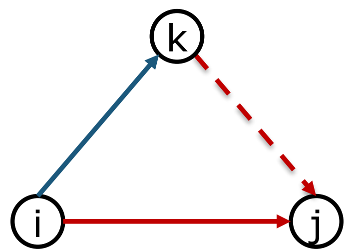
Chander: *Why would I be mad?*

Ross: *Well because, you know, there are certain rules about this kind of stuff. You don't, uh, you don't fool around with your friends' ex-girlfriends, or, uh possible girlfriends, or girls they're related to.*

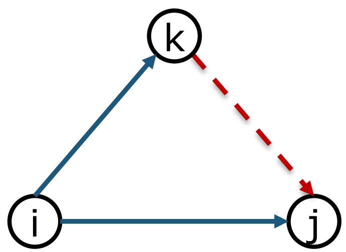
TABU: Don’t date your ex-partner’s friend.



TABU: Don’t date your friend’s ex-partner

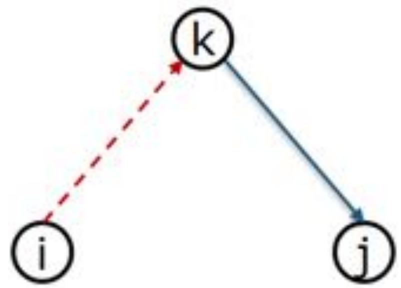


TABU: Don't befriend his/her friend's ex-partner



	Mode1 (H1)		Model 2 (H2)		Model 3 (H3)		Model 4 (H4)		Model 5 (H5)		Model 6 (H6)	
	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.
ROMANTIC NETWORK												
Basic Rate Parameter	0.334***	0.04	0.334***	0.04	0.226***	0.03	0.233***	0.03	0.334***	0.00	0.227***	0.03
Individual Attributes												
Attractiveness of Alter	0.09	0.15	0.092	0.15	0.187	0.17	0.175	0.16	0.09	0.52	0.177	0.17
BMI of Alter	-0.084x	0.05	-0.084x	0.04	-0.13*	0.06	-0.125*	0.05	-0.083x	0.06	-0.126*	0.05
Physical Maturity of Alter	-0.231	0.20	-0.241	0.18	-0.268	0.20	-0.288	0.18	-0.232	0.18	-0.273	0.18
Physical Maturity of Ego	1.114	1.29	1.071	1.53	-0.092	0.74	-0.079	0.63	1.162	0.44	-0.092	0.81
Dyadic Attributes												
Jock to Cheerleader Ties	0.918	0.59	0.898	0.66	1.557*	0.67	1.485*	0.59	0.898	0.15	1.527*	0.63
Age Dominance	0.564	0.53	0.55	0.54	0.031	0.45	0.067	0.45	0.566	0.25	0.041	0.50
Co-Affiliation	0.664*	0.29	0.664*	0.29	0.746*	0.37	0.74*	0.33	0.663x	0.10	0.743x	0.38
Structural Effects												
Reciprocity	11.941x	6.46	11.598*	5.44	10.223	6.50	19.36*	7.95	11.983x	0.05	10.269	8.31
Indegree - Popularity (sqrt)	0.6	0.47	0.592	0.43	0.407	0.69	0.371	0.53	0.582	0.23	0.476	0.66
Outdegree - Popularity (sqrt)	1.768***	0.47	1.733**	0.54	2.592***	0.63	2.414***	0.59	1.758***	0.00	2.517***	0.57
Rec.degree - Popularity (sqrt)	-0.455	0.42	-0.436	0.53	-1.072x	0.56	-0.834x	0.46	-0.434	0.44	-1.005	0.65
Rec.degree^(1/2) - Activity	-3.139	3.90	-2.98	3.53	-0.702	3.05	-6.224x	3.25	-3.177	0.41	-0.82	3.79
Outdegree-trunc(1)	-8.081***	1.10	-7.984***	0.90	-8.748***	0.96	-8.708***	0.73	-8.045***	0.00	-8.77***	0.99
Taboo Ties Hypothesis												
Reciprocity with Friendship	-0.598	1.70	-0.832	1.42	1.334	1.50	1.388	1.41	-0.646	0.62	1.359	1.70
Friendship Entrainment	5.362***	0.75	5.098***	0.80	4.866***	1.22	5.044***	0.88	5.407***	0.00	5.068***	1.01
XWX Closure of Friendship			1.023*	0.46								
Friendship to Agreement					1.707	1.04						
FRIENDSHIP NETWORK												
Basic Rate Parameter	19.171***	1.11	19.154***	1.17	17.996***	0.83	17.897***	0.70	19.012***	0.00	17.972***	0.78
Individual Attributes												
Jock Alter	-0.054	0.05	-0.054	0.05	0	0.04	-0.005	0.04	-0.06	0.24	-0.003	0.05
Jock Ego	-0.069	0.05	-0.07	0.05	0.029	0.05	0.026	0.04	-0.073	0.13	0.028	0.05
Jock Homophily	0.077	0.05	0.078	0.06	0.077	0.05	0.077x	0.05	0.075	0.16	0.074	0.06
Cheerleader Alter	0.157	0.18	0.163	0.17	0.156	0.17	0.165	0.14	0.165	0.35	0.156	0.16
Cheerleader Ego	0.258	0.17	0.263	0.18	0.188	0.17	0.19	0.15	0.27	0.14	0.185	0.17
Cheerleader Homophily	0.245	0.19	0.252	0.20	0.276	0.20	0.293x	0.16	0.264	0.21	0.283	0.19
Gender Alter	-0.05	0.04	-0.05	0.04	-0.066	0.04	-0.065*	0.03	-0.049	0.21	-0.068x	0.04
Gender Ego	-0.051	0.05	-0.052	0.04	-0.006	0.04	-0.006	0.04	-0.049	0.21	-0.006	0.04
Gender Homophily	0.148***	0.04	0.147***	0.04	0.165***	0.04	0.155***	0.04	0.139**	0.00	0.171***	0.04
Dyadic Attributes												
Same Grade	0.482***	0.04	0.483***	0.04	0.557***	0.03	0.542***	0.04	0.473***	0.00	0.557***	0.04
Co-Affiliation	0.191***	0.04	0.19***	0.04	0.208***	0.04	0.204***	0.03	0.192***	0.00	0.209***	0.04
Structural Effects												
Outdegree(density)	-3.745***	0.06	-3.747***	0.08	-3.724***	0.07	-3.728***	0.06	-3.739***	0.00	-3.724***	0.07
Reciprocity	2.418***	0.11	2.418***	0.10	2.566***	0.09	2.57***	0.07	2.424***	0.00	2.566***	0.08
GWESP I -> K -> J (69)	1.616***	0.07	1.616***	0.10	1.888***	0.07	1.934***	0.06	1.655***	0.00	1.896***	0.06
GWESP I <- K <- J (69)	-0.128	0.10	-0.128	0.12	-0.41***	0.08	-0.397***	0.07	-0.118	0.24	-0.408***	0.09
Indegree - Popularity (sqrt)	0.11	0.07	0.109	0.11	0.307***	0.03	0.31***	0.03	0.114	0.21	0.307***	0.04
Outdegree - Popularity (sqrt)	0.008	0.09	0.01	0.13	-0.382***	0.05	-0.381***	0.05	0	1.00	-0.381***	0.05
Taboo Ties Hypothesis												
Romantic Entrainment	0.564***	0.15	0.563**	0.18	0.568***	0.16	0.618***	0.15	0.617***	0.00	0.583***	0.16
Romantic to Agreement							-0.652***	0.18				
XWX Closure of Romantic									-0.264**	0.01		
Closure of Romantic											-1.329*	0.67
Convergence Ratio	0.26			0.19			0.13			0.23		0.22

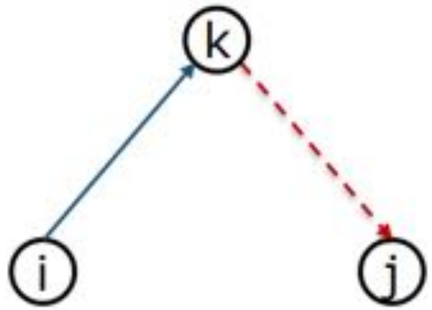
TABU: Don't date your ex-partner's friend.



RESULT: preference for dating ex-partner's friend.

	Mode1 (H1)		Model 2 (H2)		Model 3 (H3)		Model 4 (H4)		Model 5 (H5)		Model 6 (H6)	
	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.	Est	S.E.
ROMANTIC NETWORK												
Basic Rate Parameter	0.334***	0.04	0.334***	0.04	0.226***	0.03	0.233***	0.03	0.334***	0.00	0.227***	0.03
Individual Attributes												
Attractiveness of Alter	0.09	0.15	0.092	0.15	0.187	0.17	0.175	0.16	0.09	0.52	0.177	0.17
BMI of Alter	-0.084x	0.05	-0.084x	0.04	-0.13*	0.06	-0.125*	0.05	-0.083x	0.06	-0.126*	0.05
Physical Maturity of Alter	-0.231	0.20	-0.241	0.18	-0.268	0.20	-0.288	0.18	-0.232	0.18	-0.273	0.18
Physical Maturity of Ego	1.114	1.29	1.071	1.53	-0.092	0.74	-0.079	0.63	1.162	0.44	-0.092	0.81
Dyadic Attributes												
Jock to Cheerleader Ties	0.918	0.59	0.898	0.66	1.557*	0.67	1.485*	0.59	0.898	0.15	1.527*	0.63
Age Dominance	0.564	0.53	0.55	0.54	0.031	0.45	0.067	0.45	0.566	0.25	0.041	0.50
Co-Affiliation	0.664*	0.29	0.664*	0.29	0.746*	0.37	0.74*	0.33	0.663x	0.10	0.743x	0.38
Structural Effects												
Reciprocity	11.941x	6.46	11.598*	5.44	10.223	6.50	19.36*	7.95	11.983x	0.05	10.269	8.31
Indegree - Popularity (sqrt)	0.6	0.47	0.592	0.43	0.407	0.69	0.371	0.53	0.582	0.23	0.476	0.66
Outdegree - Popularity (sqrt)	1.768***	0.47	1.733**	0.54	2.592***	0.63	2.414***	0.59	1.758***	0.00	2.517***	0.57
Rec.degree - Popularity (sqrt)	-0.455	0.42	-0.436	0.53	-1.072x	0.56	-0.834x	0.46	-0.434	0.44	-1.005	0.65
Rec.degree^(1/2) - Activity	-3.139	3.90	-2.98	3.53	-0.702	3.05	-6.224x	3.25	-3.177	0.41	-0.82	3.79
Outdegree-trunc(1)	-8.081***	1.10	-7.984***	0.90	-8.748***	0.96	-8.708***	0.73	-8.045***	0.00	-8.77***	0.99
Taboo Ties Hypothesis												
Reciprocity with Friendship	-0.598	1.70	-0.832	1.42	1.334	1.50	1.388	1.41	-0.646	0.62	1.359	1.70
Friendship Entrainment	5.362***	0.75	5.098***	0.80	4.866***	1.22	5.044***	0.88	5.407***	0.00	5.068***	1.01
XWX Closure of Friendship			1.023*	0.46								
Friendship to Agreement					1.707	1.04						
FRIENDSHIP NETWORK												
Basic Rate Parameter	19.171***	1.11	19.154***	1.17	17.996***	0.83	17.897***	0.70	19.012***	0.00	17.972***	0.78
Individual Attributes												
Jock Alter	-0.054	0.05	-0.054	0.05	0	0.04	-0.005	0.04	-0.06	0.24	-0.003	0.05
Jock Ego	-0.069	0.05	-0.07	0.05	0.029	0.05	0.026	0.04	-0.073	0.13	0.028	0.05
Jock Homophily	0.077	0.05	0.078	0.06	0.077	0.05	0.077x	0.05	0.075	0.16	0.074	0.06
Cheerleader Alter	0.157	0.18	0.163	0.17	0.156	0.17	0.165	0.14	0.165	0.35	0.156	0.16
Cheerleader Ego	0.258	0.17	0.263	0.18	0.188	0.17	0.19	0.15	0.27	0.14	0.185	0.17
Cheerleader Homophily	0.245	0.19	0.252	0.20	0.276	0.20	0.293x	0.16	0.264	0.21	0.283	0.19
Gender Alter	-0.05	0.04	-0.05	0.04	-0.066	0.04	-0.065*	0.03	-0.049	0.21	-0.068x	0.04
Gender Ego	-0.051	0.05	-0.052	0.04	-0.006	0.04	-0.006	0.04	-0.049	0.21	-0.006	0.04
Gender Homophily	0.148***	0.04	0.147***	0.04	0.165***	0.04	0.155***	0.04	0.139**	0.00	0.171***	0.04
Dyadic Attributes												
Same Grade	0.482***	0.04	0.483***	0.04	0.557***	0.03	0.542***	0.04	0.473***	0.00	0.557***	0.04
Co-Affiliation	0.191***	0.04	0.19***	0.04	0.208***	0.04	0.204***	0.03	0.192***	0.00	0.209***	0.04
Structural Effects												
Outdegree(density)	-3.745***	0.06	-3.747***	0.08	-3.724***	0.07	-3.728***	0.06	-3.739***	0.00	-3.724***	0.07
Reciprocity	2.418***	0.11	2.418***	0.10	2.566***	0.09	2.57***	0.07	2.424***	0.00	2.566***	0.08
GWESP I -> K -> J (69)	1.616***	0.07	1.616***	0.10	1.888***	0.07	1.934***	0.06	1.655***	0.00	1.896***	0.06
GWESP I <- K <- J (69)	-0.128	0.10	-0.128	0.12	-0.41***	0.08	-0.397***	0.07	-0.118	0.24	-0.408***	0.09
Indegree - Popularity (sqrt)	0.11	0.07	0.109	0.11	0.307***	0.03	0.31***	0.03	0.114	0.21	0.307***	0.04
Outdegree - Popularity (sqrt)	0.008	0.09	0.01	0.13	-0.382***	0.05	-0.381***	0.05	0	1.00	-0.381***	0.05
Taboo Ties Hypothesis												
Romantic Entrainment	0.564***	0.15	0.563**	0.18	0.568***	0.16	0.618***	0.15	0.617***	0.00	0.583***	0.16
Romantic to Agreement							-0.652***	0.18				
XWX Closure of Romantic									-0.264**	0.01		
Closure of Romantic											-1.329*	0.67
Convergence Ratio	0.26			0.19			0.13			0.23		0.22

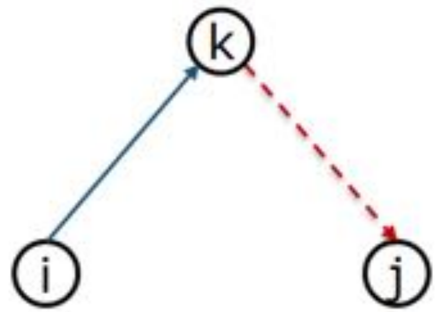
TABU: Don't date your friend's ex-partner



RESULT: preference for dating friend's ex-partner

TABU: Don’t befriend his/her friend’s ex-partner

	Mode1 (H1) Est	S.E.	Model 2 (H2) Est	S.E.	Model 3 (H3) Est	S.E.	Model 4 (H4) Est	S.E.	Model 5 (H5) Est	S.E.	Model 6 (H6) Est	S.E.
ROMANTIC NETWORK												
Basic Rate Parameter	0.334***	0.04	0.334***	0.04	0.226***	0.03	0.233***	0.03	0.334***	0.00	0.227***	0.03
Individual Attributes												
Attractiveness of Alter	0.09	0.15	0.092	0.15	0.187	0.17	0.175	0.16	0.09	0.52	0.177	0.17
BMI of Alter	-0.084x	0.05	-0.084x	0.04	-0.13*	0.06	-0.125*	0.05	-0.083x	0.06	-0.126*	0.05
Physical Maturity of Alter	-0.231	0.20	-0.241	0.18	-0.268	0.20	-0.288	0.18	-0.232	0.18	-0.273	0.18
Physical Maturity of Ego	1.114	1.29	1.071	1.53	-0.092	0.74	-0.079	0.63	1.162	0.44	-0.092	0.81
Dyadic Attributes												
Jock to Cheerleader Ties	0.918	0.59	0.898	0.66	1.557*	0.67	1.485*	0.59	0.898	0.15	1.527*	0.63
Age Dominance	0.564	0.53	0.55	0.54	0.031	0.45	0.067	0.45	0.566	0.25	0.041	0.50
Co-Affiliation	0.664*	0.29	0.664*	0.29	0.746*	0.37	0.74*	0.33	0.663x	0.10	0.743x	0.38
Structural Effects												
Reciprocity	11.941x	6.46	11.598*	5.44	10.223	6.50	19.36*	7.95	11.983x	0.05	10.269	8.31
Indegree - Popularity (sqrt)	0.6	0.47	0.592	0.43	0.407	0.69	0.371	0.53	0.582	0.23	0.476	0.66
Outdegree - Popularity (sqrt)	1.768***	0.47	1.733**	0.54	2.592***	0.63	2.414***	0.59	1.758***	0.00	2.517***	0.57
Rec.degree - Popularity (sqrt)	-0.455	0.42	-0.436	0.53	-1.072x	0.56	-0.834x	0.46	-0.434	0.44	-1.005	0.65
Rec.degree^(1/2) - Activity	-3.139	3.90	-2.98	3.53	-0.702	3.05	-6.224x	3.25	-3.177	0.41	-0.82	3.79
Outdegree-trunc(1)	-8.081***	1.10	-7.984***	0.90	-8.748***	0.96	-8.708***	0.73	-8.045***	0.00	-8.77***	0.99
Taboo Ties Hypothesis												
Reciprocity with Friendship	-0.598	1.70	-0.832	1.42	1.334	1.50	1.388	1.41	-0.646	0.62	1.359	1.70
Friendship Entrainment	5.362***	0.75	5.098***	0.80	4.866***	1.22	5.044***	0.88	5.407***	0.00	5.068***	1.01
XWX Closure of Friendship			1.023*	0.46								
Friendship to Agreement					1.707	1.04						
FRIENDSHIP NETWORK												
Basic Rate Parameter	19.171***	1.11	19.154***	1.17	17.996***	0.83	17.897***	0.70	19.012***	0.00	17.972***	0.78
Individual Attributes												
Jock Alter	-0.054	0.05	-0.054	0.05	0	0.04	-0.005	0.04	-0.06	0.24	-0.003	0.05
Jock Ego	-0.069	0.05	-0.07	0.05	0.029	0.05	0.026	0.04	-0.073	0.13	0.028	0.05
Jock Homophily	0.077	0.05	0.078	0.06	0.077	0.05	0.077x	0.05	0.075	0.16	0.074	0.06
Cheerleader Alter	0.157	0.18	0.163	0.17	0.156	0.17	0.165	0.14	0.165	0.35	0.156	0.16
Cheerleader Ego	0.258	0.17	0.263	0.18	0.188	0.17	0.19	0.15	0.27	0.14	0.185	0.17
Cheerleader Homophily	0.245	0.19	0.252	0.20	0.276	0.20	0.293x	0.16	0.264	0.21	0.283	0.19
Gender Alter	-0.05	0.04	-0.05	0.04	-0.066	0.04	-0.065*	0.03	-0.049	0.21	-0.068x	0.04
Gender Ego	-0.051	0.05	-0.052	0.04	-0.006	0.04	-0.006	0.04	-0.049	0.21	-0.006	0.04
Gender Homophily	0.148***	0.04	0.147***	0.04	0.165***	0.04	0.155***	0.04	0.139**	0.00	0.171***	0.04
Dyadic Attributes												
Same Grade	0.482***	0.04	0.483***	0.04	0.557***	0.03	0.542***	0.04	0.473***	0.00	0.557***	0.04
Co-Affiliation	0.191***	0.04	0.19***	0.04	0.208***	0.04	0.204***	0.03	0.192***	0.00	0.209***	0.04
Structural Effects												
Outdegree(density)	-3.745***	0.06	-3.747***	0.08	-3.724***	0.07	-3.728***	0.06	-3.739***	0.00	-3.724***	0.07
Reciprocity	2.418***	0.11	2.418***	0.10	2.566***	0.09	2.57***	0.07	2.424***	0.00	2.566***	0.08
GWESP I -> K -> J (69)	1.616***	0.07	1.616***	0.10	1.888***	0.07	1.934***	0.06	1.655***	0.00	1.896***	0.06
GWESP I <- K <- J (69)	-0.128	0.10	-0.128	0.12	-0.41***	0.08	-0.397***	0.07	-0.118	0.24	-0.408***	0.09
Indegree - Popularity (sqrt)	0.11	0.07	0.109	0.11	0.307***	0.03	0.31***	0.03	0.114	0.21	0.307***	0.04
Outdegree - Popularity (sqrt)	0.008	0.09	0.01	0.13	-0.382***	0.05	-0.381***	0.05	0	1.00	-0.381***	0.05
Taboo Ties Hypothesis												
Romantic Entrainment	0.564***	0.15	0.563**	0.18	0.568***	0.16	0.618***	0.15	0.617***	0.00	0.583***	0.16
Romantic to Agreement							-0.652***	0.18				
XWX Closure of Romantic									-0.264**	0.01		
Closure of Romantic											-1.329*	0.67
Convergence Ratio	0.26			0.19			0.13			0.23		0.22



RESULT: confirmed

People do not form ties at random

Random graph models for networks – two approaches

- Compare observed networks with random networks
  - people do not form ties at random
  - in what ways is tie-formation different from random
  - some (surprising) network features are random – pathlengths
  - we need to rule out mathematical constraints and random artifacts
- Model the actual tie-formation process
  - in this course only an orientation

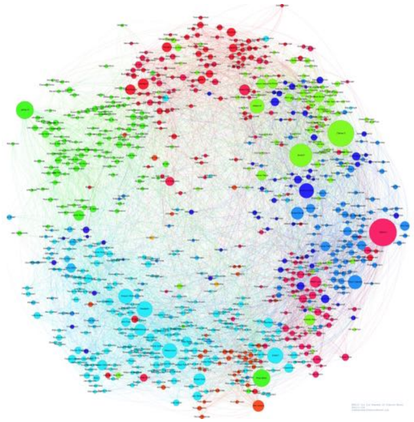
Agenda

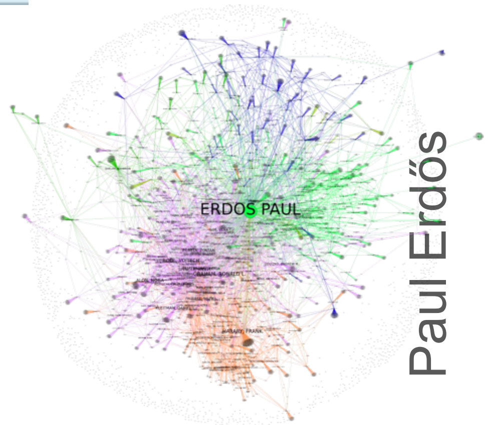
- Random networks are efficient in creating short path-lengths
- Network configurations, random graphs, and underlying mechanisms
- Configurations
  - degree-based
  - closure
  - reciprocation
- Random graphs
  - Bernoulli (Erdös-Reyni)
  - Uniform (density)
  - Uniform (degrees)
  - U|MAN
  - ERGM

Random graphs and small worlds

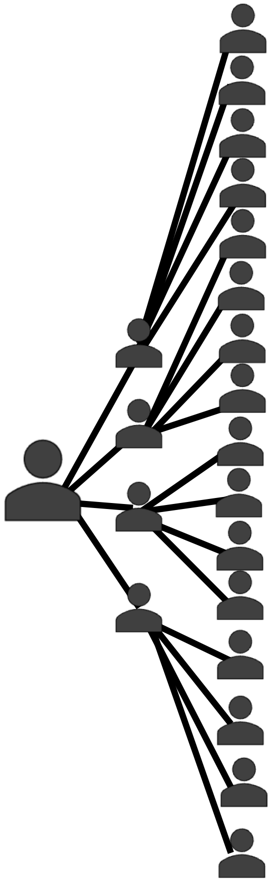
Stanley Milgram

Kevin Bacon





Paul Erdős



You can reach a lot of people in a small number of steps - even with small degree

Combinatorial explosion

Step 0: 1
Step 1: $1 + d^1$
Step 2: $1 + d^1 + (d-1)^2$
Step 3: $1 + d^1 + (d-1)^2 + (d-1)^3$
…

$s = a^0+a^1+\dots+a^k$
$as-s=\dots$

Step k: $1+(d-1)^0+(d-1)^1+(d-1)^2 \dots + (d-1)^k = 1 + \sum_i^k (d-1)^i$

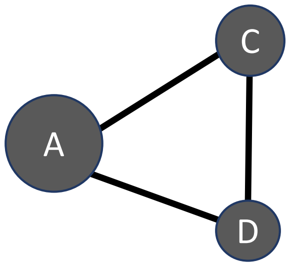
$$1 + \sum_i^k (d-1)^i = 1 + [1-(d-1)^{k+1}]/(2-d) = (3-d-(d-1)^{k+1})/(2-d)$$

d \ k	1	2	3	6	10
2	3	5	7	13	21
3	4	8	16	128	2048
5	6	22	86	5,462	1,398,102
7	8	44	260	55,988	72,559,412

Even if there is no pattern to connections, mathematically, there will be short path lengths (even for small degrees)

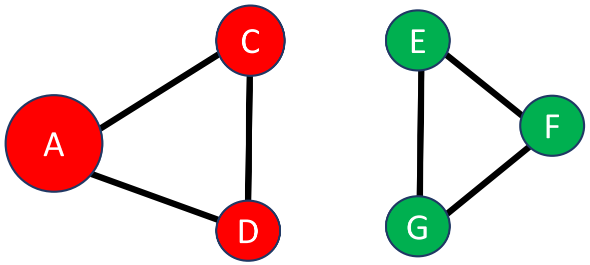
But there are patterns

Triadic closure



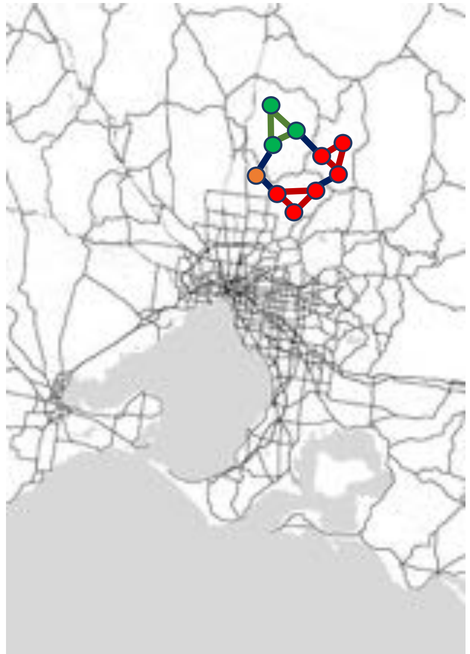
C and D met through A

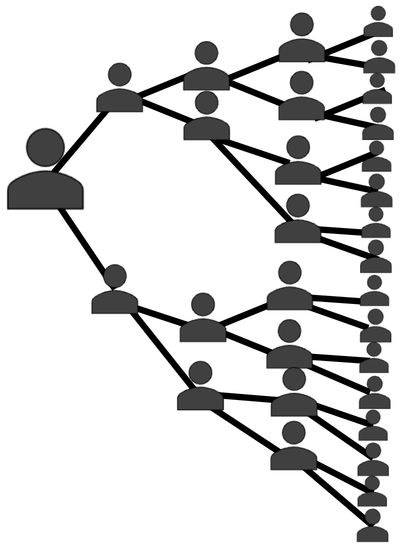
Homophily



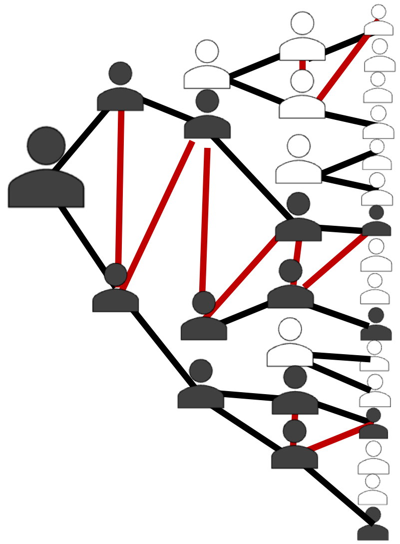
A, C, and D like sports *E, G, and F like the ballet*

Propinquity



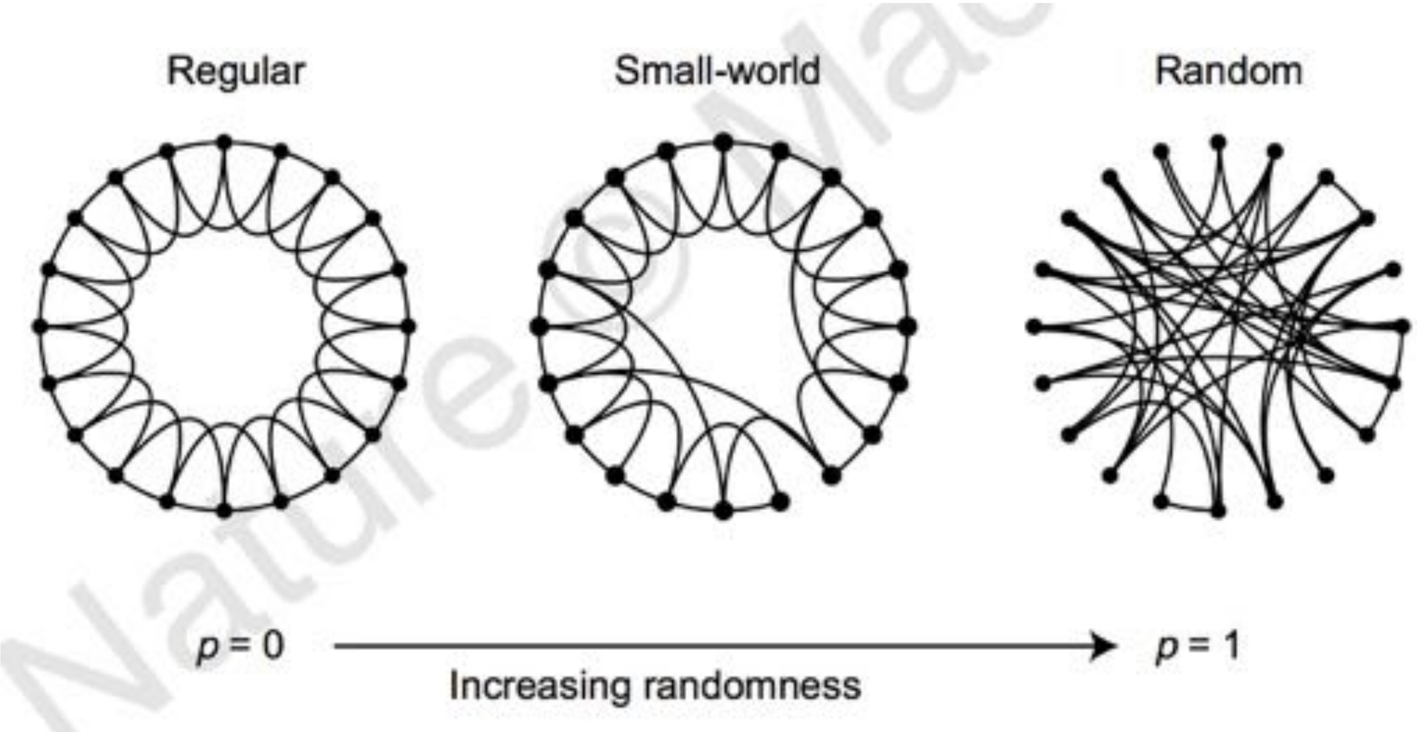


Every generation unique people



Some redundant ties close triads

Granovetter: how can a network be connected (cohesive) if there is strong triadic closure?


Regular
Small-world
Random
p = 0
Increasing randomness
p = 1

Brokers: High-degree nodes

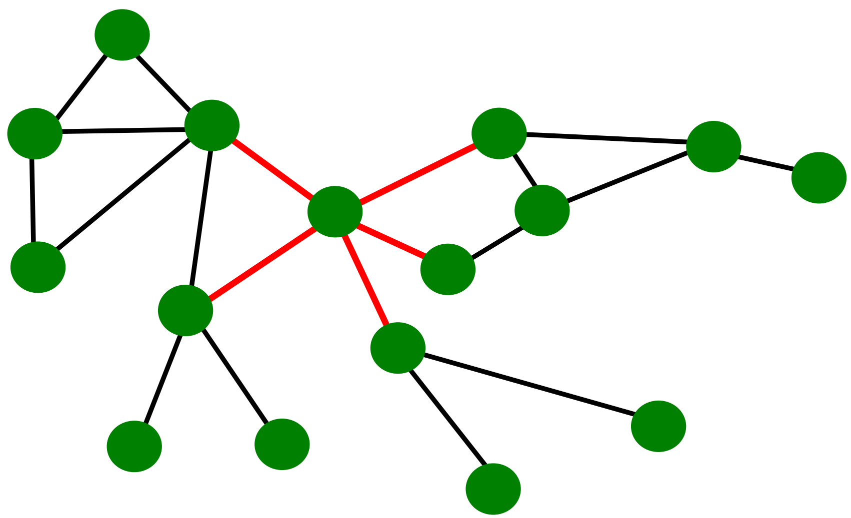
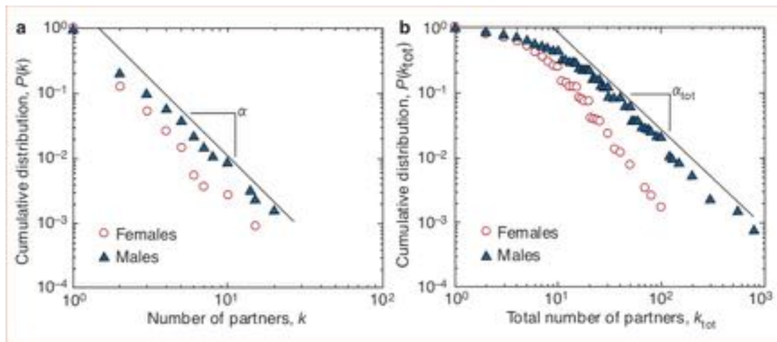
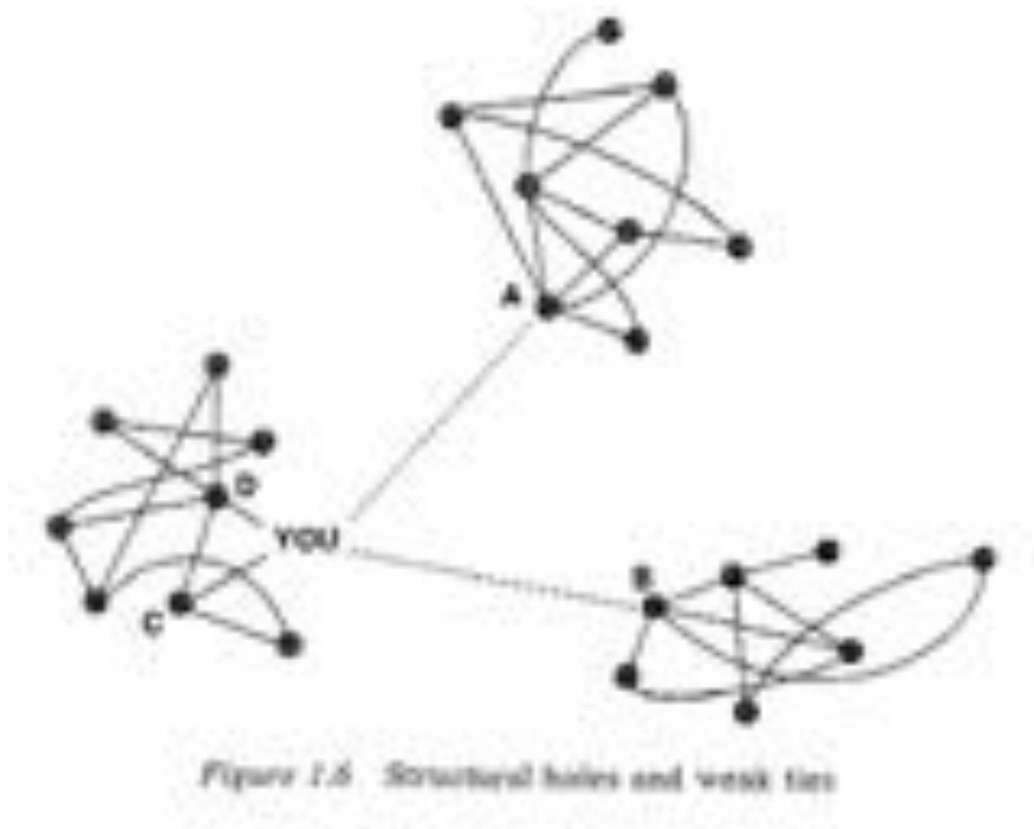


Figure 1 It's a small world: social networks have small average path lengths between connections and show a large degree of clustering. Painting by Idahlia Stanley.



Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Åberg, Y. (2001). The web of human sexual contacts. *Nature*, *411*(6840), 907-908.

Brokers: strategically placed



Burt, R. S. (2009). Structural holes: The social structure of competition. Harvard university press.

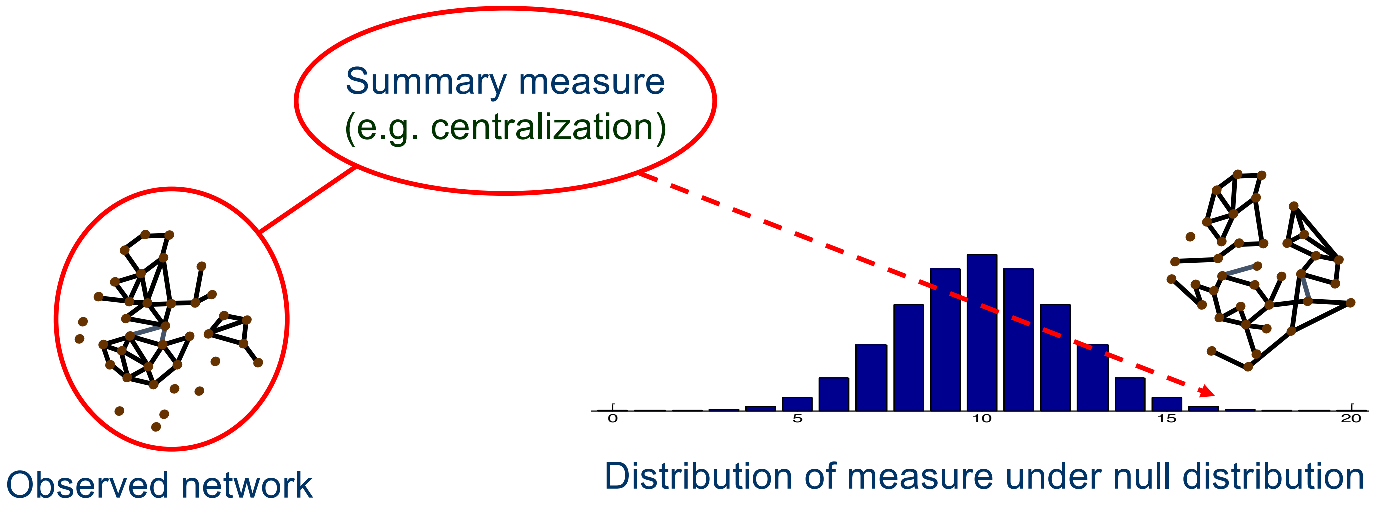
Ties are not random yet there is structure

- In what way
  - There might be a tendency for high-degree nodes to appear
  - There might be a tendency for triads to close
- How do identify these tendencies from data
- What are the socio/behavioural mechanisms at play?

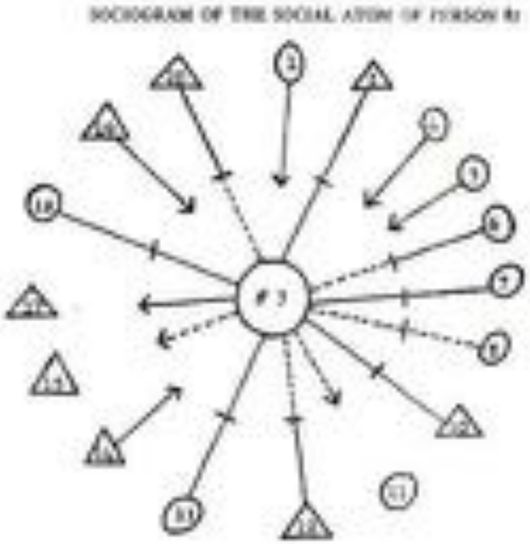
Identifying patterns in data

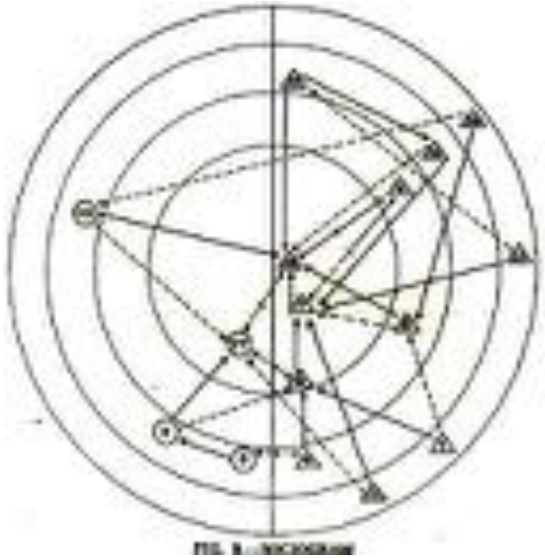
Comparing networks against random graphs

"the **first generation** of research dealt with the distribution of various network statistics, under a variety of null models" (Wasserman and Pattison, 1996)



Moreno (1934) – Sociometrics





Jacob L Moreno

Aim: to create a social equivalent of Psychometrics and map out the forces that affect the individual

Network configurations (Moreno & Jennings, 1938)

TABLE 1
Chance Experiments with the Sociometric Test
Statistical Analysis of the Choices

No. of Choices	0	1	2	3	4	5	6	7	8	9
Chance Balloting 1	2	4	4	4	8	2	2	...	...	...
Chance Balloting 2	2	3	6	3	8	3	...	1	...	...
Chance Balloting 3	1	1	10	5	4	4	1	...	...	...
Chance Balloting 4	...	3	10	5	2	4	2	...	...	...
Chance Balloting 5	3	5	2	9	2	3	2	1	...	...
Chance Balloting 6	1	3	8	5	5	1	2	1	...	...
Chance Balloting 7	2	2	5	8	5	2	2	...	...	...
Total	10	21	45	39	34	19	11	3	...	...
Average	1.4	3.0	6.3	5.6	4.9	2.7	1.6	.4	...	...

random graph degree distribution

TABLE 2
Actual Sociometric Test
Statistical Analysis of the Choices

No. of Choices	0	1	2	3	4	5	6	7	8	9	10	11
Test 1	4	7	4	3	...	2	2	2	1	...	1	...
Test 2	6	3	4	3	2	4	1	1	1	1	...	...
Test 3	5	4	3	4	4	1	2	1	2	...	...	...
Test 4	3	5	4	6	3	1	...	3	...	1	...	...
Test 5	7	3	5	1	2	4	...	2	...	1	...	1
Test 6	3	2	5	8	3	2	2	...	1	...	...	...
Test 7	7	5	5	1	2	...	1	1	1	1	...	2
Total	35	29	30	26	16	14	8	10	6	4	1	3
Average	5.0	4.1	4.3	3.7	2.3	2.0	1.1	1.4	.9	.6	.1	.4

observed degree distribution

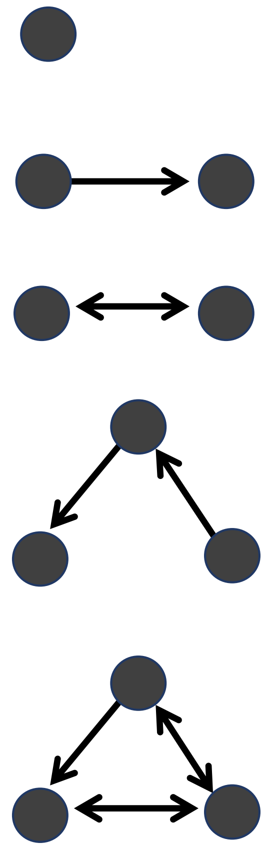
TABLE 3

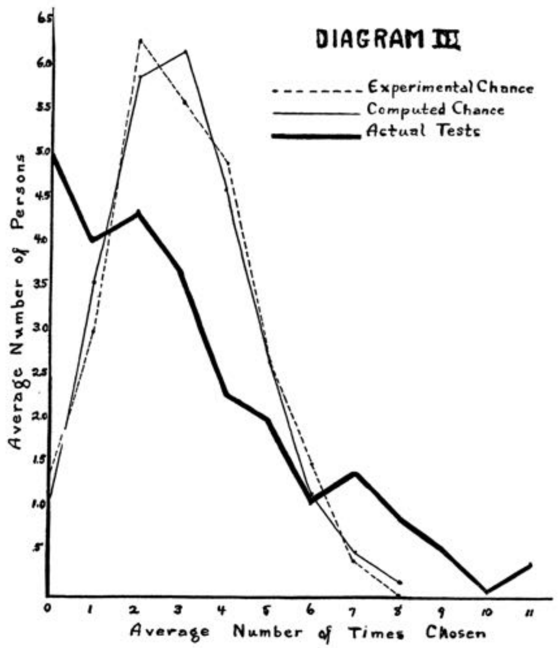
Statistical Analysis of Configurations Occurring in Chance

	Isolated	Unreciprocated	Mutual	Chain Relations	Closed Structures (triangles, etc.)	Leader Structures
Chance Balloting 1	2	68	5	..	..	4
Chance Balloting 2	2	74	2	..	..	4
Chance Balloting 3	1	64	7	2	..	5
Chance Balloting 4	..	72	3	..	..	6
Chance Balloting 5	2	68	5	2	..	6
Chance Balloting 6	1	70	4	1	..	4
Chance Balloting 7	2	70	4	1	..	4
Total	10	486	30	6	0	33
Average	1.4	69.4	4.3	0.9	0	4.7

Statistical Analysis of Configurations Occurring in Actual Sociometric Tests

	Isolated	Unreciprocated	Mutual	Chain Relations	Closed Structures (triangles, etc.)	Leader Structures
Test 1	4	54	12	4	1	8
Test 2	6	48	15	1	1	8
Test 3	5	56	11	4	..	6
Test 4	3	46	16	2	2	5
Test 5	7	48	15	1	2	8
Test 6	3	44	17	2	1	5
Test 7	7	62	8	2	...	6
Total	35	358	94	16	7	46
Average	5	51.1	13.4	2.3	1	6.6





A greater concentration of many choices upon few individuals and of a weak concentration of few choices upon many individuals skews the distribution of the sampling still further than takes place in the chance experiments, and in a direction it need not necessarily take by chance. This feature of the distribution is an expression of the phenomenon which has been called the *socio-dynamic effect.* The chance distribution seen as a whole is also

Degree distributions (post 1998)

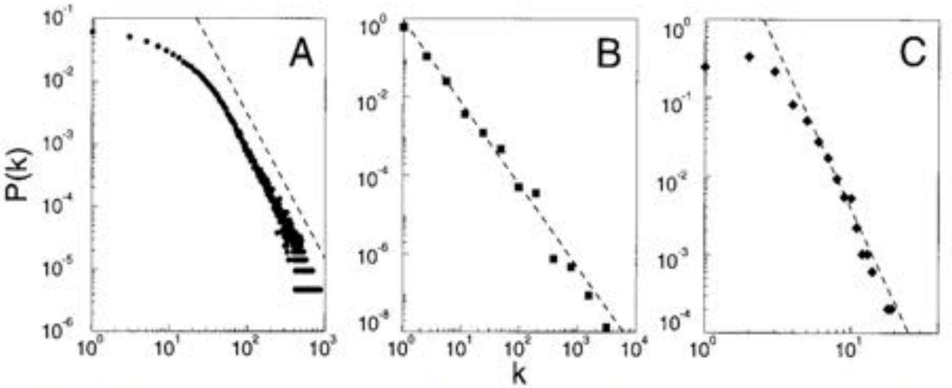
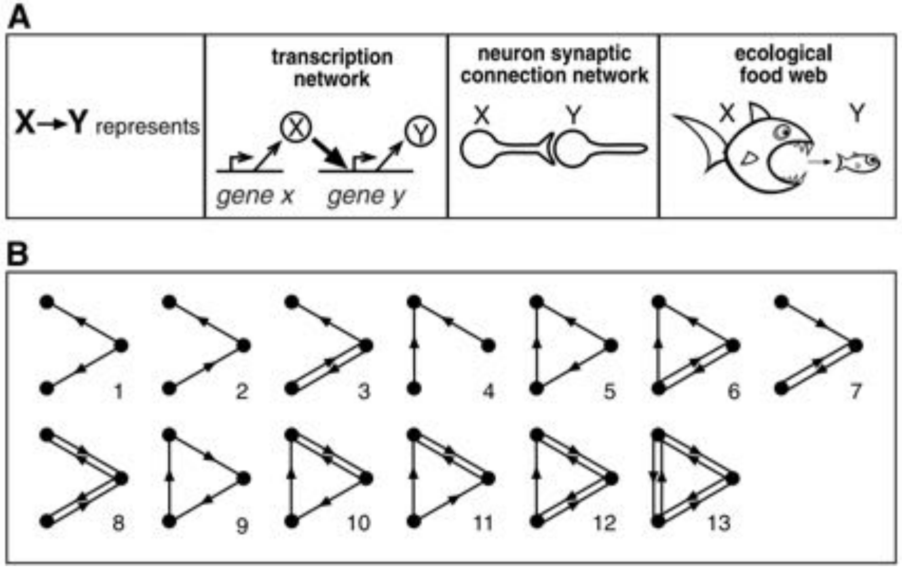


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with N = 212,250 vertices and average connectivity $\langle k \rangle$ = 28.78. **(B)** WWW, N = 325,729, $\langle k \rangle$ = 5.46 (6). **(C)** Power grid data, N = 4941, $\langle k \rangle$ = 2.67. The dashed lines have slopes (A) γ_{actor} = 2.3, (B) γ_{www} = 2.1 and (C) γ_{power} = 4.

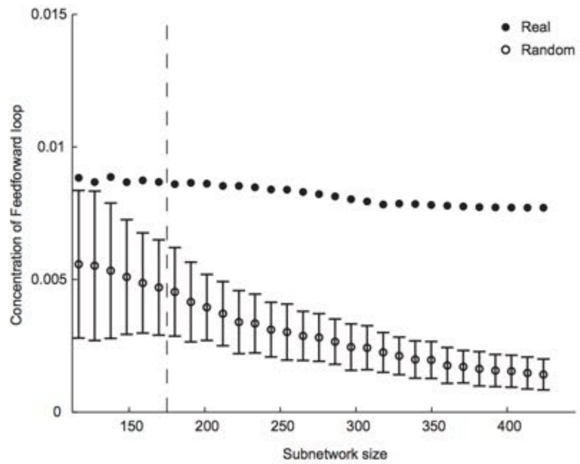
Barabasi & Albert (1999)

Milo et al. (2002): renames subgraphs ‘network motifs’



Milo et al. (2002)

Network	Nodes	Edges	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	Z score	N_{real}	$N_{rand} \pm$ SD	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000



random graph

Network 'motifs'

The question may be raised whether all structures of which a configuration is composed have to be determined or whether a minimum of crucial structures can be a reliable index of their measure. If only the isolates in each configuration were counted up, this would be an insufficient basis of comparison. It would not be known if the remainder consists of chosen but unreciprocated persons or whether it consists of pairs. If, on the other hand, only the number of mutual pairs were counted up, this also would be an unreliable basis of comparison. It would not be known whether the remainder of the configuration consists of entirely unchosen ones because their choices go to those who form the pairs, or whether the individuals who form the pairs are practically isolated from the rest because they choose each other but are cut off from others. As discussed elsewhere, the number of chain-relations, squares, triangles, etc., seems to depend upon the number of mutual pairs. This needs some explanation. There may be many mutual pairs in a structure and no chain-relations or more complex structures. But if there are many complex structures, then a relatively large number of

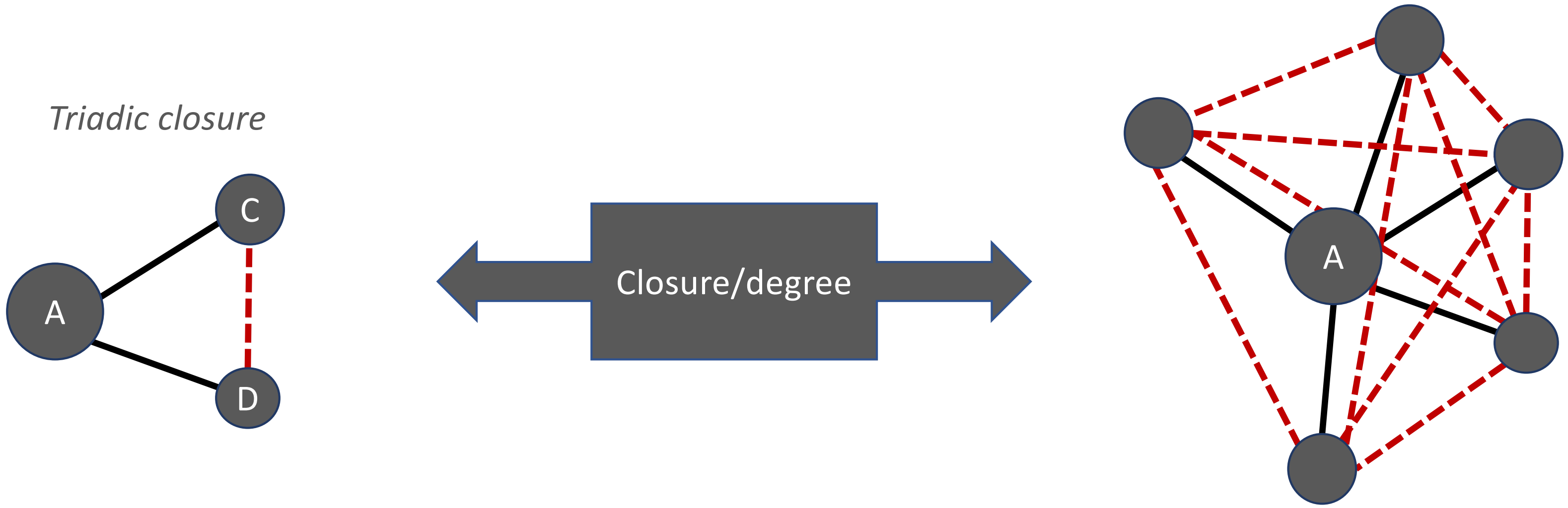
Take-homes

- Configurations are interconnected/nested
- What is a random network and why would anything look random?

Nested configurations

$$\binom{d_A}{2} = \frac{d_A(d_A - 1)}{2}$$

Triadic closure

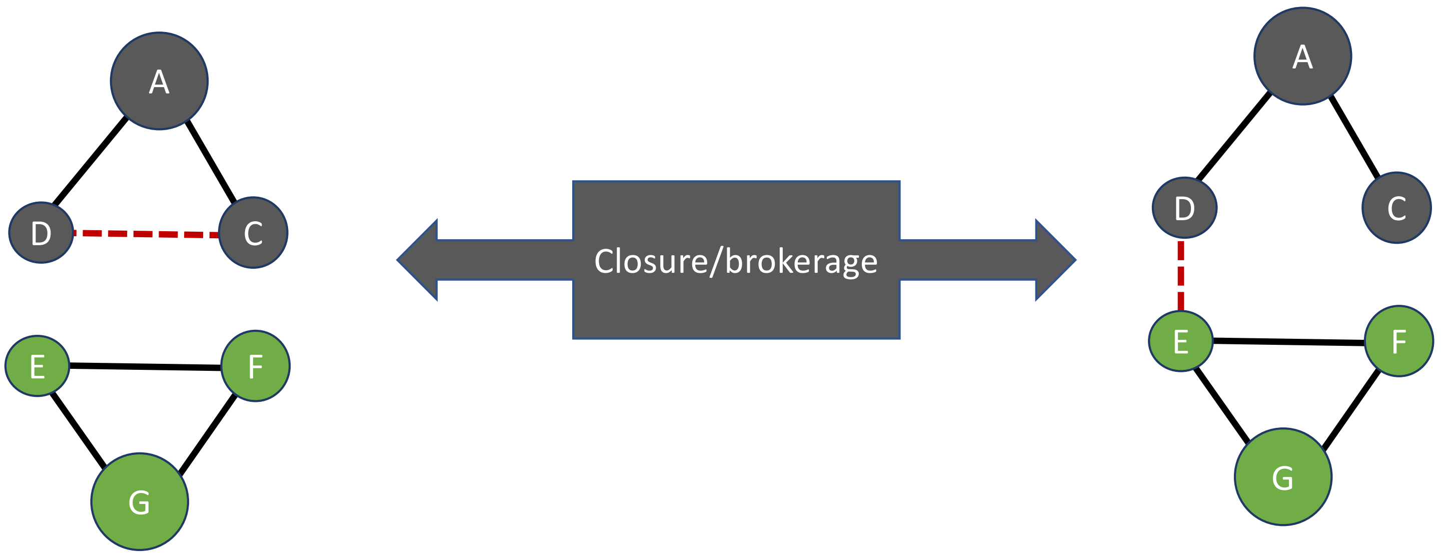


C and D can close the triad

High-degree nodes provide many opportunities for closure

Nested configurations

Triadic closure

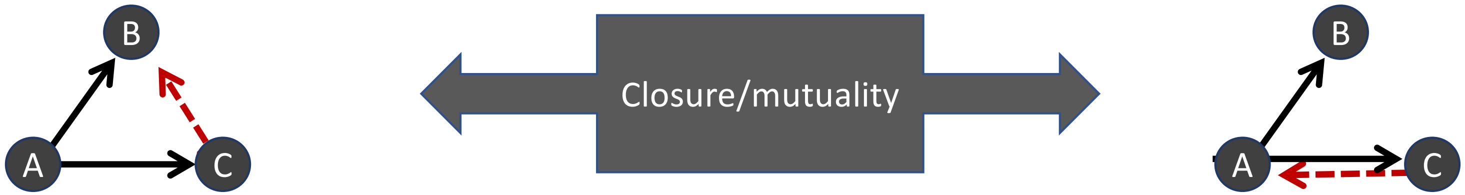


C and D can close the triad

D have opportunity to broker between groups

Nested configurations

Triadic closure



C can close transitive triad

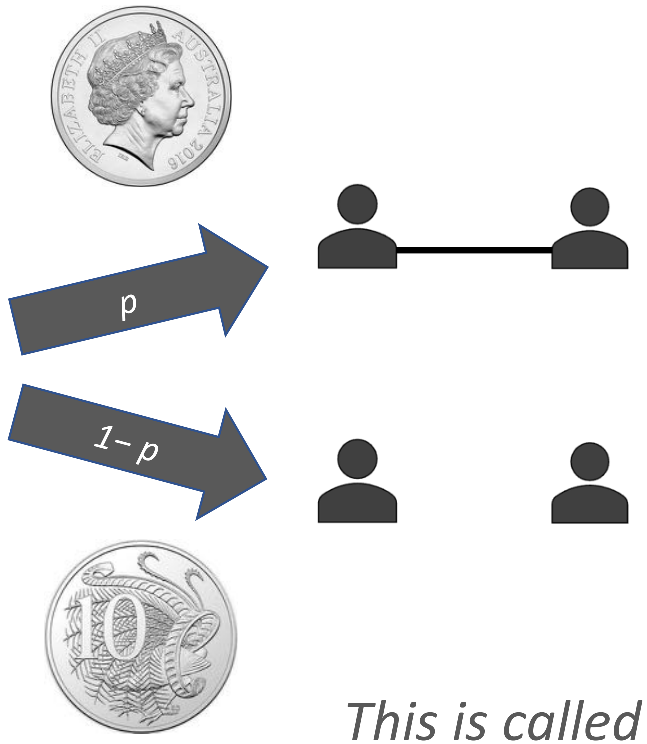
C can reciprocate A

Random graphs

Bernoulli (Erdős- Rényi) random graph

Independently for each dyad

flip a p-coin



This is called a Bernoulli trial

Bernoulli (Erdős- Rényi) random graph

Independently for each dyad:

$$X_{ij} \sim Bernoulli(p),\ E(X_{ij}) = \mathrm{p}$$

$$\begin{pmatrix} - & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & - & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & - & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & - & X_{45} \\ X_{51} & X_{52} & X_{53} & X_{54} & - \end{pmatrix}$$

$$d_i = \sum_j X_{ij} \sim Bin(n-1, p)$$

… but d_i not independent of d_j

$$E\left(\frac{\sum_{i<j} X_{ij}}{n(n\text{-}1)/2}\right) = E\left(\frac{\sum_{i,j} X_{ij}}{n(n\text{-}1)}\right) = \frac{\sum_{i,j} E(X_{ij})}{n(n\text{-}1)} = \frac{n(n\text{-}1)p}{n(n\text{-}1)} = p$$

On average the density will be equal to p

Bernoulli (Erdős- Rényi) random graph

Independently for each dyad:

$$X_{ij} \sim Bernoulli(p),\ E(X_{ij}) = p$$

$$\begin{pmatrix} - & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & - & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & - & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & - & X_{45} \\ X_{51} & X_{52} & X_{53} & X_{54} & - \end{pmatrix}$$

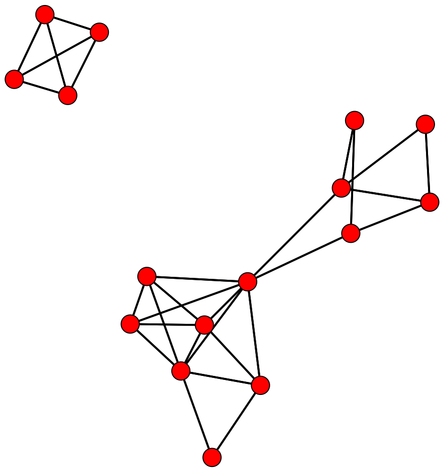
- Actors form ties completely at random
- The tie-probability is the same as the expected density
- X_{ij} is independent of X_{kh} for all other dyads
- The degrees of every node follow the same distribution

$$d_i = \sum_j X_{ij} \sim Bin(n-1,p)$$

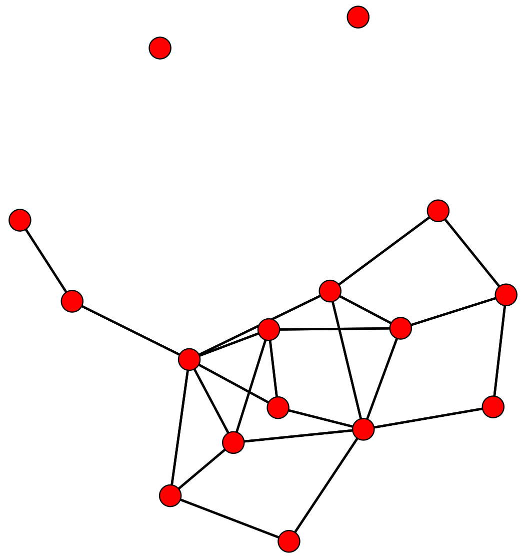
Bernoulli (Erdős- Rényi) random graph

```
g <- rgraph(dim(tribesPos)[1],
            tprob = gden(tribesPos ),
            mode ='graph')
```

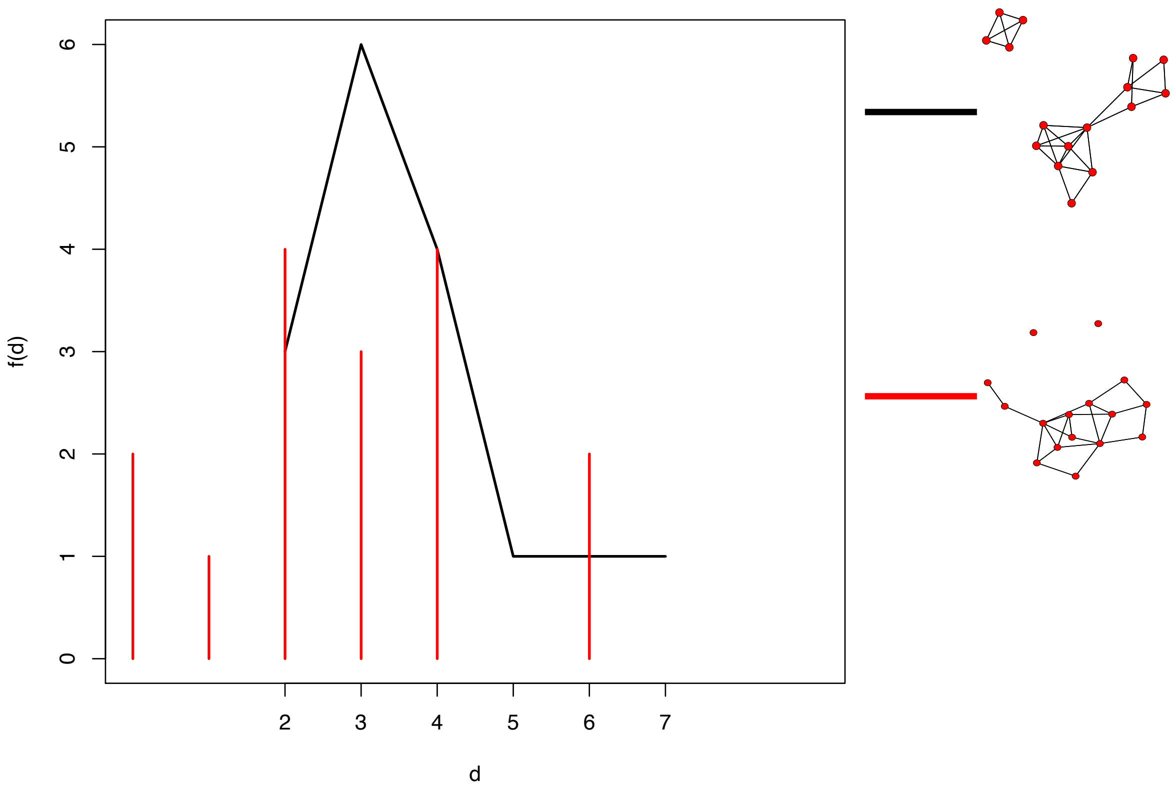
p

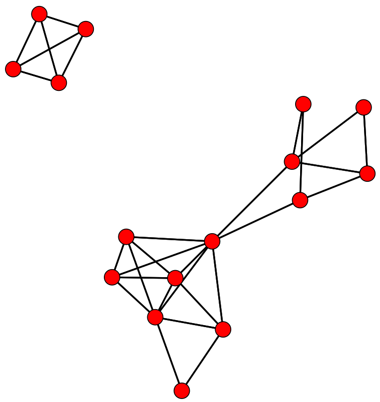


```
> gden(g)
[1] 0.1916667
> gden(tribesPos)
[1] 0.2416667
```

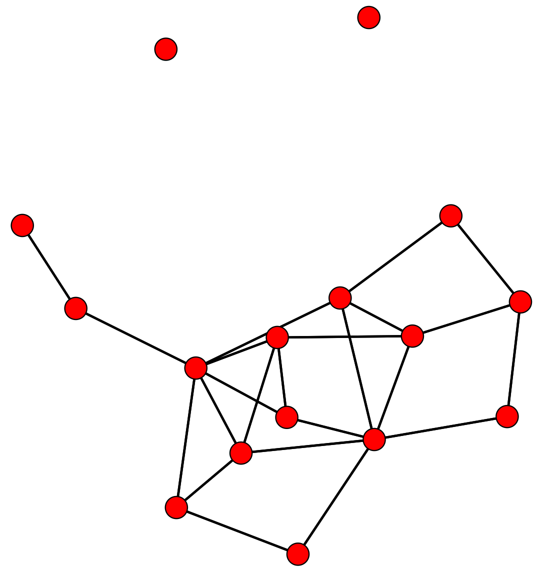


Bernoulli (Erdős- Rényi) random graph





```
> triad.census( tribesPos , mode='graph')
       0   1  2  3
[1,] 226 281 34 19
> triad.census( g , mode='graph')
       0   1  2 3
[1,] 301 200 55 4
```



Bernoulli (Erdős- Rényi) random graph

Independently for each dyad: $X_{ij} \sim Bernoulli(p)$, $E(X_{ij}) = p$

$$\begin{pmatrix} - & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & - & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & - & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & - & X_{45} \\ X_{51} & X_{52} & X_{53} & X_{54} & - \end{pmatrix}$$

Implies a distribution over all graphs

$$\Pr(X = x)$$

Since all tie-variables are X_{ij} *indepedent*

$$\Pr(X = x) = \prod_{i<j} p^{xij}(1-p)^{1-xij}$$

$$= p^{L(x)}(1-p)^{n(n-1)/2-L(x)}$$

Where L(x) is the number of ties

Density-conditioned uniform graph

What if we want exactly L ties?

$$
\begin{pmatrix}
- & X_{12} & X_{13} & X_{14} & X_{15} \\
X_{21} & - & X_{23} & X_{24} & X_{25} \\
X_{31} & X_{32} & - & X_{34} & X_{35} \\
X_{41} & X_{42} & X_{43} & - & X_{45} \\
X_{51} & X_{52} & X_{53} & X_{54} & -
\end{pmatrix}
\qquad
\begin{pmatrix}
- & & & 1 & \\
 & - & 1 & & 1 \\
 & 1 & - & & \\
1 & & & - & \\
 & 1 & & & -
\end{pmatrix}
\xrightarrow{\text{shuffle}}
\begin{pmatrix}
- & & 1 & & 1 \\
 & - & & & \\
1 & & - & 1 & \\
 & & 1 & - & \\
1 & & & & -
\end{pmatrix}
$$

Out of the $\binom{n}{2} = \frac{n(n-1)}{2}$ *tie-variables, we need to pick L*

Density-conditioned uniform graph

Out of the $\binom{n}{2} = \dfrac{n(n-1)}{2}$ *tie-variables, we need to pick* L

write $M = n(n-1)/2$

$$\binom{M}{L} = \frac{M!}{L!(M-L)!}$$

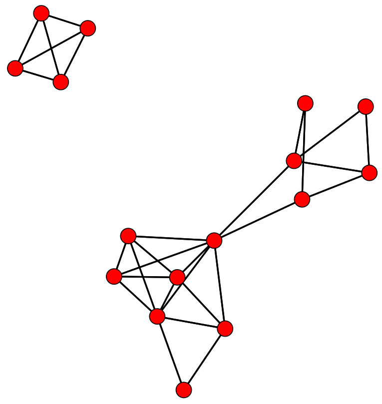
Density-conditioned uniform graph

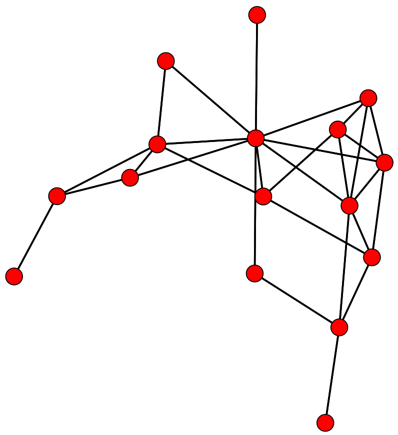
We say that $X \sim \mathrm{U} \mid L(X)=k$, *meaning* X *is conditionally uniform conditional on the density*

$$\begin{pmatrix} - & & & 1 & \\ & - & 1 & & 1 \\ & 1 & - & & \\ 1 & & & - & \\ & 1 & & & - \end{pmatrix} \xrightarrow{\text{shuffle}} \begin{pmatrix} - & & 1 & & 1 \\ & - & & & \\ 1 & & - & 1 & \\ & & 1 & - & \\ 1 & & & & - \end{pmatrix}$$

$$\Pr(X = x) = \begin{cases} \dfrac{1}{M!/(k!(M-k)!)} & \text{, if } L(x)=k \\ 0 & \text{, otherwise} \end{cases}$$

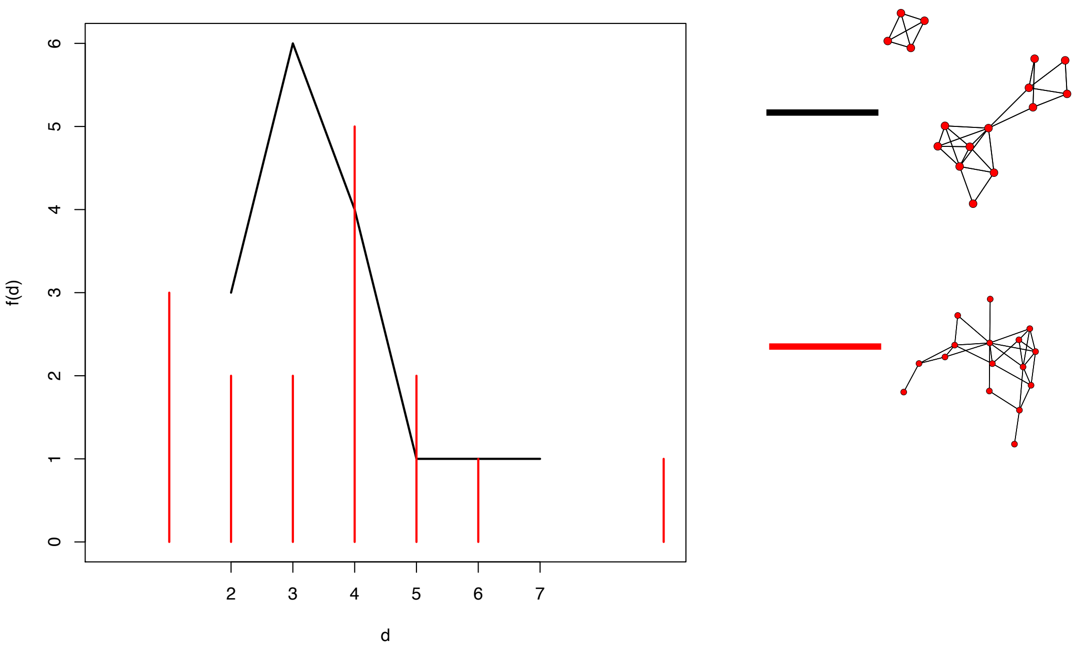
Density-conditioned uniform graph



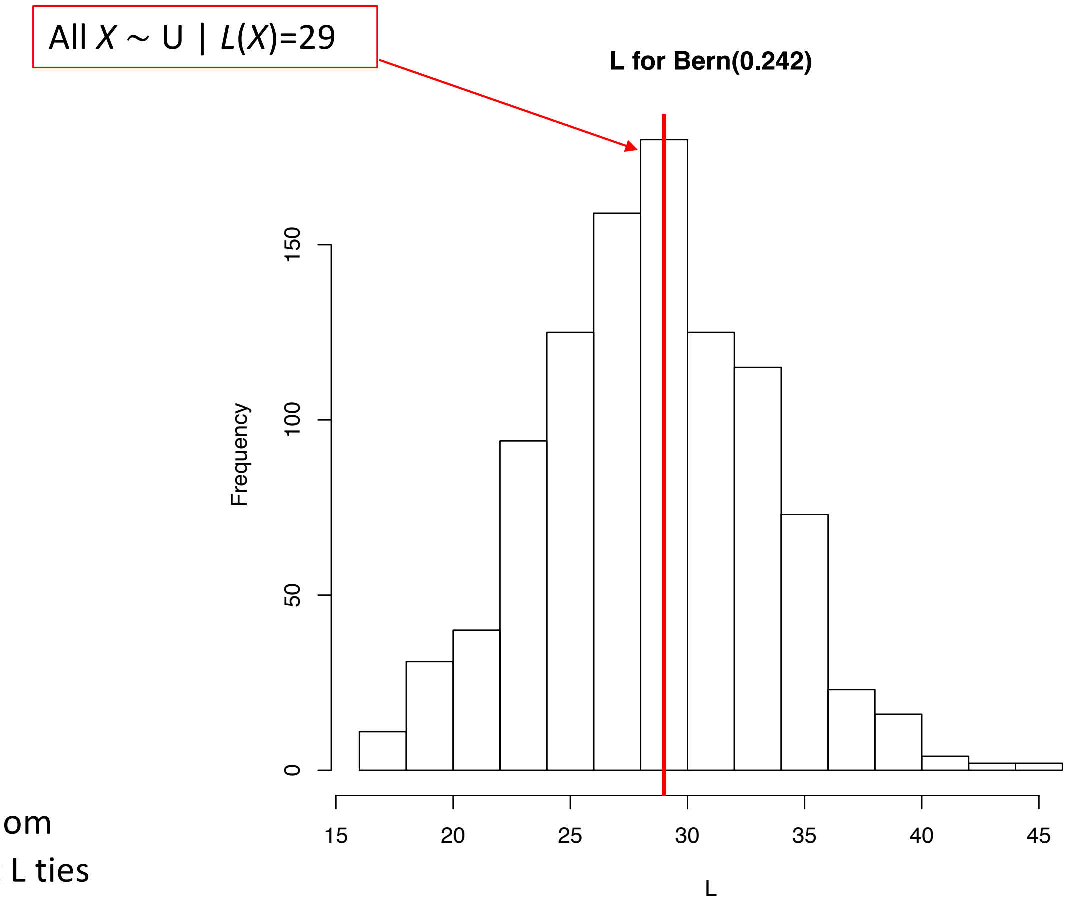


```
g <- rgnm( n = 1,# number of networks to generate
               nv = dim(tribesPos)[1], # the size of the network
               m = sum(tribesPos)/2, # match the number of ties
               mode='graph') # make
```

Density-conditioned uniform graph



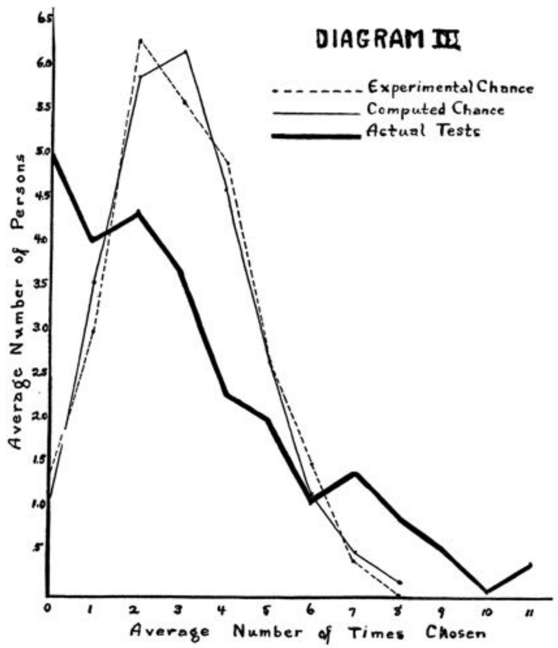
Bernoulli v conditional uniform

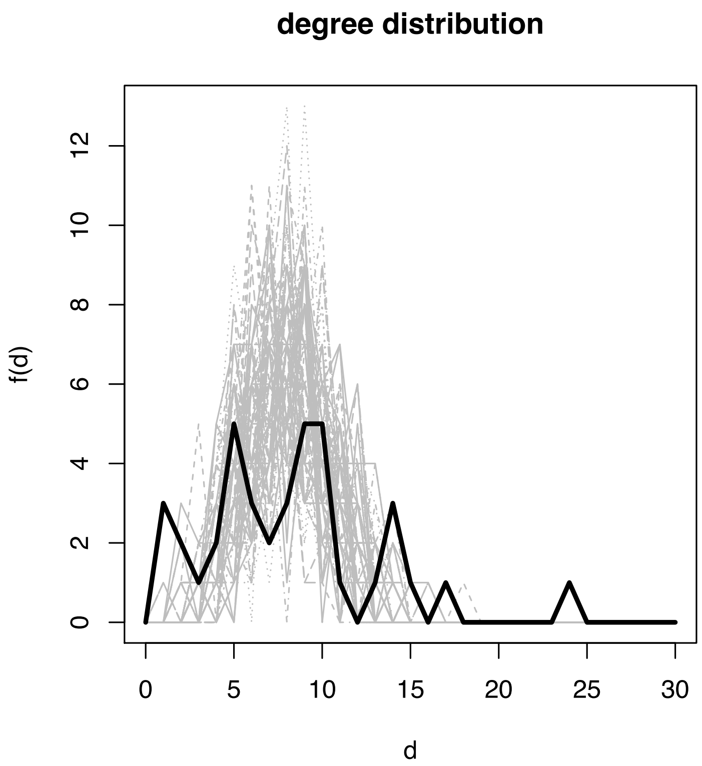


- U|L just more restrictive
- Bernoulli: nodes formed ties at random
- U|L: what if we randomly threw out L ties

Degree distribution

Kapferer's (1972) tailors in Zambia n=39
100 networks Bernoulli

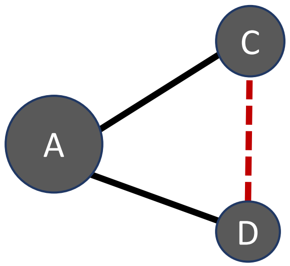




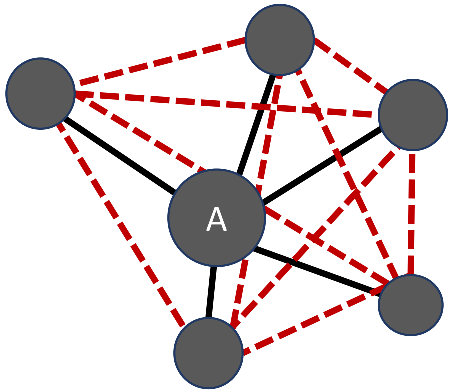
Chance and Bernoulli graphs

- The Bernoulli graph provides us with a null-distribution with realistic density
- We can test if the degrees are evenly distributed
- We can test if there is more closure than if ties were independent, but…

{C,D} higher prob. If {A,C} and {A,D} exists?



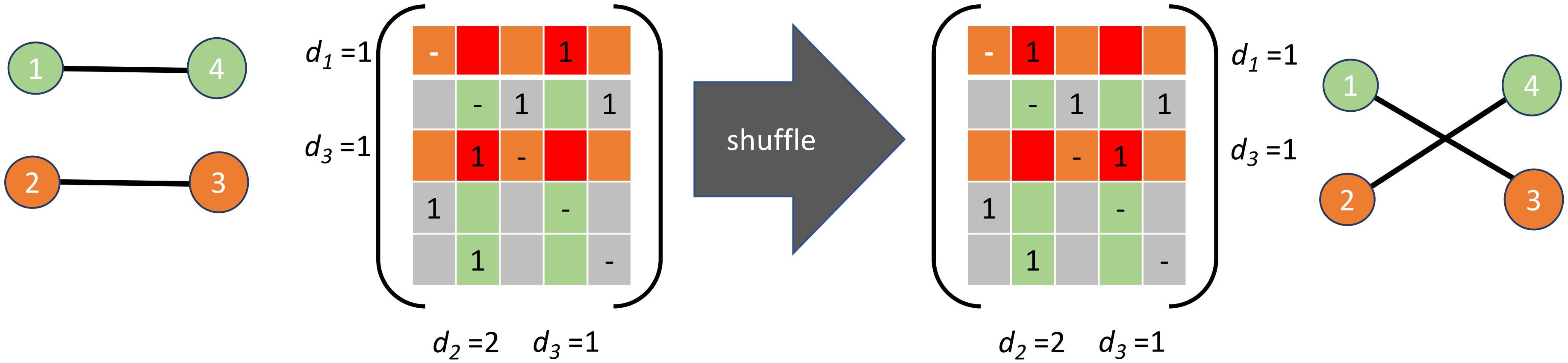
$\Pr(X_{CD}=1 \mid X_{AD}=1, X_{Ac}=1) = p$?



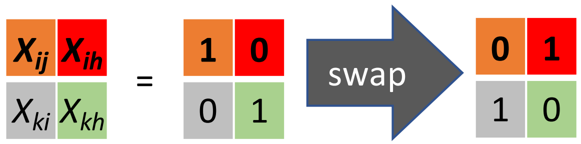
$\Pr(X_{ij}=1,\ldots, X_{kh}=1 \mid X_{Ai}=1, X_{Ac}=1,.., X_{Ak}=1) = p^{10}$?

Uniform distribution conditional on degrees

What if we want each node ***i*** *to have degree* $\boldsymbol{d_i}$



Randomly picking i,j,k, and h



We will randomize ties but keep degree distribution

Uniform distribution conditional on degrees

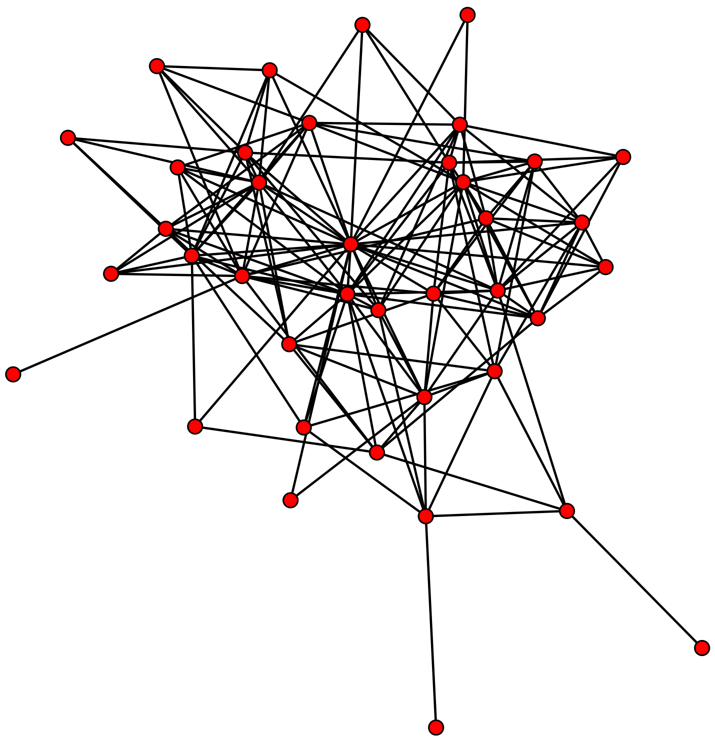
We say that $X \sim \text{U} \mid d$, *meaning* X *is conditionally uniform conditional on the degree distribution*

$$\begin{pmatrix} - & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & - & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & - & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & - & X_{45} \\ X_{51} & X_{52} & X_{53} & X_{54} & - \end{pmatrix} \quad \begin{matrix} \boldsymbol{d_1} \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{matrix}$$

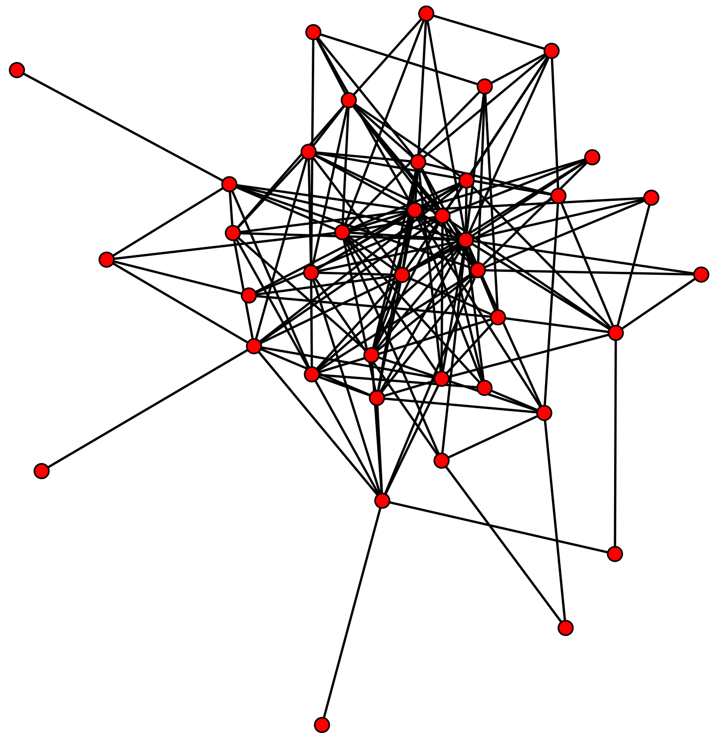
We can use $X \sim \text{U} \mid d$, *to investigate features net of any degree-based effects*

Uniform distribution conditional on degrees

Kapferer's (1972) tailors

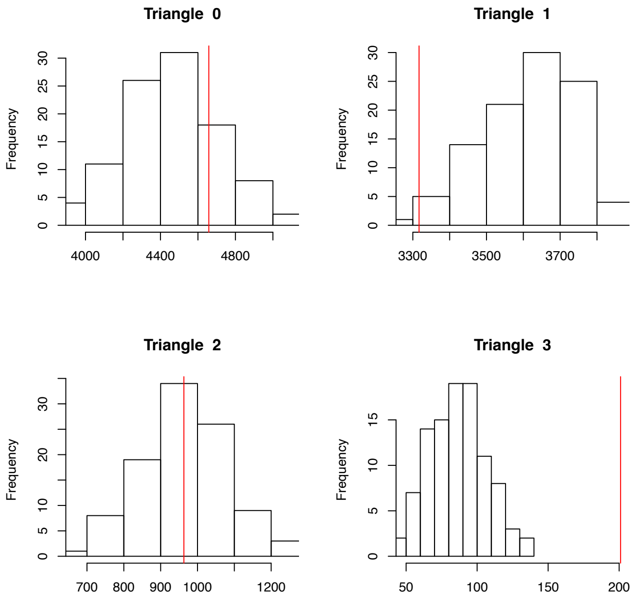


$X \sim \mathrm{U} \mid d$

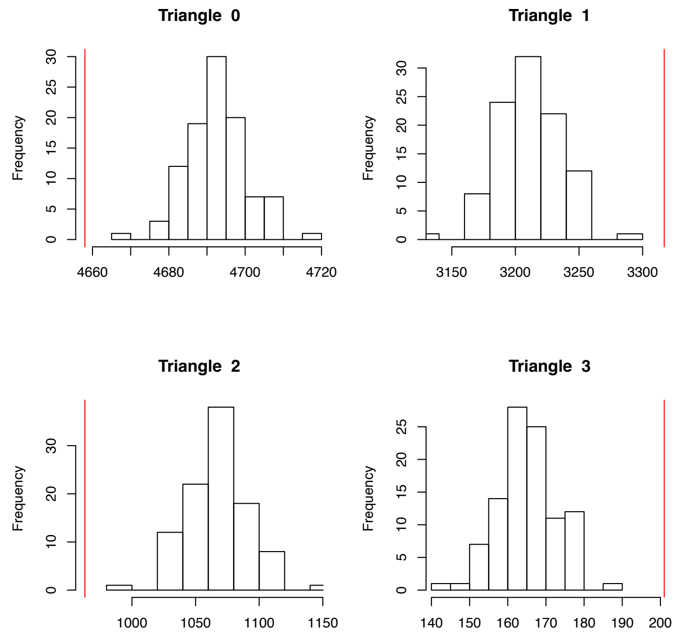


Uniform distribution conditional on degrees

$X \sim \text{Bern(p)}$

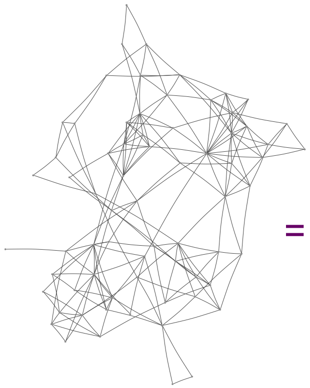
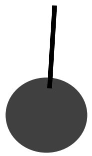
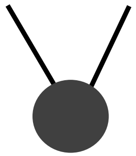
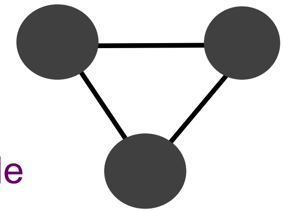


$X \sim \text{U} \mid d$



Does the degree distribution explain triangles?

Exponential Random Graph Model (ERGM)

 = Coef.edges  +Coef.2stars  +Coef.triangle 

Generate graphs that the same number of edges, two-stars, and triangles as a network on average

- positive coefficient means more of that configutation
- For example, a positive triangle parameter that we will generate networks with more triangles

Exponential Random Graph Model (ERGM)

Parameters/weights

Normalising constant

$$\Pr(X=x) = \exp\{\sum_s \overbrace{\theta_s}^{\text{Parameters/weights}} \underbrace{z_s}_{\text{Statistics/configurarions}}(x) - \overbrace{\psi(\theta)}^{\text{Normalising constant}}\}$$

Statistics/configurarions

Generate graphs that the same number of edges, two-stars, and triangles as a network on average

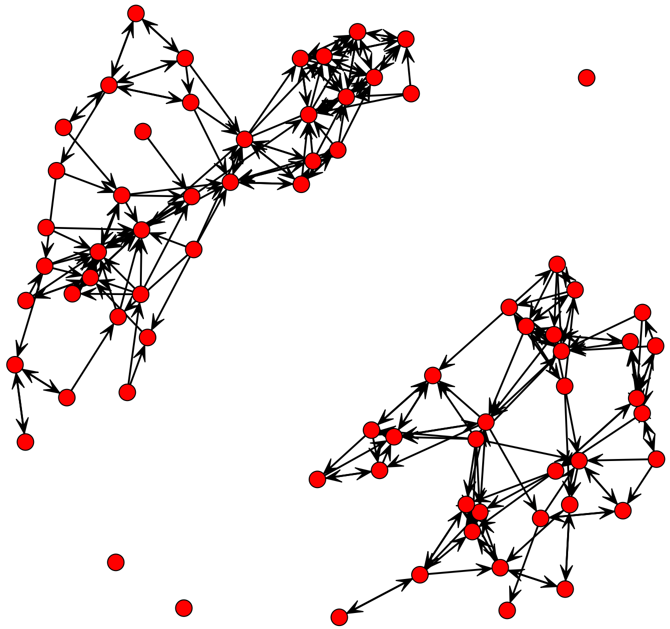
- positive coefficient means more of that configutation
- For example, a positive triangle parameter that we will generate networks with more triangles

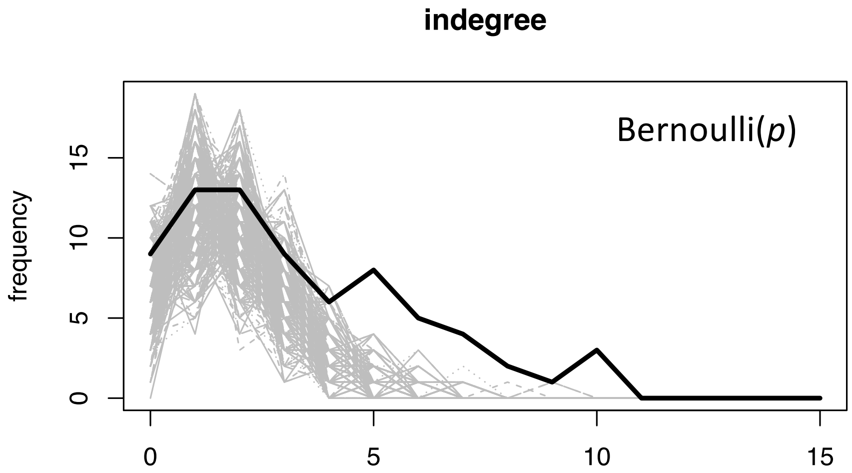
Random graphs for directed networks

Random graphs for directed networks

- Bernoulli(p) – the outdegrees are now independent of eachother
- $X \sim \mathrm{U} \mid L(X)$
- $X \sim \mathrm{U} \mid d_{out}$: we can fix outdegree only – what does activity explain
- $X \sim \mathrm{U} \mid d_{in}$: we can fix indegree only – what does popularity explain
- $X \sim \mathrm{U} \mid d_{out}, d_{in}$: we can fix outdegree AND indegree
- $X \sim \mathrm{U} \mid MAN$: we can fix the dyad census

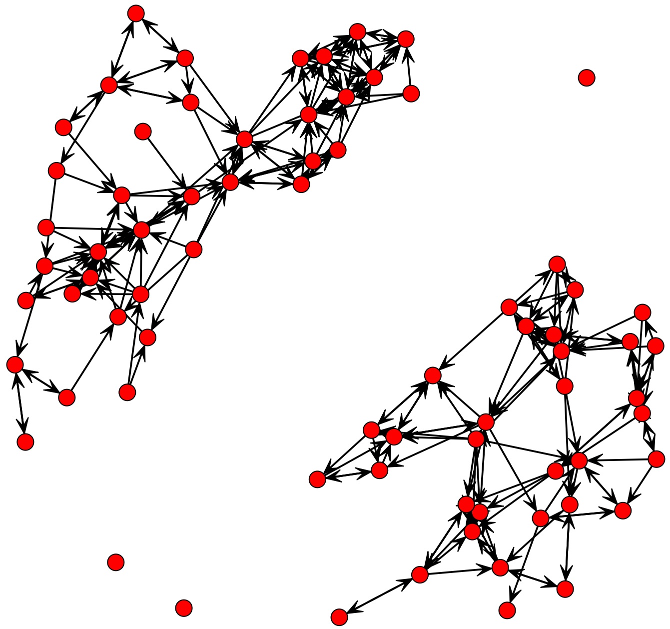
Coleman’s freshmen students

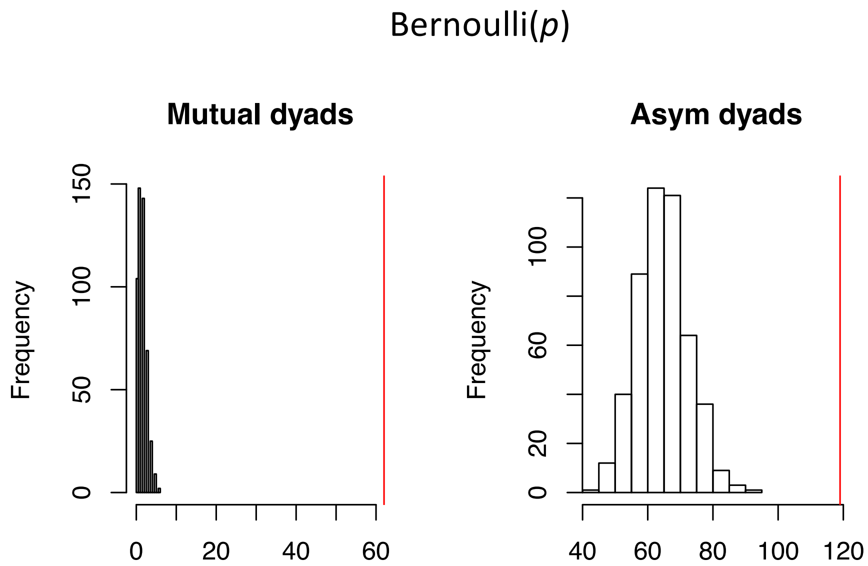




James Coleman (1964) reports research on self-reported friendship ties among 73 boys in a small high school in Illinois over the 1957-1958 academic year.

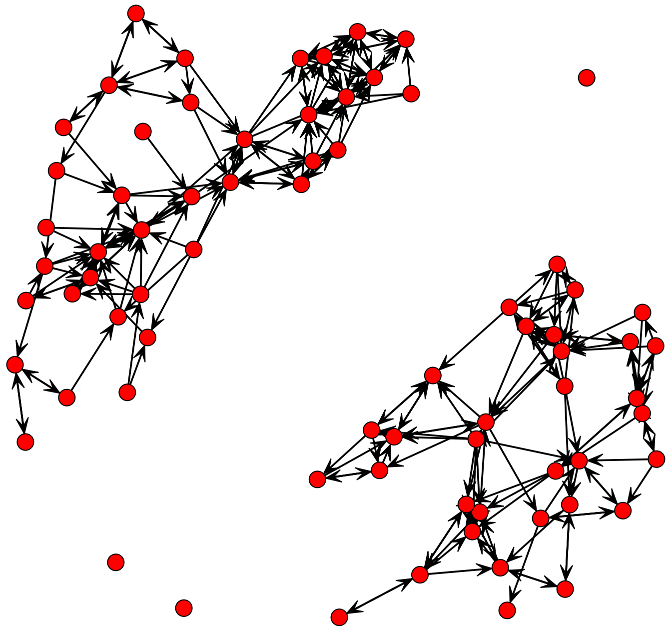
Coleman's freshmen students

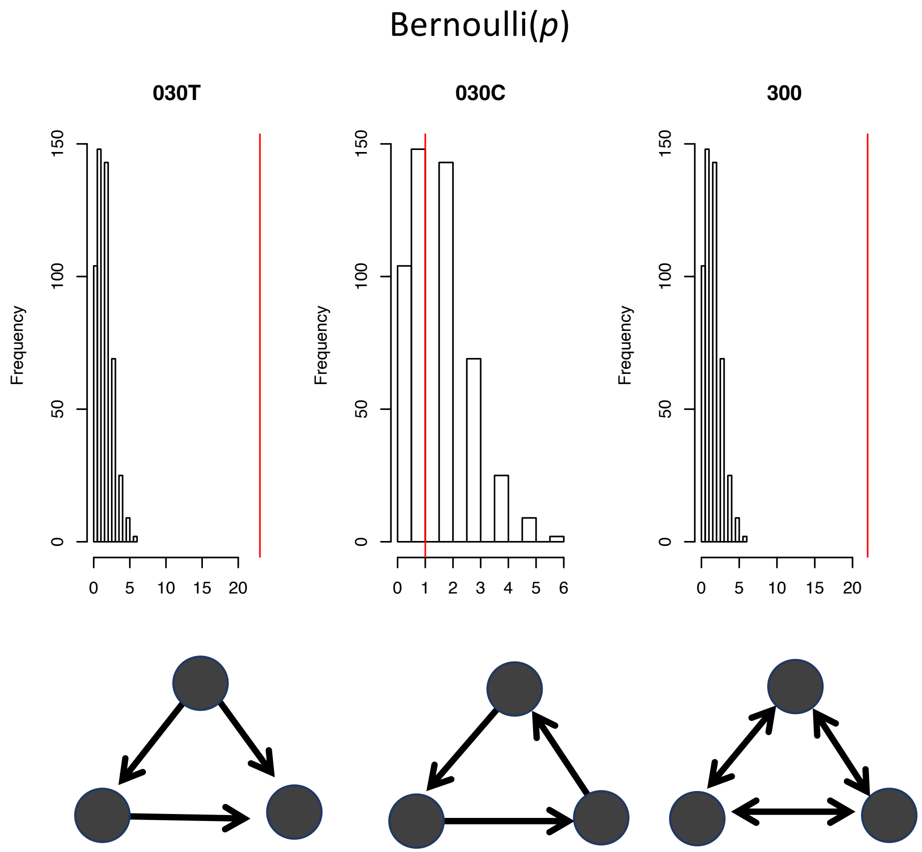




James Coleman (1964) reports research on self-reported friendship ties among 73 boys in a small high school in Illinois over the 1957-1958 academic year.

Coleman's freshmen students





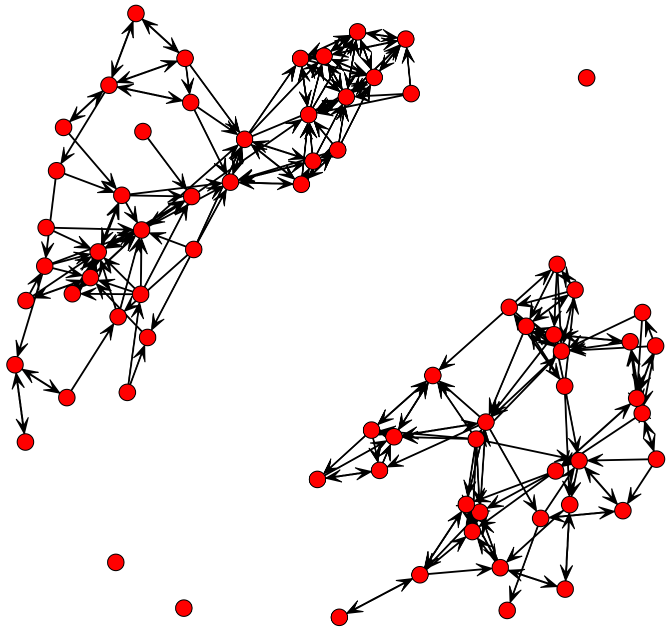
Conditional U |MAN

Similar to U | L, we know that we need
M mutual dyads
A asymmetric dyads, and
N Null dyads

Out of the $\binom{n}{2} = \frac{n(n-1)}{2}$ *total dyads*

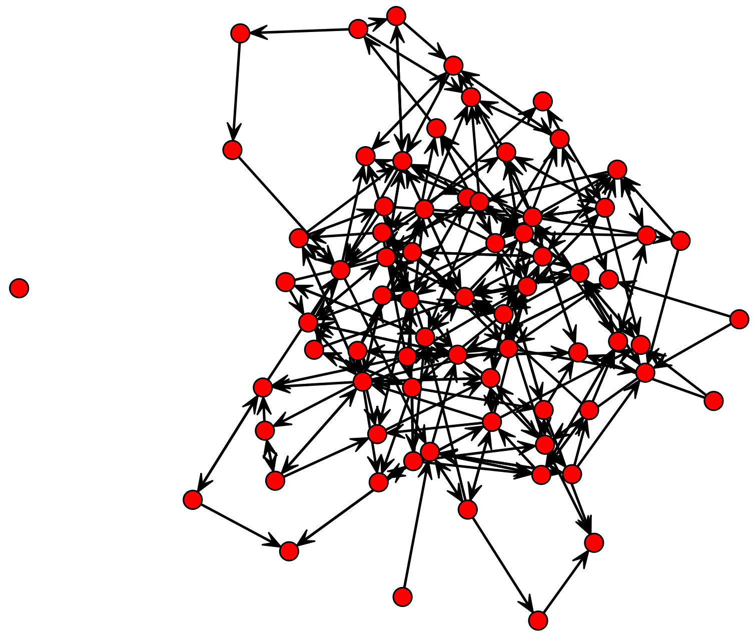
You can randomly select M dyads and set them to mutual, A dyads and set them to asymmetric, and leave the rest Null

Coleman’s freshmen students

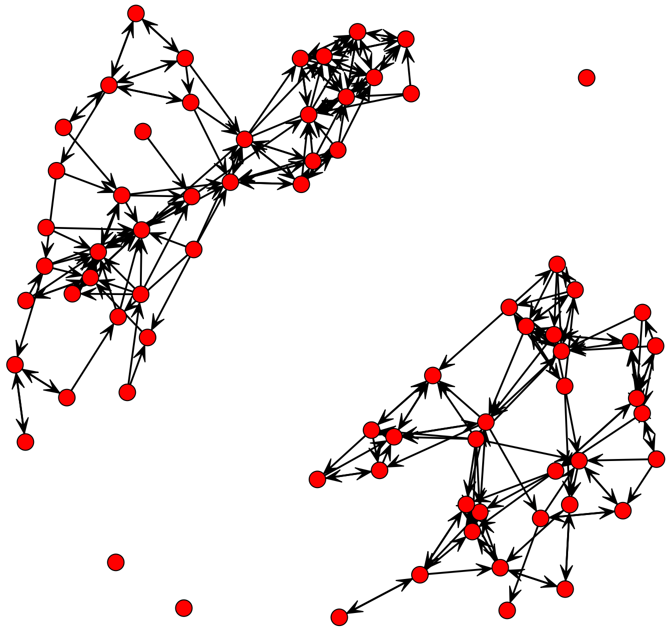


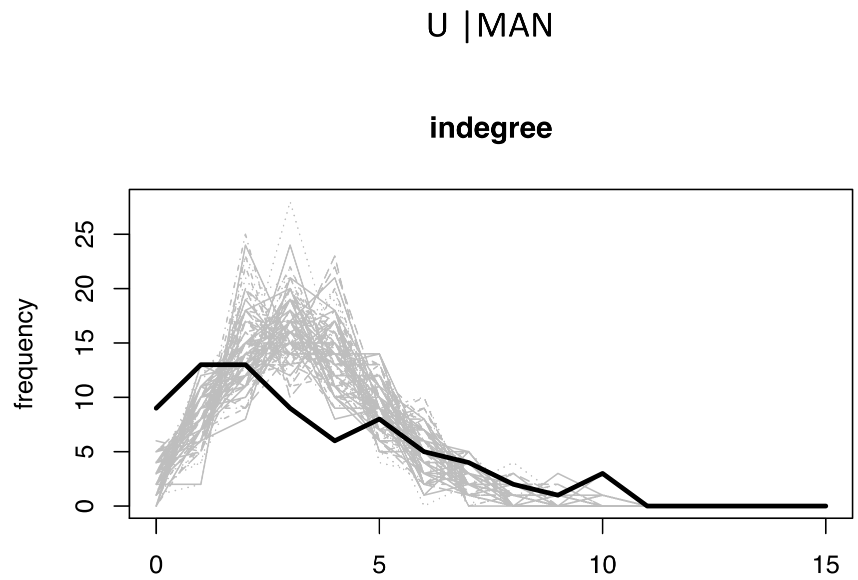
```
> dyad.census(X)
     Mut Asym Null
[1,]  62  119 2447
```

```
Xsim <- rguman( n = 100,# generate 500 random networks # match network size
                nv = 73, # the size of the networks
                mut = 62,# the number of mutual dyads
                asym = 119,# the number of asymetric dyads
                null = 2447,# the number of null dyads
                method='exact') # make sure these are undirected graphs
```



Coleman's freshmen students

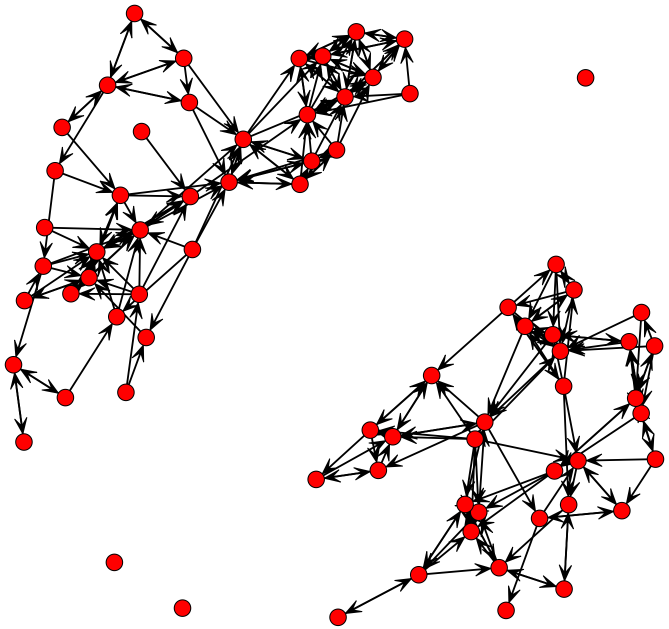




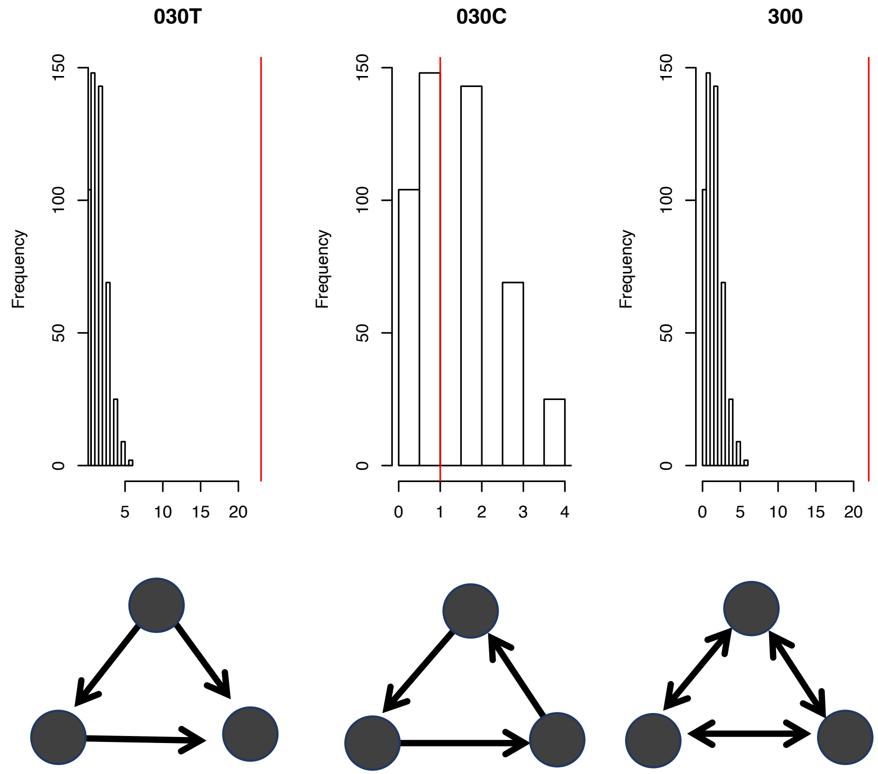

```
> dyad.census(X)
     Mut Asym Null
[1,]  62  119 2447
```

Does reciprocity explain some of the heterogeneity in popularity?

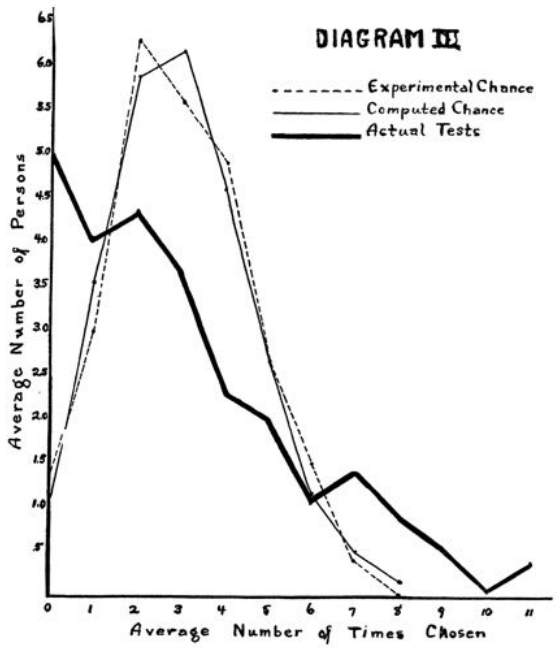
Coleman’s freshmen students



U |MAN



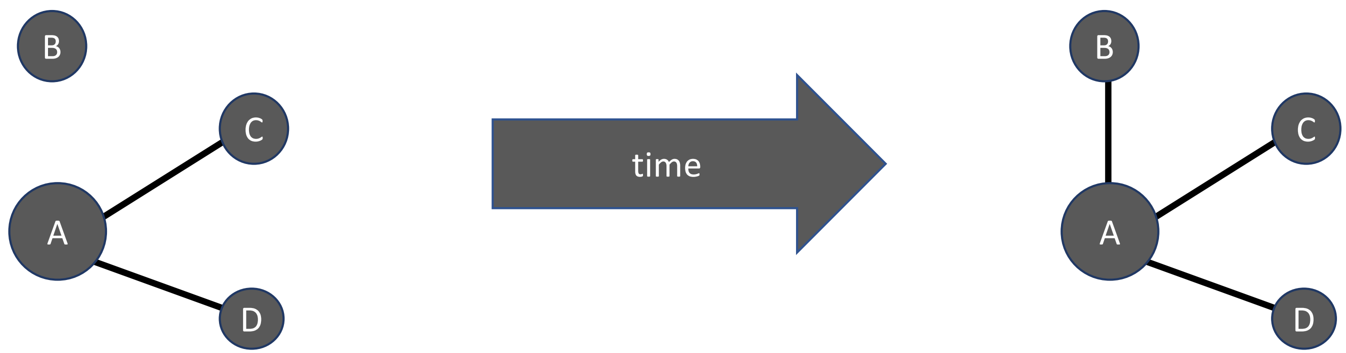
Network mechanisms



A greater concentration of many choices upon few individuals and of a weak concentration of few choices upon many individuals skews the distribution of the sampling still further than takes place in the chance experiments, and in a direction it need not necessarily take by chance. This feature of the distribution is an expression of the phenomenon which has been called the *socio-dynamic effect*. The chance distribution seen as a whole is also

Endogenous network processes

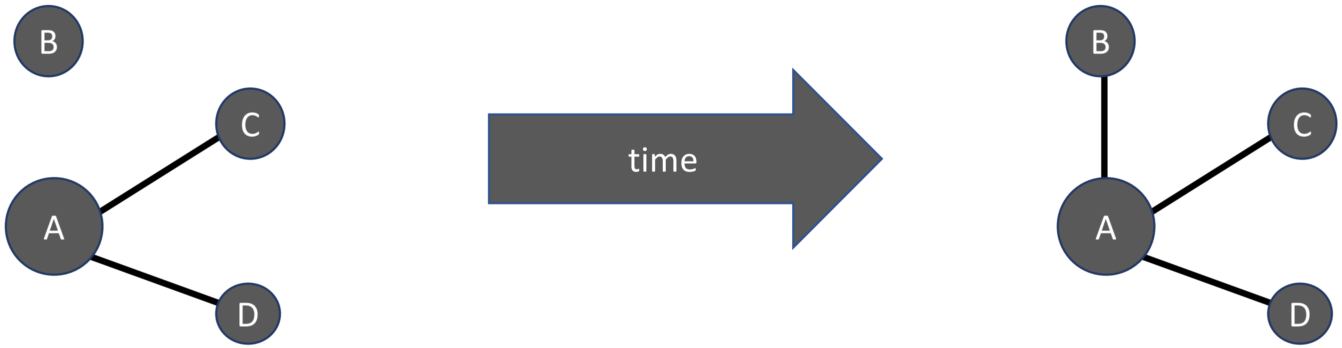
- The Matthew Effect: Merton (1968) cumulative advantage
- de Solla Price (1976)
- Albert & Barabási (2002) preferential attachment leads to power-laws (fpr degree-distribution)



Higher degree node more likely to be the target of the next tie

Endogenous network processes

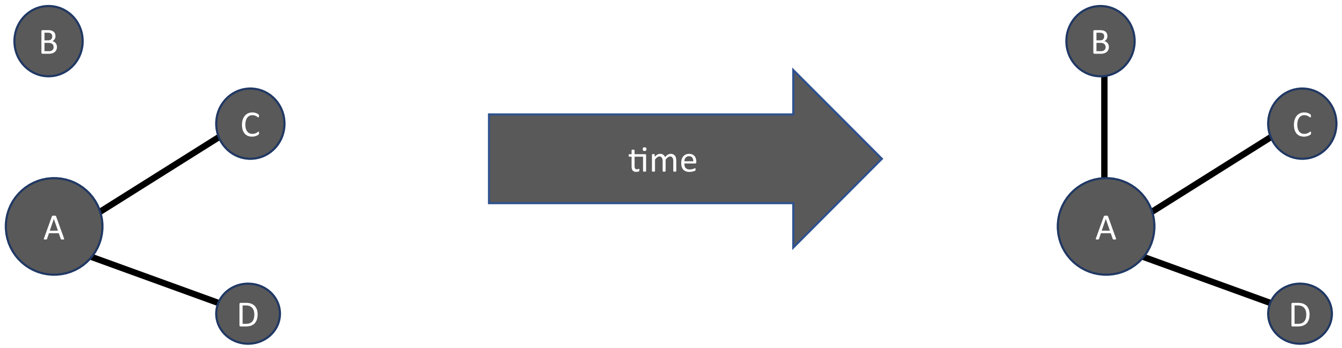
- Popular nodes more visible
- Popular nodes may be popular for a reason - signal
- People want to be friends with the popular guy
- People that have many ties have demonstrated that they are capable of having many ties



Higher degree node more likely to be the target of the next tie

Endogenous network processes

But why isn't everyone popular?

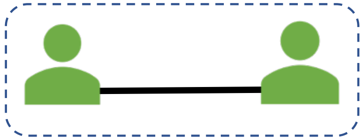


Higher degree node more likely to be the target of the next tie

What does it mean to be or not be in a triad?

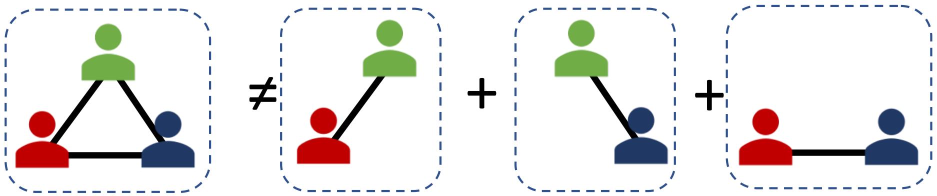
Triangle: the smallest group

The Web of Group Affiliations (Simmel, 1922):



Dyad

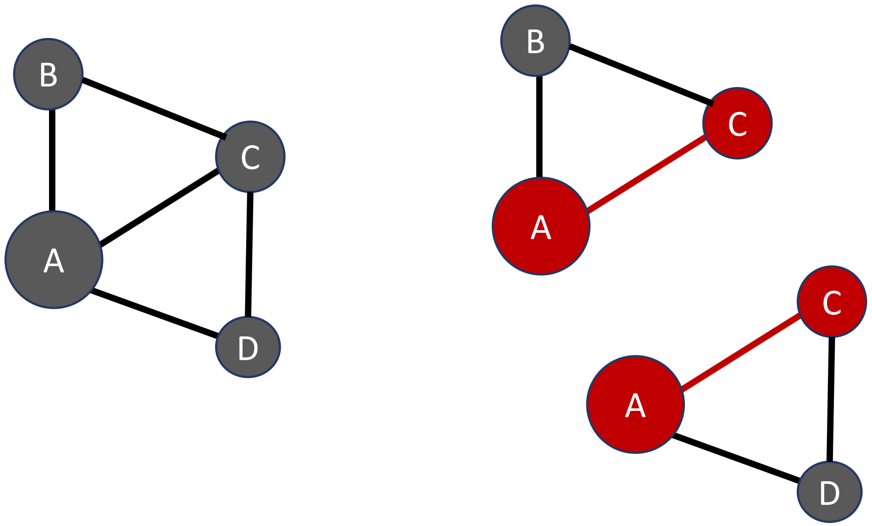
Triad



Smallest collection of individuals where there is a majority

Triads are overlapping

You can decompose the network into distinct dyads

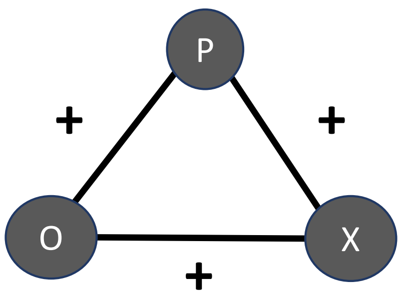


But triangles share edges

Balance theory

Blau (1964) exchange theory (costs and benefits) explain groups

Heider, F (1958) P-O-X: Person, Other, Object X

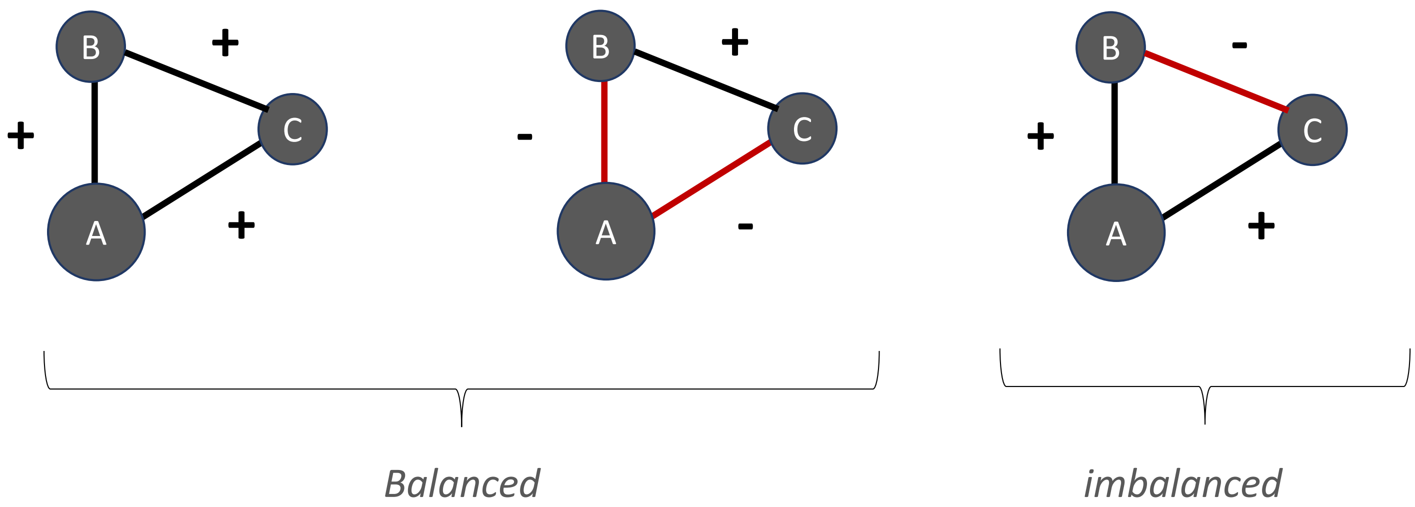


Balance theory and social ties

- Cartwright and Harary (1956) generalised Heider's (2013) balance theory to (signed) networks
- Davis (1967) relaxed the conditions under which balanced (signed) triangles lead to a polarised graph
- Holland and Leinhardt (1971), for directed (non-signed) graphs, balance corresponds to transitive triangles. This work has been cited a lot in the literature to support the notion that transitive triangles should be common and cyclic triangles less common.

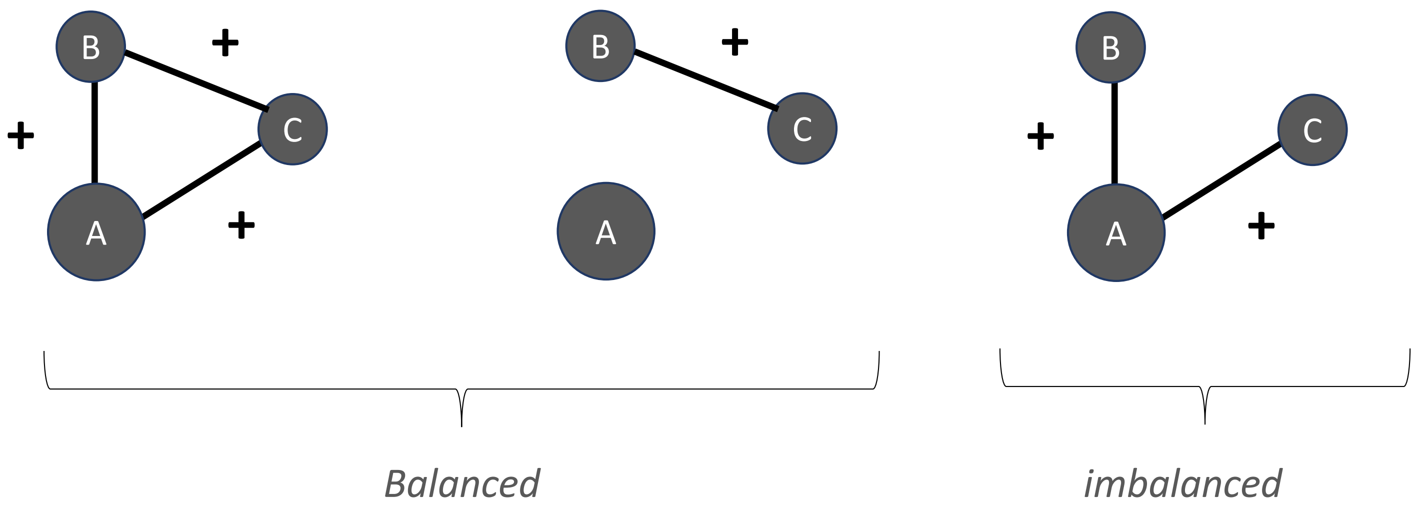
Triadic closure (Granovetter)

Davis (1970) dyadic properties combine through mechanism of balance to create triangles



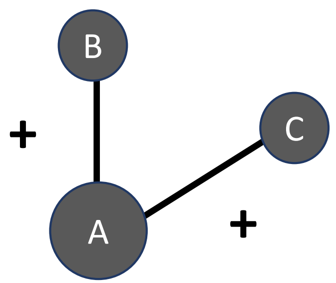
Triadic closure (Granovetter)

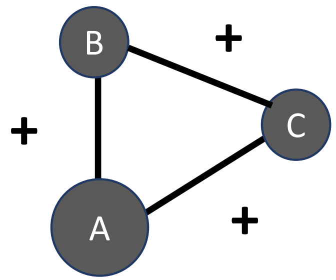
Take negative tie to mean absence of positive tie



Triadic closure (Granovetter)

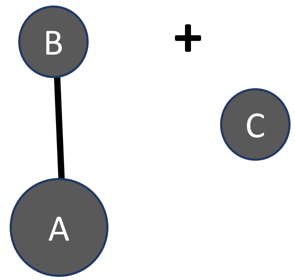
Take negative tie to mean absence of positive tie





Resolved by either
A bringing B and C together, or
B and C getting to know each other

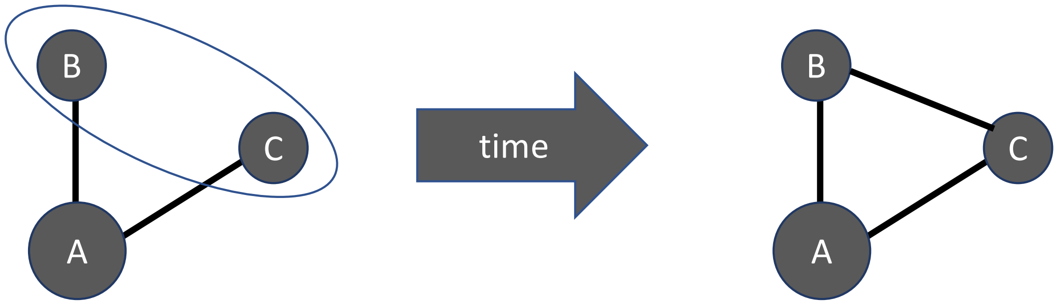
Imbalanced has to be resolved



Resolved by either
A bringing cutting ties to C, or
C cutting ties to A

Triadic closure (Granovetter)

Simmel (1955) friendship transitivity implies a social mechanism for closure; tension between dyad and triad



A provides a social setting for B and C to get to know each other

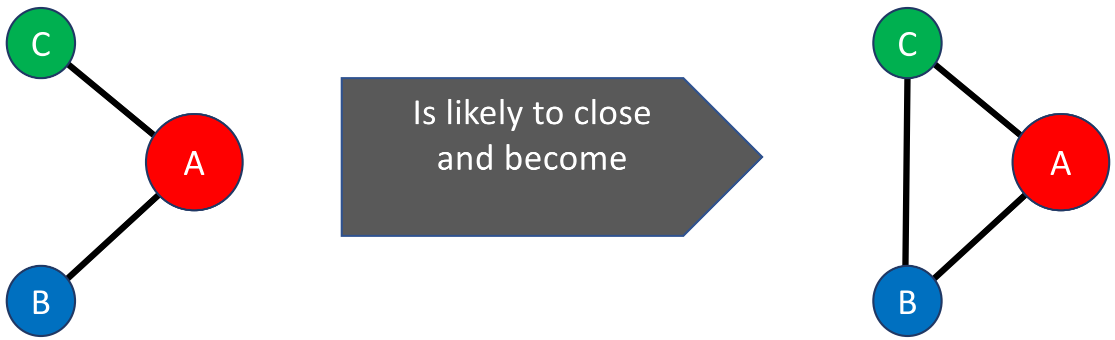
"It is well-known fact that the likely contacts of two individuals who are closely acquainted tend to be more overlapping than those of two arbitrarily selected individuals" (Rapoport, 1954, p.75)

Granovetter and the forbidden triad and network cohesion

Strength of weak ties (Granovetter, 1973)

If a person A has a strong tie (e.g., a close friendship) with persons B and C, then B and C are themselves likely to become friends. (Triadic Closure or clustering)

A two-path A triangle

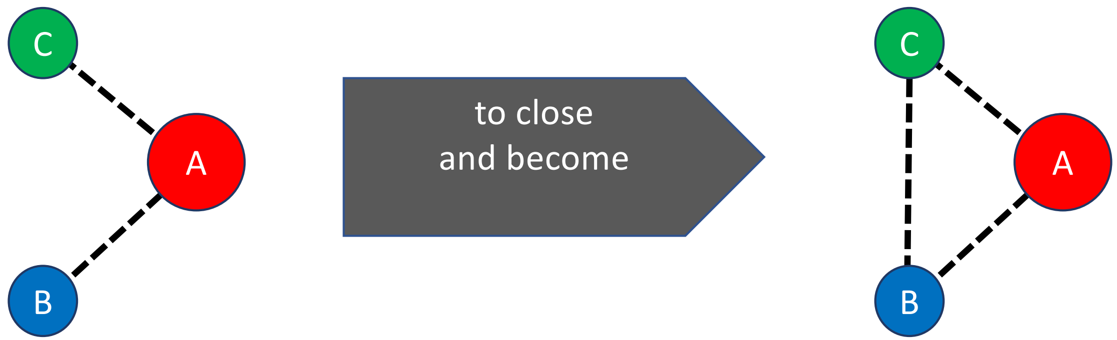


Strength of weak ties (Granovetter, 1973)

The same tendency does not apply to weak ties.

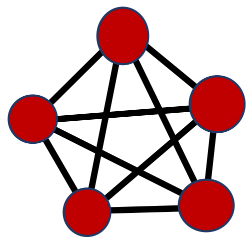
There is no tendency for
a weak two-path

A triangle

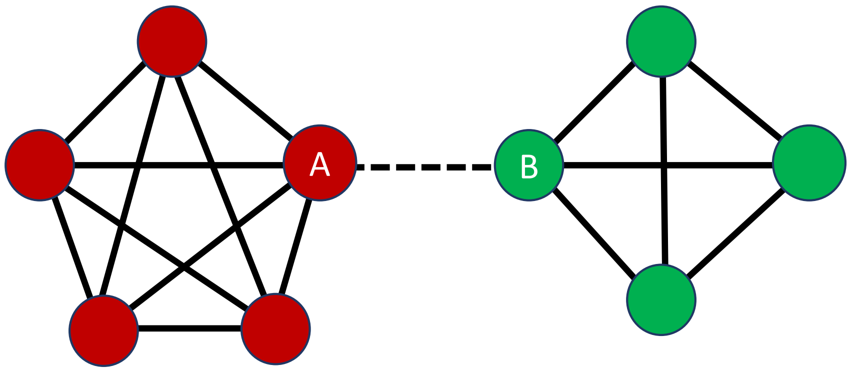


Strength of weak ties (Granovetter, 1973)

The result is that strong ties will tend to *cluster* into *cliques*, whereas weak ties will not.



The global structure will tend to be of cliques of strong ties, connected by weak ties

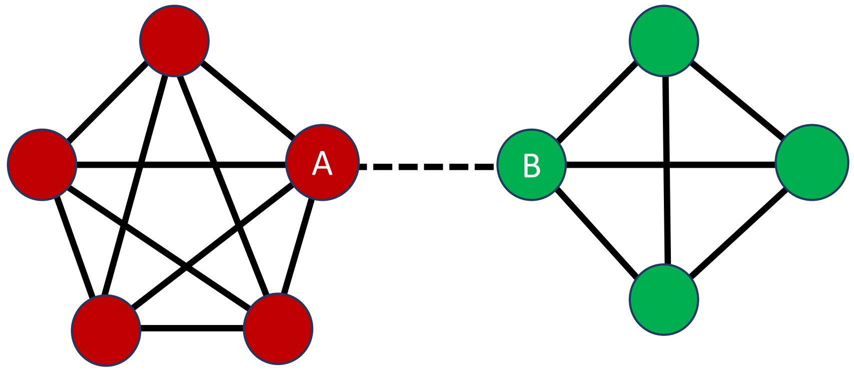


The A-B tie is a brige

Strength of weak ties (Granovetter, 1973)

Without the weak ties, the network is disconnected, information does not flow freely, and innovations are difficult to spread.

Within a clique of strong ties information spreads freely, so that such cliques cannot be a source of new information



Weak ties enable new information to come from outside the clique.

Individuals tend to use their weak ties to obtain new or different information. Weak ties are a form of *social capital.* Hence, *the strength of weak ties.*

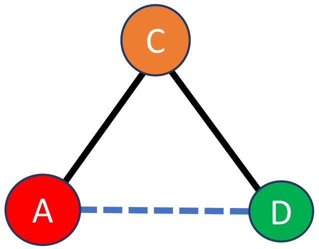
Trust

Triads and trust

Third party vouching facilitate trust

Trust is the opposite of assurance

High trust means I am comfortable with my friends sharing my secrets



Closure indicate lack of trust

Evidence of others interaction strengthens my trust (social learning)

Closed triangle makes being untrustworthy impossible

Bearman (1997): generalized exchange as a triadic pattern of network ties that does not involve immediate needs for reciprocation – instead of me returning a favour, a third party 'that I owe a favour' provides you with something.

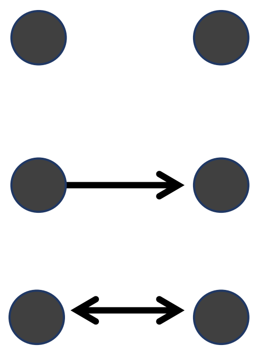
Closure can be seen as a means of enforcing norms and to enforce sanctions against antinormative behaviour (Coleman, 1988, Burt, 1995, 2000) – 'there is always a third party monitoring our interaction' - Coleman said: "reputation cannot arise in an open structure" (1988, S107).

Ties to common third parties also promote adherence to norms by promoting trust and by facilitating social monitoring and sanctioning of opportunistic behaviour (Burt & Knez, 1995).

Directed networks

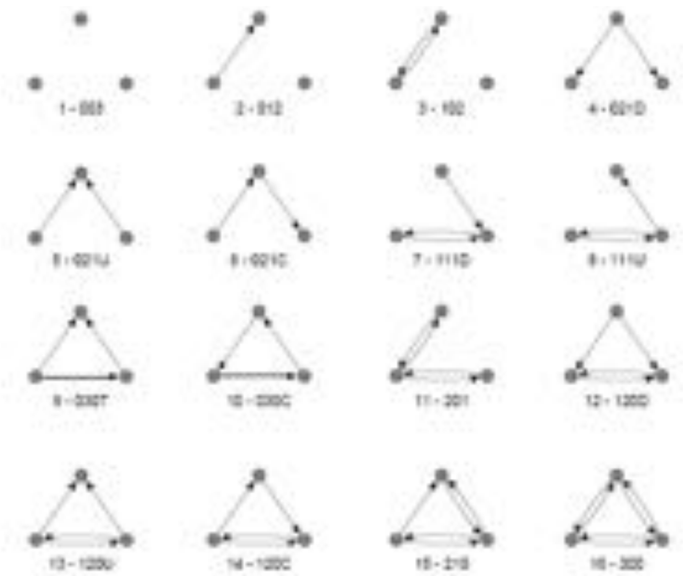
Dyadic recoprocity

Gouldner (1969): norm as a mechanism rather than reciprocity implied by function or complimentarity



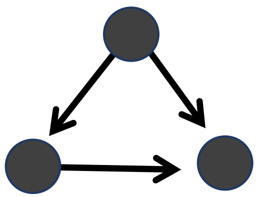
Katz & Wilson (1956) develop tests for detecting reciprocity relative to null-distribution

Triads in directed networks

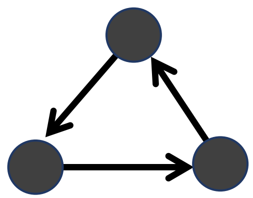


Triads in directed networks: hierarchy

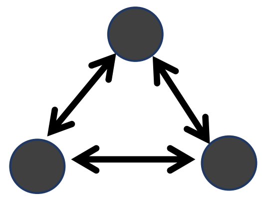
What does the direction tell us?



Transitive



cyclic

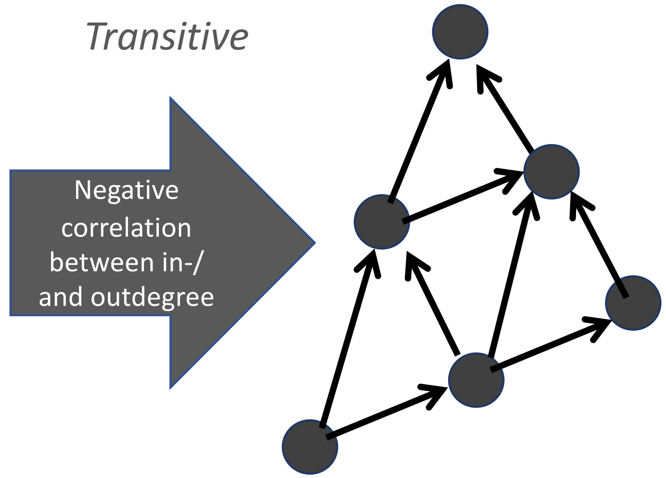


Simmilean tie (Krackhardt)

Admiration
Giving orders/managing
Best friend

Triads in directed networks: hierarchy

What does the direction tell us?



Admiration
Giving orders/managing
Best friend

People do not form ties at random

How do we test the prevalence of a mechanism?

Assume that there is no mechanism A

Let $T(X)$ be a statistic that is sensitive to A (e.g. larger the stronger A is)

Let $p(X)$ be a distribution that does not have A

Test

H_0: tie-formation is not driven by *A,* against

H_1: tie-formation is driven by A

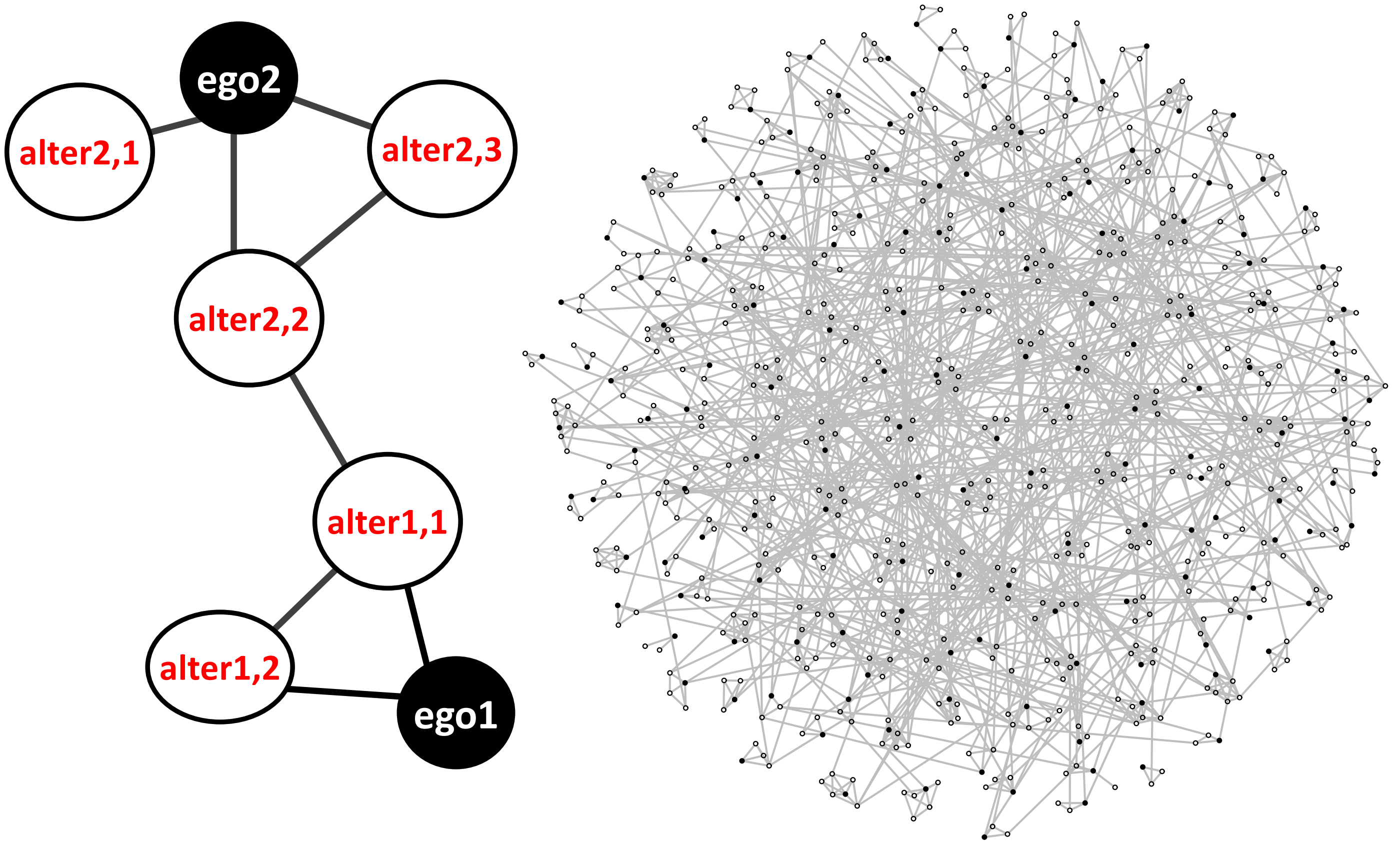
If $X \sim \mathrm{p}(X)$, H_0 is true, we can calculate the probability

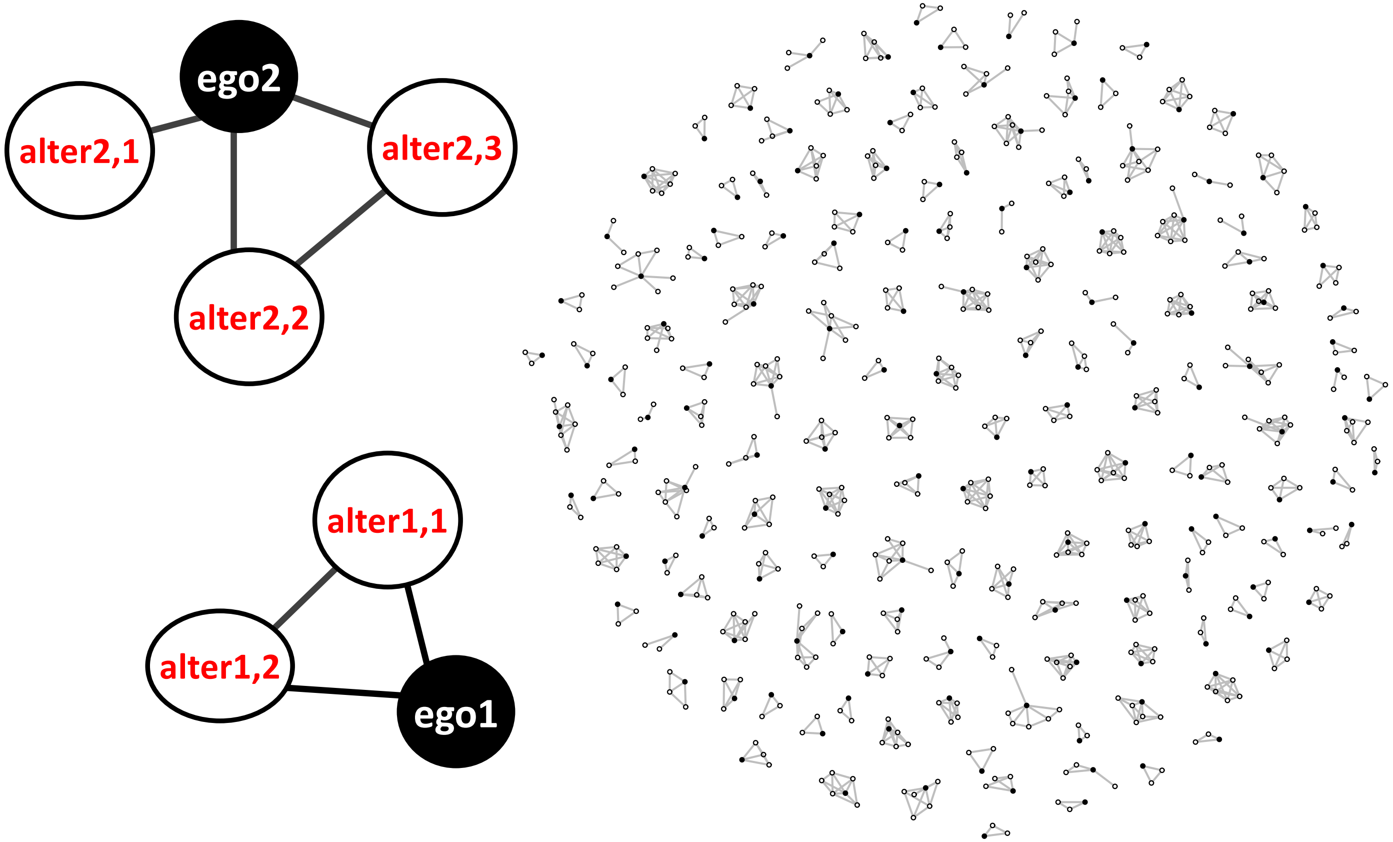
$$\Pr(T(X) \geq k)$$

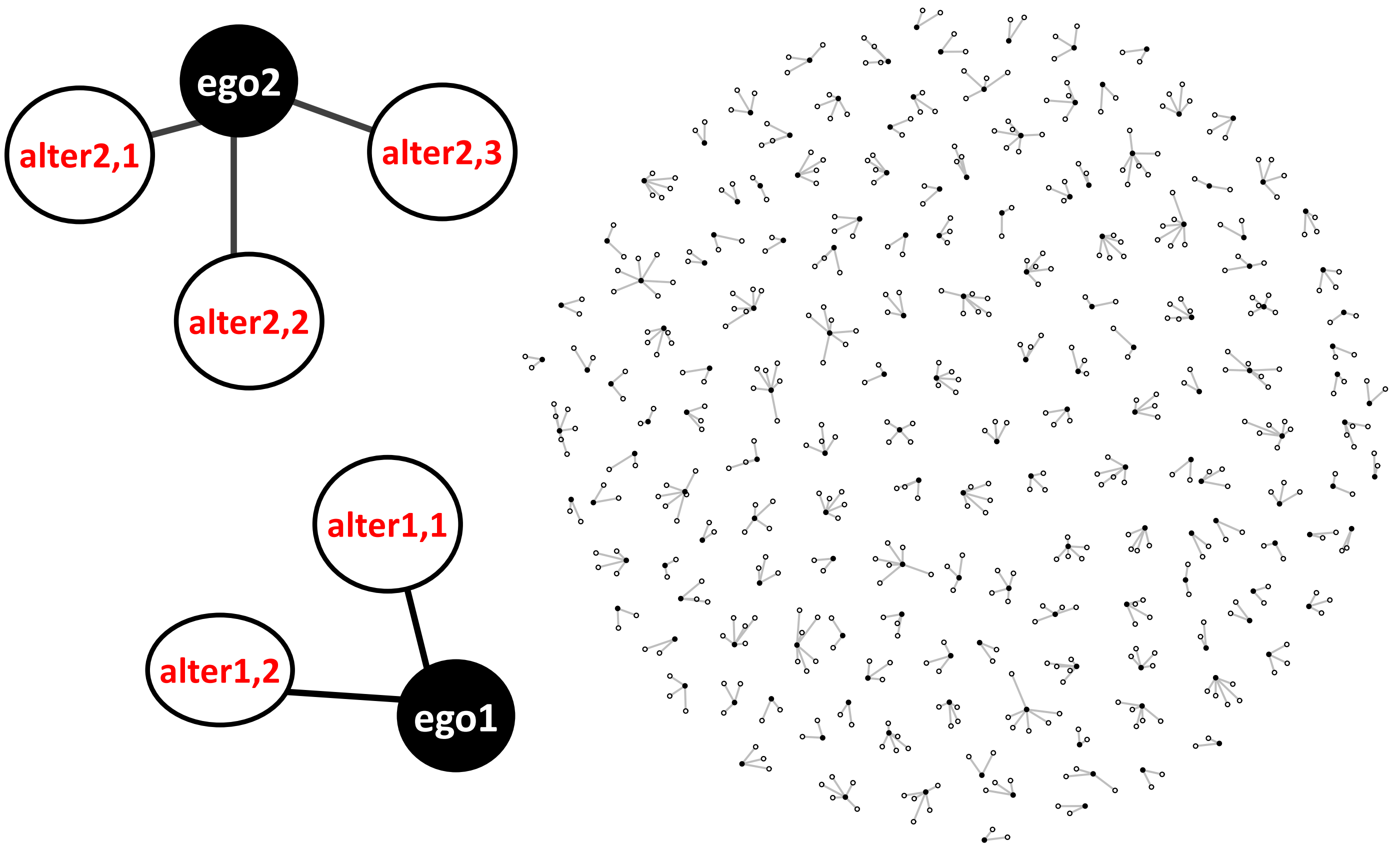
Reject H_0 on the α-level if

$$\Pr(T(X) \geq T(X_{\text{obs}})) < \alpha$$

Egonets


ego2
alter2,1
alter2,3
alter2,2
alter1,1
alter1,2
ego1

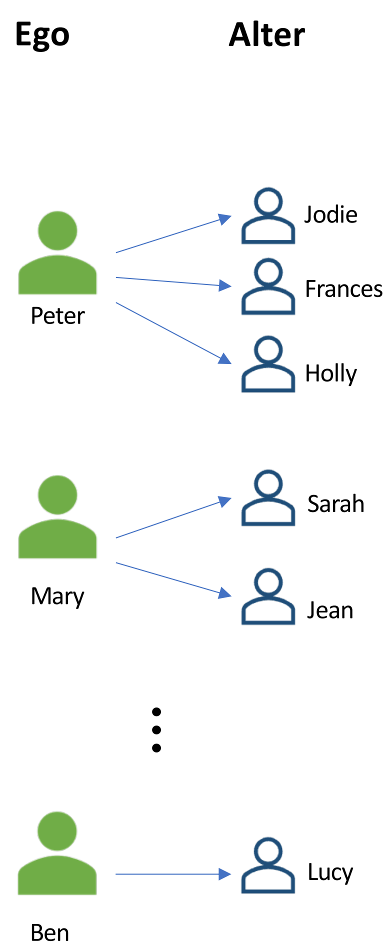

ego2
alter2,1
alter2,3
alter2,2
alter1,1
alter1,2
ego1


ego2
alter2,1
alter2,3
alter2,2
alter1,1
alter1,2
ego1

What is an egocentric network?

- We have n independent **respondents**
- We call each respondent **Ego**
- Each Ego ‘nominates’ network partners/contacts
- We call the partner of an Ego the **Alter** of Ego.

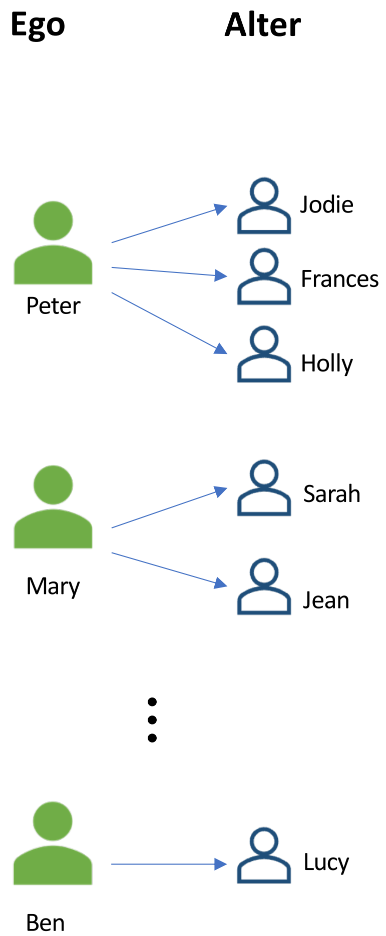
Id	Ego	Alter 1	Alter 2	Alter 3	…	Alter N
1	Peter	Jodie	Frances	Holly	…	
2	Mary	Sarah	Jean		…	
⋮	⋮	⋮	⋮	⋮	⋱	⋮
n	Ben	Lucy			…	



What can be say about the network?

What people connect the network?

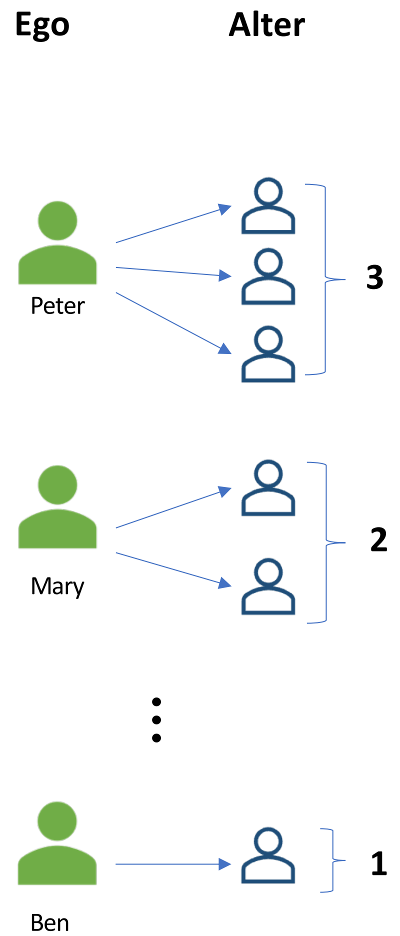
- Degree centrality
- As dependent variable
  - Do certain actors have more ‘friends’?
- As independent variable
  - Do actors with more social support have better mental health outcomes?

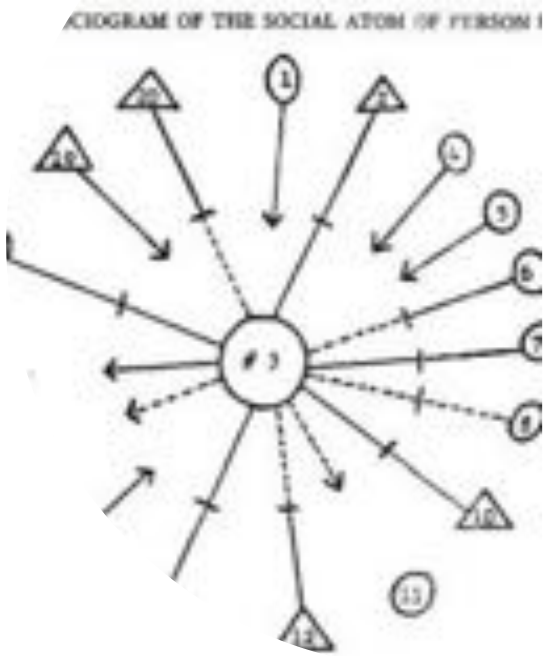


Analysing degree centrality

- For degree centrality we only need number of alters

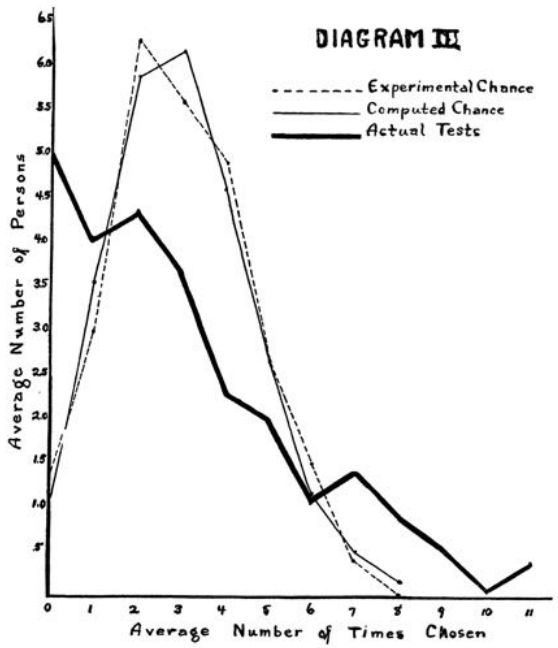
Id	Ego	Age ego	Neuroticism	SMI	Sex	degree
1	Peter	19	5	0	M	3
2	Mary	25	3	1	F	2
⋮	⋮	⋮	⋮	⋮	⋱	⋮
n	Ben	30	2	0	M	1





Skewed degree distributions

Moreno and Jennings (1938)



A greater concentration of many choices upon few individuals and of a weak concentration of few choices upon many individuals skews the distribution of the sampling still further than takes place in the chance experiments, and in a direction it need not necessarily take by chance. This feature of the distribution is an expression of the phenomenon which has been called the *socio-dynamic effect*. The chance distribution seen as a whole is also

Network science - Barabasi & Albert (1999)

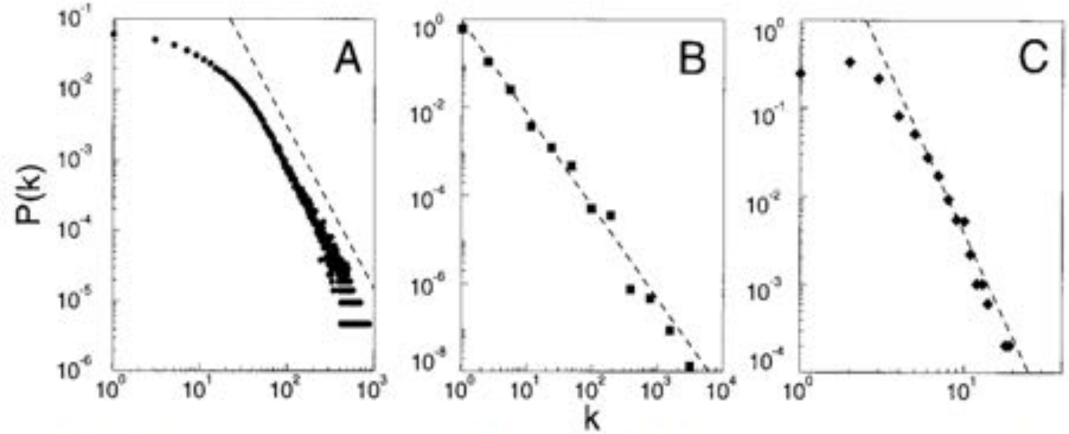


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with N = 212,250 vertices and average connectivity $\langle k \rangle$ = 28.78. (**B**) WWW, N = 325,729, $\langle k \rangle$ = 5.46 (6). (**C**) Power grid data, N = 4941, $\langle k \rangle$ = 2.67. The dashed lines have slopes (A) γ_{actor} = 2.3, (B) γ_{www} = 2.1 and (C) γ_{power} = 4.

Does a skewed degree distribution reflect network processes

- Moreno and Jennings (1938) – socio dynamic effect
- Merton (1968) - The Matthew Effect
- de Solla Price (1976) - cumulative advantage
- Albert & Barabási (2002) - preferential attachment leads to power-law

Critiques

- Pattison, Robins, Koskinen, 2008
- "Here again, the power-law statistics tells nothing about the critical or noncritical nature of the underlying system. ... More generally, these results also put caution on the interpretation of power-law relations found in nature." *Power-law statistics and universal scaling in the absence of criticality, Touboul and Destexhe, Phys. Rev. E 95, 012413 – Published 31 January 2017*
- "A central claim in modern network science is that real-world networks are typically 'scale free,' meaning that the fraction of nodes with degree k follows a power law, decaying like k^lambda, often with 2" *Broido, A. D., & Clauset, A. (2018). Scale-free networks are rare. arXiv preprint arXiv:1801.03400*

Despite critique power-laws pervasive

Implications for sexually transmitted diseases

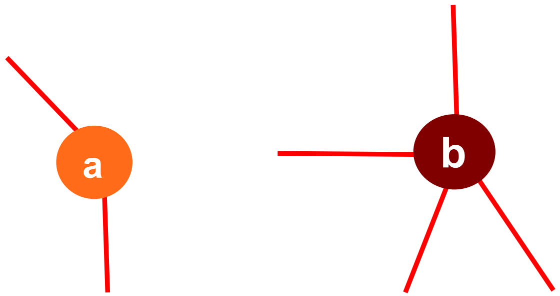
Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Åberg, Y. (2001). The web of human sexual contacts. *Nature*, *411*(6840), 907-908.

Brief review of social networks for disease transmission

Koskinen, J.H. (2018), Discussion of "Optimal treatment allocations in space and time for on-line control of an emerging infectious disease" by Laber, N. J. Meyer, B. J. Reich, K. Pacifici, J. A. Collazo and J. Drake, J.R.Statist.Soc. C 67, 779. (pre-print: arXiv:2006.16527)

Figure 1 It's a small world: social networks have small average path lengths between connections and show a large degree of clustering. Painting by Idahlia Stanley.

If we are only interested in degree distributions



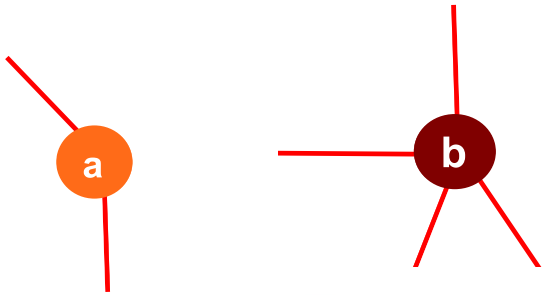
We only only need survey data:

a -> # nominations

b -> # nominations

…

If we are only interested in degree distributions



Pr(D=k) ∝ $k^{-\gamma}$

for k > c

…

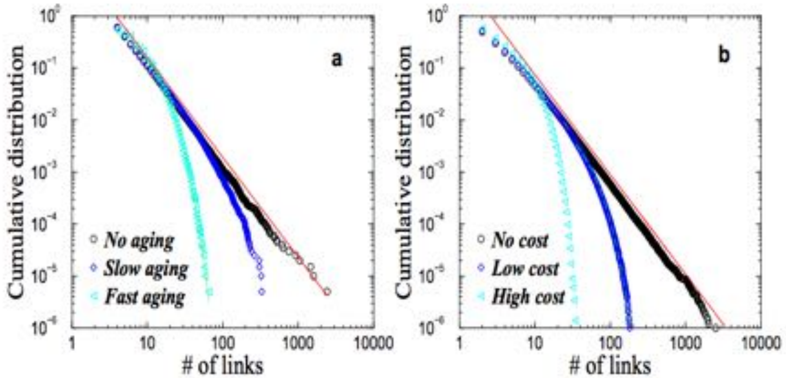
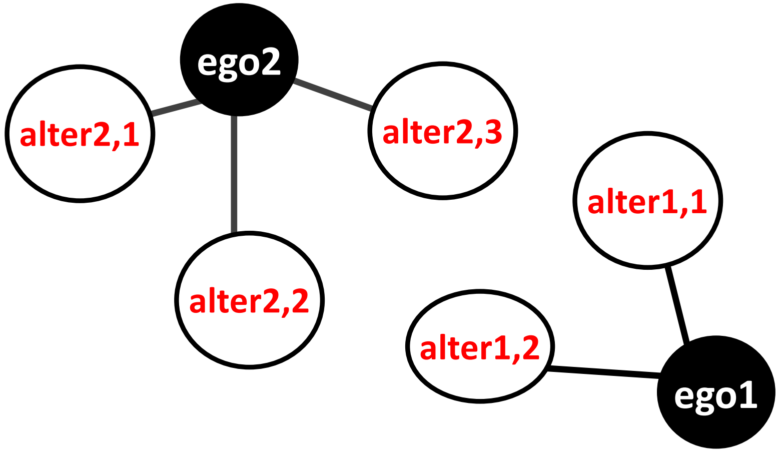


FIG. 27. Deviation from a power-law of the degree distribution due to adding age (a) and capacity (b) constraints to the SF model. The constraints result in cutoffs of the power-law scaling. After Amaral *et al.* (2000).

If we are only interested in degree distributions

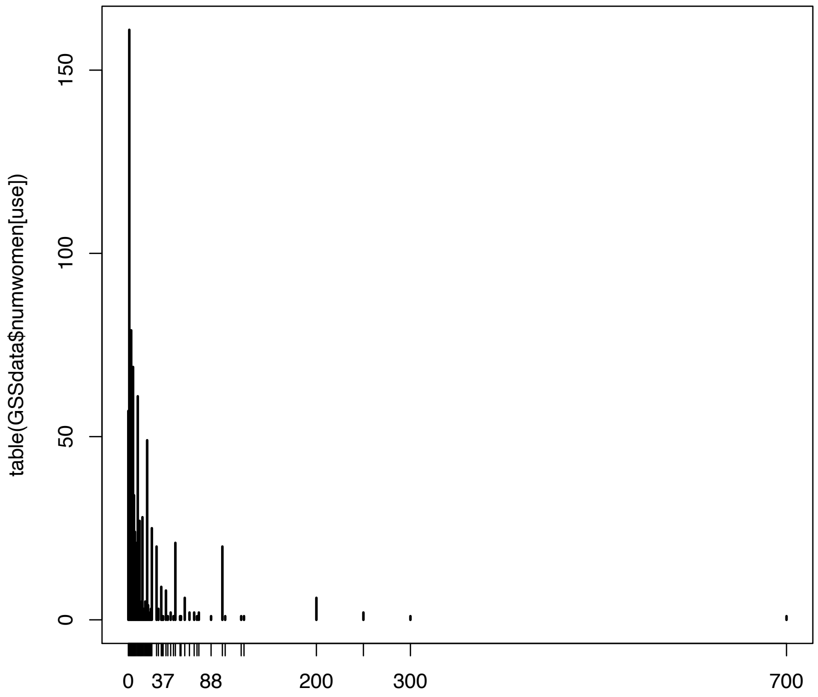


numwomen

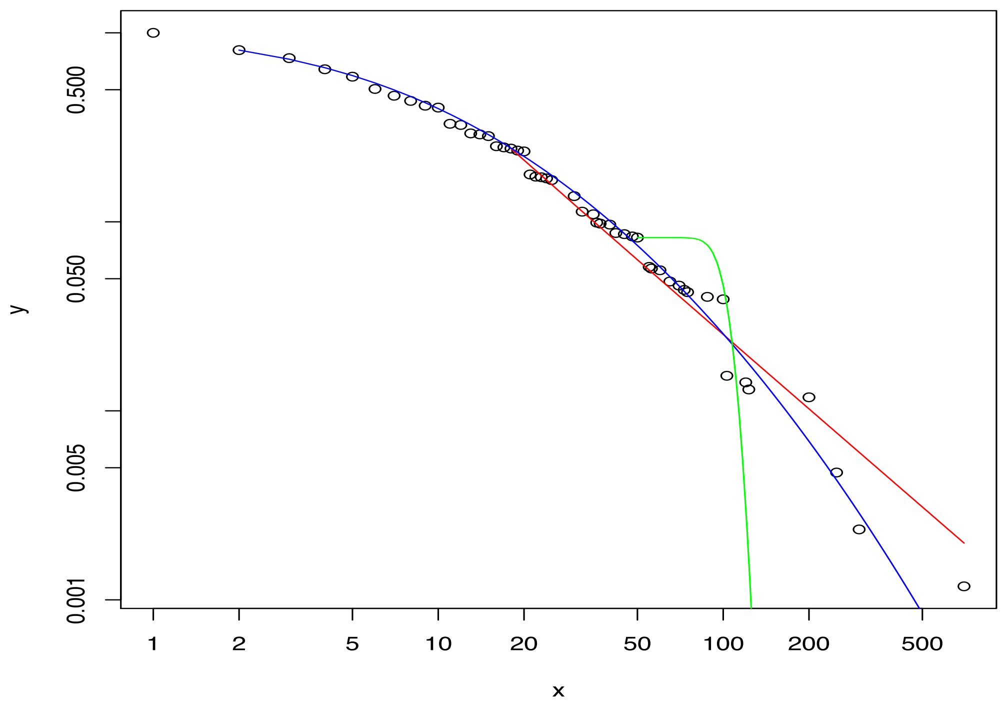
http://www.thearda.com/Archive/Files/Codebooks/GSS2004_CB.asp

Figure 1 It's a small world: social networks have small average path lengths between connections and show a large degree of clustering. Painting by Idahlia Stanley.

Look at the (reported) number of sex-partners in the General Social Survey



Different fitted distributions (which one is which?)

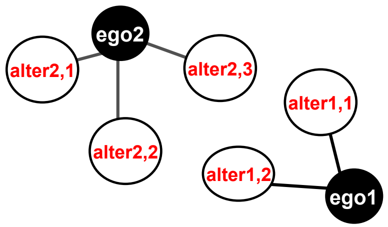



Alternative to power-law? Normal
Blue line says ‘lognormal’
Let’s do it ourselves
start with normal

Number of sex-partners on attitude to premarital sex

If we are only interested in degree distributions

```
lm( GSSdata$numwomen[use] ~ GSSdata$premarsx[use])
```



numwomen

$$Y_i | x_i \sim N(\alpha + \beta^{\top} x_i, \sigma^2)$$

or equivalently

$$Y_i = \alpha + \beta x_i + \varepsilon_i,$$

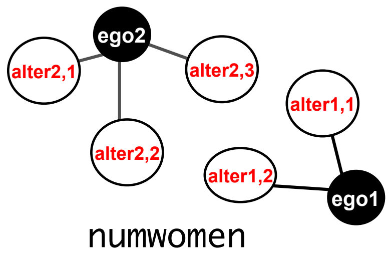
independently

$$\varepsilon_i \sim N(0, \sigma^2)$$

Regression

If we are only interested in degree distributions

```
lm( log(GSSdata$numwomen[use & grThanz]) ~
GSSdata$premarsx[use & grThanz])
```



If Y is log-normal

log(Y) is Normal

Alternative to power-law? Poisson

The **Poission** distribution has one parameter: $\lambda>0$

$$P(Y=y) = e^{-\lambda} \frac{\lambda^{y}}{y!}$$

parameter:

$E(Y) = \lambda$

$V(Y) = \lambda$

Regressing mean on covariates:

$$\log\{E(Y_i|x_i)\} = \log(\lambda_i) = \alpha + \beta^{\top}x_i$$

Alternative to power-law? Poisson

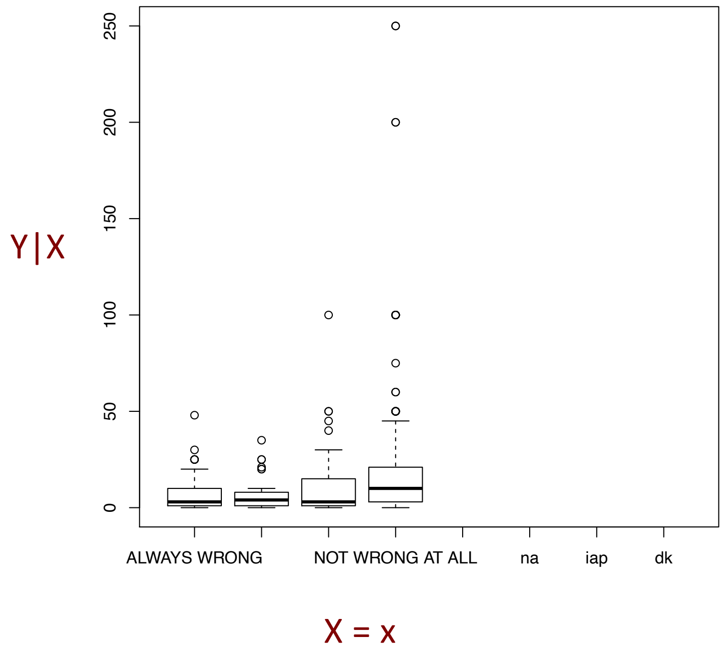
The **Poission** distribution has one parameter: $\lambda>0$

$$P(Y=y) = e^{-\lambda} \frac{\lambda^{y}}{y!}$$

parameter:

$$E(Y|X) = \lambda$$

$$V(Y|X) = \lambda$$



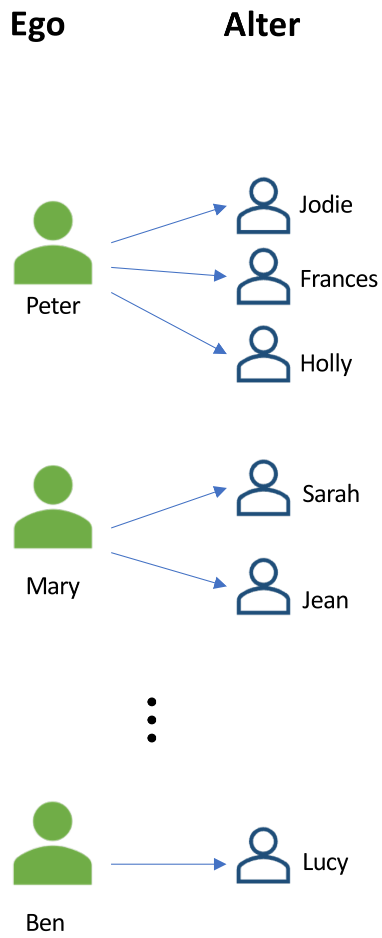
Ego-alter ties

Wide format – long format

Id	Ego	Alter 1	Alter 2	Alter 3	...	Alter N
1	Peter	Jodie	Frances	Holly	...	
2	Mary	Sarah	Jean		...	
⋮	⋮	⋮	⋮	⋮	⋱	⋮
n	Ben	Lucy			...	

Wide format

Each alter is a ‘variable’ of ego

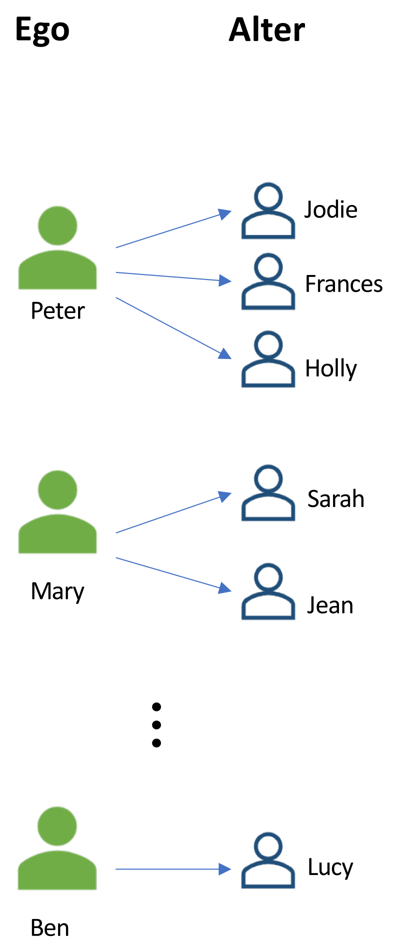


Wide format – long format

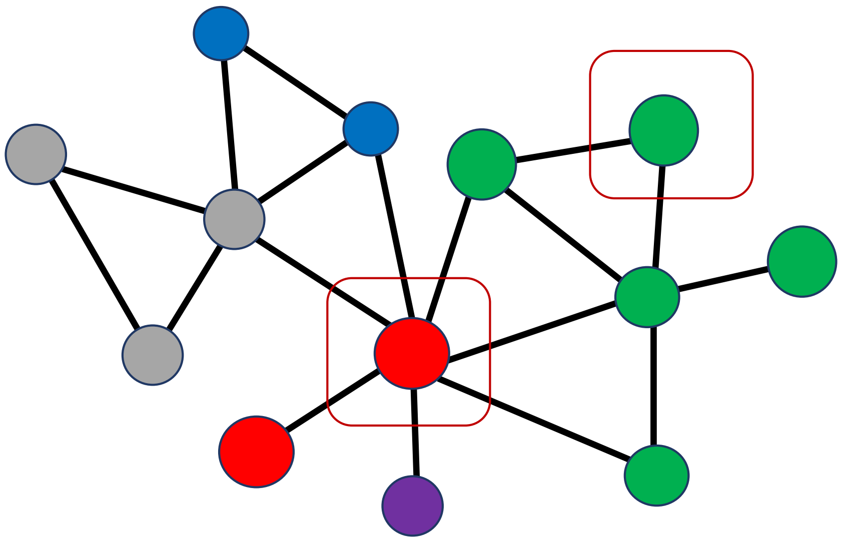
Id	Ego	Alter	Ego Age	Alter Age
1	Peter	Jodie	19	19
1	Peter	Frances	19	23
1	Peter	Holly	19	21
2	Mary	Sarah	25	27
2	Mary	Jean	25	35
⋮	⋮	⋮	⋮	⋮
n	Ben	Lucy	30	29

Long format

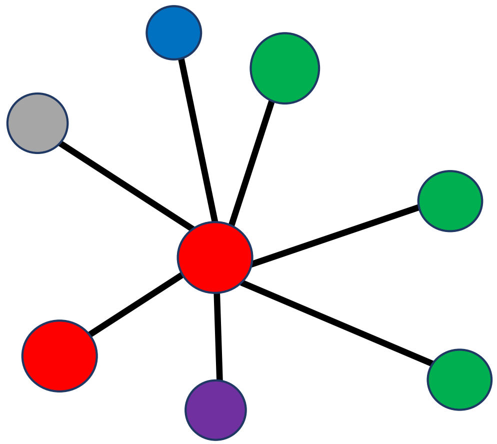
Each ego-alter pair is an observation



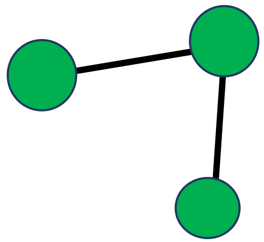
Homophily and heterophily



An actor that has a varied network is more likely to connect different people



Than an actor that only knows similar others



Analysing homophily in wide and long formats

Id	Ego	Alter	Ego Age	Alter Age
1	Peter	Jodie	19	19
1	Peter	Frances	19	23
1	Peter	Holly	19	21
2	Mary	Sarah	25	27
2	Mary	Jean	25	35
⋮	⋮	⋮	⋮	⋮
n	Ben	Lucy	30	29

Y_{ij} : age Alter j of ego i

Ego-alter ties are nested within egos

- Variation across egos
- Variation within egos

Y_i : average age of alters of ego i

Id	Ego		Age A1	Age A2	Age A3	…	Age AN
1	Peter	…	19	23	21	…	
2	Mary	…	27	35		…	
⋮	⋮	⋱	⋮	⋮	⋮	⋱	⋮
n	Ben	…	39			…	

Ave age
21
31
⋮
39

Analysing homophily in wide and long formats

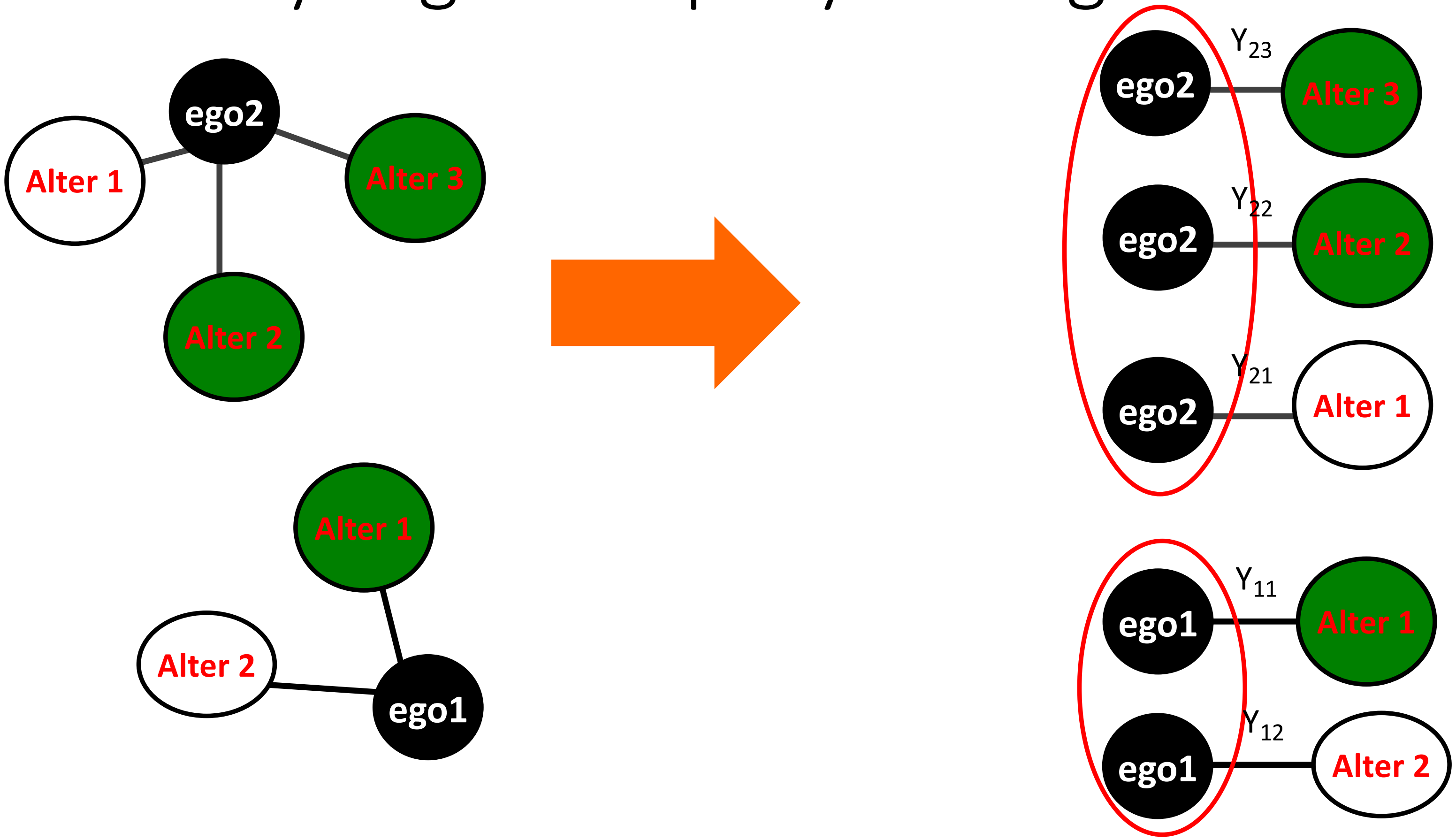
Id	Ego	Alter	Ego Age	Alter Age
1	Peter	Jodie	19	19
1	Peter	Frances	19	23
1	Peter	Holly	19	21
2	Mary	Sarah	25	27
2	Mary	Jean	25	35
⋮	⋮	⋮	⋮	⋮
n	Ben	Lucy	30	29

Y_{ij} : age Alter j of ego i

Ego-alter ties are nested within egos

- Variation across egos
- Variation within egos

Analysing homophily in long format



Ego-alter ties are nested within egos

- Variation across egos
- Variation within egos

Analysing homophily in long format

Ego-alter ties are nested within egos

Regression

$$Y_{ij} = \alpha + u_i + \beta x_i + \varepsilon_{ij}$$

Variation across egos

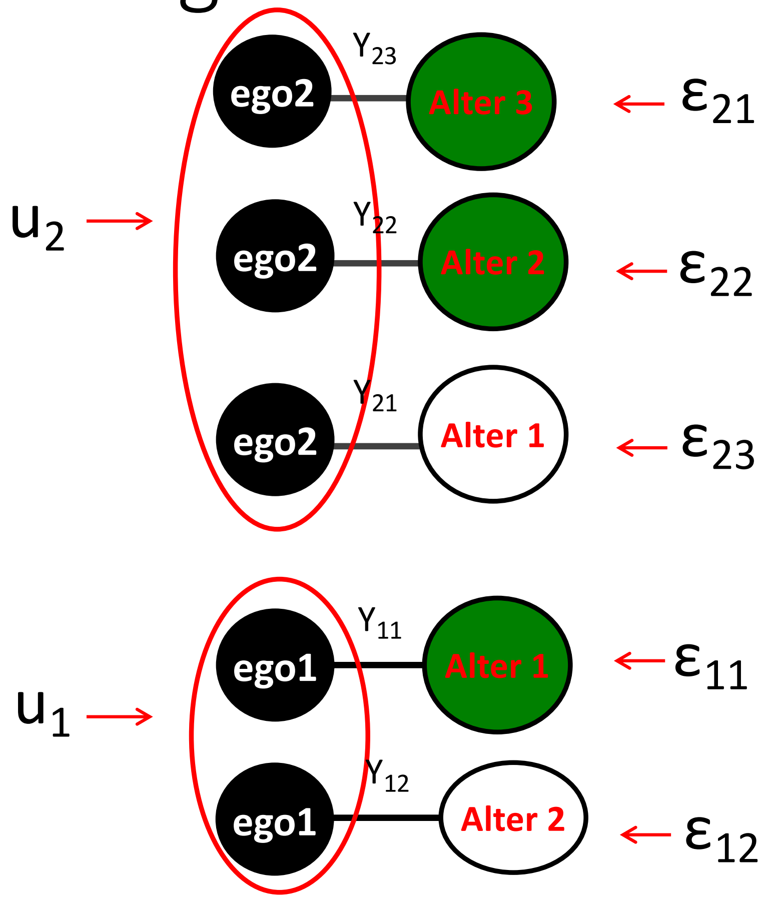
Variation within egos

Ego-specific error/variation

$$u_i \sim N(0,\tau^2)$$

alter-specific error/variation

$$\varepsilon_{ij} \sim N(0,\sigma^2)$$



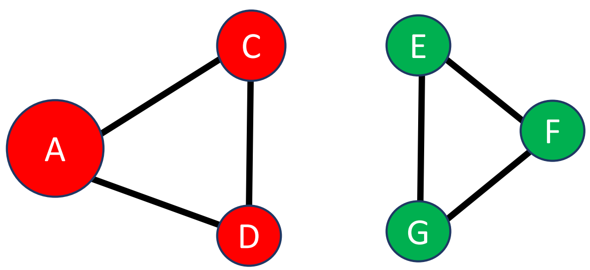
Network regression

Homophily

“Birds of a feather flock together”

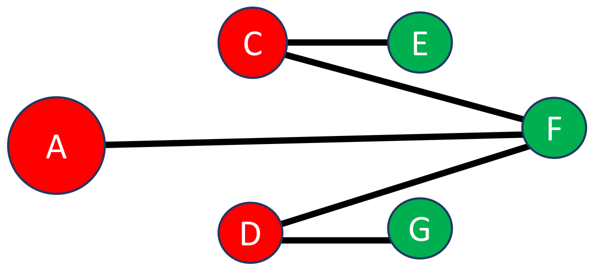
Homophilous/heterophilous dyads

Homophily



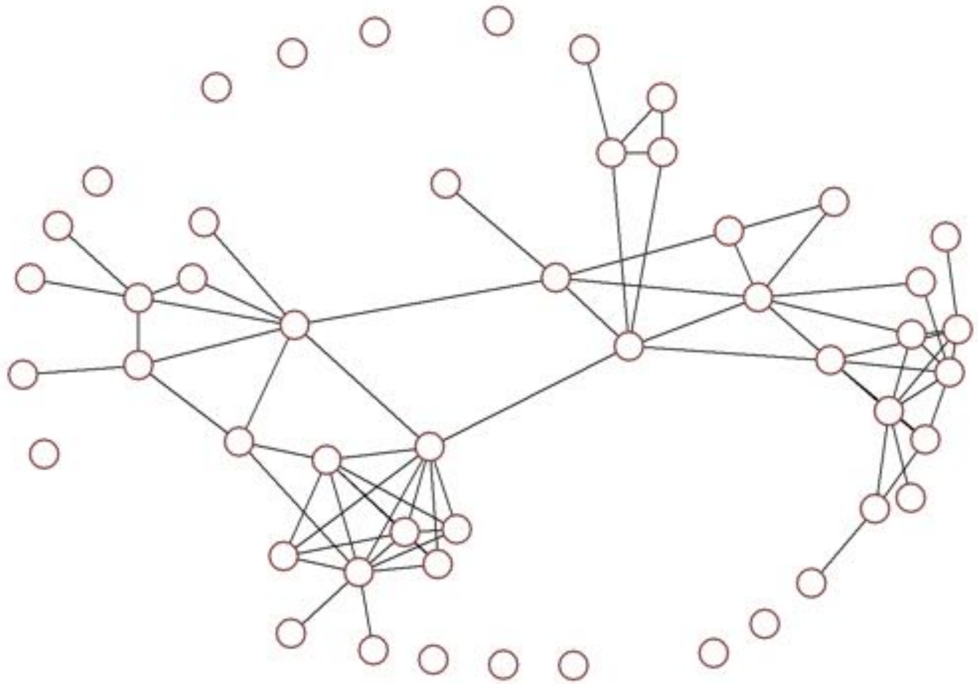
A, C, and D like sports *E, G, and F like the ballet*

Heterophily

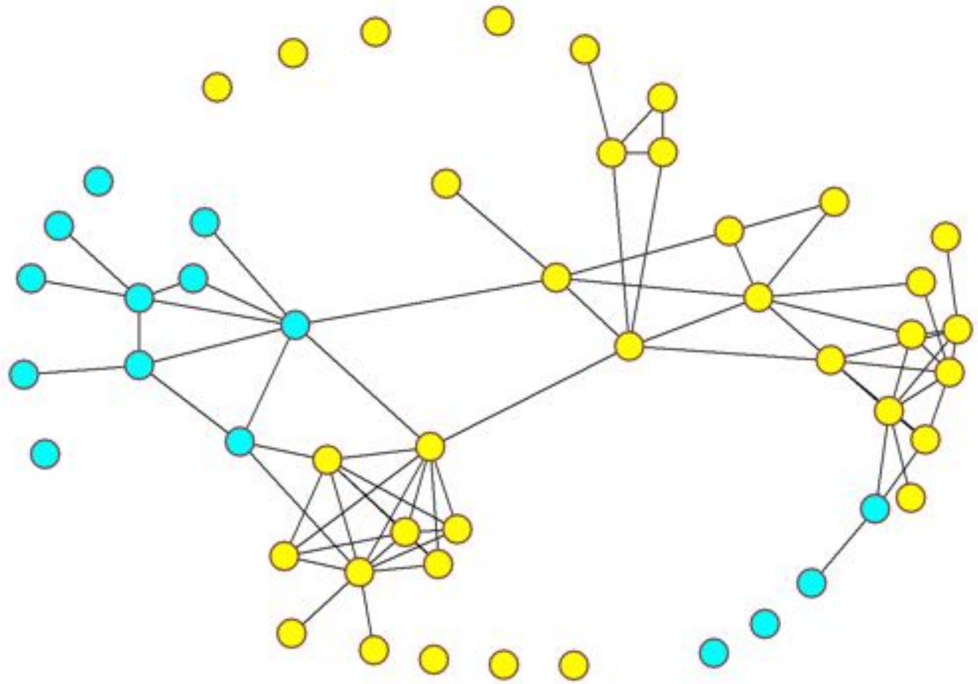


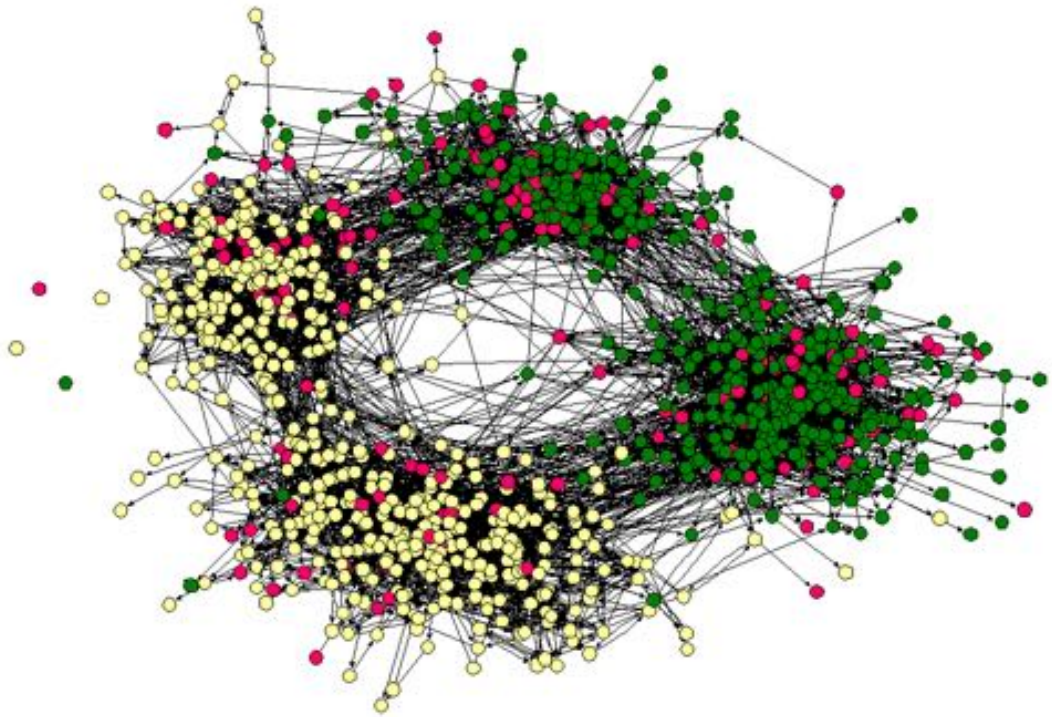
A, C, and D produce *E, G, and F consume*

Homophilous/heterophilous dyad



Homophilous/heterophilous dyad

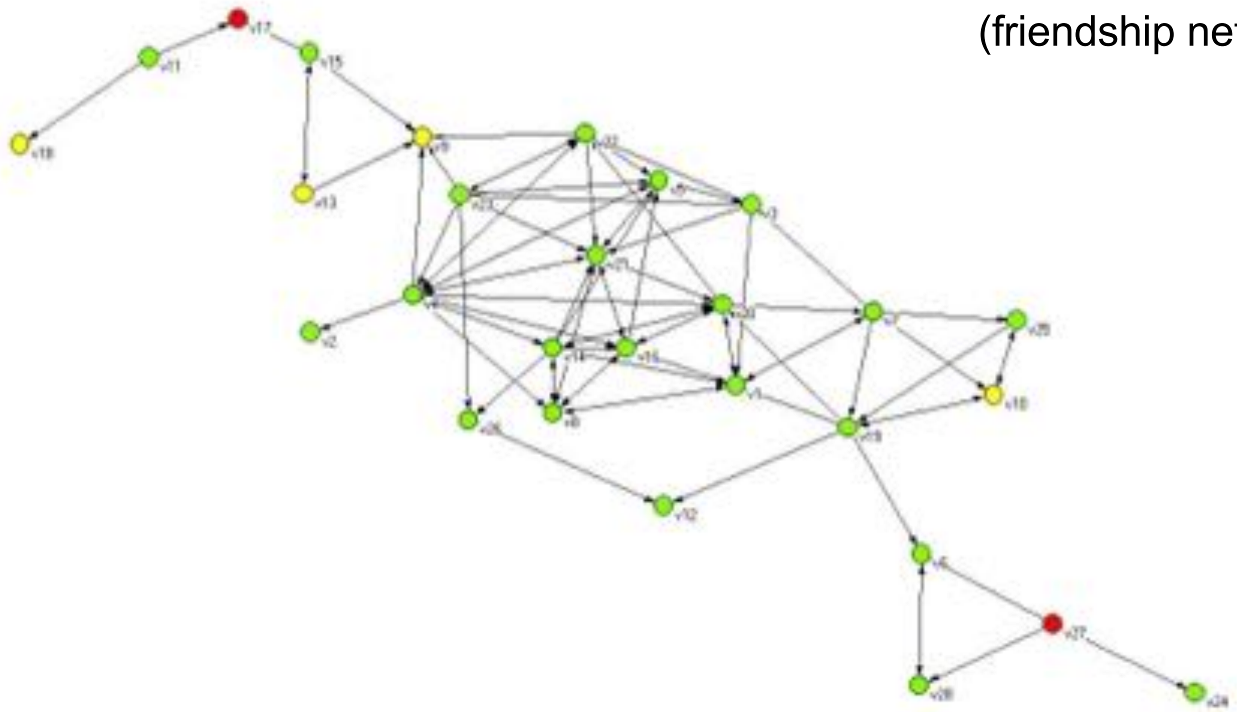




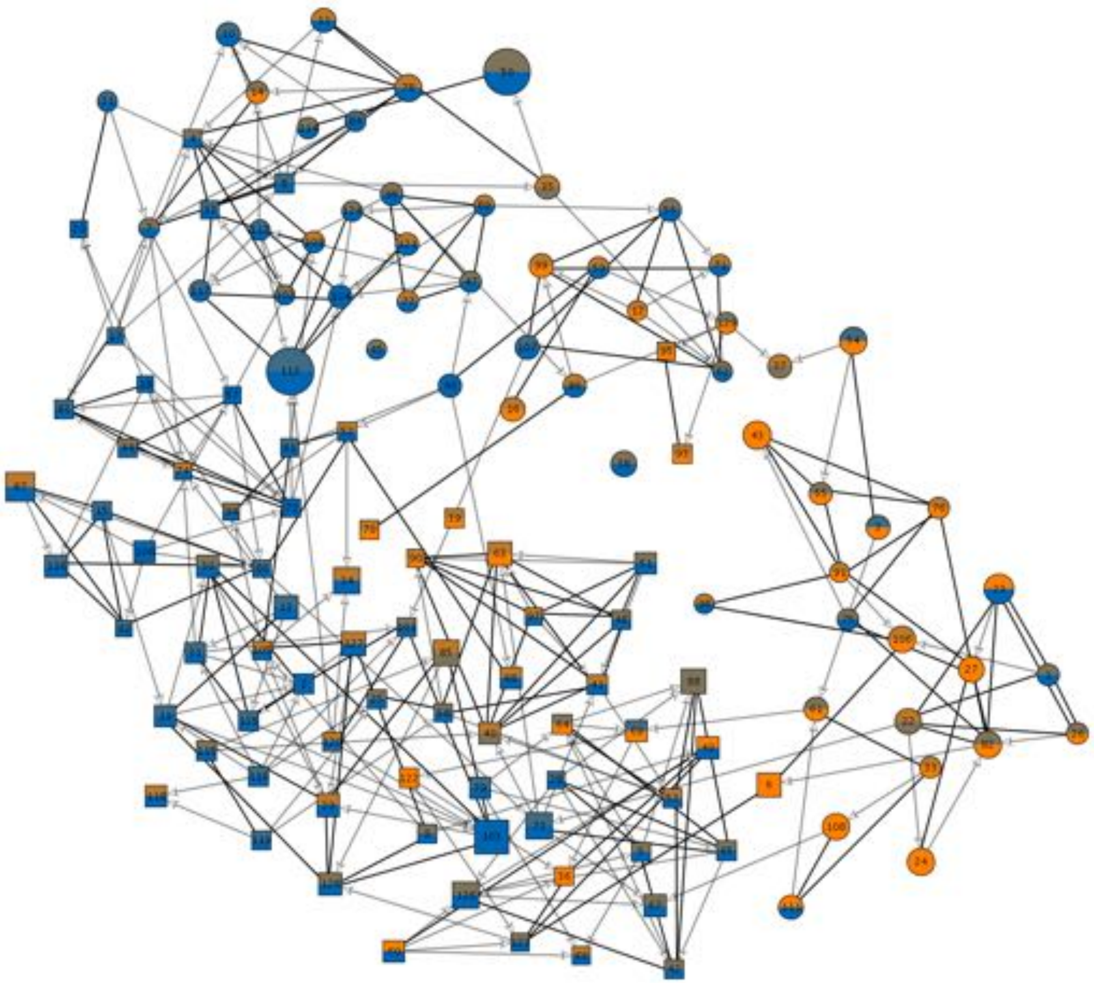
white black other

Team structures in training squads (Pane, 2003)

(friendship network in 12th week of training)



detached team oriented positive



example: friendship and substance abuse adolescents in Glasgow

EDGES

friend →

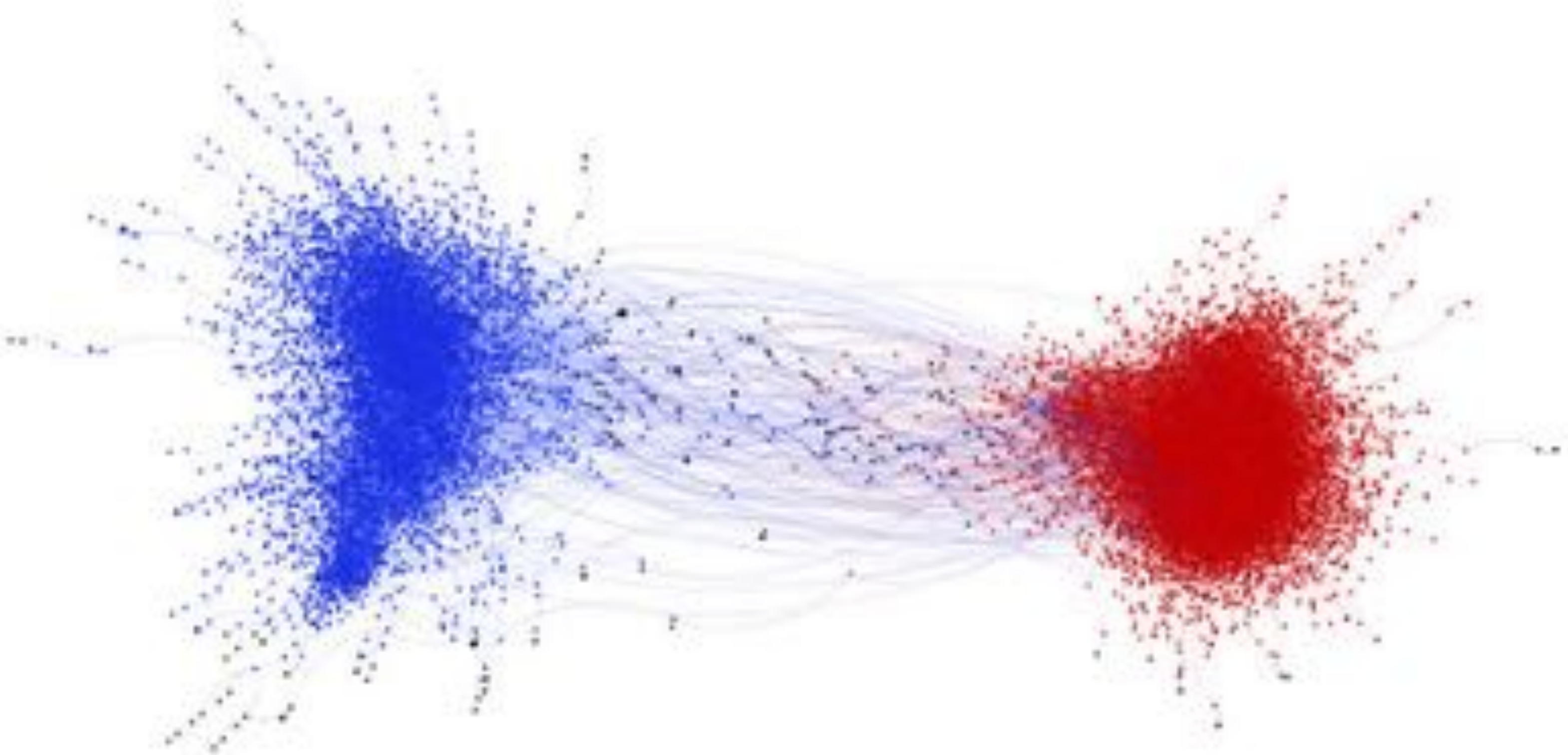
reciprocal —

NODES

girl ○

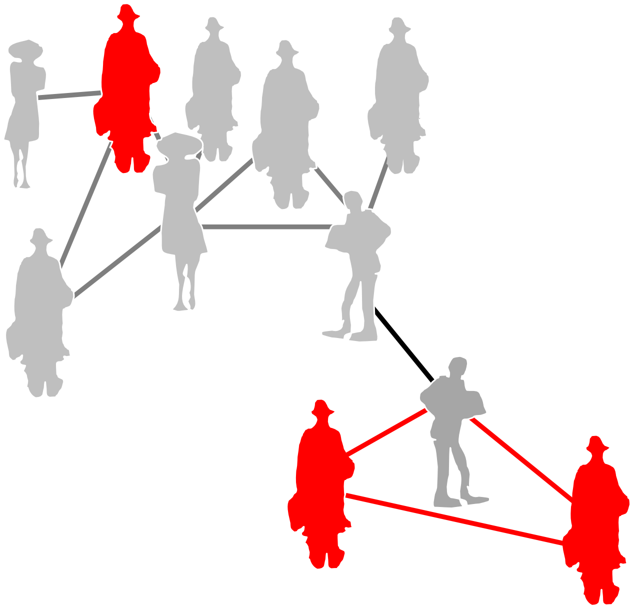
Boy □

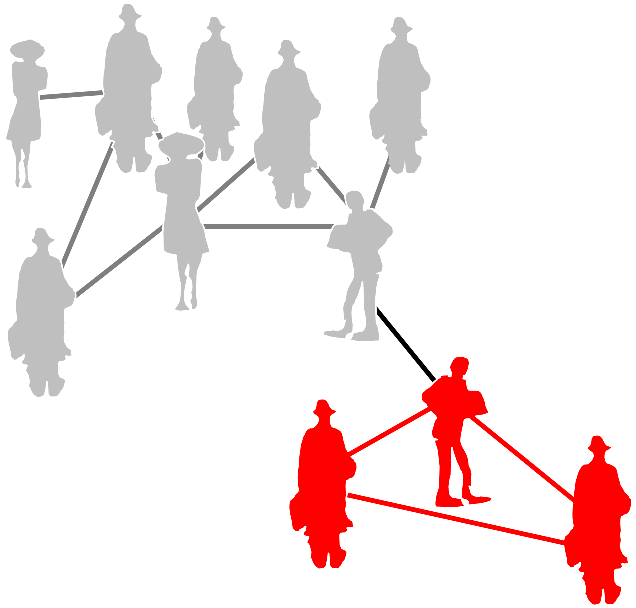
	Little	Lot
smoke		
drink		



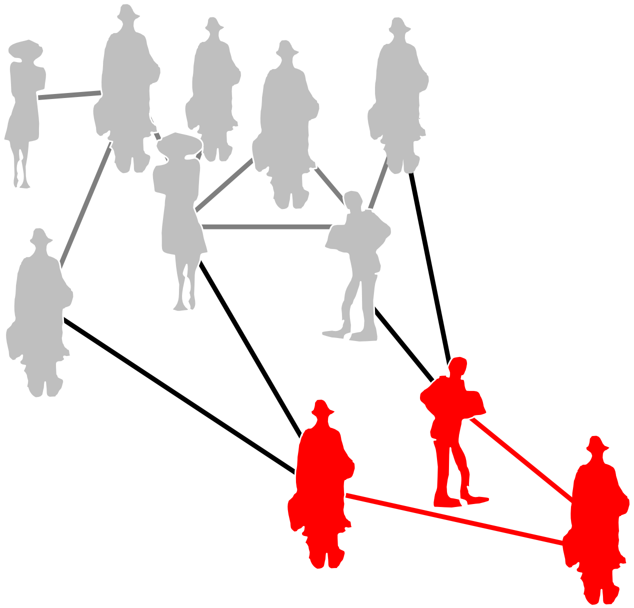
Brady et al.(2017), PNAS

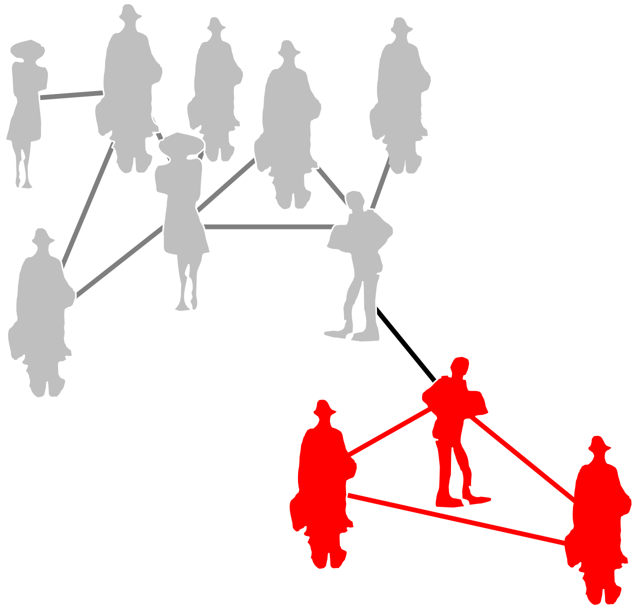
Social influence





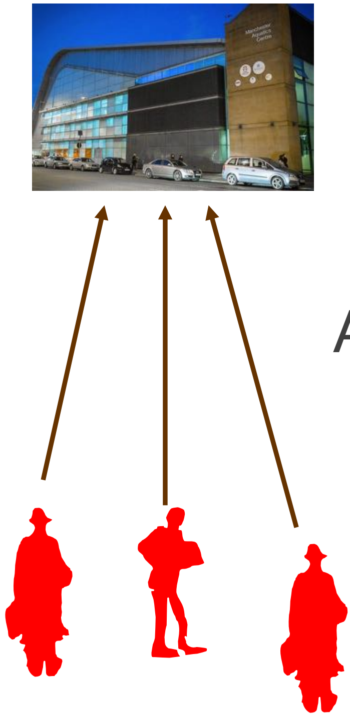
Social selection





Propinquity, context, and foci

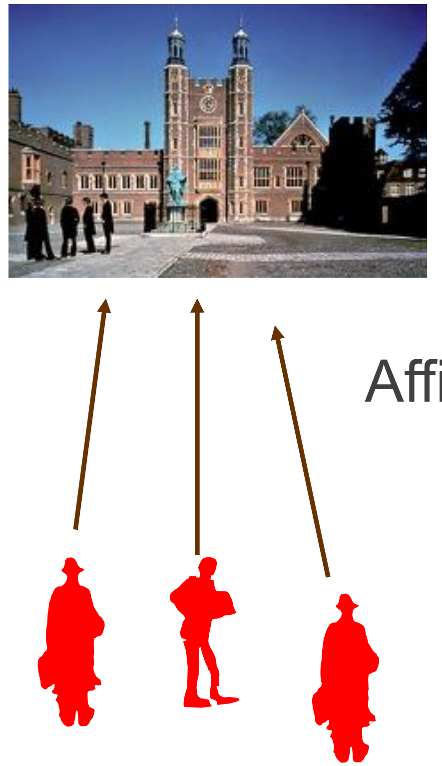
“Feld’s (1981, 1982, 1984) argument that focused activity puts people into contact with one another to foster the formation of personal relationships“
MCPHERSON et al (2001:431)



Affiliated with same things

Same attributes, same activities, form ties

Propinquity, context, and foci

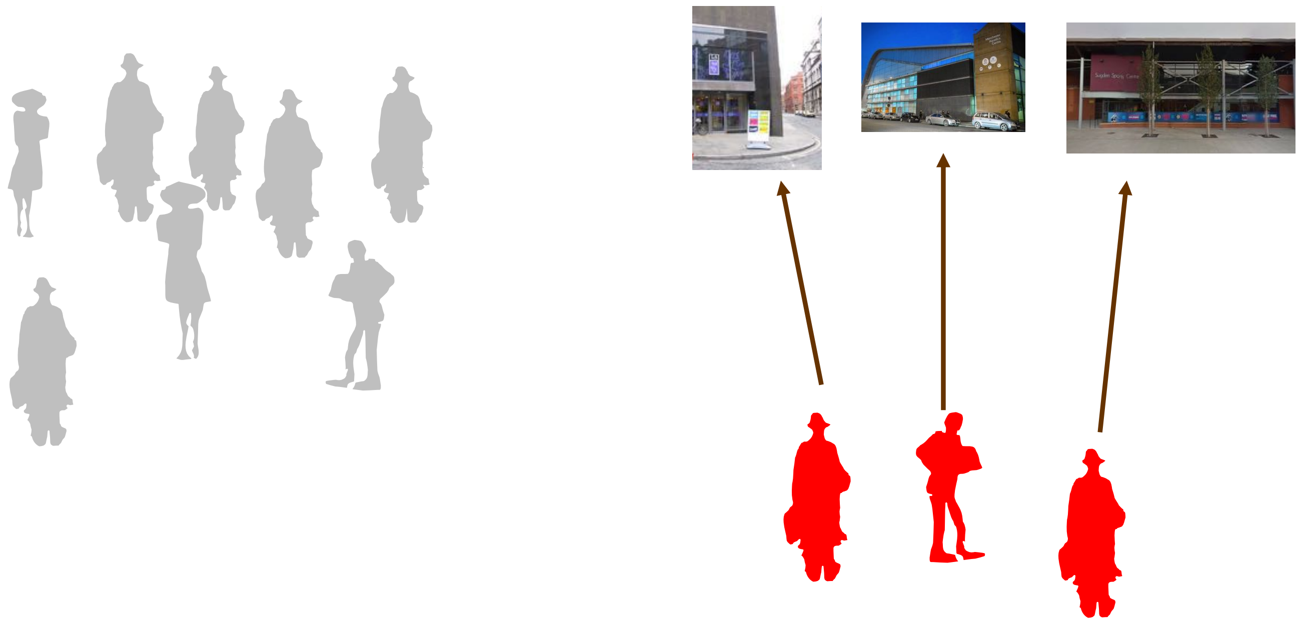


Affiliated with same things

Same attributes, same activities, form ties

Propinquity, context, and foci

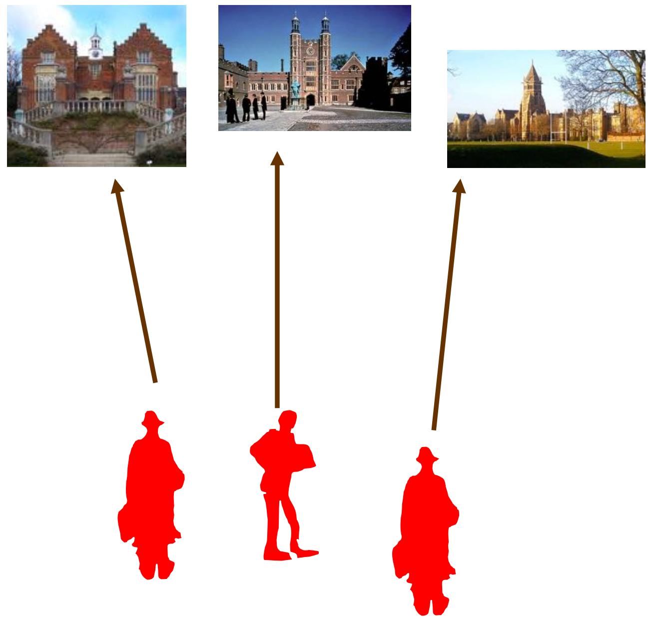
Affiliated with similar (not the same) things



same type of activities leads to same attributes form ties

Propinquity, context, and foci

Affiliated with similar (not the same) things



same type of activities leads to same attributes form ties


LET THERE BE EGGS!


LET THERE BE CHICKENS!

Mechanisms behind social influence

Social influence	**factors**
• Persuasion	✓ Active persuasion?
• Imitation	✓ Who is being influenced
• Diffusion	✓ Is the attribute salient or hidden?
• Adoption	✓ Is it an absorbing state?
• Contagion	✓ Number of contacts or majority of contacts
• Learning	✓ Is variable a single thing or many?
• etc	✓ do you seek it out or is it simply transmitted?

History


Jacob Moreno

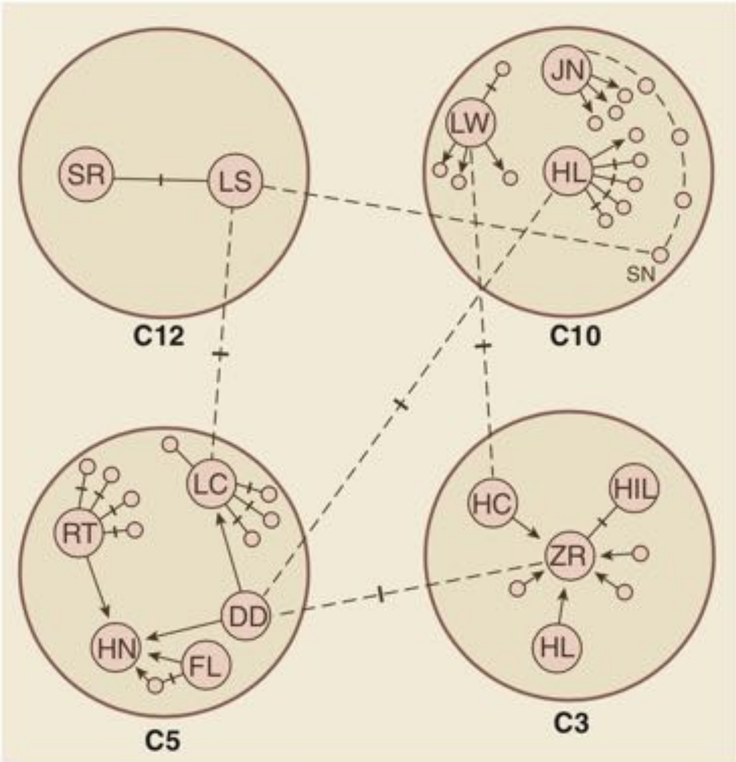
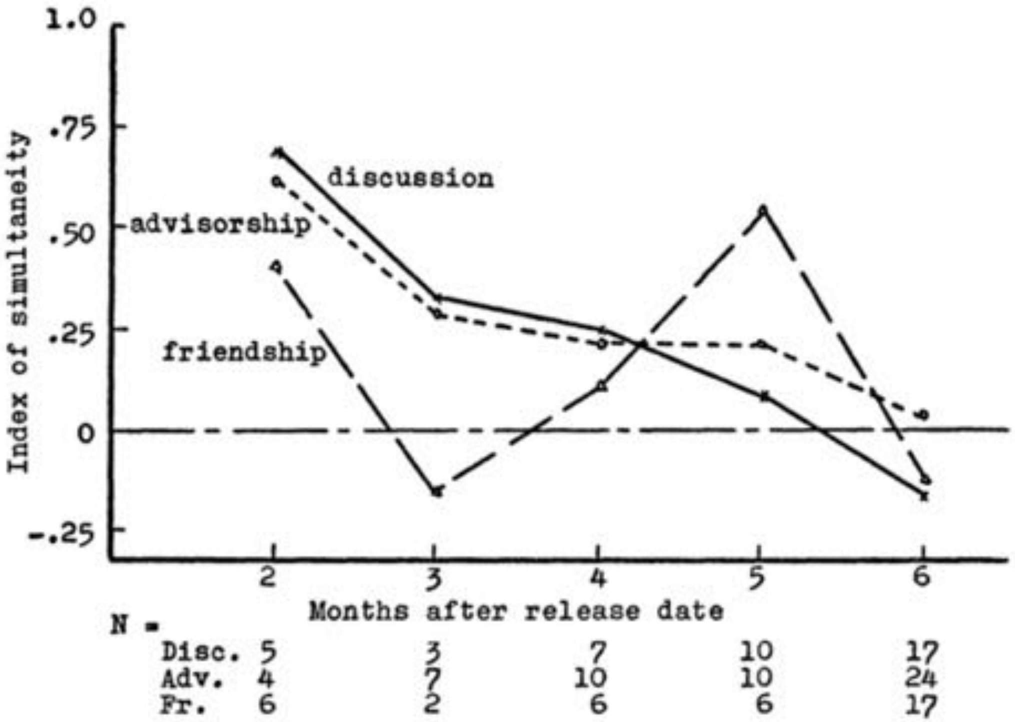


Fig. 1. Moreno's network of runaways. The four largest circles (C12, C10, C5, C3) represent cottages in which the girls lived. Each of the circles within the cottages represents an individual girl. The 14 runaways are identified by initials (e.g., SR). All nondirected lines between a pair of individuals represent feelings of mutual attraction. Directed lines represent one-way feelings of attraction.

feelings toward one another (Fig. 1). The links in this social network, Moreno argued, provided channels for the flow of social influence and ideas among the girls. In a way that even the girls themselves may not have been conscious of, it was their location in the social network that determined whether and when they ran away.

Coleman, Katz, and Menzel (1957)

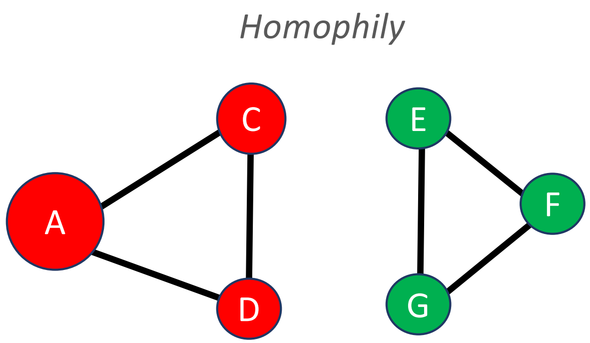


The Diffusion of an Innovation Among Physicians, Sociometry

To analyse pairs of individuals instead of single individuals may seem like only a very modest step in the direction of the analysis of networks of social relations. And so it is; it would be more satisfactory, and truer to the complexity of actual events, if it were possible to use longer chains and more ramified systems of social relations as the units of analysis. But so little developed are the methods for the analysis of social processes, that it seemed best to be content with the analysis of pair relationships”
Coleman, Katz and Menzel 1966, p 114

Investigating prevalence of influence

Statistic



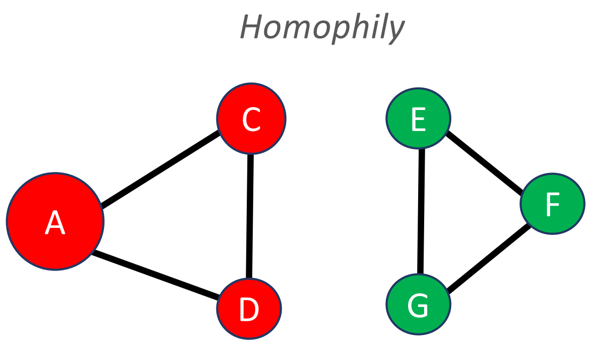
$$Y_i = \begin{cases} 1 & \text{, if } i \text{ is a smoker} \\ 0 & \text{,otherwise} \end{cases}$$

$$X_{ij} = \begin{cases} 1 & \text{, if } i \text{ has a tie to } j \\ 0 & \text{,otherwise} \end{cases}$$

Number of pairs of nodes that are connected and are smokers $\sum_j y_i y_j x_{ij}$

Number of pairs of nodes that are connected have same smoking status $\sum_j I\{y_i = y_j\} x_{ij}$

Statistic



Y_i : *Trust in government (on a scale)*

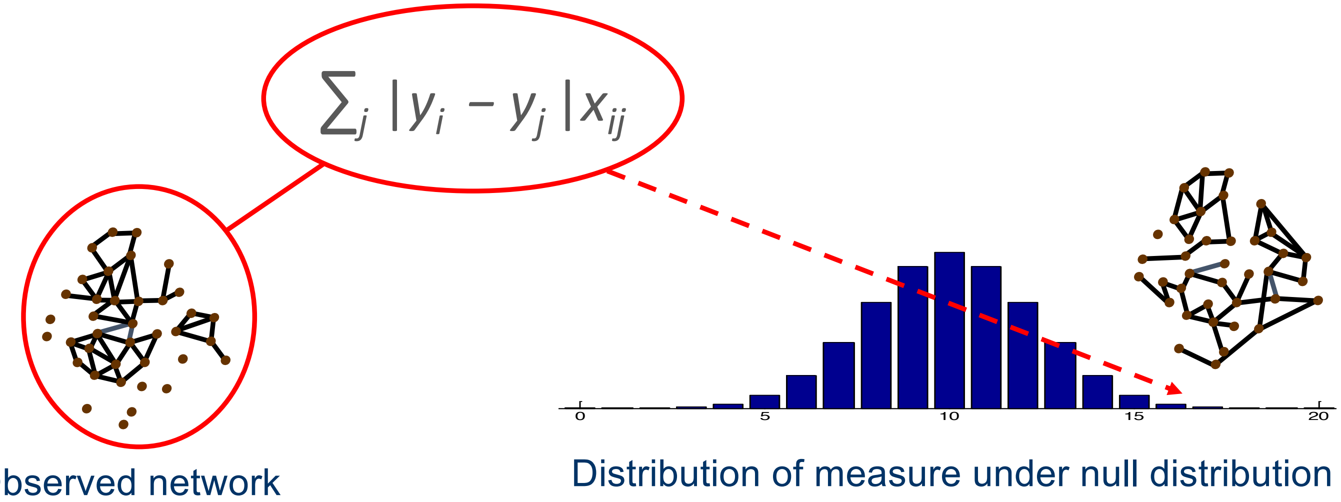
$$X_{ij} = \begin{cases} 1 & \text{, if } i \text{ has a tie to } j \\ 0 & \text{,otherwise} \end{cases}$$

Tied partners with high values contribute more $\sum_j y_i y_j x_{ij}$

Distance between tied partners in trust $\sum_j |y_i - y_j| x_{ij}$

Similarity between tied partners in trust $\sum_j (1 - |y_i - y_j|/\mathrm{d}_{max})x_{ij}$

How test?


$\sum_j |y_i - y_j| x_{ij}$
0
5
10
15
20
Observed network
Distribution of measure under null distribution

Randomise networks?

What are we testing

- What is the counterfactual?
- Drawing inference to network with same properties
- There are n(n-1)/2 dyads but only n outcomes

$X \sim$ Bern(p): the outdegrees are now independent of each other

$X \sim$ U | $L(X)$:

$X \sim$ U | d_{out} : we can fix outdegree only – what does activity explain

$X \sim$ U | d_{in} : we can fix indegree only – what does popularity explain

$X \sim$ U | d_{out}, d_{in} : we can fix outdegree AND indegree

$X \sim$ U | MAN: what can reciprocity explain

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix}$$

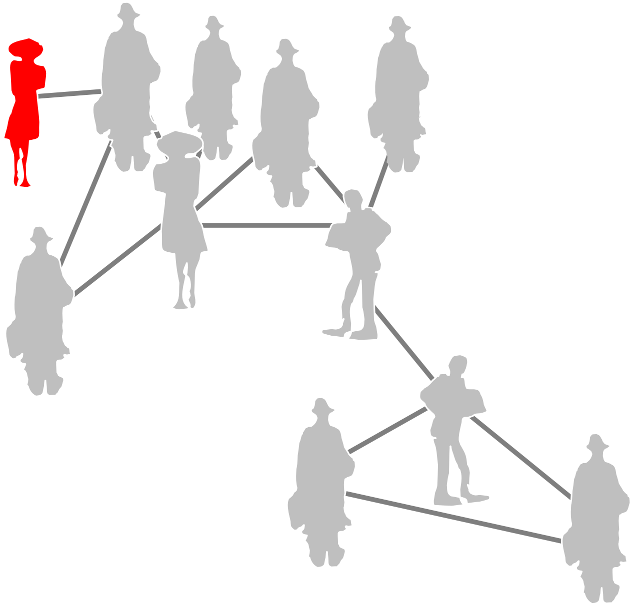
$$X = \begin{pmatrix} - & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & - & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & - & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & - & X_{45} \\ X_{51} & X_{52} & X_{53} & X_{54} & - \end{pmatrix}$$

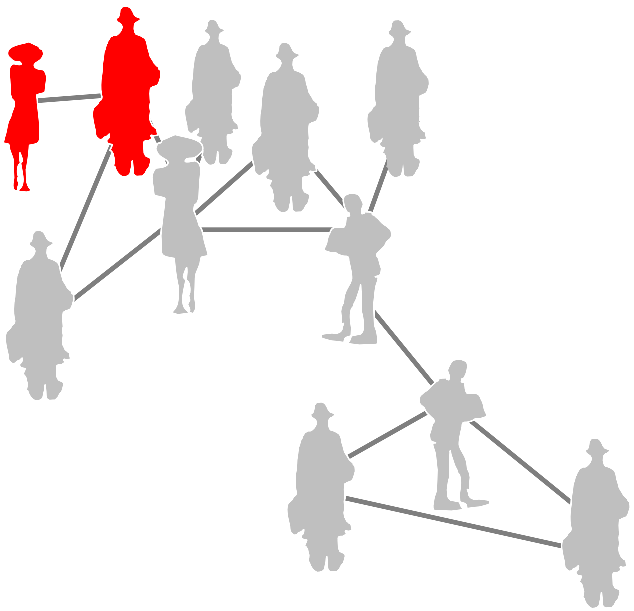
Randomise outcome?

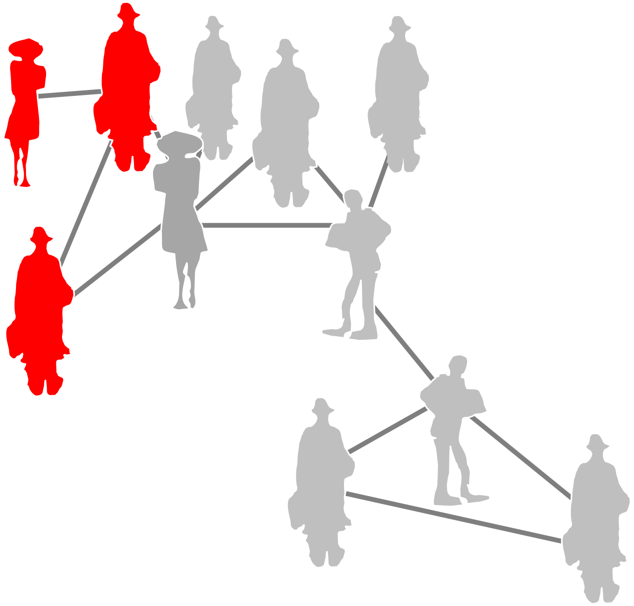
Interpretation: 'if all actors swapped positions…'
Only values in the range of what we have observed
What about other covariates 'what if this teen who loves tik-tok is all of sudden a 50-year-old man…'

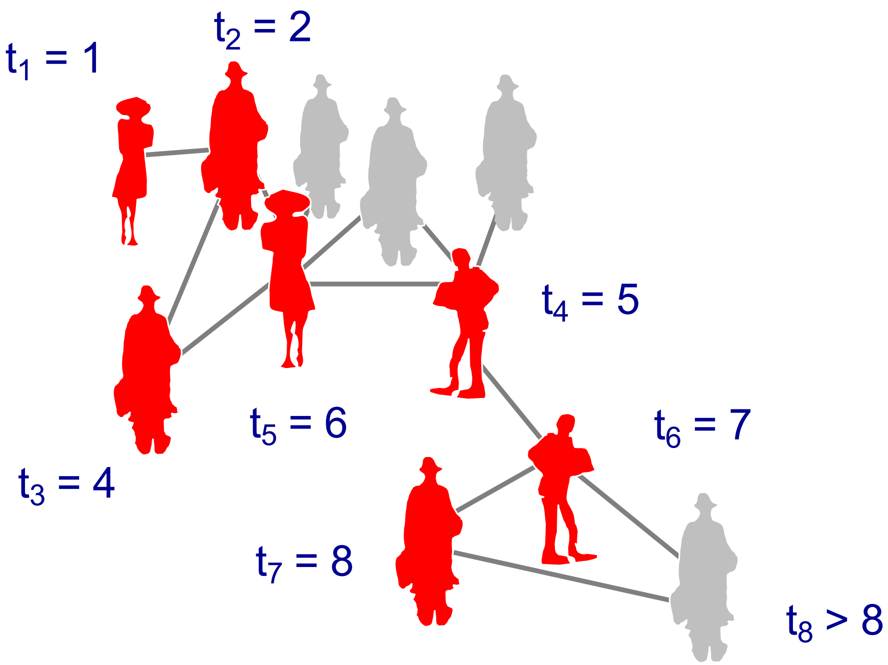
$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} \qquad X = \begin{pmatrix} - & X_{12} & X_{13} & X_{14} & X_{15} \\ X_{21} & - & X_{23} & X_{24} & X_{25} \\ X_{31} & X_{32} & - & X_{34} & X_{35} \\ X_{41} & X_{42} & X_{43} & - & X_{45} \\ X_{51} & X_{52} & X_{53} & X_{54} & - \end{pmatrix}$$

Event-history analysis



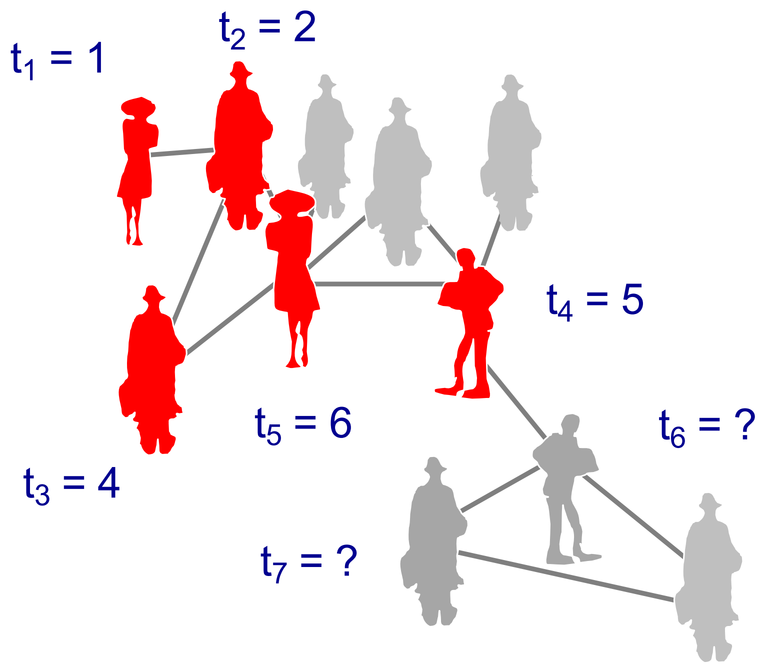






Empirical question:
Is the ***time to adoption*** shorter for individuals that have a tie to someone who has already adopted
Compared to people that do not have a tie to someone who has adopted?

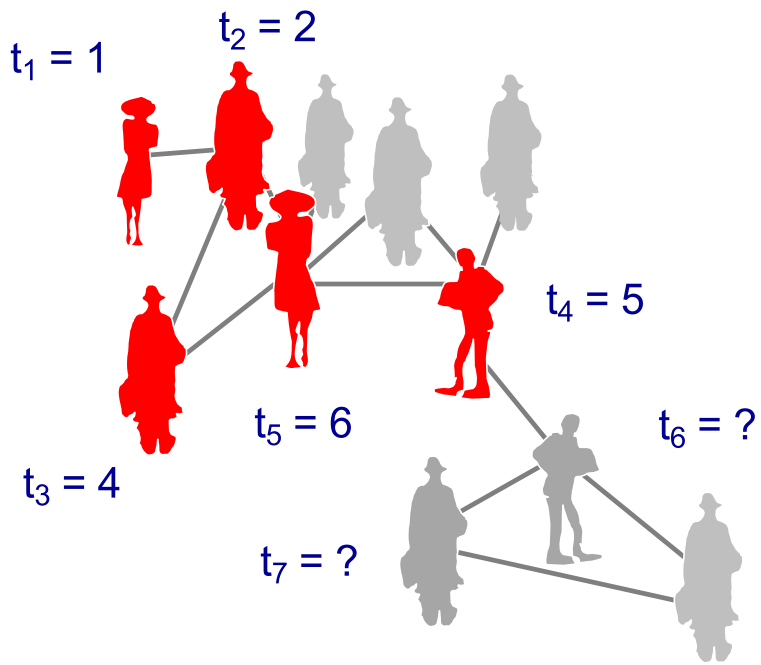
t_k = time of adoption of the k^{th} adopter or k^{th} ndividual



$$Pr(t_6 < t_7) = ?$$

Empirical question:
Is the *time to adoption* shorter for individuals that have a tie to someone who has already adopted Compared to people that do not have a tie to someone who has adopted?

Assume time t = 6:
who will adopt next?



Assume time t = 6:
who will adopt next?

If $t_i \sim \mathrm{Exp}(\lambda_i)$:

$$\Pr(\ t_i < \underbrace{\ s+t \mid t > t) = \Pr(\ t_i <\ }_{\text{Exponential distribution memoryless}} s\ )$$

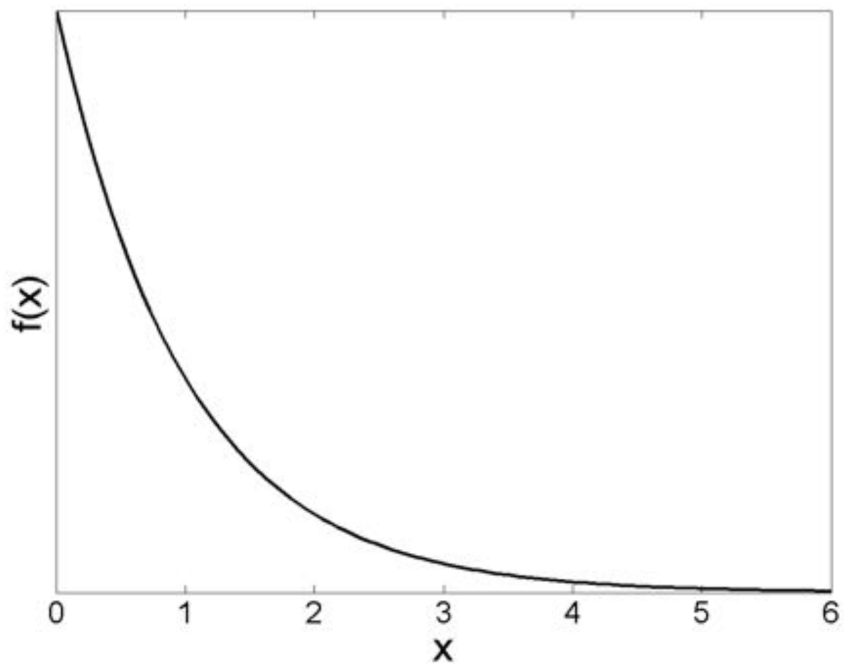
$$\Pr(\ t_6 < \ t_7) = \Pr(\ (t_6 - 6) < \ (t_7 - 6)\ )$$

Assume that we want to model something positive that gets increasingly less probably the larger it gets

- The time till divorce
- The time spent as unemployed

$$f(x) = \lambda e^{-\lambda x}$$

probability density function (pdf)

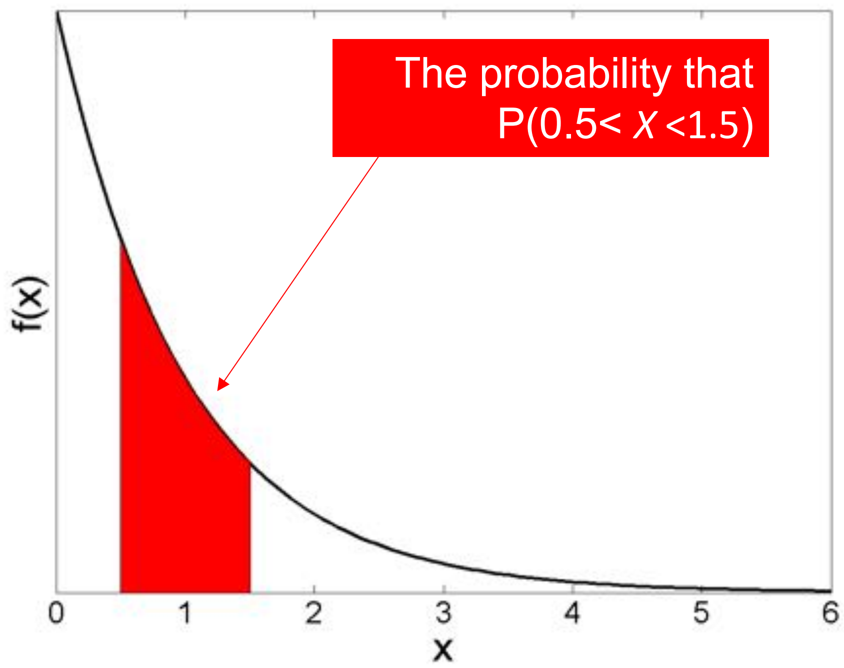


Assume that we want to model something positive that gets increasingly less probably the larger it gets

- The time till divorce
- The time spent as unemployed

$$f(x) = \lambda e^{-\lambda x}$$

probability density function (pdf)

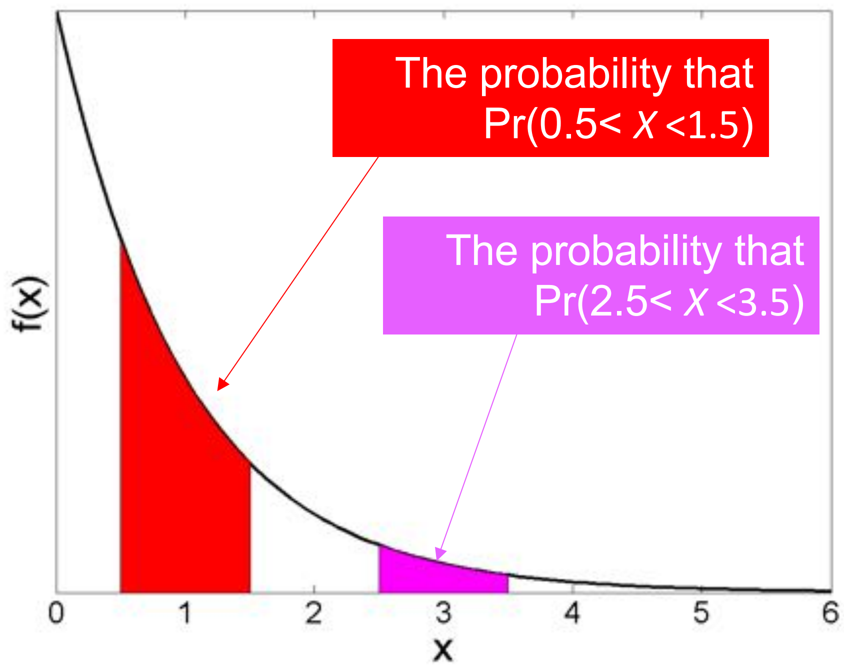


Assume that we want to model something positive that gets increasingly less probably the larger it gets

- The time till divorce
- The time spent as unemployed

$$f(x) = \lambda e^{-\lambda x}$$

probability density function (pdf)



The higher the rate

λ

The quicker the change

The parameter

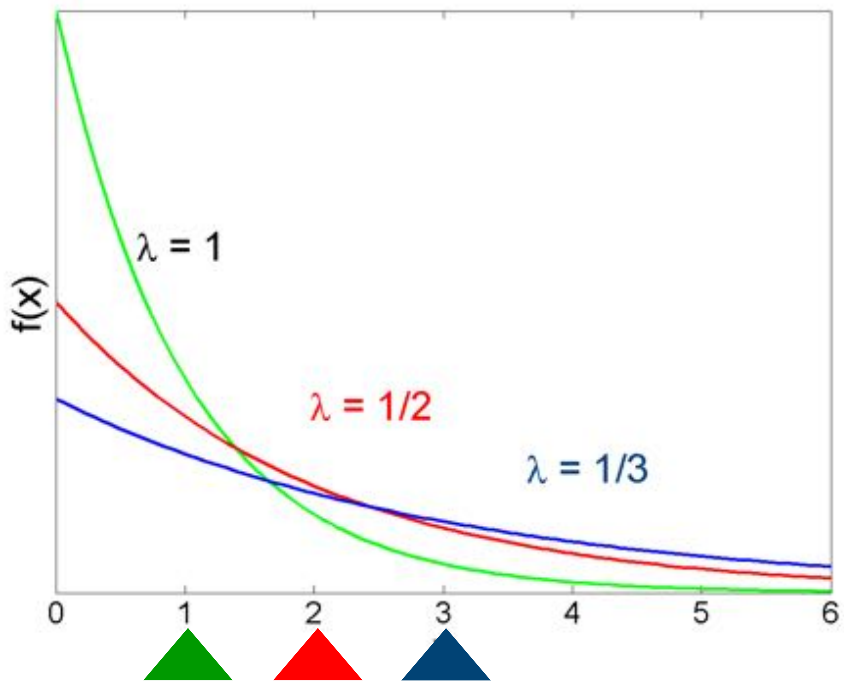
λ

Describes the location:

$E(X) = 1/\lambda$

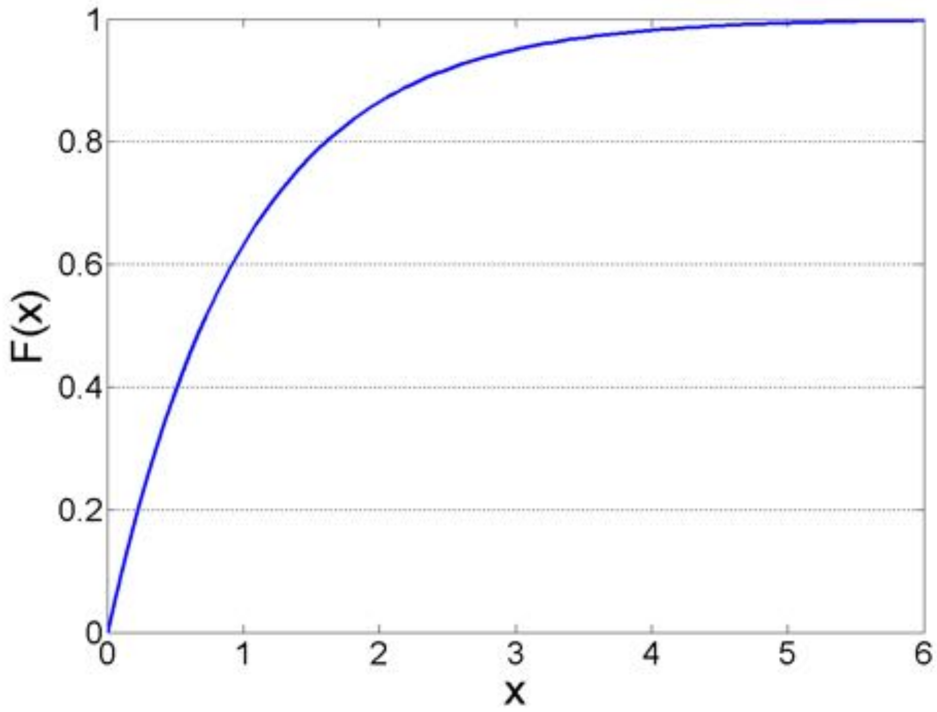
And the spread

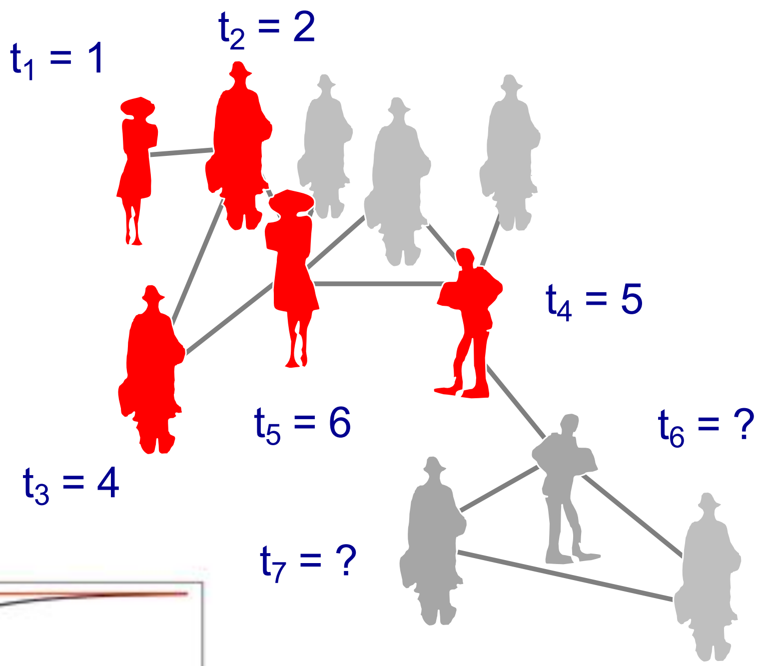
$V(X) = 1/\lambda^2$

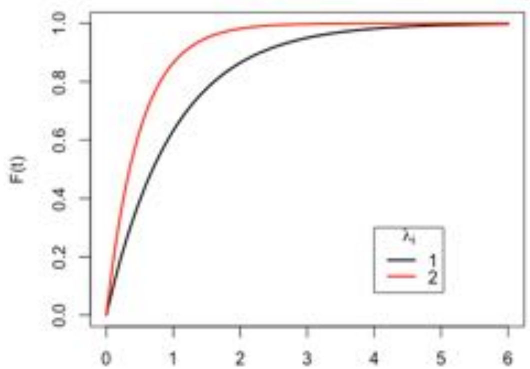


A useful property is that we can write down the cdf

$$\Pr(X < x) = F(x) = 1 - e^{-\lambda x}$$





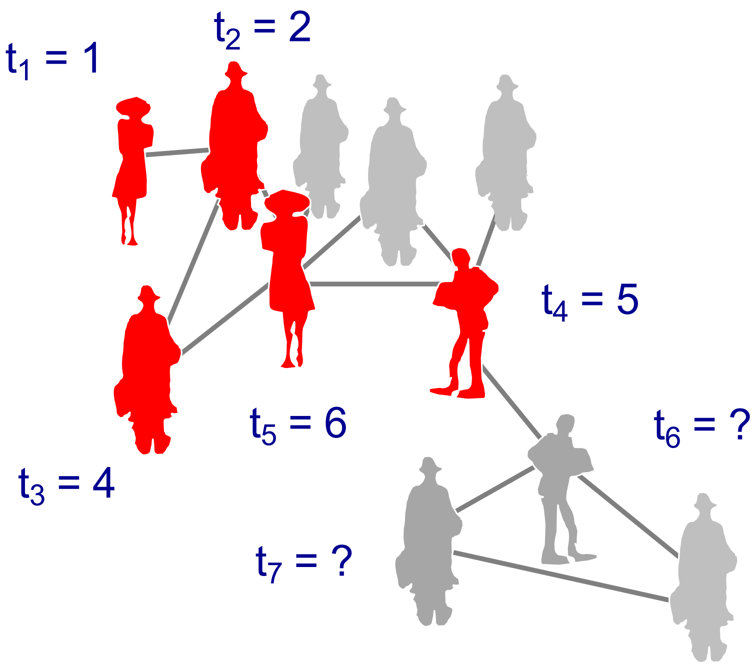


For diffusion we want

$t_6 \sim \text{Exp}(\lambda_6)$
$t_7 \sim \text{Exp}(\lambda_7)$

So that $\Pr(t_6 < t_7)$ is large

If we set $\lambda_6 > \lambda_7$, then
$E(t_6) = 1/\lambda_6 < 1/\lambda_7 = E(t_7)$
and
$\Pr(t_6 < t) > \Pr(t_7 < t)$



Define for each *i*

$$Y_i(t) = \begin{cases} 1 & \text{,if } i \text{ has adopted at time } t \\ 0 & \text{,otherwise} \end{cases}$$

For any point in time t_0

$$\Pr(t_i < t_0 + s \mid Y_i(t_0) = 0)$$
$$= \Pr(t_i < s \mid Y_i(t_0) = 0)$$

$$t_i - t_0 \sim \text{Exp}\{\ \lambda_i(Y(t_0), X)\ \}$$

A rate that depends on all $Y_j(t)$ and network X

Piecewise constant rate functions

rate must be positive

function linking current state to i's rate of change

$$\lambda_i(Y(t), X(t)) = \exp\{\alpha + \beta a_i(Y(t), X(t))\}$$

rate specific to *i*

A rate that depends on all $Y_j(t)$ and network $X(t)$

Statistical parameter for influence strength

Note: if a change occur to $Y(t)$, rates are updated

Piecewise constant rate functions

$$\lambda_i(Y(t), X(t)) = \exp\{\alpha + \beta a_i(Y(t), X(t))\}$$

Positive and large β – stronger dependence on current state

If β>0 the greater a_i, the quicker the change

Note: if a change occur to $Y(t)$, rates are updated

Meyers (2000): rates depend on important adopters

$$a_i(y, x) = \sum_j y_j s_j(x)$$

y_j = 1 if other person j has adopted

The adoption $y_j = 1$ carries more weight if importance $s_j(x)$ of j is great

The more 'important' nodes that have adopted, the more other people will adopt

Valente (2005): rates depend on important adopters that you are connected to

The adoption $y_j = 1$ carries more weight if importance $s_j(x)$ of j is great

$$a_i(y, x) = \sum_j y_j x_{ij} s_j(x)$$

$y_j = 1$ if other person j has adopted

$x_{ij} = 1$ if i has a tie to j

The more 'important' nodes that have adopted, the more other people will adopt

Total exposure

Count of how many people that i is connected to have already adopted

$$a_i(y, x) = \sum_j \underbrace{y_j x_{ij}}$$

y_j = 1 if other person *j* has adopted

x_{ij} = 1 if *i* has a tie to *j*

The more people you know that have adopted, the quicker you will adopt yourself

Average exposure

proportion of how many people that i is connected to have already adopted

$$a_i(y, x) = \frac{\overbrace{\sum_j y_j x_{ij}}}{\sum_j x_{ij}}$$

degree of i

The greater the proportion of people you know that have adopted, the quicker you will adopt yourself

Average v total exposure

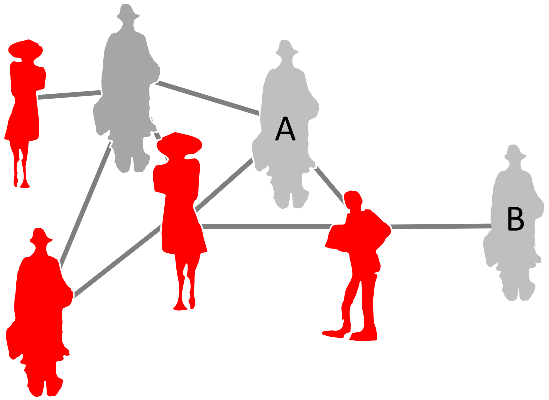
$$a_i(y, x) = \frac{\sum_j y_j x_{ij}}{\sum_j x_{ij}}$$

$$a_A(y, x) = \frac{2}{3}$$

$$a_B(y, x) = \frac{1}{1} = 1$$

RED: $y_j = 1$

GREY: $y_j = 0$



$$a_i(y, x) = \sum_j y_j x_{ij}$$

$$a_B(y, x) = 1$$

$$a_A(y, x) = 2$$

When is the total number important and when is it the proportion – examples?

Further implications of exponential times

In the process, all (susceptible) nodes draw a time $T_i \sim \text{Exp}\{ \lambda_i(Y(t),X(t)) \}$
The node with the shortest time, gets to change – we call that the winner

What is the probability that *i* is the winner?

$$\Pr(T_i = \text{argmin}\{T_1 ,\ldots, T_n\}) = \frac{\lambda_i(Y(t),X(t))}{\sum_j \lambda_j(Y(t),X(t))}$$

A property of the exponential distribution

Further implications of exponential times

In the process, all (susceptible) nodes draw a time $T_i \sim \text{Exp}\{ \lambda_i(Y(t),X(t)) \}$
The node with the shortest time, gets to change – we call that the winner

What is distribution of the winning time?

$$\min\{T_1 ,\dots, T_n\} \sim \text{Exp}\Big(\textstyle\sum_j \lambda_j(Y(t),X(t)) \Big)$$

A property of the exponential distribution

Simulation of the diffusion process

At time t

Pick a winner i with probability

$$\frac{\lambda_i(Y(t),X(t))}{\sum_j \lambda_j(Y(t),X(t))}$$

Draw winning time

$$s \sim \text{Exp}\left(\sum_j \lambda_j(Y(t),X(t)) \right)$$

Update status of winner

$$Y(t+s) := Y(t)+1$$

Increment time

$$t := t+s$$

Estimation of the diffusion process

Based on the network *X*, and the times of adoption

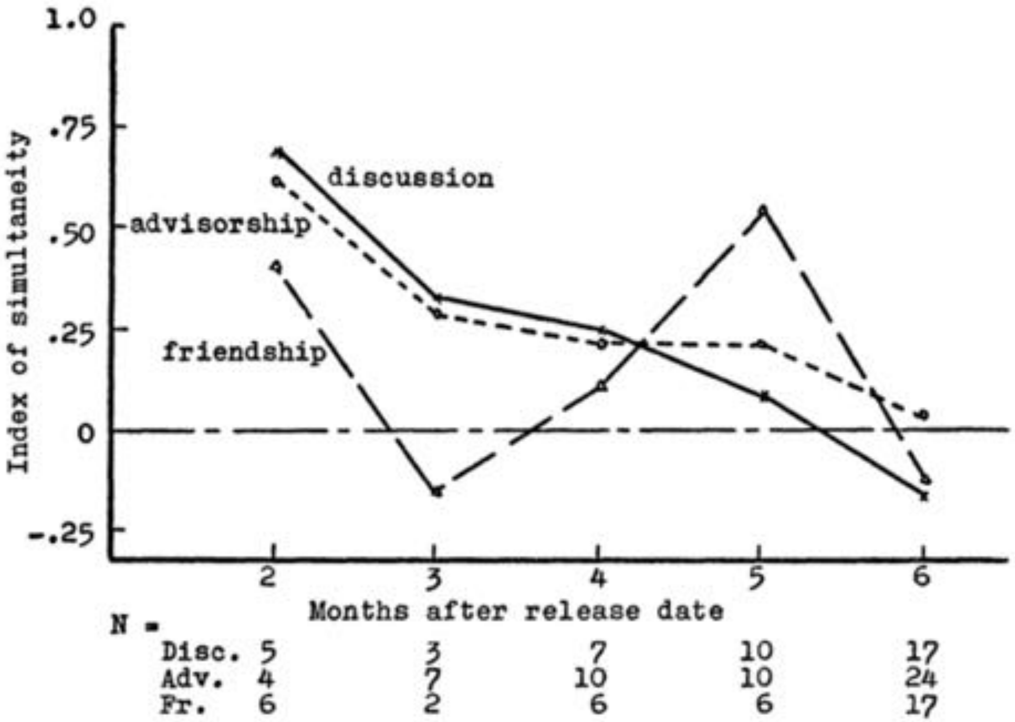
$$t_1 ,\dots, t_k$$

The likelihood is just a product of exponential pdf's and cdf's (for non-adopters)

$$\prod \lambda_j(Y(t_j),X(t_j)) \exp\left\{ -\textstyle\sum_j (t_j - t_{j-1})\lambda_j(Y(t_j),X(t_j)) \right\}$$
$$\times \underbrace{\prod \exp\{ -(t_j - t_{j-1})\lambda_h(Y(t_j),X(t_j)) \}}_{\boxed{Pr(T>t)\text{: censoring for those who didn't adopt}}}$$

Parameters can be estimated using Maximum likelihood or Bayes

EXAMPLE: Coleman, Katz, and Menzel (1957)

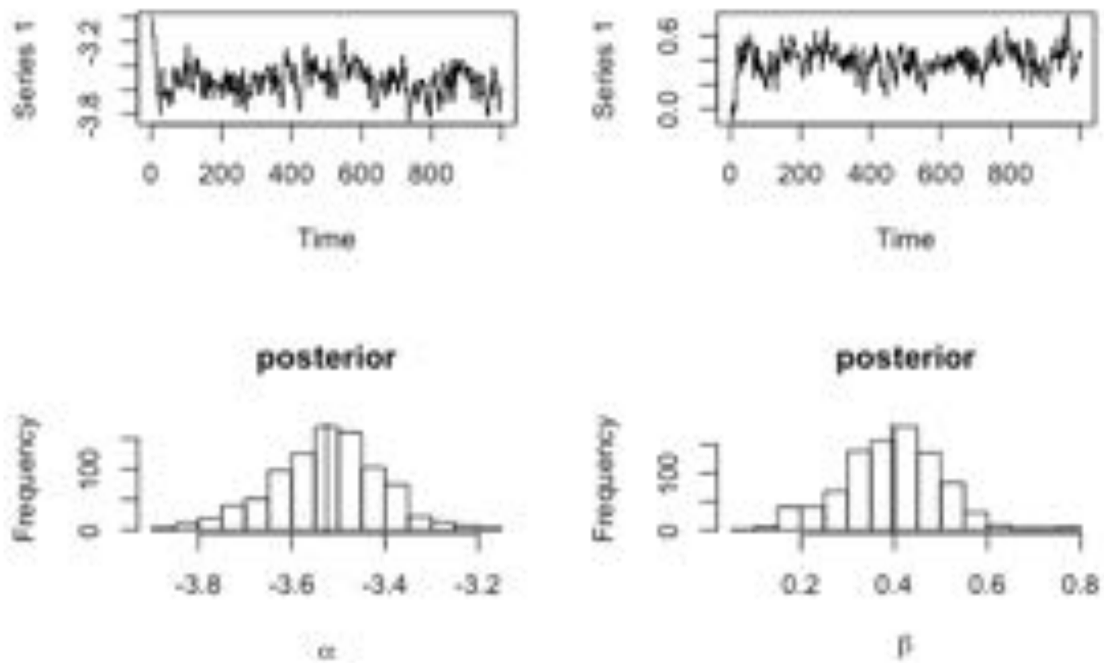


The Diffusion of an Innovation Among Physicians, Sociometry

To analyse pairs of individuals instead of single individuals may seem like only a very modest step in the direction of the analysis of networks of social relations. And so it is; it would be more satisfactory, and truer to the complexity of actual events, if it were possible to use longer chains and more ramified systems of social relations as the units of analysis. But so little developed are the methods for the analysis of social processes, that it seemed best to be content with the analysis of pair relationships"
Coleman, Katz and Menzel 1966, p 114

difusion.html

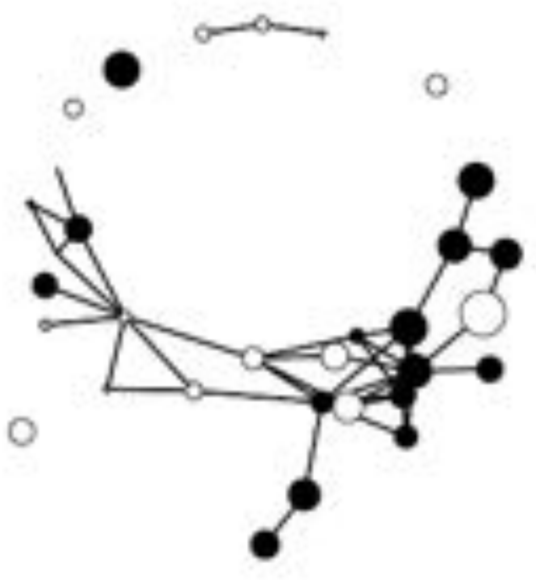
EXAMPLE:
Coleman,
Katz, and
Menzel
(1957)



The Diffusion of an Innovation Among Physicians, Sociometry

Social influence in cross-sectional data

A dataset of academics

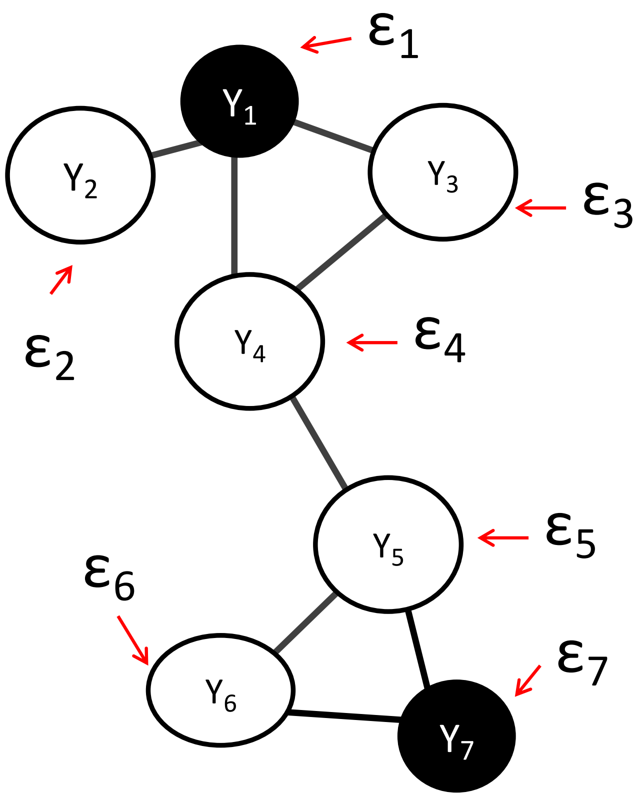


Y_i : *Log citation of researcher i*

$$X_{ij} = \begin{cases} 1 & \text{, if } i \text{ collaborates with } j \\ 0 & \text{,otherwise} \end{cases}$$

$$M_i : \begin{cases} 1 & \text{, if } i \text{ sociologist} \\ 0 & \text{,otherwise} \end{cases}$$

Ordinary least squares regression



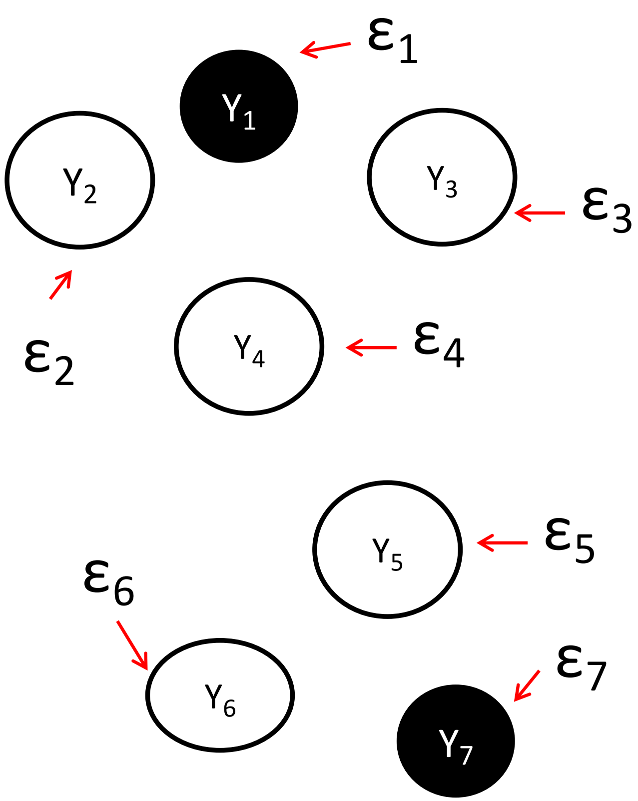
$$Y_i = \alpha + \beta M_i + \varepsilon_i$$

where

$$\varepsilon_i \sim N(0,\sigma^2)$$

INDEPENDENT for all *i*

Ordinary least squares regression



$$Y_i = \alpha + \beta M_i + \varepsilon_i$$

where

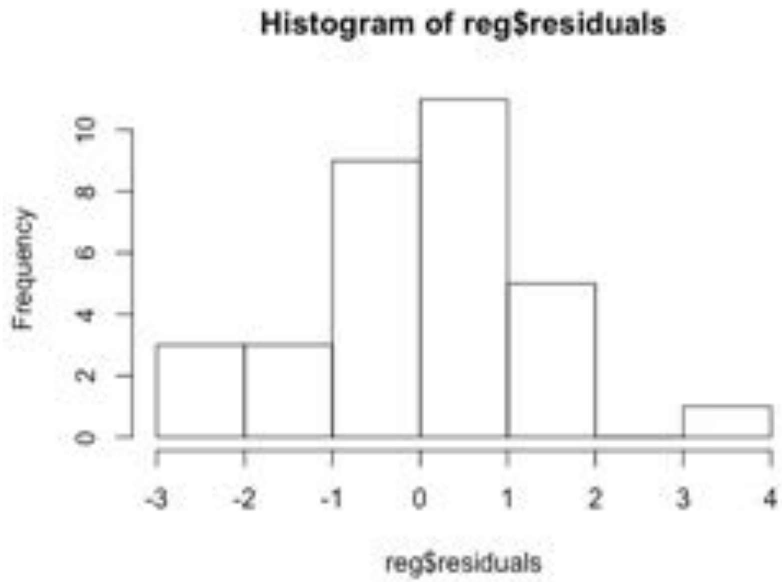
$$\varepsilon_i \sim N(0,\sigma^2)$$

INDEPENDENT for all *i*

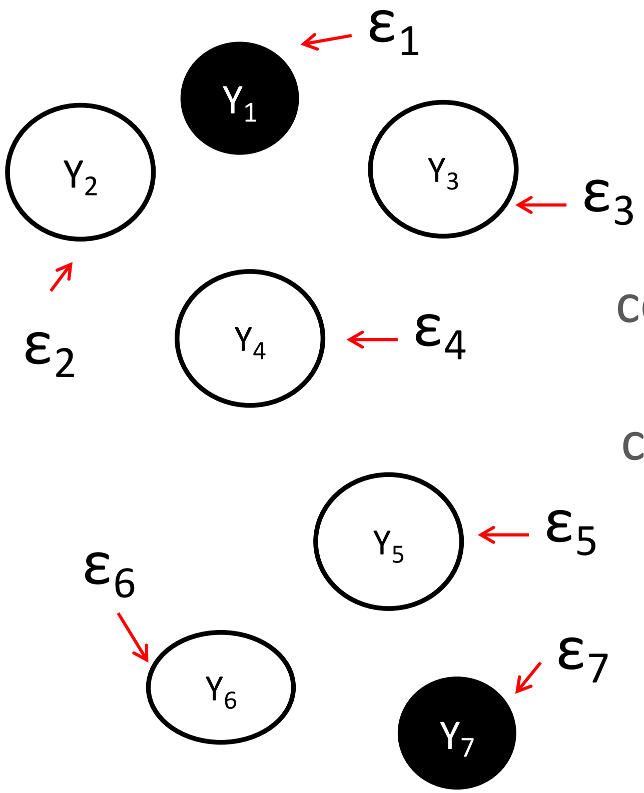
OLS

$\varepsilon_i \sim N(0,\sigma^2)$?

```
Call:
lm(formula = logcite ~ covariates[, 2])

Residuals:
     Min       1Q   Median       3Q      Max
-2.87901 -0.47427  0.03839  0.81261  3.11374

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       2.0279     0.3170   6.397 4.62e-07 ***
covariates[, 2]   0.8511     0.4349   1.957   0.0597 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.228 on 30 degrees of freedom
Multiple R-squared:  0.1132,    Adjusted R-squared:  0.08364
F-statistic: 3.829 on 1 and 30 DF,  p-value: 0.05973
```



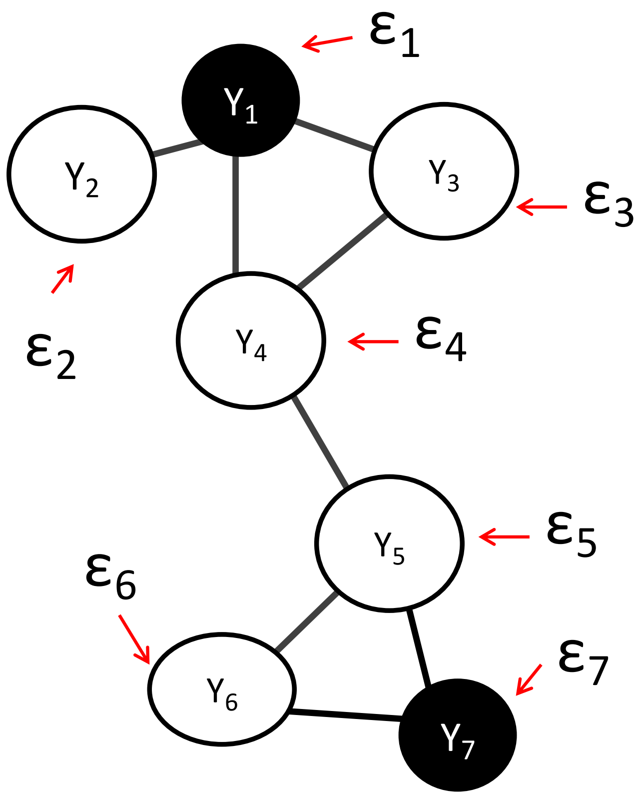
Independent error terms?



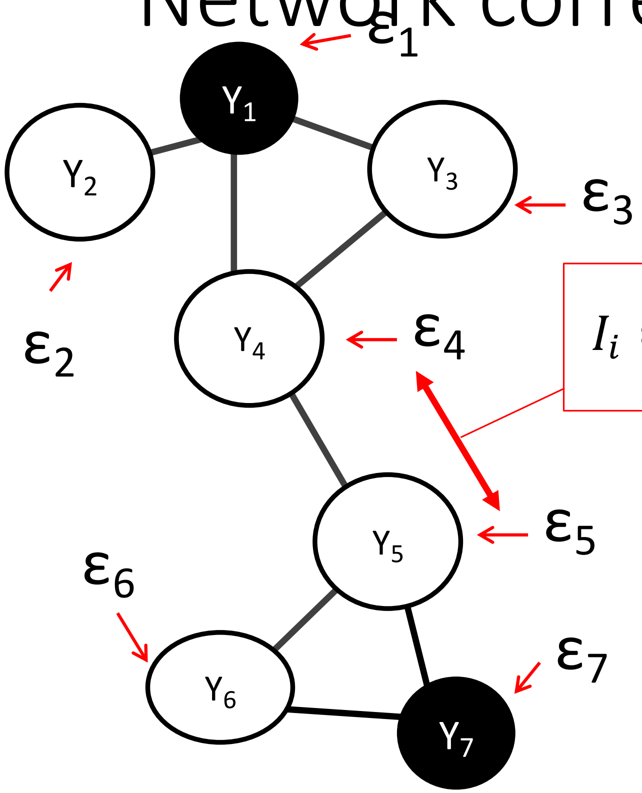
$\mathcal{E}_i$ INDEPENDENT for all i ?

$cor(\varepsilon_5 , \varepsilon_2) = ?$

$cor(\varepsilon_5 , \varepsilon_7) = ?$



Network correlation – Moran's I



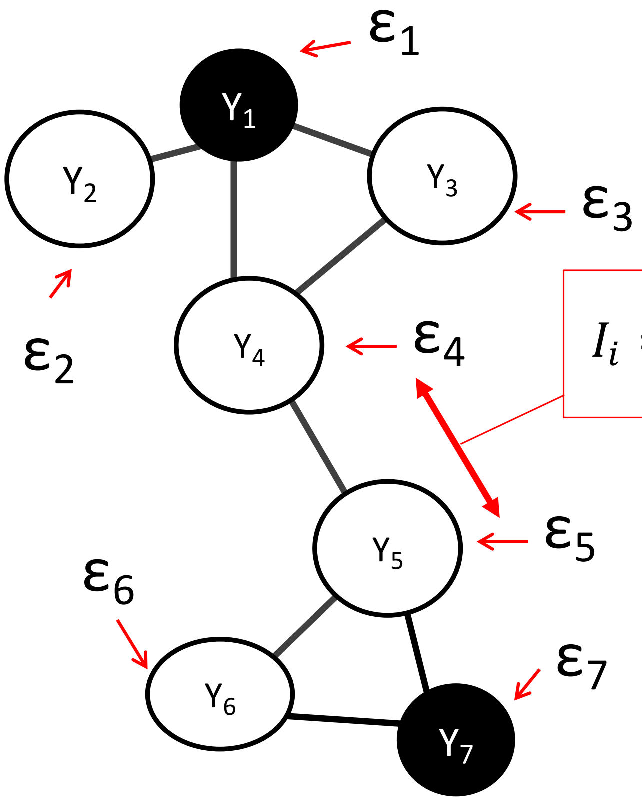
cov(ε_j , ε_k) but only for X_{jk} = 1

$$I_i = \frac{n\sum_{j=1}^{n}\sum_{k=1}^{n}(y_j-\bar{y})(y_k-\bar{y})A_{ijk}}{\sum_{j,k}A_{ijk}\sum_{j=1}^{n}y_j^2}$$

We test directly if there are departures from independence

Alternative:

$$C_i = \frac{(n-10\sum_{j=1}^{n}\sum_{k=1}^{n}(y_j-y_k)^2A_{ijk}}{2\sum_{j,k}A_{ijk}\sum_{j=1}^{n}(y_j-\bar{y})^2}$$



cov(ε_j , ε_k) but only for X_{jk} = 1

$$\boxed{I_i = \frac{n\sum_{j=1}^{n}\sum_{k=1}^{n}(y_j-\bar{y})(y_k-\bar{y})A_{ijk}}{\sum_{j,k}A_{ijk}\sum_{j=1}^{n}y_j^2}}$$

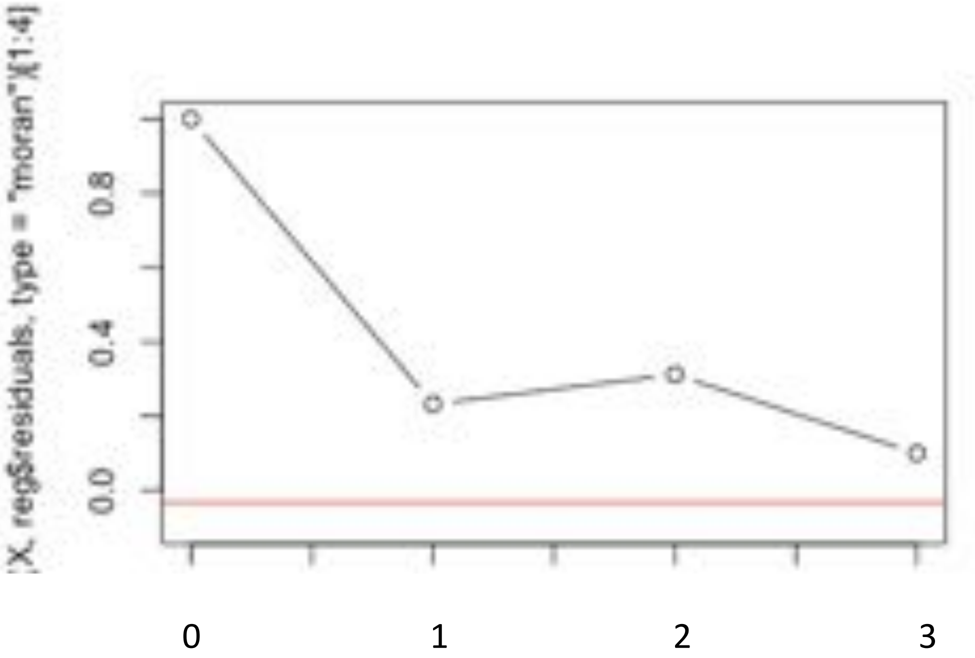
If I_i >0 what does it mean?

Residual for connected nodes

$$y_j - \hat{y}_j > 0$$

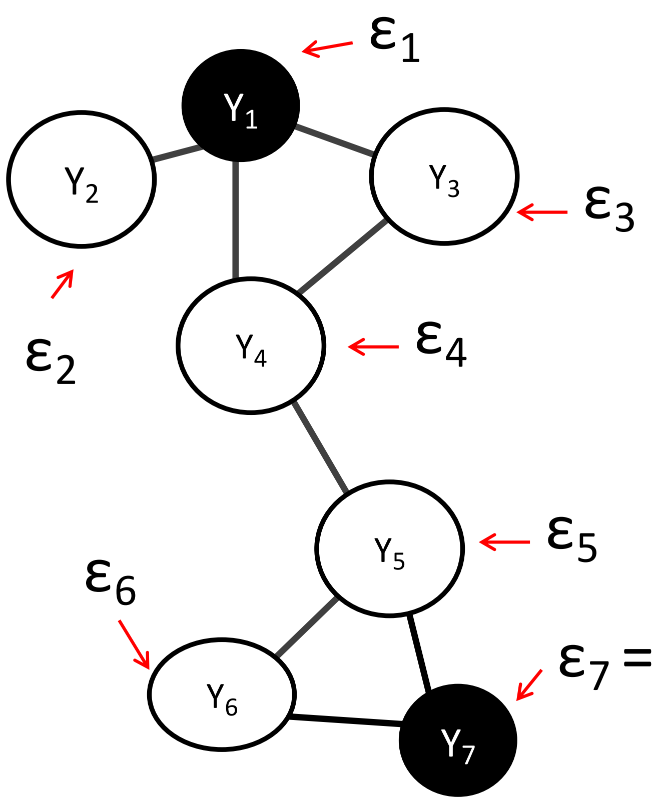
$$y_k - \hat{y}_k > 0$$

Are errors uncorrelated?




```r
plot(nacf(X,reg$residuals,type="moran")[1:4],
     type='b',ylim=c(-0.1,1))
abline(h=-1/(n-1),col='red')
```

Network auto-correlation model



$\varepsilon_7 = \rho\varepsilon_5 + \rho\varepsilon_6 + \xi_7$

$$Y_i = \alpha + \beta M_i + \varepsilon_i$$

where

$$\varepsilon_i = \rho \sum_j X_{ij}\varepsilon_j + \xi_i$$

and

$$\xi_i \sim N(0, \sigma^2)$$

INDEPENDENT for all *i*

Network auto-correlation model

```
Call:
lnam(y = logcite, x = covariates, W2 = X)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6925 -0.5811  0.1828  0.9824  3.0069

Coefficients:
            Estimate Std. Error Z value Pr(>|z|)
cons         2.13474    0.33909   6.296 3.06e-10 ***
sociologist  0.55778    0.47269   1.180   0.2380
rho2.1       0.10050    0.05868   1.713   0.0868 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Estimate Std. Error
Sigma     1.13      0.021
```

$$Y_i = \alpha+\beta M_i+\varepsilon_i$$

where

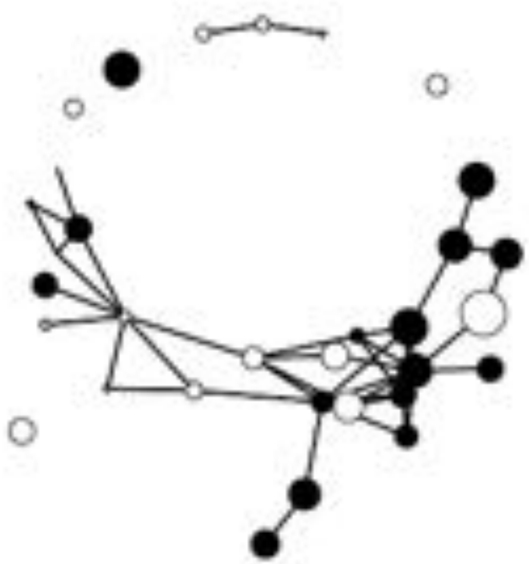
$$\varepsilon_i = \rho \sum_j X_{ij}\varepsilon_j + \xi_i$$

and

$$\xi_i \sim \mathrm{N}(0,\sigma^2)$$

INDEPENDENT for all *i*

Weight matrix



Since

$$\varepsilon_i = \rho \textstyle\sum_j X_{ij}\varepsilon_j + \xi_i$$

Higher degree nodes
'get more errors'
What about scaling ties

$$\varepsilon_i = \rho \frac{\sum_j X_{ij}\varepsilon_j}{\sum_j X_{ij}} + \xi_i$$

Weight matrix

```
Call:
lnam(y = logcite, x = covariates, W2 = W)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6855 -0.7442  0.1371  0.8724  2.8438

Coefficients:
            Estimate Std. Error Z value Pr(>|z|)
cons          2.2978     0.3810   6.032 1.62e-09 ***
sociologist   0.3877     0.4640   0.836  0.40341
rho2.1        0.9271     0.3596   2.578  0.00993 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Estimate Std. Error
Sigma    1.066      0.019
```

$\sum_j W_{ij} = 1$ (or 0)

Since

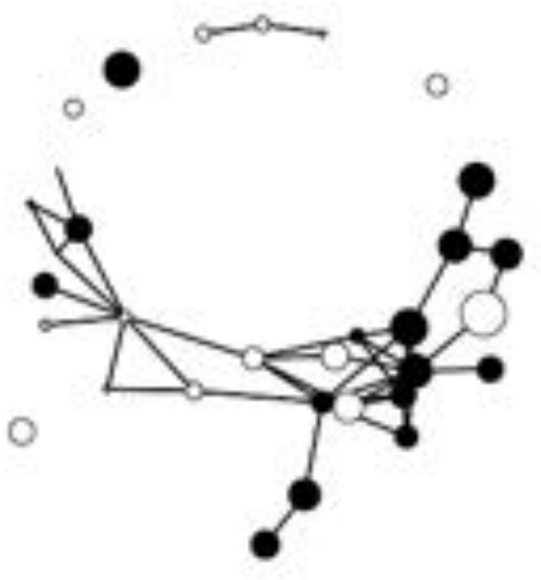
$$\varepsilon_i = \rho \sum_j X_{ij}\varepsilon_j + \xi_i$$

Higher degree nodes
'get more errors'
What about scaling ties

$$\varepsilon_i = \rho \frac{\sum_j X_{ij}\varepsilon_j}{\sum_j X_{ij}} + \xi_i$$

$$= \rho \sum_j W_{ij}\varepsilon_j + \xi_i$$

Interpretation

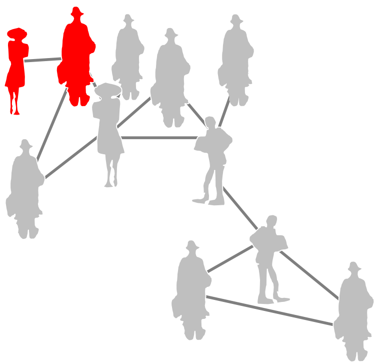


Researchers tend to be cited more than average if their collaborators are cited more than average

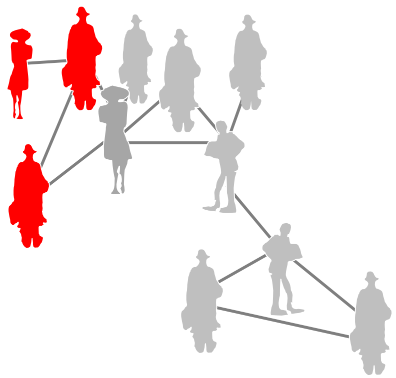
Correlation through errors mean that we do not predict with the network only explaining unobserved heterogenity

Network effects model

What if there was a process of social influence



Collaborating with highly cited researcher makes you more cited



$$Y_{i,t} = \rho \sum_j X_{ij} \, Y_{j,t-1} \, \alpha + \beta M_i + \varepsilon_{i,t}$$

Citation at time *t*

Citation of all others at time *t*-1

Network effects model

What if there was a process of social influence

$$Y_{i,t} = \rho \textstyle\sum_j X_{ij}\, Y_{j,t-1} + \alpha + \beta M_i + \varepsilon_{i,t}$$

Citation at time *t*

Citation of all others at time *t*-1

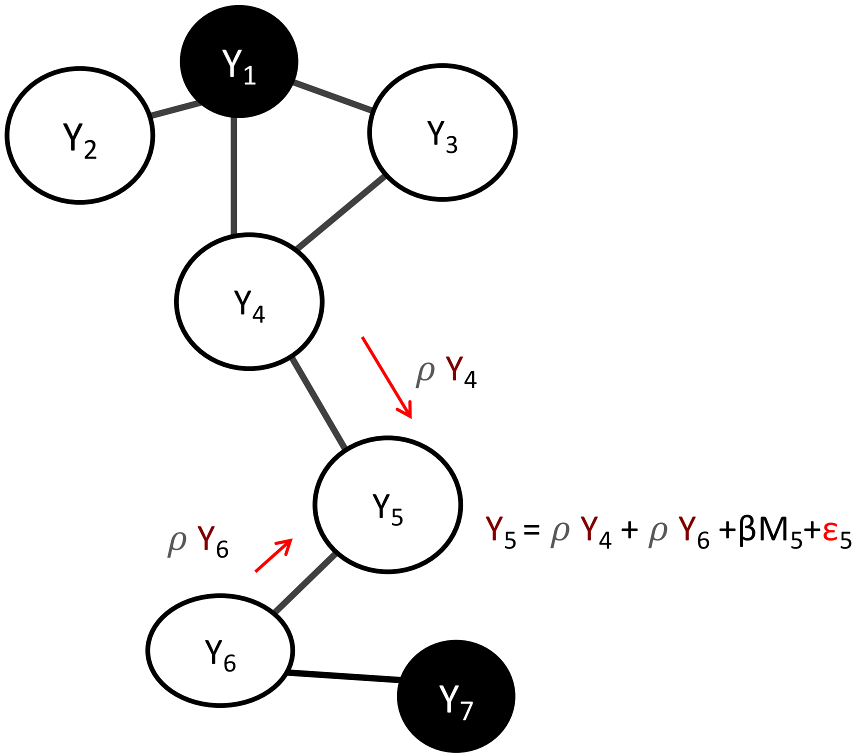
The equilibrium (when t tends to infinity):

$$Y_i = \rho \textstyle\sum_j X_{ij}\, Y_j + \alpha + \beta M_i + \varepsilon_i$$

Citations

Citation of all others

Network effects model



The network effect model (element-wise):

$$Y_i = \rho \sum_j X_{ij}\, Y_j \ + \alpha + \beta M_i + \varepsilon_i$$

Citations

Citation of all others

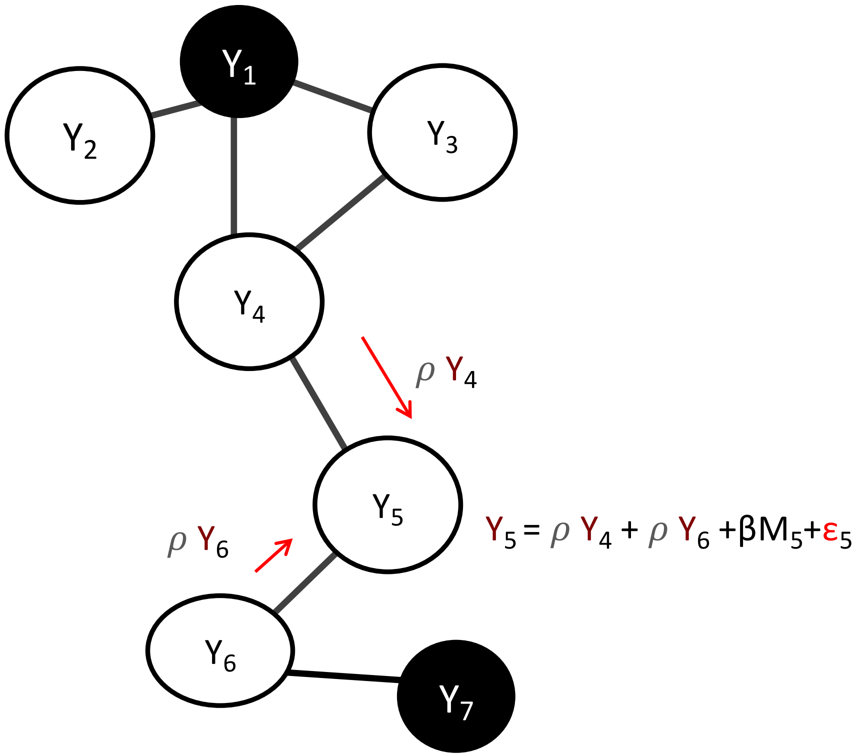
The network effect model (simulataneous):

$$Y = \rho X Y \ + \alpha + M\beta + \varepsilon$$

Vector of citations

Citation of all others

Network effects model



The network effect model (simultaneous):

$$Y = \rho XY + \alpha + M\beta + \varepsilon$$

```
Call:
lnam(y = logcite, x = covariates, W1 = X)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6665 -0.8386  0.1060  0.8026  3.0625

Coefficients:
            Estimate Std. Error Z value Pr(>|z|)
cons         1.87066    0.33036   5.663 1.49e-08 ***
sociologist  0.69711    0.43357   1.608    0.108
rho1.1       0.03397    0.02969   1.144    0.253
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Estimate Std. Error
Sigma    1.163      0.021
```

Weight matrix

```
Call:
lnam(y = logcite, x = covariates, W1 = W)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8873 -0.5815  0.1188  0.8561  2.8750

Coefficients:
            Estimate Std. Error Z value Pr(>|z|)
cons          1.7051     0.3926   4.343  1.4e-05 ***
sociologist   0.6244     0.4465   1.398    0.162
rho1.1        0.3945     0.3124   1.263    0.207
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      Estimate Std. Error
Sigma    1.154      0.021
```

Is everyone equally important?

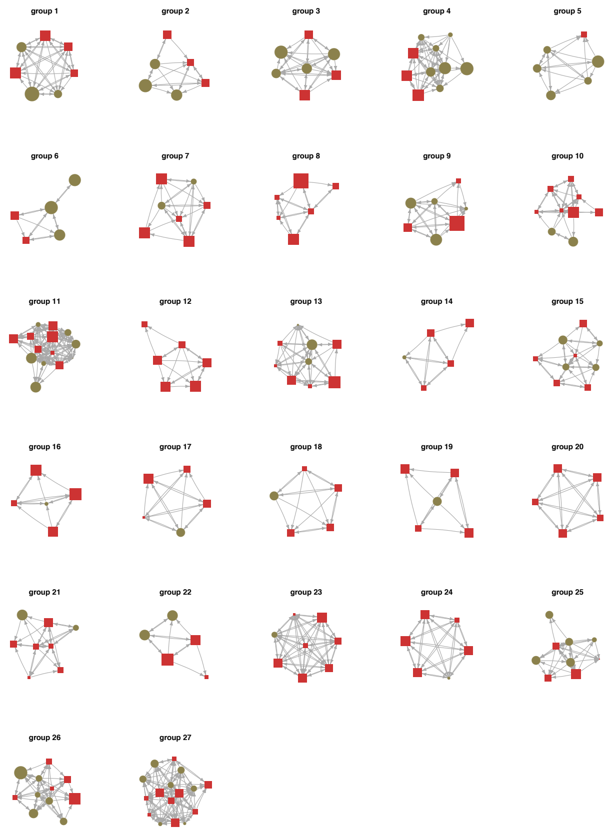
"I have two friends and they both like the government"

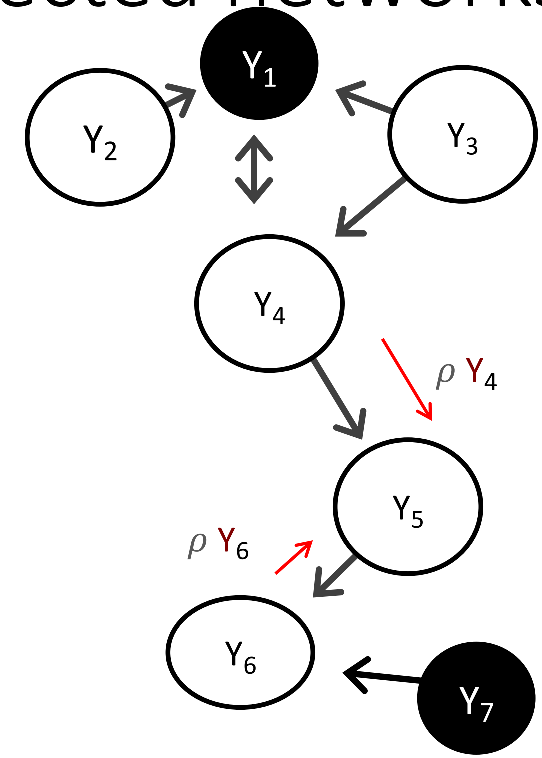
"I have two fifty friends and two like the government"

$$Y_i = \rho \frac{\sum_j X_{ij} Y_j}{\sum_j X_{ij}} + \varepsilon_i$$

$$Y_i = \rho \sum_j W_{ij} Y_j + \varepsilon_i$$

Network Effects model for directed networks



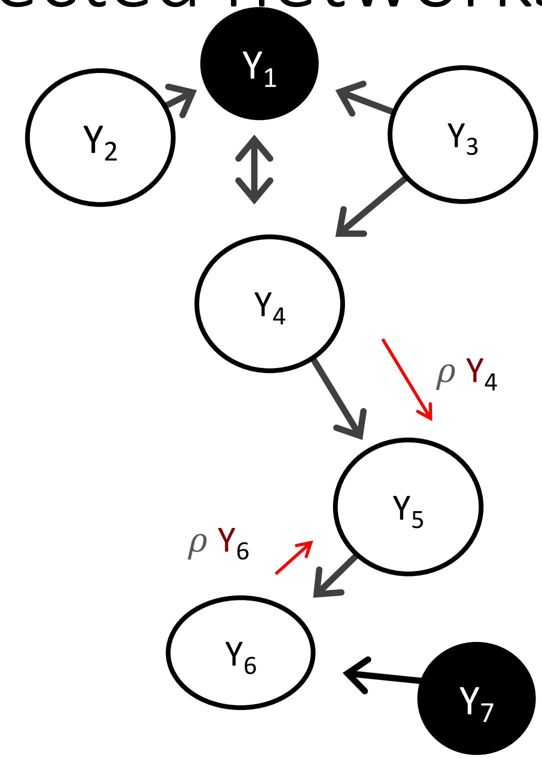


In-ties $Y_5 = \rho\ Y_4 + \beta M_5 + \varepsilon_5$

Out-ties $Y_5 = \rho\ Y_6 + \beta M_5 + \varepsilon_5$

Network Effects model for directed networks

Consider the differences between:
'I become similar to the people I like'
'I influence the people that I chose'



In-ties $Y_5 = \rho\ Y_4 + \beta M_5 + \varepsilon_5$
Out-ties $Y_5 = \rho\ Y_6 + \beta M_5 + \varepsilon_5$

… but what about selection