Introduction to SNA

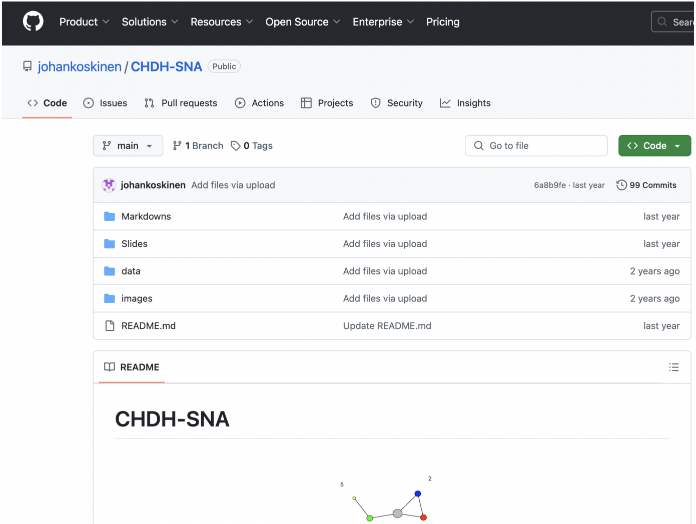
Introduction to analysing networks in R

Johan Koskinen

Department of Statistics
Stockholm University
University of Melbourne

February 23, 2025

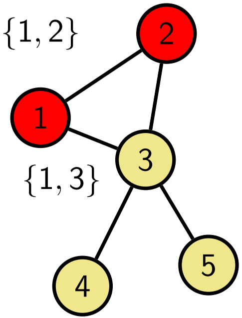
Course site


Product Solutions Resources Open Source Enterprise Pricing
johankoskinen / CHDH-SNA Public
Code Issues Pull requests Actions Projects Security Insights
main 1 Branch 0 Tags
Go to file
Code
johankoskinen Add files via upload 6a8b9fe · last year 99 Commits
Markdowns Add files via upload last year
Slides Add files via upload last year
data Add files via upload 2 years ago
images Add files via upload 2 years ago
README.md Update README.md last year
README
CHDH-SNA
5
2

Preamble

- All material is on the workshop repository https://github.com/johankoskinen/CHDH-SNA
  - Download the RMarkdown file CHDH-SNA-1.Rmd
- In order to run the Markdown you need
  - The R-package
  - The RStudio interface
- We will predominantly use the packages
  - sna
  - network

Graphs



Graph:

$$G(V, E)$$

Is a collection of **Nodes**

$$V = \{1, 2, \ldots, n\}$$

with **Edges**(lines)

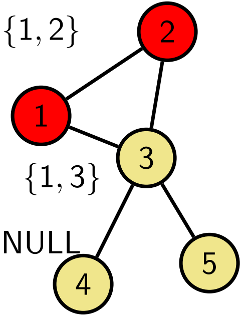
$$E \subseteq \binom{V}{2}$$

Example:

$$V = \{1, 2, 3, 4, 5\}$$

$$E = \{\{1,2\}, \{1,3\}, \{2,3\}, \{3,4\}, \{3,5\}\}$$

Graphs: Edge list (1)

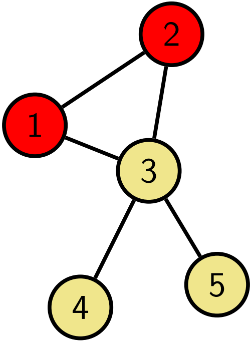


We can represent the graph in an **edge list**

i	j	value
1	2	1
1	3	1
1	4	0
1	5	0
2	3	1
2	4	0
2	5	0
3	4	1
3	5	1
4	5	0

With NULL-ties

Graphs: Edge list (2)

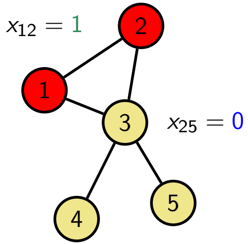


We can represent the graph in an **edge list**

i	j	value
1	2	1
1	3	1
2	3	1
3	4	1
3	5	1

Without NULL-ties

Network data: Adjacency matrix



Tie-variables:

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

Adjacency matrix

$$\mathbf{X} = \begin{bmatrix} \cdot & x_{12} & x_{13} & x_{14} & x_{15} \\ x_{21} & \cdot & x_{23} & x_{24} & x_{25} \\ x_{31} & x_{32} & \cdot & x_{34} & x_{35} \\ x_{41} & x_{42} & x_{43} & \cdot & x_{45} \\ x_{51} & x_{52} & x_{53} & x_{54} & \cdot \end{bmatrix} = \begin{bmatrix} \cdot & 1 & 1 & 0 & 0 \\ 1 & \cdot & 1 & 0 & 0 \\ 1 & 1 & \cdot & 1 & 1 \\ 0 & 0 & 1 & \cdot & 0 \\ 0 & 0 & 1 & 0 & \cdot \end{bmatrix}$$

Where do ties come from?

- Ethnographic
  - Kapferer (1972)
- Archival
  - Padgett and Ansell (Marriage and business records)
  - Bright, Koskinen, Malm (court records)
- Name generator
- Resource generator
- Position generator

Modes of data collection

- Roster method
- Free recall
- Participant-aided sociograms

Lazega's (2001) lawfirm partners

Rooster method with following tie-definitions
"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

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			

US 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

US 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 (Nan Lin and co)

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 (van Der Gaag)

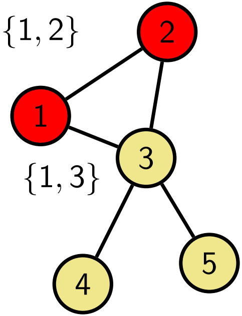
Table B: The SSND Resource Generator and responses: percentage of sample who mentioned at least one alter per resource item in any relationship, and strongest relationship when known (Survey on the Social Networks of the Dutch (SSND), 1999-2000; N=1,004).

		% '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	e
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	e
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	e
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 land	[illegible]	[illegible]	[illegible]	[illegible]	

Network metrics

Network metrics: degree

The degree of a node



The degree of a node $i \in V$:
The number of edges of the node
Example:
Node 1 has two ties

$$\{1,2\},\{1,3\}$$

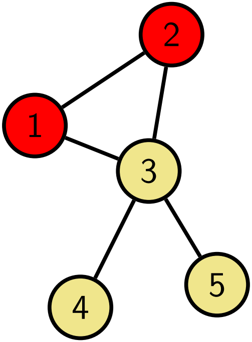
Node 1 has degree 2
Easily calculated as:

$$d_1 = x_{12} + x_{13} + x_{14} + x_{15} = 2$$

Node 3 has degree

$$d_3 = x_{31} + x_{32} + x_{34} + x_{35} = 1+1+1+1 = 4$$

The degree distribution



We can tabulate degrees

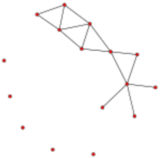
i	d_i
1	2
2	2
3	4
4	1
5	1

And we can tabulate the frequencies

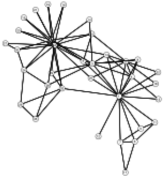
k	0	1	2	3	4
$D_k = \sharp\{i : d_i = k\}$	0	2	2	0	1

This is called the **degree distribution**

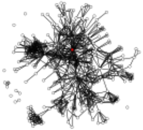
Density and average degree of a graph



$$\frac{1}{n}\sum_{i,j} x_{ij} = 1.88, \ \frac{1}{n(n-1)}\sum_{i<j} x_{ij} = 0.125$$

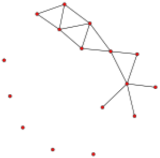


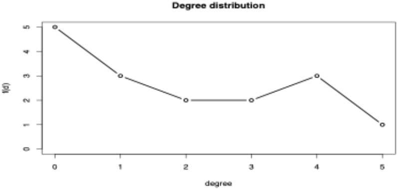
$$\frac{1}{n}\sum_{i,j} x_{ij} = 4.59 \ \frac{1}{n(n-1)}\sum_{i<j} x_{ij} = 0.14$$

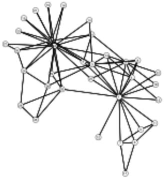


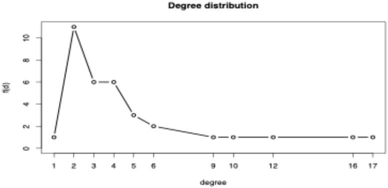
$$\frac{1}{n}\sum_{i,j} x_{ij} = 6.44 \ \frac{1}{n(n-1)}\sum_{i<j} x_{ij} = 0.018$$

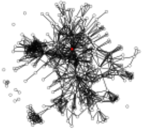
Density and average degree of a graph

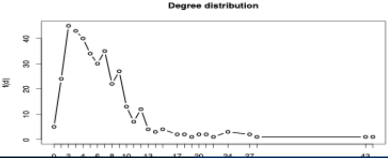




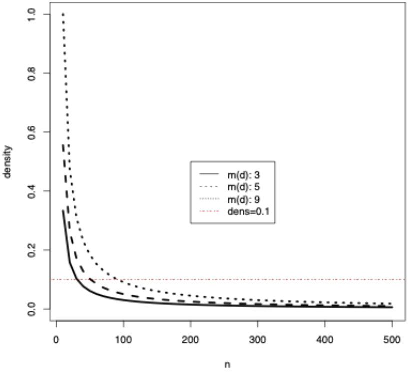






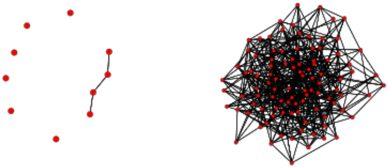


How does density scale?

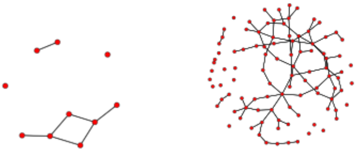


Density as a function of n
for different average degree

Density 10%



Average degree 2

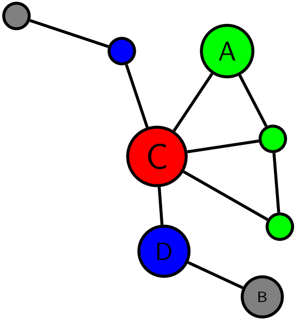


Network metrics

Network metrics: distance measures and reach

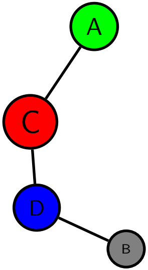
Path

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



For example e_1, e_2, e_3

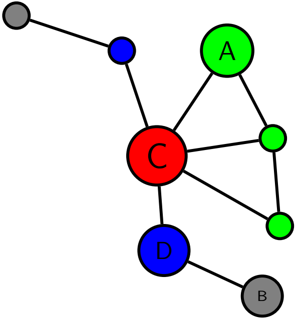
$e_1 = \{A, C\}, e_2 = \{C, D\}, e_3 = \{D, B\},$



Node sequence: A, C, D, B

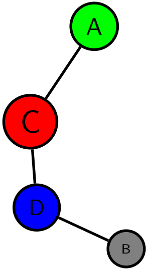
Length of a Path (1)

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



For example e_1, e_2, e_3

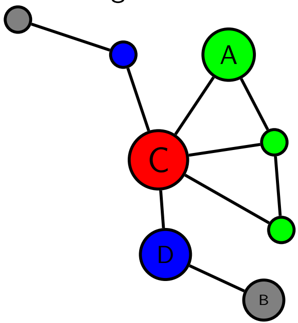
$e_1 = \{A, C\}, e_2 = \{C, D\}, e_3 = \{D, B\},$



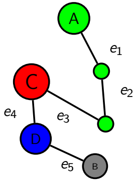
Length: $|\{e_1, e_2, e_3\}| = 3$

Length of a Path (2)

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



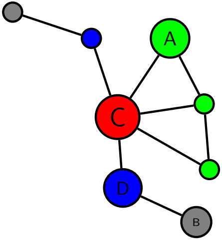
A path e_1, e_2, e_3, e_4, e_5 from A to B



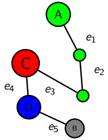
Length: $|\{e_1, e_2, e_3, e_4, e_5\}| = 5$

Geodesic Path

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

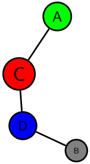


A path e_1, e_2, e_3, e_4, e_5 from A to B



Length: $|\{e_1, e_2, e_3, e_4, e_5\}| = 5$

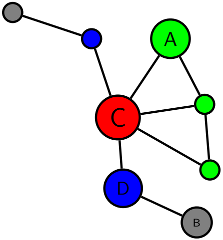
A path e_1, e_2, e_3 from A to B



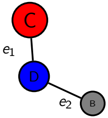
Length: $|\{e_1, e_2, e_3\}| = 3$

Geodesic Distance (1)

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

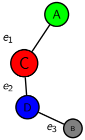


$d(C, B) = 2$



Length: $|\{e_1, e_2\}| = 2$

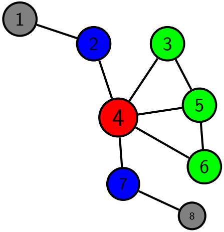
$d(A, B) = 3$



Length: $|\{e_1, e_2, e_3\}| = 3$

Geodesic Distance (2)

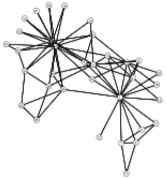
The **geodesic distance** between two nodes is the length of the geodesic

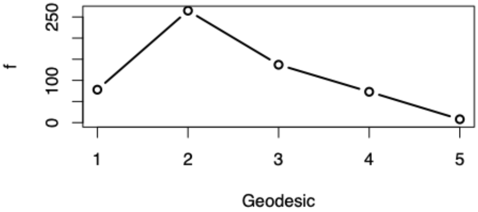


$i \backslash j$	1	2	3	4	5	6	7	8
1	-	1	3	2	3	3	3	4
2	-	-	2	1	2	2	2	3
3	-	-	-	1	1	2	2	3
4	-	-	-	-	1	1	1	2
5	-	-	-	-	-	1	2	3
6	-	-	-	-	-	-	2	3
7	-	-	-	-	-	-	-	1

Geodesic Distance (3)

Zackary's karate club

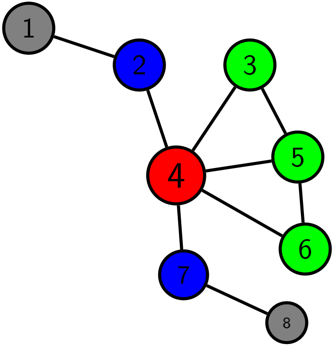




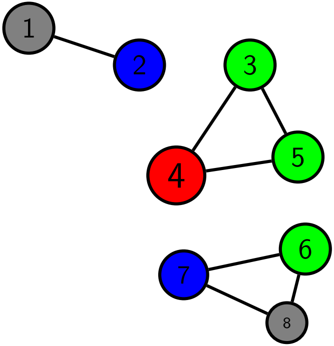
Connectedness

A graph is **connected** if there is path between any two nodes (every node is *reachable* from every other node)

Connected

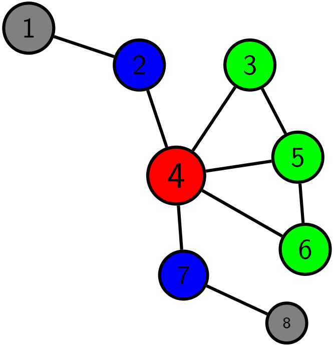


Not Connected

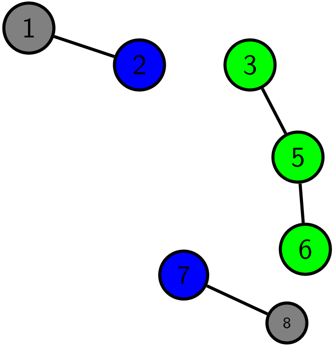


Cutpoint

A node is a **cutpoint** if the removal of the node disconnects the network

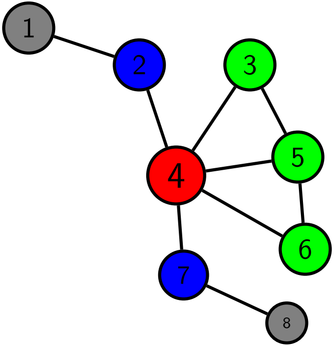


Remove 4:

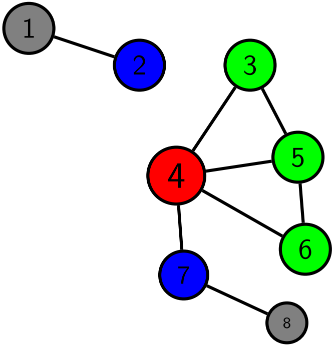


Bridge

An *edges* is a **bridge** if the removal of the edge disconnects the network



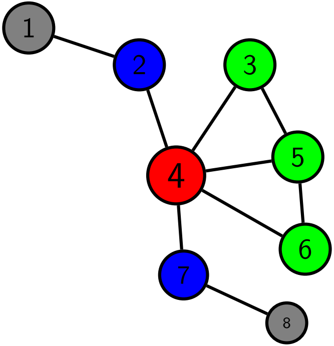
Remove $\{2, 4\}$:



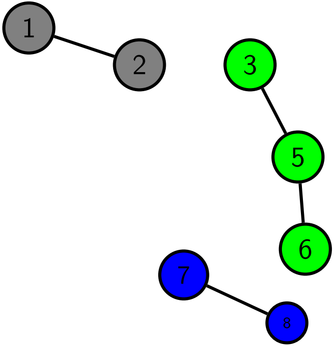
Component

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

One component



Three components

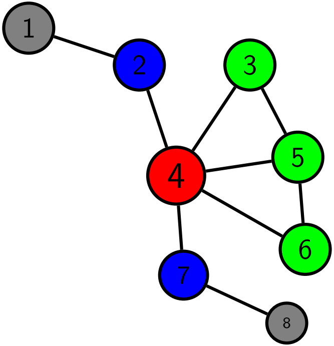


Network metrics

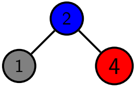
Network metrics: clustering and triads

Triad

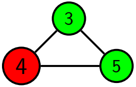
A **triad** is the induced *subgraph* that is induced by three (3) nodes



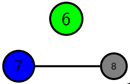
Examples:



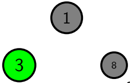
$x_{12} + x_{14} + x_{24} = 2$



$x_{34} + x_{35} + x_{45} = 3$



$x_{67} + x_{68} + x_{78} = 1$

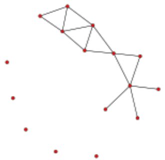


$x_{13} + x_{18} + x_{38} = 0$

The **triad census**: frequencies of triads labelled by their number of ties

Triad census: example

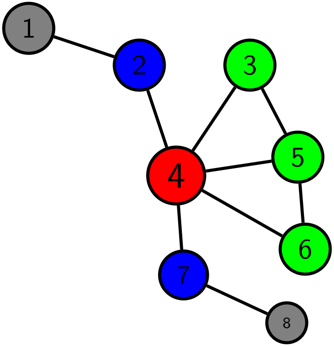
Padgetts Florentine families



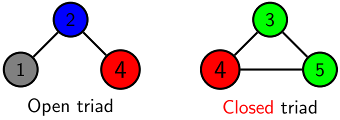
```
> triad.census(padgettbus, mode ='graph')
       0   1  2 3
[1,] 381 153 21 5
```

Triad closure (1)

A **closed triad** is a *triangle*



Examples:

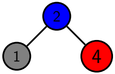


Triad closure (2)

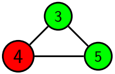
The **Clustering coefficient**

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

i.e. the proportion of open and closed triads that are closed



Open triad

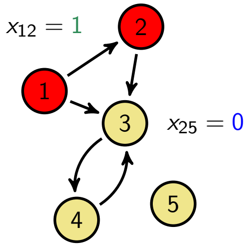


Closed triad

Directed networks

Generalisations for directed networks

Directed network data: Adjacency matrix



Tie-variables:

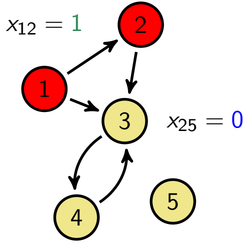
$$X_{ij} = \begin{cases} 1, & \text{if tie from } i \text{ to } j \\ 0, & \text{else} \end{cases}$$

Adjacency matrix

$$\mathbf{X} = \begin{bmatrix} \cdot & 1 & 1 & 0 & 0 \\ 0 & \cdot & 1 & 0 & 0 \\ 0 & 0 & \cdot & 1 & 0 \\ 0 & 0 & 1 & \cdot & 0 \\ 0 & 0 & 0 & 0 & \cdot \end{bmatrix}$$

Matrix **no longer symmetric**

Directed network data: degree distributions



We can tabulate degrees

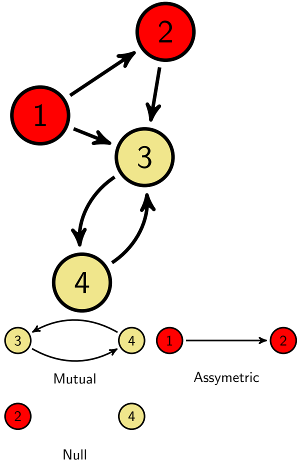
i	$d_i^{(out)}$	$d_i^{(in)}$
1	2	0
2	1	1
3	1	3
4	1	1
5	0	0

And we can tabulate the frequencies

k	0	1	2	3	4
$\sharp\{i : d_i^{(out)} = k\}$	1	3	1	0	0
$\sharp\{i : d_i^{(in)} = k\}$	2	2	0	3	0

Matrix **no longer symmetric**

Dyad census



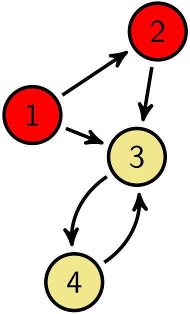
We have different types of dyads

i	j	x_{ij}	x_{ji}	type
1	2	1	0	A
1	3	1	0	A
1	4	0	0	N
2	3	1	0	A
2	4	0	0	N
3	4	1	1	M

And we can tabulate the frequencies

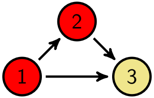
	M	A	N
Census	1	3	2

Directed triads

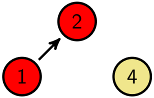


We have different types of Triads

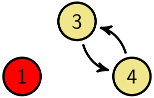
We lable using the MAN count



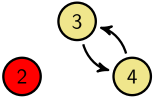
030T



012

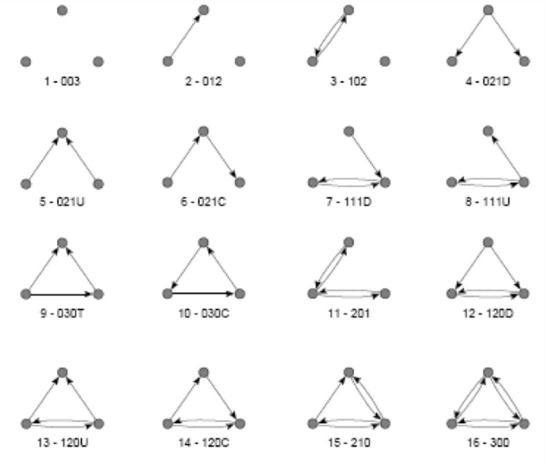


102



102

Triad census for directed graphs



Triad census: a count of all 16 types of triads (MAN)

Halfway summary

- 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

Non-parametric models

Comparing metrics to random networks

Non-parametric models

If we calculated a metric for a network, we might want to know if it is larger/smaller than we would expect by chance
Simple models of chance

- Bernoulli graphs
- Conditional uniform graphs
  - Conditional on density
  - Conditional on degree distribution (-s)
  - Dyad census (only directed)

How model tie-variables

Since binary

$$X_{12}, X_{13}, \ldots, X_{n(n-1)} \overset{i.i.d}{\sim} Bernoulli(p)$$

Independence:

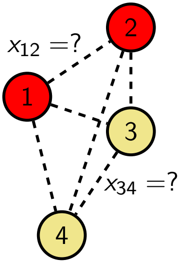
- Easy to estimate
- Can model p using logit or probit

If

$$Y = \sum_{i<j} X_{ij} \mapsto Y \sim Bin\left(\frac{1}{2}n(n-1), p\right)$$

We call X a **Bernoulli graph**

Bernoulli graph



Tie-variables:

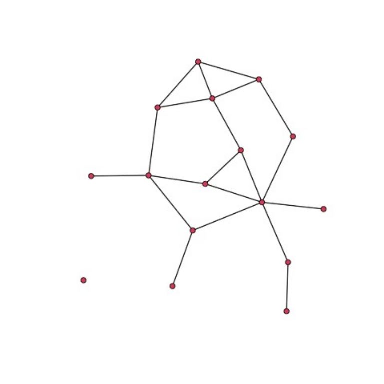
$$\Pr(X_{ij} = 1) = p$$

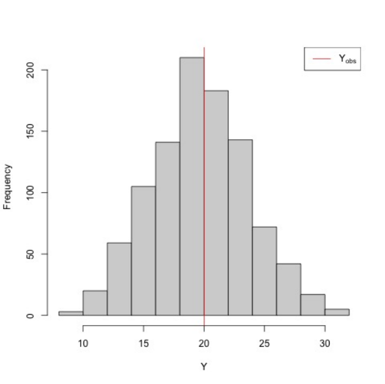
Probabilities

$$\begin{bmatrix} \cdot & p & p & p & p \\ - & \cdot & p & p & p \\ - & - & \cdot & p & p \\ - & - & - & \cdot & p \\ - & - & - & - & \cdot \end{bmatrix}$$

Example (1)

Example (Padgett Florentine Families Marriage ties, $n = 16$)

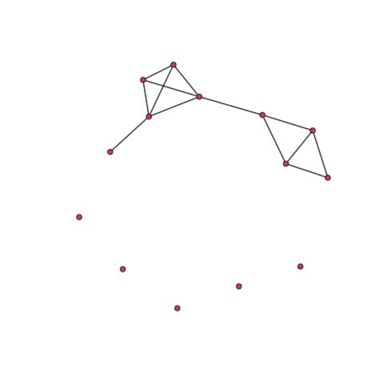


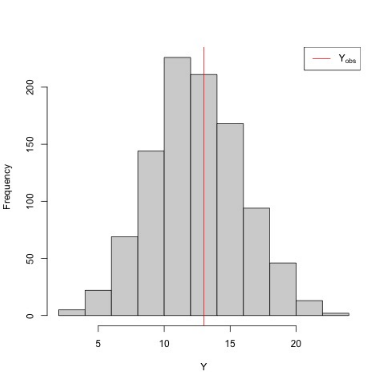


Comparing Y from Bernoulli graph and observed count Y_{obs}

Example (2)

Example (Red Highland Tribes, $n = 14$)

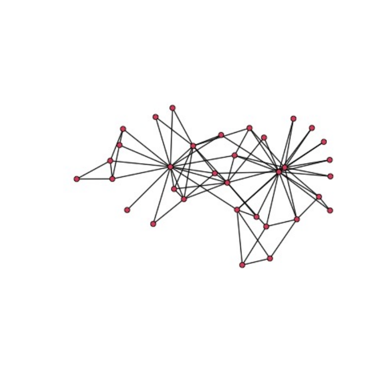


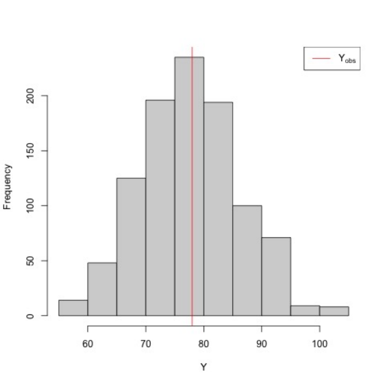


Comparing Y from Bernoulli graph and observed count Y_{obs}

Example (3)

Example (Zackary Karate Club, $n = 34$)

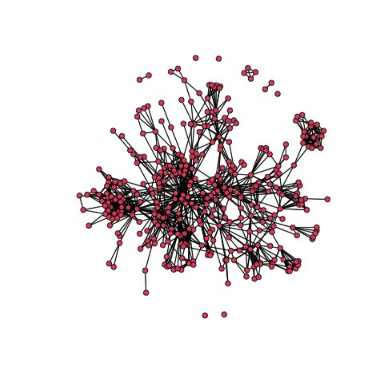


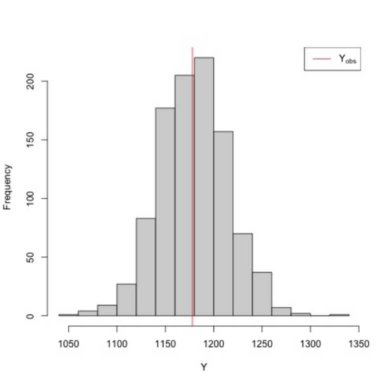


Comparing Y from Bernoulli graph and observed count Y_{obs}

Example (4)

Example (Sageman's Al-Qaeda datset, $n = 366$)

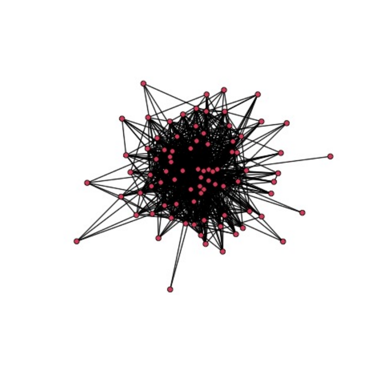


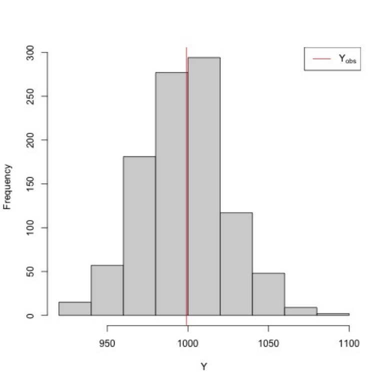


Comparing Y from Bernoulli graph and observed count Y_{obs}

Example (5)

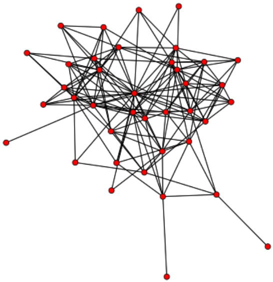
Example (Palotti Lazio Hospital transfers, $n = 85$)





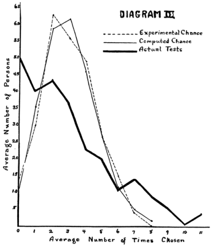
Comparing Y from Bernoulli graph and observed count Y_{obs}

Kapferer's (1972) taylors ($n = 39$)

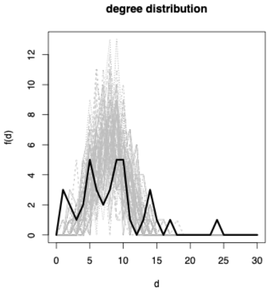


Kapferer's (1972) taylors ($n = 39$)

We can also look at DEGREE DISTRIBUTION



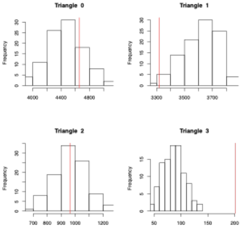
Moreno and Jennings (1934)



Observed degree distribution (black) vs 100 simulated (grey) from Bernoulli

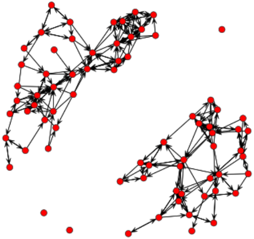
Kapferer's (1972) taylors ($n = 39$)

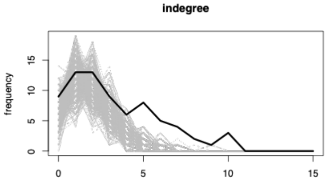
We can also look at TRIAD CENSUS



Directed network: Coleman's freshmen students ($n = 73$)(1)

Degree distribution

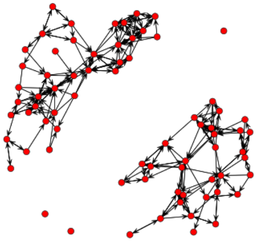


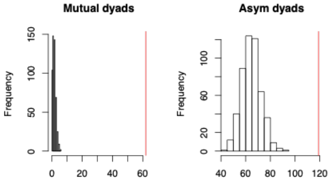


Bernoulli

Directed network: Coleman's freshmen students ($n = 73$)(2)

Dyad census

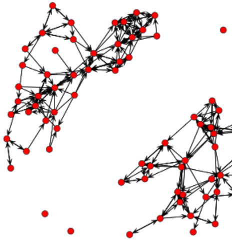


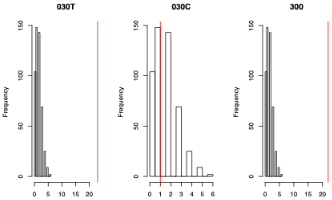


Bernoulli

Directed network: Coleman's freshmen students ($n = 73$) (3)

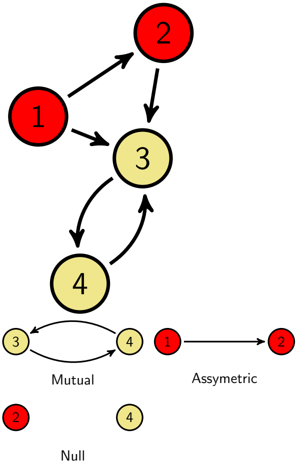
Triad census





Bernoulli

Conditional $U \mid MAN$



We have different types of dyads

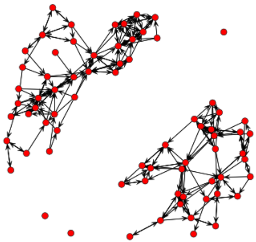
i	j	type
1	2	A
1	3	A
1	4	N
2	3	A
2	4	N
3	4	M

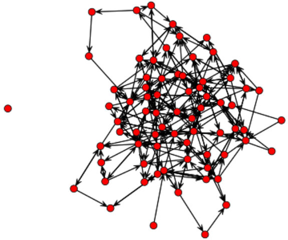
Randomize the observed DYADS Which preserves the dyad census

	M	A	N
Census	1	3	2

Coleman's freshmen - randomised dyads (1)

M: 62; A: 119; N: 2447

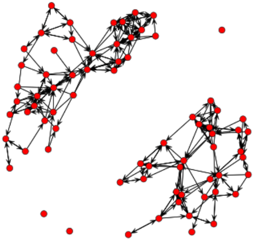


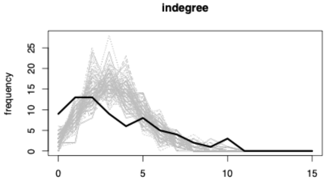


Random $U \mid MAN$

Coleman's freshmen - randomised dyads (2)

M: 62; A: 119; N: 2447

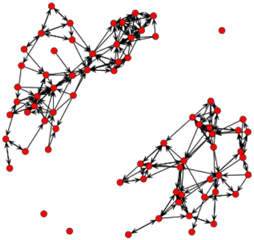


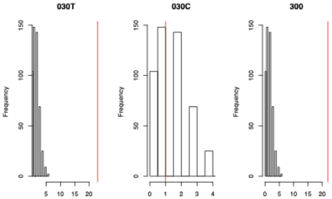


Random $U \mid MAN$

Coleman's freshmen - randomised dyads (3)

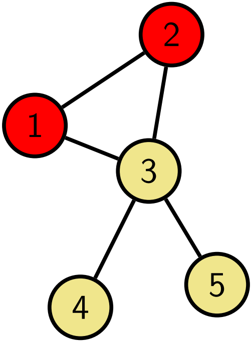
M: 62; A: 119; N: 2447





Random $U \mid MAN$

Conditionally uniform conditional on degrees



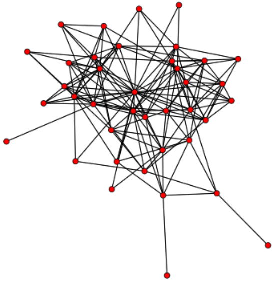
We can take the tabulate degrees

i	d_i
1	2
2	2
3	4
4	1
5	1

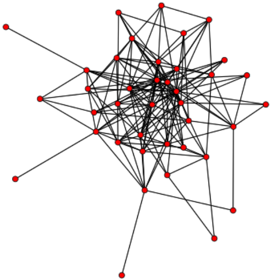
And generate random graphs **Y** with the exact same degree distribution

$$\sum_{j \neq i} y_{ij} = d_i, \; i = 1, \ldots, 5$$

Kapferer's (1972) taylors ($n = 39$)

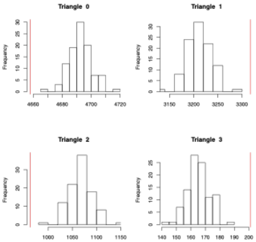


Observed



Random conditional on degrees

Kapferer's (1972) taylors ($n = 39$)



Random conditional on degrees

Does the degree distribution explain the clustering?

Halfway summary

- We can take any metrics
  - Representation: Graph, sets, edge list, adjacency matrix
  - Degree: density, degree, degree distribution
  - Reach: path, geodesic, distance, diameter
  - Clustering: clique, triads, closure
- And see if these differ from chance, given some constraints
  - Density: Bernoulli and $U \mid L$
  - Degree distribution: $U \mid d_1, \ldots, d_n$ (directed: either in- and out-degrees, or both)
  - Reciprocity: $U \mid MAN$
  - labels of nodes: QAP[1]

[1] We have not talked about these - the network is unchanged but nodes are randomized

Exponential random graph models

ERGM: modelling dependence

Exponential random graph models

None of the non-parametric models

$$Bern(p),\ U \mid MAN,\ U \mid d_1, \dots, d_n$$

get dyadic features (density, reciprocity), reach (distances), degree distribution, or clustering (triads), right
Is there a way of geting these ‘exactly’ **right** . . . ?

Fit of model assuming independence

Is the independence assumption true?

Independence

	x_{ik}		
x_{ij}	1	0	marginal
1	0.09	0.21	0.3
0	0.21	0.49	0.7
marginal	0.3	0.7	1

Because

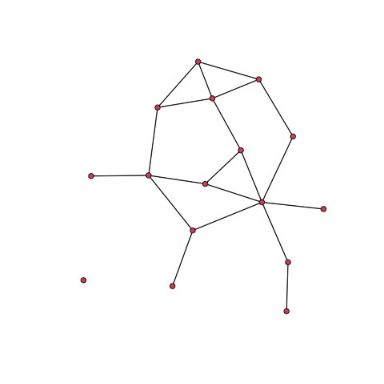
$$\Pr(X_{ij} = 1, X_{ik} = 1) = \Pr(X_{ij} = 1)\Pr(X_{ik} = 1)$$

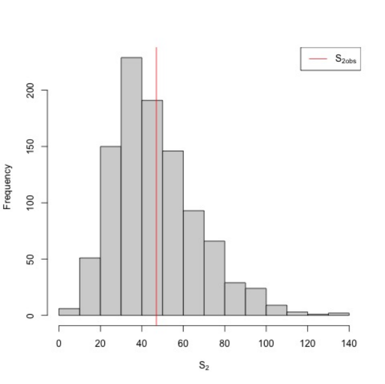
Let us look at the distribution

$$S_2 = \sum_{i} \sum_{j<k} X_{ij} X_{ik}$$

Example (1)

Example (Padgett Florentine Families Marriage ties, $n = 16$)

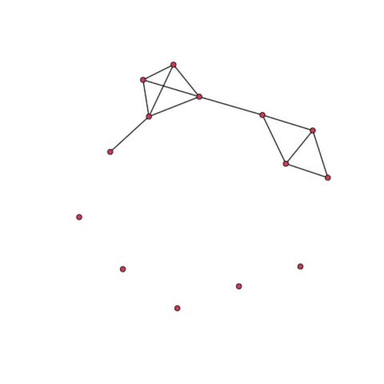


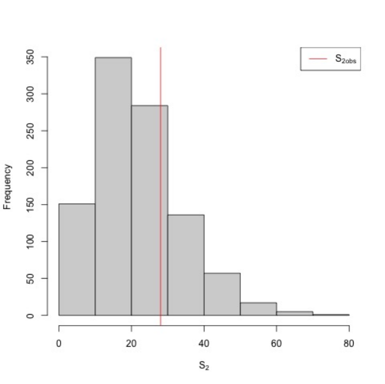


Comparing S_2 from Bernoulli graph and observed count $S_{2\mathrm{obs}}$

Example (2)

Example (Red Highland Tribes, $n = 14$)

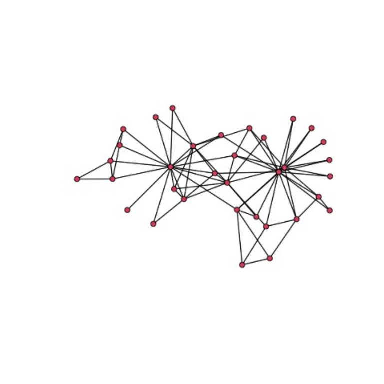


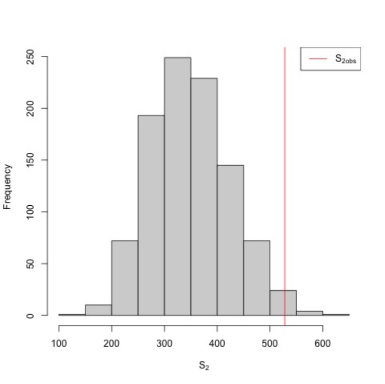


Comparing S_2 from Bernoulli graph and observed count $S_{2\text{obs}}$

Example (3)

Example (Zackary Karate Club, $n = 34$)

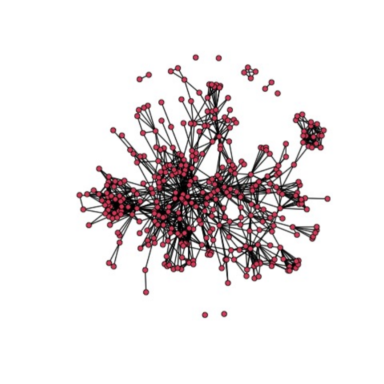


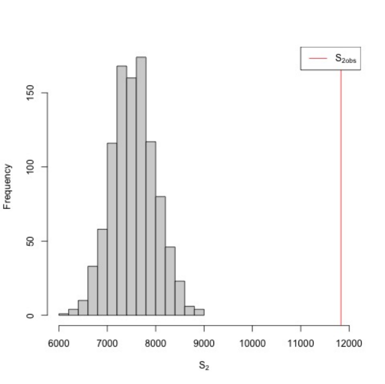


Comparing S_2 from Bernoulli graph and observed count $S_{2\mathrm{obs}}$

Example (4)

Example (Sageman's Al-Qaeda datset, $n = 366$)

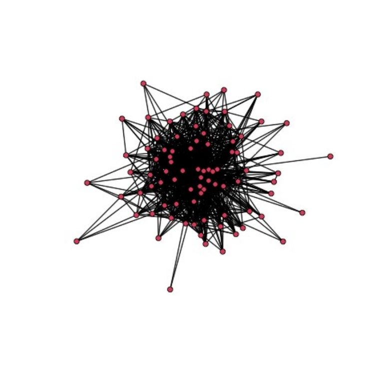


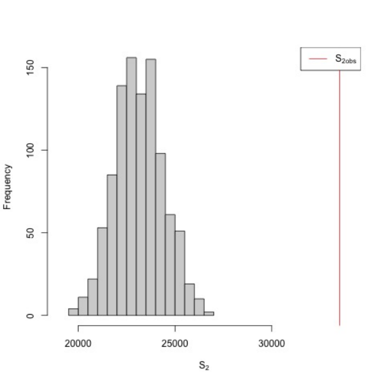


Comparing S_2 from Bernoulli graph and observed count $S_{2\mathrm{obs}}$

Example (5)

Example (Palotti Lazio Hospital transfers, $n = 85$)





Comparing S_2 from Bernoulli graph and observed count $S_{2\text{obs}}$

Interactions: Loglinear interpretation

Independence

	x_{ik} = 1	x_{ik} = 0	marginal
x_{ij} = 1	.12	.18	0.3
x_{ij} = 0	.28	.42	0.7
marginal	0.4	0.6	1

Non-Independence

	x_{ik} = 1	x_{ik} = 0	marginal
x_{ij} = 1	.29	.01	0.3
x_{ij} = 0	.11	.59	0.7
marginal	0.4	0.6	1

Joint probability completely determined by

- Marginal effects
  - $\theta_{ij} x_{ij}$
  - $\theta_{ik} x_{ik}$

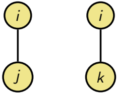
Joint probability requires

- Marginal effects
  - $\theta_{ij} x_{ij}$
  - $\theta_{ik} x_{ik}$
- and interaction effects
  - $\theta_{ij,ik} x_{ij} x_{ik}$

Interactions: Loglinear interpretation

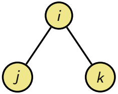
- Marginal effects
  - $\theta_{ij}x_{ij}$
  - $\theta_{ik}x_{ik}$

May account for different tie-frequencies



- and interaction effects
  - $\theta_{ij,ik}x_{ij}x_{ik}$

May account for *joint* occurrences
For example dependencies through cross-classification by nodes



What interactions non-zero $\iff$ what variables (cond.) dependent
(In a Besag (1974) sense)

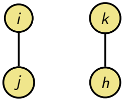
Markov dependence - dependence through nodes

Assume that tie-variables that *do not* share a node are conditionally independent

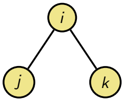
Markov dependence assumption (Frank and Strauss, 1986)

Two tie-variables X_{ij} and X_{kh} are conditionally independent (given the rest) if

$$\{i, j\} \cap \{k, h\} = \emptyset$$



$\{i, j\} \cap \{k, h\} = \emptyset$



$\{i, j\} \cap \{i, k\} = \{i\}$

Markov dependence - dependence graph

The Markov dependence implies a **dependence graph** D, whose node set is

$$X_{12}, X_{13}, \ldots, X_{n(n-1)}$$

and there is an *edge*

$$\{ij, kh\} \in D$$

in D if the variables X_{ij} and X_{kh} are conditionally dependent.

Dependence graphs and pmfs

Theorem (Frank and Strauss (1986))

The probability mass function for a graph $\mathbf{X}$ *with dependence graph* D

$$p(\mathbf{X}) = c^{-1} \exp\left\{ \sum_{A} \alpha_A \prod_{ij \in A} x_{ij} \right\} \quad (1)$$

where the sum is over subsets A *of tie-variables,*

- α_A *is non-zero if* A *is a clique in* D
- α_A *is zero otherwise*

and c *is a normalising constant*

Remark: D tells us exactly what *interactions* we need

Markov graphs

Theorem (Frank and Strauss (1986))

A graph $\mathbf{X}$ *with dependence graph* D *given by Markov dependence has probability*

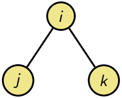
$$p(\mathbf{X}) = c^{-1} \exp \left\{ \sum_{i,j,k} \tau_{ijk} T_{ijk} + \sum_{k=1}^{n-1} \sum_{i_1,i_2,\ldots,i_k} \sigma_{i_1,i_2,\ldots,i_k} S_{i_1,i_2,\ldots,i_k} \right\} \quad (2)$$

with sufficient statistics,

- $T_{ijk} = x_{ij} x_{ik} x_{jk}$, *and*
- $S_{i_1,i_2,\ldots,i_k} = x_{i_1 i_2} x_{i_1 i_3} \cdots x_{i_1 i_k}$

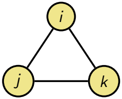
Markov graphs - sufficient statistics

Stars of order k



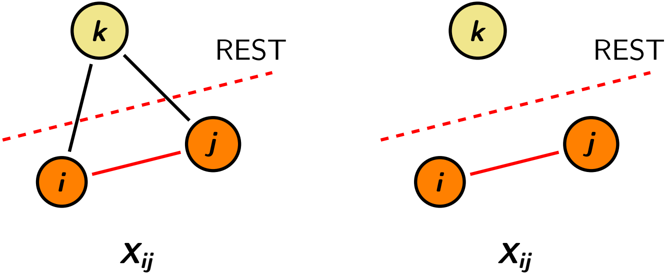
$$S_{i,j,k} = x_{ij}x_{ik}$$

Triangles



$$T_{i,j,k} = x_{ij}x_{ik}x_{jk}$$

Assuming independence (say logistic regression)
we cannot afford ‘I scratch your back’, ‘friends of my friends’, etc



With the Markov dependence assumption we can

Likelihood

ERGM with additional dependence assumptions (Snijders et al., 2006; Pattison & Robins, 2002)

defines a distribution on $\mathbf{X} \in \mathcal{X} = \{0,1\}^{\binom{V}{2}}$

ERGM pmf

$$p_{\theta}(\mathbf{X}) = \exp\{\theta^{\top} z(\mathbf{X}) - \psi(\theta)\}$$

where

- $\theta \in \mathrm{IR}^{p}$ $p \times 1$ vector of parameters
- z a $p \times 1$ vector of graph statistics
- $\psi(\theta) = \log \underbrace{\sum_{\mathbf{X} \in \mathcal{X}} \exp\{\theta^{\top} z(\mathbf{X})\}}_{\text{really many terms}}$ is a normalising constant

Degree

Why are degree-related metrics/configurations interesting?

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

- 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

Triadic closure

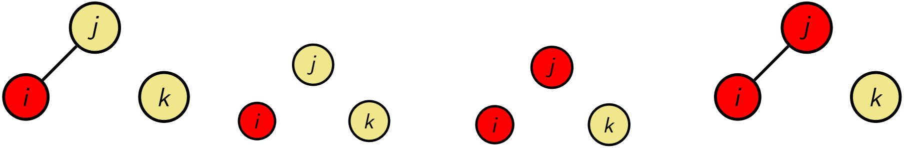
Why are triangles interesting?

- The triangle is the smallest group (in which there is a majority) (The Web of Group Affiliations, Simmel, 1922)
- Simmel (1955) friendship transitivity implies a social mechanism for closure; tension between dyad and triad
- "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)
- 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)
- Blau (1964) exchange theory (costs and benefits) explain groups
- Heider (1958) balance theory

Social influence v social selection

Homophilous ties provide evidence for **social influence**

Time t_0 Heterophily: $y_i \neq y_j$ $\quad t_0 < t < t_1 \quad \longmapsto \quad t' \in (t, t_1) \quad \longmapsto$ Time t_1 Homophily: $y_i = y_j$



but homophilous ties may be the result of **social selection**