The Social Statistics Discipline Area, School of Social Sciences

Introduction to statistical analysis of Social Networks

Johan Koskinen

http://www.ccsr.ac.uk/staff/jk.htm

johan.koskinen@manchester.ac.uk

Mitchell Centre for Network Analysis

Jan 29 2013, Statistical Analysis of Social Networks, YES Workshop on Statistics for Complex and High Dimensional Systems, Eindhoven

Statistics and networks?

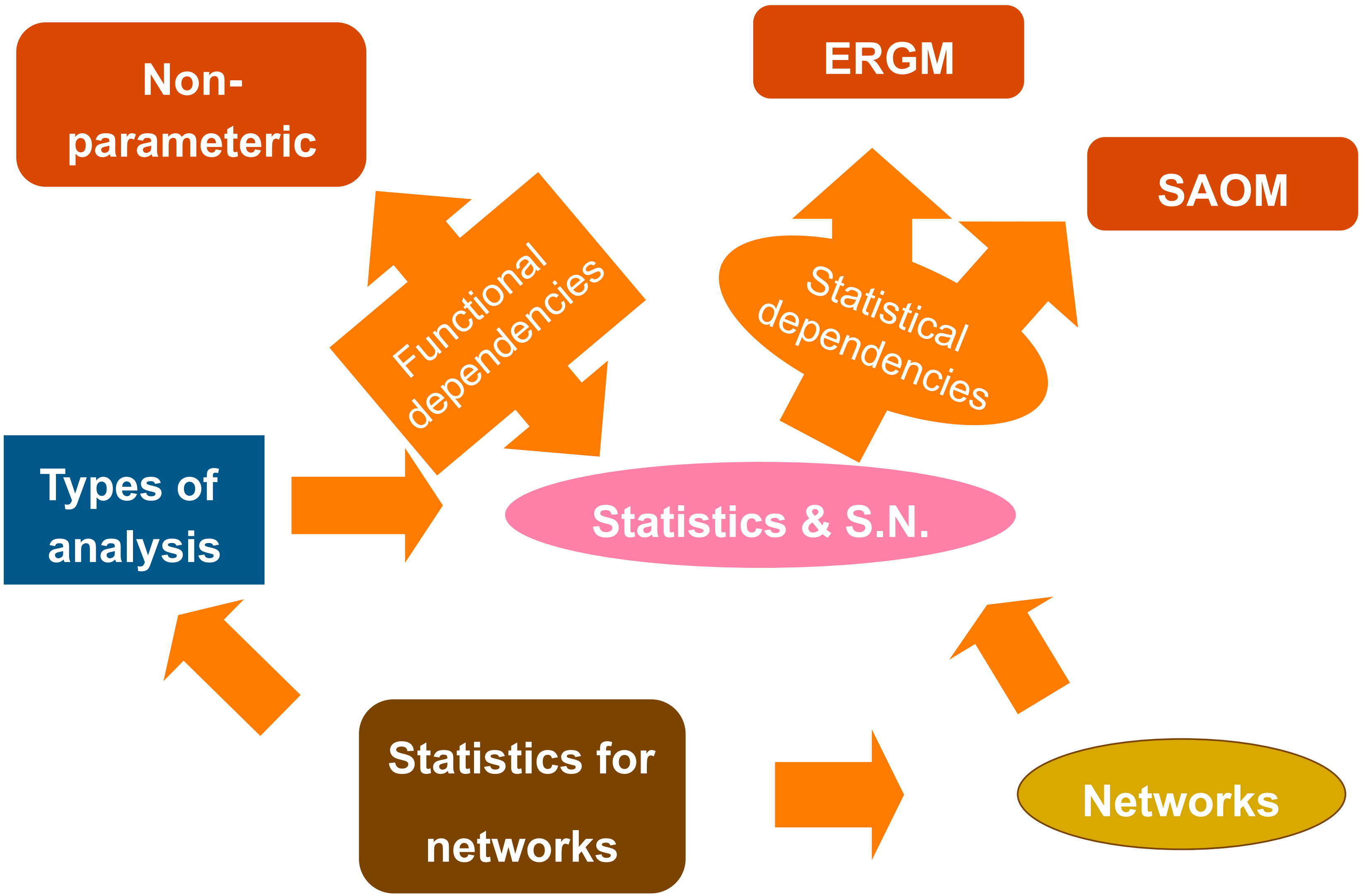
Why statistics?

- Is the network a unique narrative?
- Numbers in lieu of ethnography?

Possible answers

- Detecting systematic tendencies
- Social mechanisms
- Why not?

Outline

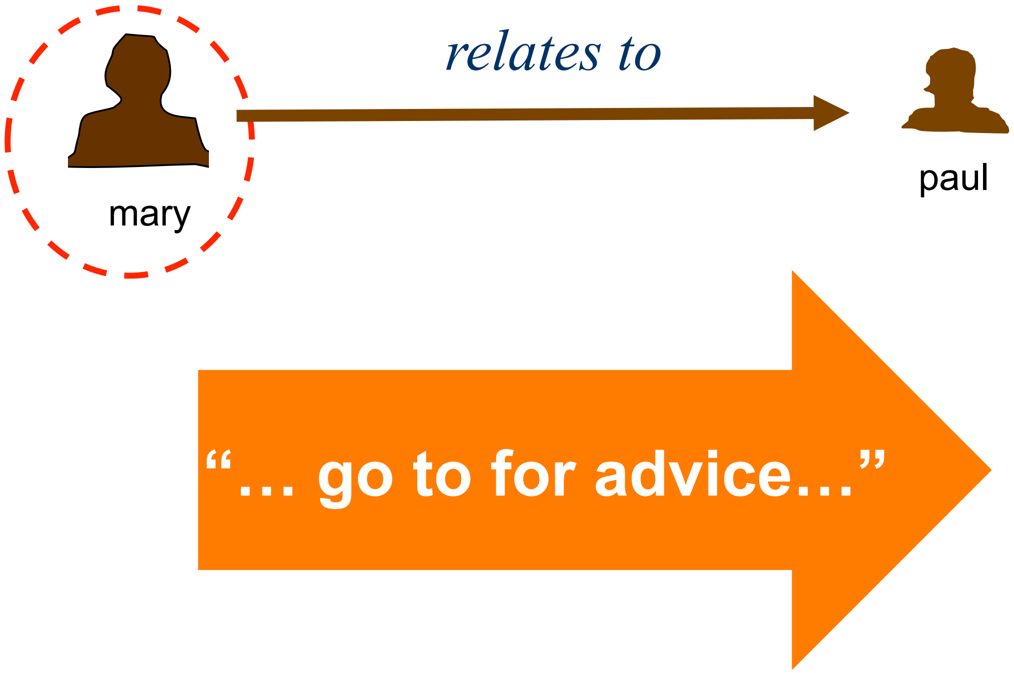

Non-parametric
ERGM
SAOM
Functional dependencies
Statistical dependencies
Types of analysis
Statistics & S.N.
Statistics for networks
Networks

Part 1

Social network data?

Social networks

We conceive of a network as a **Relation** defined on a collection of **individuals**



Social networks

We conceive of a network as a **Relation** defined on a collection of **individuals**

"… consider a friend…"

Social networks

We conceive of a network as a **Relation** defined on a collection of **individuals**

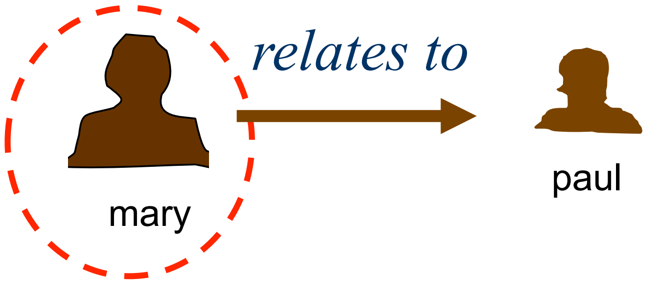
Generally binary

on → **Tie present**

off $\not\rightarrow$ **Tie absent**

Social networks

We conceive of a network as a **Graph**: $G(V,E)$, on



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

Relation: $E \subseteq \{(i,j) : i,j \in V\}$

Generally binary

on $\longrightarrow$ **Tie present**

off $\nrightarrow$ **Tie absent**

Social networks

We conceive of a network as a **Graph**: $G(V,E)$, on

$$i \xrightarrow{(i\,,\,j)} j$$

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

Relation: $E \subseteq \{(i,j) : i,j \in V\}$

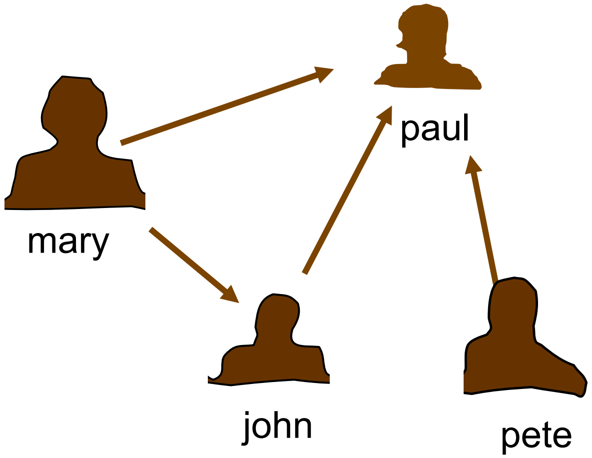
Generally binary

on $\longrightarrow$ **Tie present**

off $\not\longrightarrow$ **Tie absent**

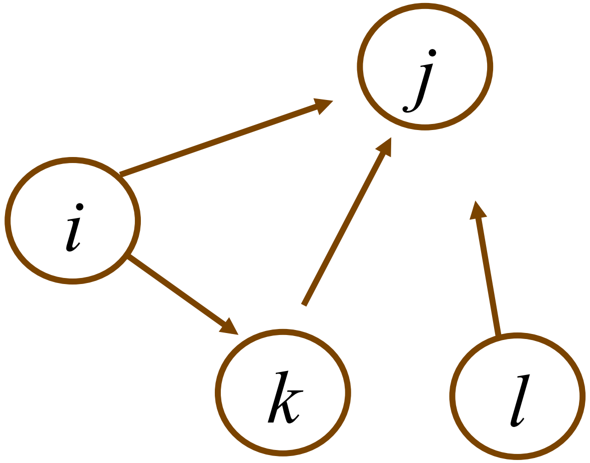
Social networks

We conceive of a network as a **Graph**: $G(V,E)$, on



Individuals: $V=\{i,j,k,l\}$

Relation: $E=\{(i,j),(i,k),(k,j),(l,j)\}$



MANCHESTER 1824 The University of Manchester

Social networks

We conceive of the **Graph** as a collection of

$$i \xrightarrow{(i\,,\,j)} j$$

Tie variables: $\{X_{ij}: i,j \in V\}$

$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$

Social networks

We conceive of the **Graph** as a collection of

$$i \xrightarrow{(i\,,\,j)} j$$

Tie variables: $\{X_{ij}: i,j \in V\}$

$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$

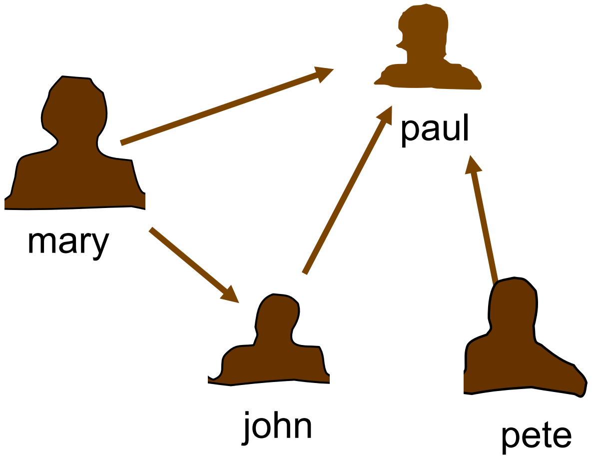
Generally binary

on $\longrightarrow$ $x_{ij} = 1$

off $\not\longrightarrow$ $x_{ij} = 0$

Social networks

We conceive of the **Graph** as a collection of



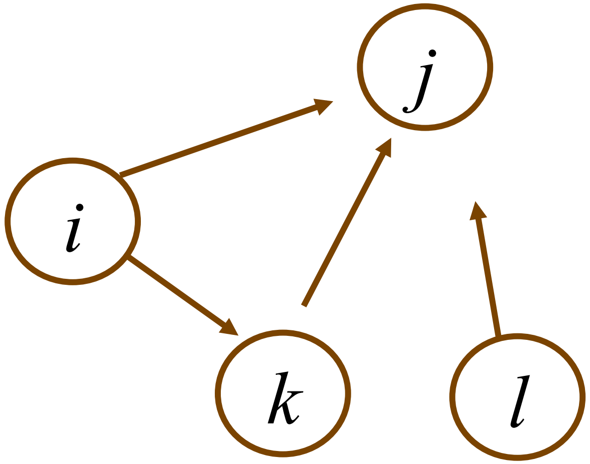
Tie variables: $\{X_{ij}: i,j \in V\}$

$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$

$$x = \begin{array}{c} i \\ j \\ k \\ l \end{array} \left(\begin{array}{|c|c|c|c|} \hline - & x_{ij} & x_{ik} & x_{il} \\ \hline x_{ji} & - & x_{jk} & x_{jl} \\ \hline x_{ki} & x_{kj} & - & x_{kl} \\ \hline x_{li} & x_{lj} & x_{lk} & - \\ \hline \end{array} \right) = \begin{array}{c} i \\ j \\ k \\ l \end{array} \left(\begin{array}{|c|c|c|c|} \hline - & 1 & 1 & 0 \\ \hline 0 & - & 0 & 0 \\ \hline 0 & 1 & - & 0 \\ \hline 0 & 1 & 0 & - \\ \hline \end{array} \right)$$

Social networks

We conceive of the **Graph** as a collection of



Tie variables: $\{X_{ij}: i,j \in V\}$

$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$

$$x = \begin{array}{c} i \\ j \\ k \\ l \end{array} \begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} = \begin{array}{c} i \\ j \\ k \\ l \end{array} \begin{pmatrix} - & 1 & 1 & 0 \\ 0 & - & 0 & 0 \\ 0 & 1 & - & 0 \\ 0 & 1 & 0 & - \end{pmatrix}$$

The **Adjacency matrix**:

The matrix of the collection **Tie var.** $\{X_{ij}: i,j \in V\}$

$$x = \begin{array}{c} i \\ j \\ k \\ l \end{array} \begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix}$$

The **Adjacency matrix**:

The matrix of the collection **Tie var.** $\{X_{ij}: i,j \in V\}$

$$x = \begin{array}{c} i \\ j \\ k \\ l \end{array} \begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix}$$

out-degree

$$x_{i+} = \sum_j x_{ij}$$

The **Adjacency matrix**:

The matrix of the collection **Tie var.** $\{X_{ij}: i,j \in V\}$

$$x = \begin{array}{c} i \\ j \\ k \\ l \end{array} \left(\begin{array}{cccc} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{array} \right)$$

In-degree

$$x_{+i} = \sum_j x_{ji}$$

The **Adjacency matrix**:

The matrix of the collection **Tie var.** $\{X_{ij}: i,j \in V\}$

$$x = \begin{array}{c} i \\ j \\ k \\ l \end{array} \begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \times \quad x_{i\bullet}x_{k\bullet}^{T} = \sum_j x_{ij}x_{kj}$$

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
```

number of nodes

density (#arcs/#possible arcs)

number of cells

$$x = \begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix}$$

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
```

number of nodes

density (#arcs/#possible arcs)

number of cells

$$x = \begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix}$$

No diagonal (self-nominations)

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
```

Print matrix to screen

```
> x
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]    0    1    0    0    0    0    0    0    0     0
 [2,]    0    0    0    0    0    0    0    0    0     0
 [3,]    1    1    0    0    1    0    1    1    1     1
 [4,]    0    0    0    0    1    1    1    0    0     0
 [5,]    1    0    1    0    0    1    0    0    0     0
 [6,]    0    1    0    1    0    0    0    0    1     0
 [7,]    1    0    0    0    1    0    0    0    1     0
 [8,]    0    0    1    0    0    0    1    0    1     1
 [9,]    1    0    1    0    0    1    0    0    0     0
[10,]    1    0    0    0    0    1    1    0    0     0
```

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
```

Print matrix to screen

To sum third row

```
> x
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]    0    1    0    0    0    0    0    0    0     0
 [2,]    0    0    0    0    0    0    0    0    0     0
 [3,]    1    1    0    0    1    0    1    1    1     1
 [4,]    0    0    0    0    1    1    1    0    0     0
 [5,]    1    0    1    0    0    1    0    0    0     0
 [6,]    0    1    0    1    0    0    0    0    1     0
 [7,]    1    0    0    0    1    0    0    0    1     0
 [8,]    0    0    1    0    0    0    1    0    1     1
 [9,]    1    0    1    0    0    1    0    0    0     0
[10,]    1    0    0    0    0    1    1    0    0     0
```

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
```

Print matrix to screen

To sum third row

```
> sum(x[3,])
[1] 7
```

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
```

To sum all rows

```
> x
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]    0    1    0    0    0    0    0    0    0     0
 [2,]    0    0    0    0    0    0    0    0    0     0
 [3,]    1    1    0    0    1    0    1    1    1     1
 [4,]    0    0    0    0    1    1    1    0    0     0
 [5,]    1    0    1    0    0    1    0    0    0     0
 [6,]    0    1    0    1    0    0    0    0    1     0
 [7,]    1    0    0    0    1    0    0    0    1     0
 [8,]    0    0    1    0    0    0    1    0    1     1
 [9,]    1    0    1    0    0    1    0    0    0     0
[10,]    1    0    0    0    0    1    1    0    0     0
```

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
```

To sum all rows

```
> rowSums(x)
 [1] 1 0 7 3 3 3 3 4 3 3
```

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
```

To sum all rows

```
> x
```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	
[1,]	0	1	0	0	0	0	0	0	0	0	1
[2,]	0	0	0	0	0	0	0	0	0	0	0
[3,]	1	1	0	0	1	0	1	1	1	1	7
[4,]	0	0	0	0	1	1	1	0	0	0	3
[5,]	1	0	1	0	0	1	[illegible]	0	0	0	3
[6,]	0	1	0	1	0	0	0	0	1	0	3
[7,]	1	0	0	0	1	0	0	0	1	0	3
[8,]	0	0	1	0	0	0	1	0	1	1	4
[9,]	1	0	1	0	0	1	0	0	0	0	3
[10,]	1	0	0	0	0	1	1	0	0	0	3

Out-degree distribution

Social networks – example in R

Let's create an **Adjacency matrix**:

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
colSums(x)
```

To sum all columns

in-degree distribution

```
> colSums(x)
 [1] 5 3 3 1 3 4 4 1 4 2
```

To draw the **Graph**

$$i \xrightarrow{(i\,,\,j)} j$$

Tie variables: $\{X_{ij}: i,j \in V\}$

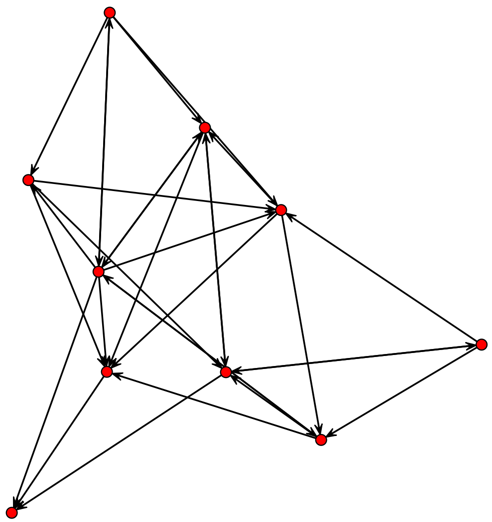
$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$

To draw the **Graph**

$$i \xrightarrow{(i\,,\,j)} j$$

Tie variables: $\{X_{ij}: i,j \in V\}$

$$x_{ij} = \begin{cases} 1 & \text{if } i \rightarrow j \\ 0 & \text{otherwise} \end{cases}$$



?

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	0	1	0	0	0	0	0	0	0	0
[2,]	0	0	0	0	0	0	0	0	0	0
[3,]	1	1	0	0	1	0	1	1	1	1
[4,]	0	0	0	0	1	1	1	0	0	0
[5,]	1	0	1	0	0	1	0	0	0	0
[6,]	0	1	0	1	0	0	0	0	1	0
[7,]	1	0	0	0	1	0	0	0	1	0
[8,]	0	0	1	0	0	0	1	0	1	1
[9,]	1	0	1	0	0	1	0	0	0	0
[10,]	1	0	0	0	0	1	1	0	0	0

To draw the **Graph**

load package “network”

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
colSums(x)
library('network')
```

To draw the **Graph**

load package “network”

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
colSums(x)
library('network')
myGraph <- as.network(x)
```

Transform the **adjacency matrix** to a “**network object**”

To draw the **Graph**

load package "network"

```
x <- matrix(rbinom(100,1,.4),10,10)
diag(x) <- 0
x
sum(x[3,])
rowSums(x)
colSums(x)
library('network')
myGraph <- as.network(x)
plot(myGraph)
```

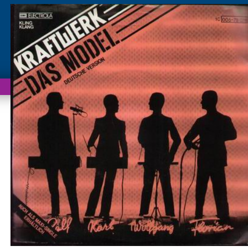
plot the new "**network object**"

Part 2

Modes of analysis of Social network data?

MANCHESTER
1824
The University of Manchester

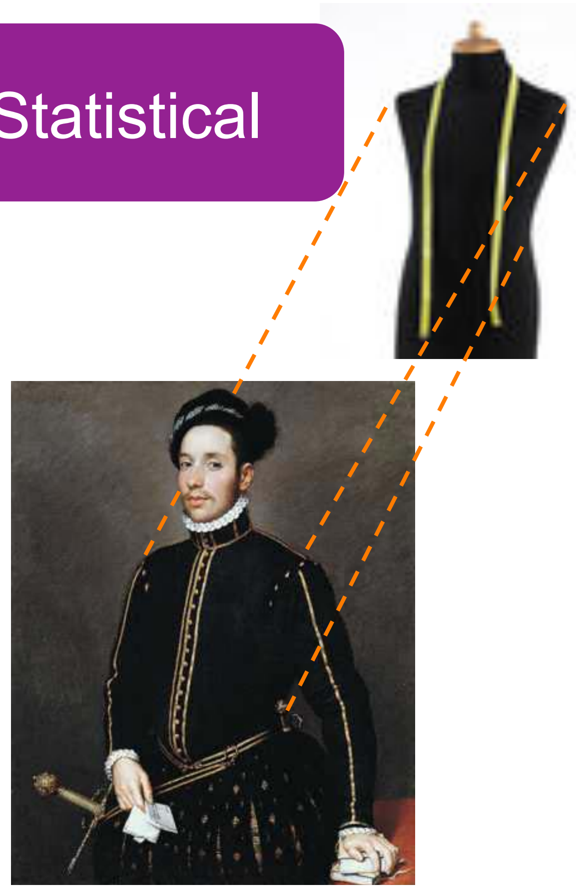
Modes of Analysis SNA



Graphical

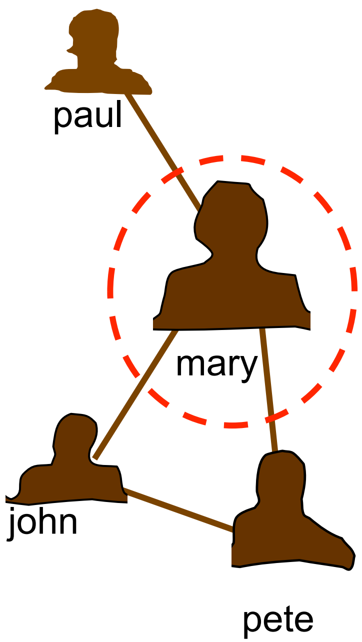
Descriptive

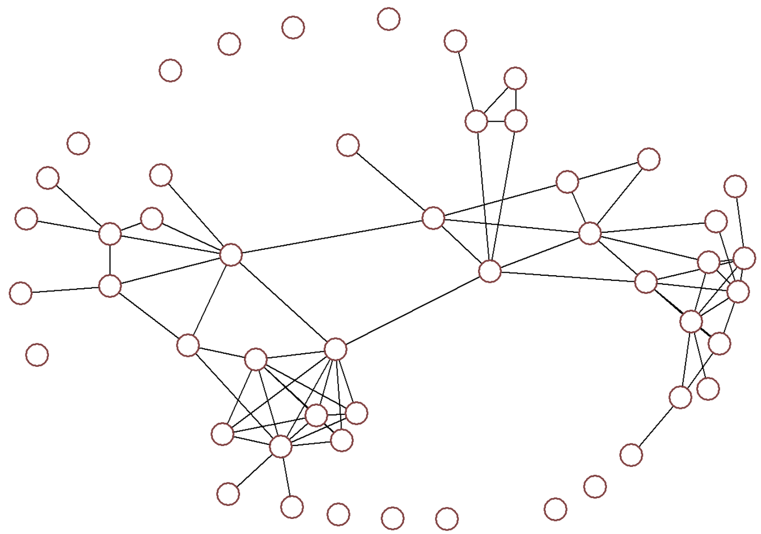
Statistical



Modes of Analysis SNA

Graphical

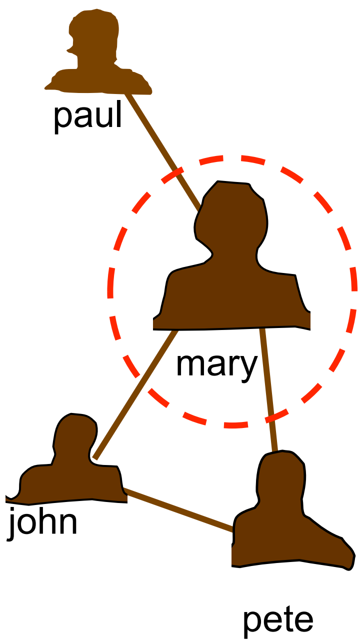


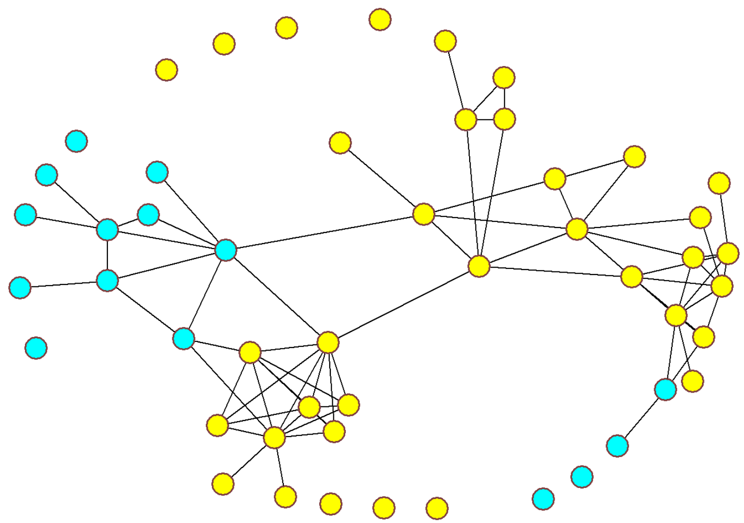


A social network of tertiary students – Kalish (2003)

Modes of Analysis SNA

Graphical





Yellow: Jewish Blue: Arab

A social network of tertiary students – Kalish (2003)

Modes of Analysis SNA

Descriptive

Centrality index

mary

john

pete

paul

Density

	arab	jew
arab	medium	low
jew		high

Modes of Analysis SNA

Statistical

“nonparametric”

Centrality index

mary

john

pete

paul

Differences in centrality may be explained by chance

Density

Differences in densities unlikely if classes assumed “equal”

	arab	jew
arab	medium	low
jew		high

Modes of Analysis SNA

Statistical

model based

Bernoulli

mary

john

pete

paul

Ties are distributed independently with parameter

$$\hat{p} = \frac{4}{6}$$

The network may be described by an

- a priori **BBM**

- **social selection ERGM** with separate effects for **clustering** and **homophily** on race

	arab	jew
arab	medium	low
jew		high

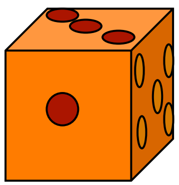
Part 2

Background: statistical analysis

Statistical analysis – why statistics?

Why statistics?

Statistics ≈ assessing whether observed measured quantities are ”big”
≈ reject chance or not



six in 50 out of 51:
balanced dice?

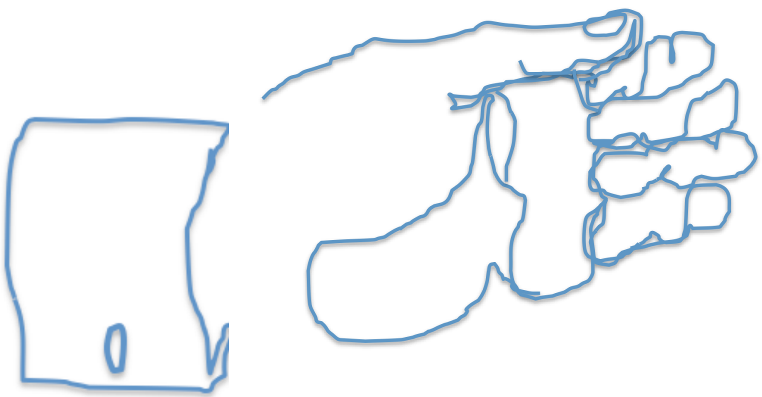
Networks not as easy

- Good model for chance in SNA?
- Model to capture systematic patterns (the typical)

Can’t we simply do t-tests?

Consider testing:

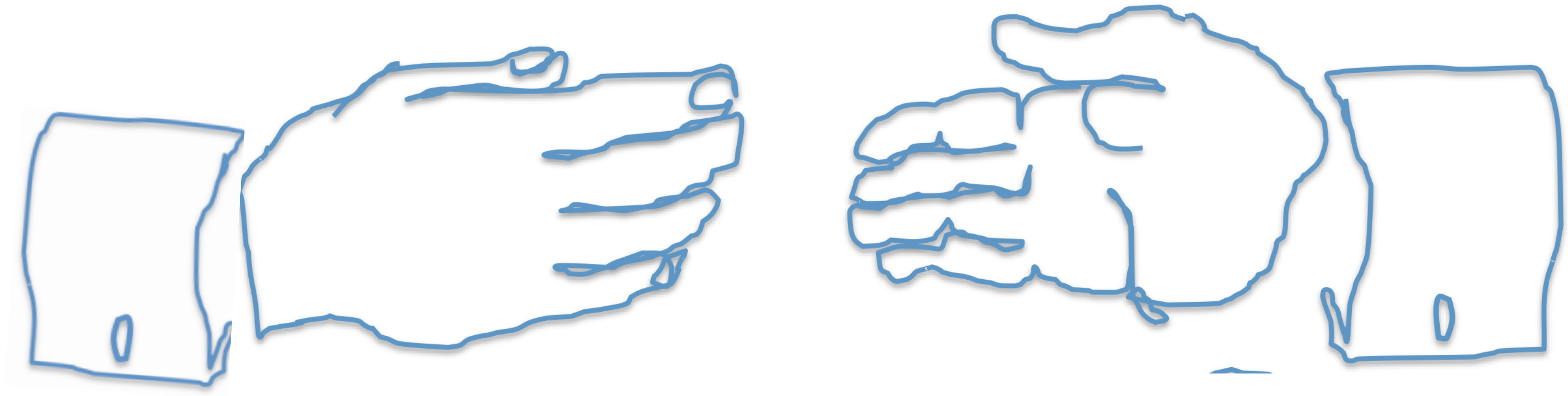
“I give advice…”



Can't we simply do t-tests?

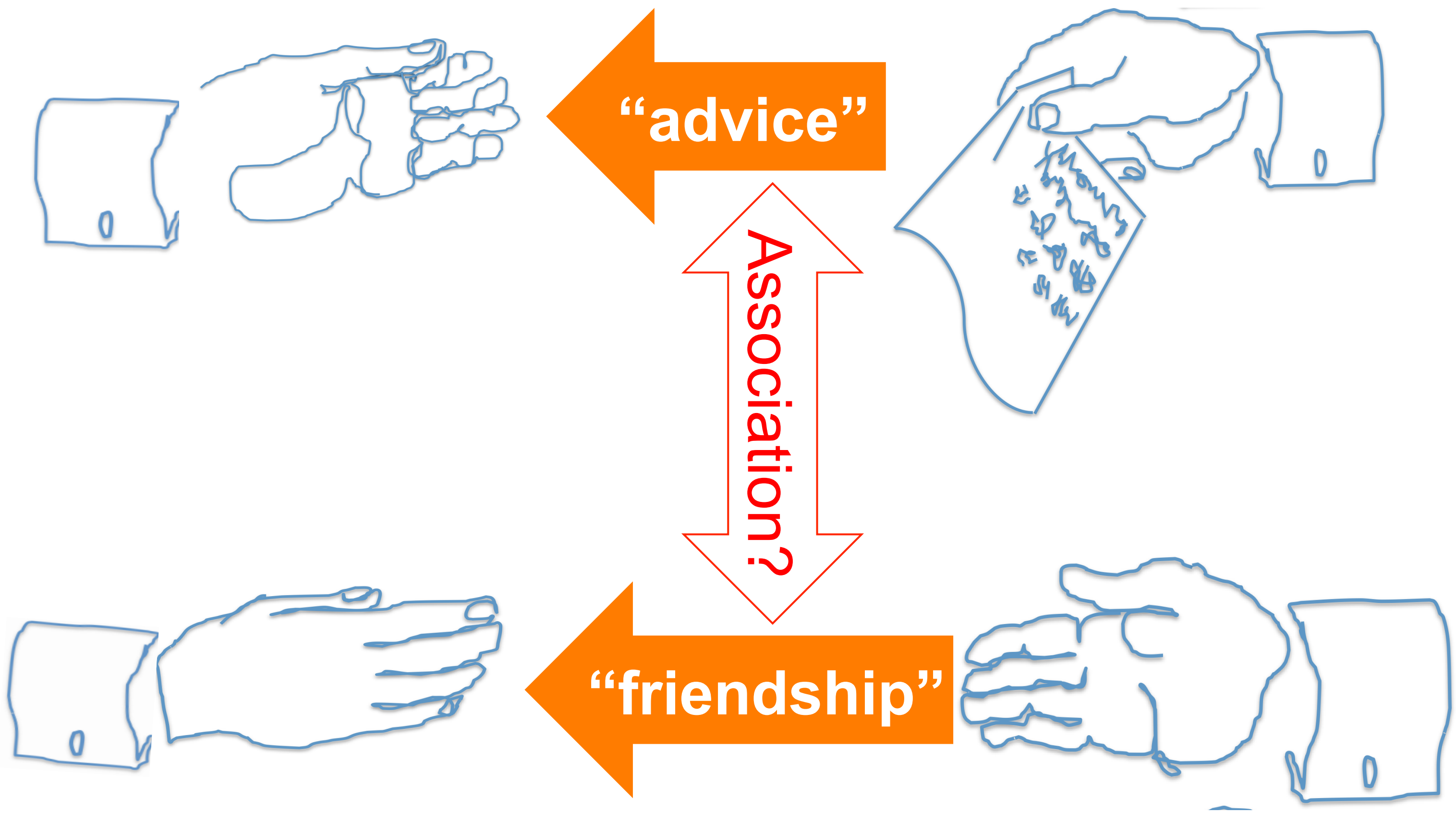
Consider testing:

"… to people I consider my friends"



Can't we simply do t-tests?

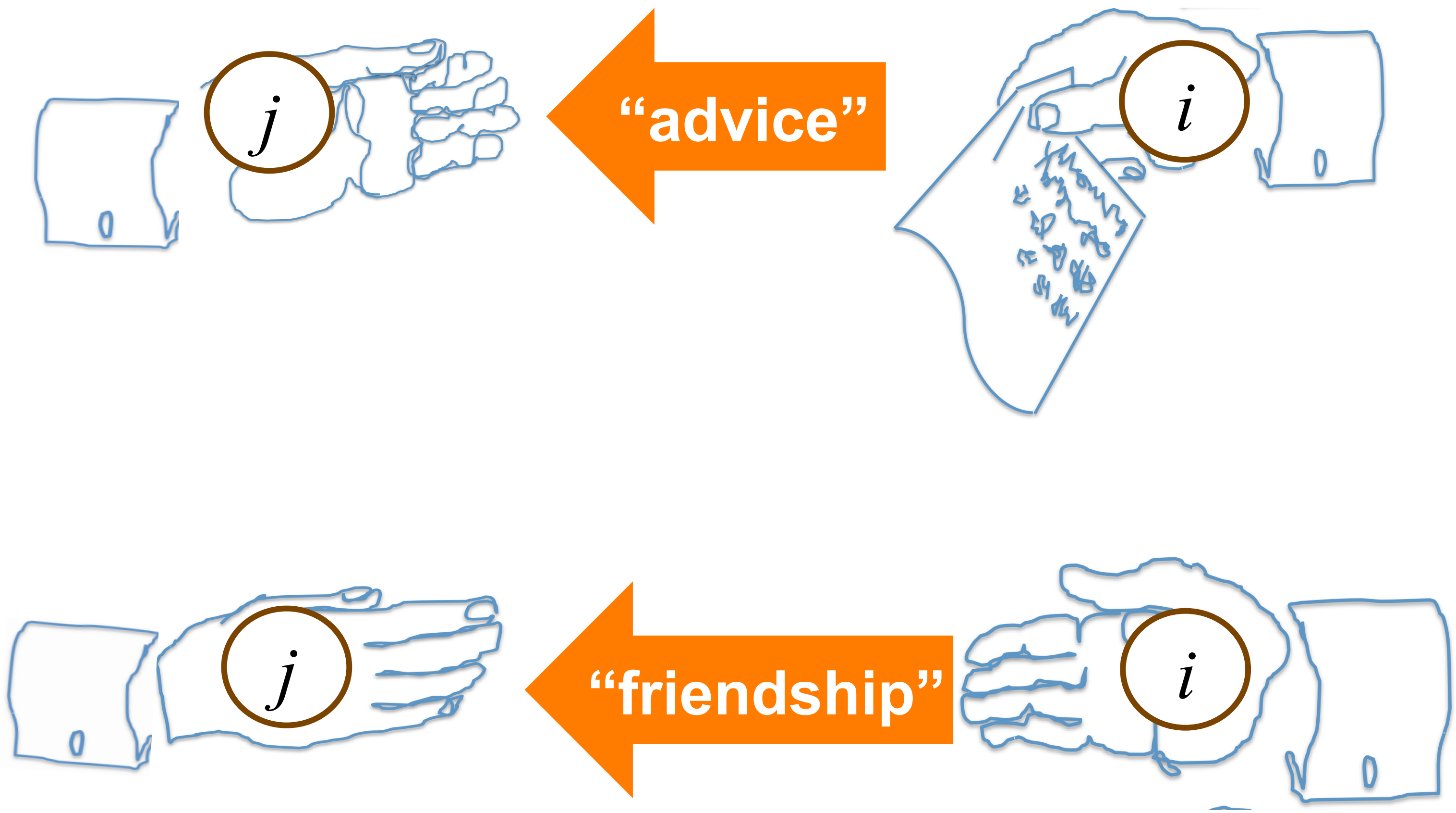
Consider testing:



Can't we simply do t-tests?

Consider testing:

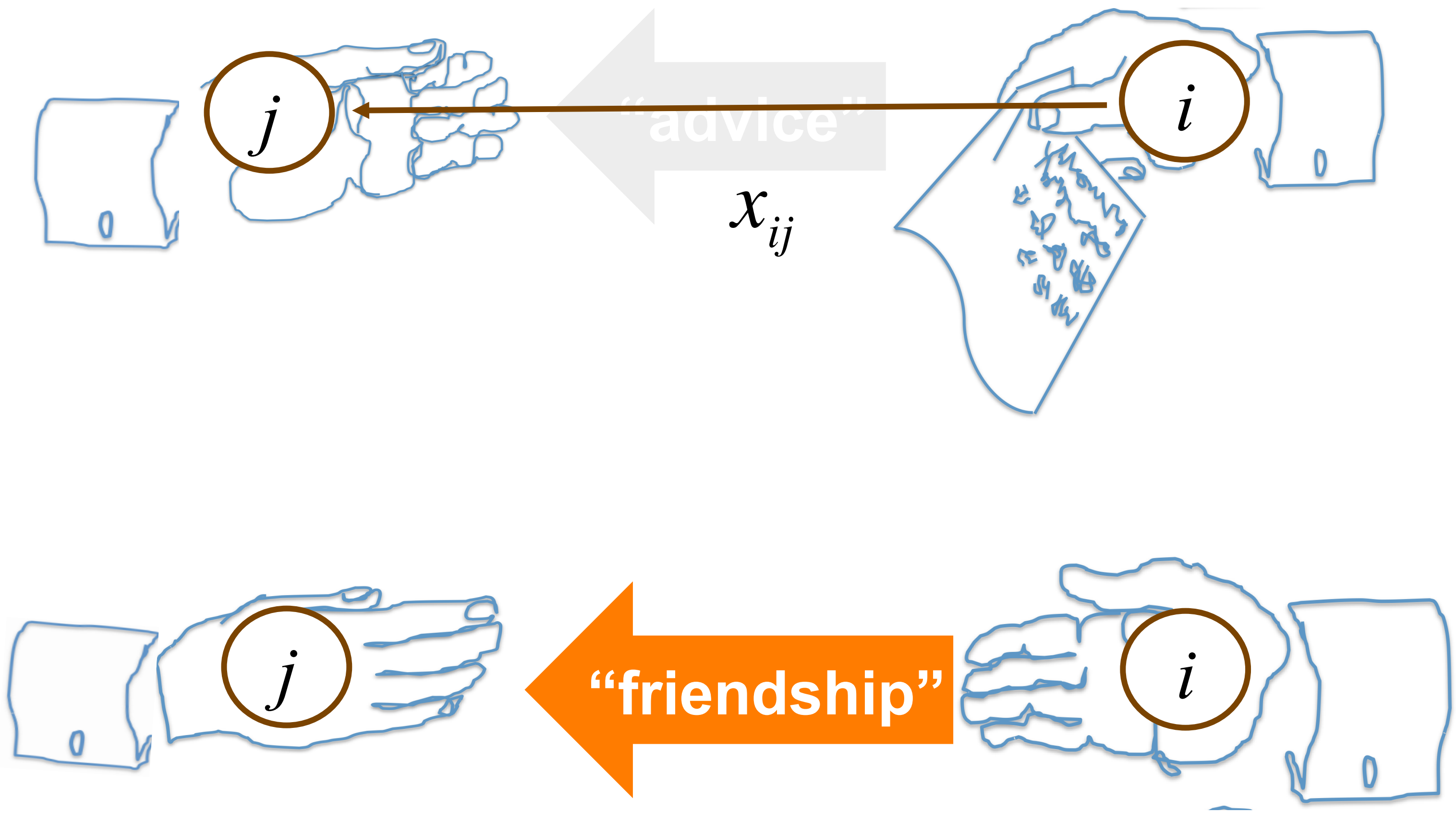
Correlate advice x with friendship y?



Can’t we simply do t-tests?

Consider testing:

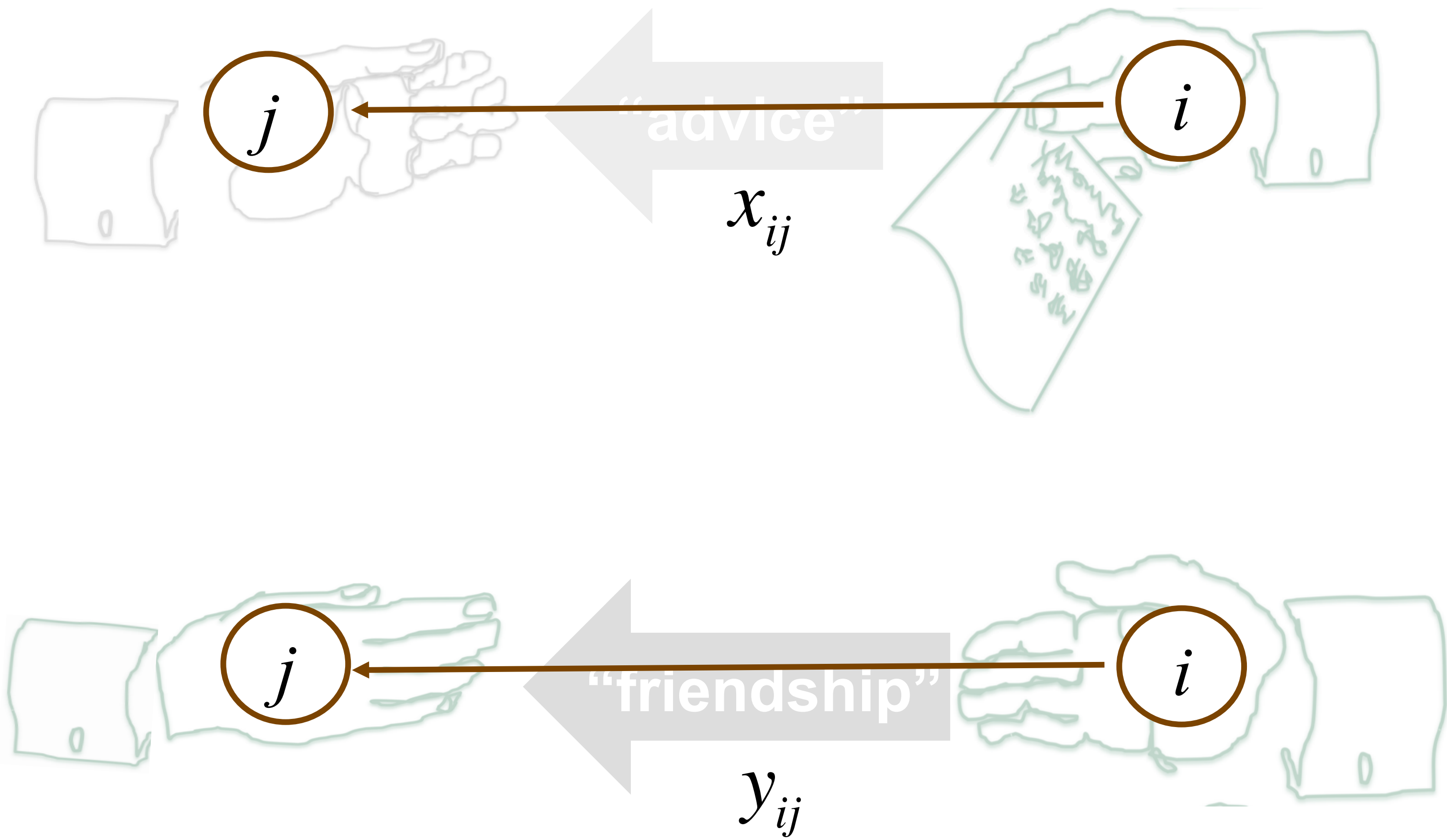
Correlate advice x with friendship y?



Can’t we simply do t-tests?

Consider testing:

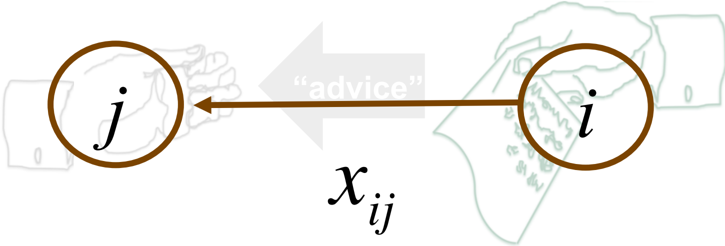
Correlate advice x with friendship y?

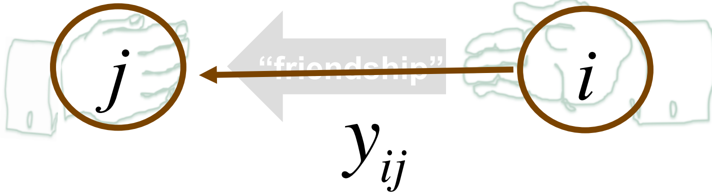


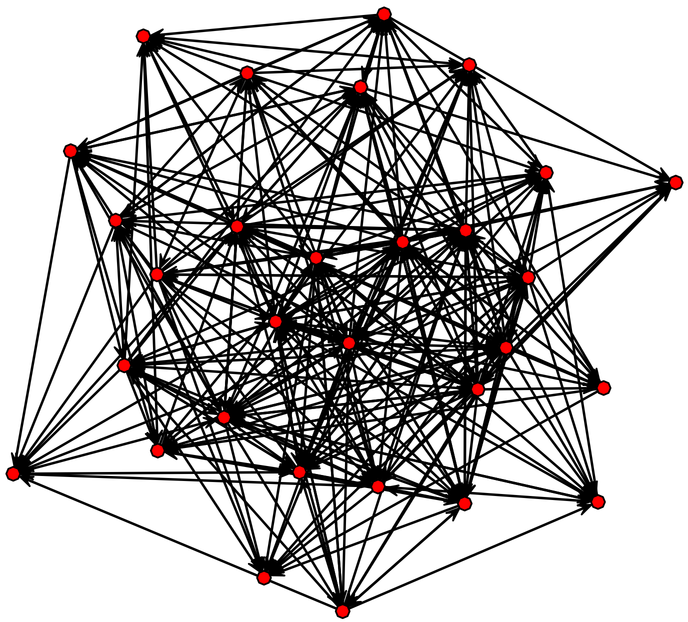
Can’t we simply do t-tests?

Consider testing:

Correlate advice x with friendship y?



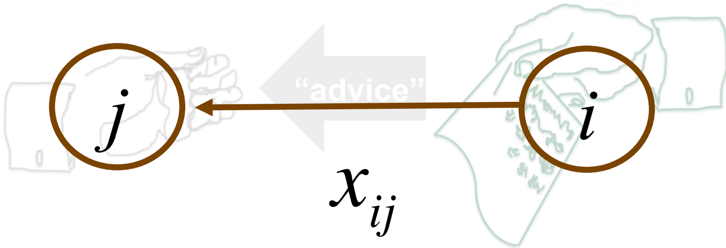


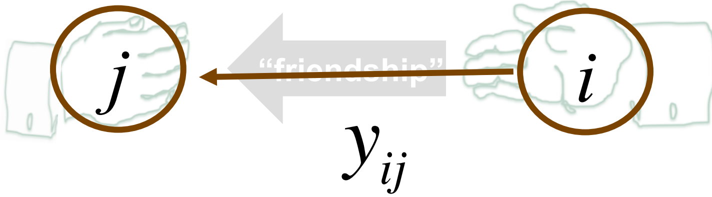


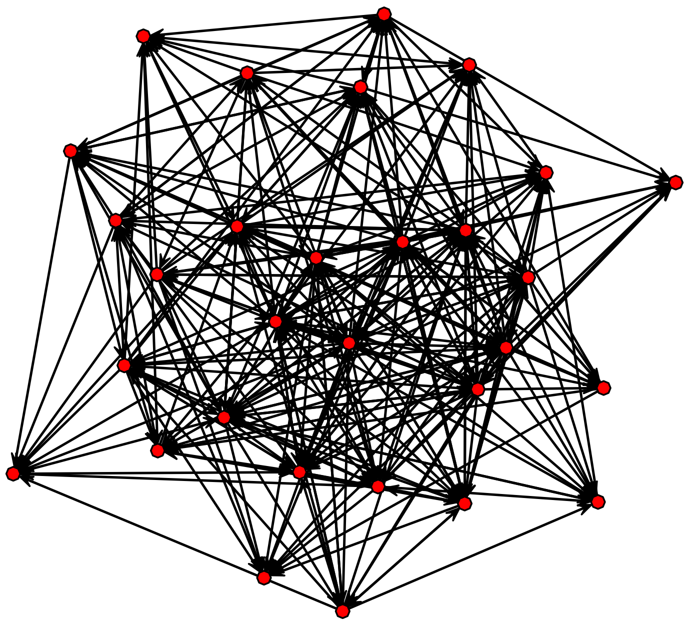
Can’t we simply do t-tests?

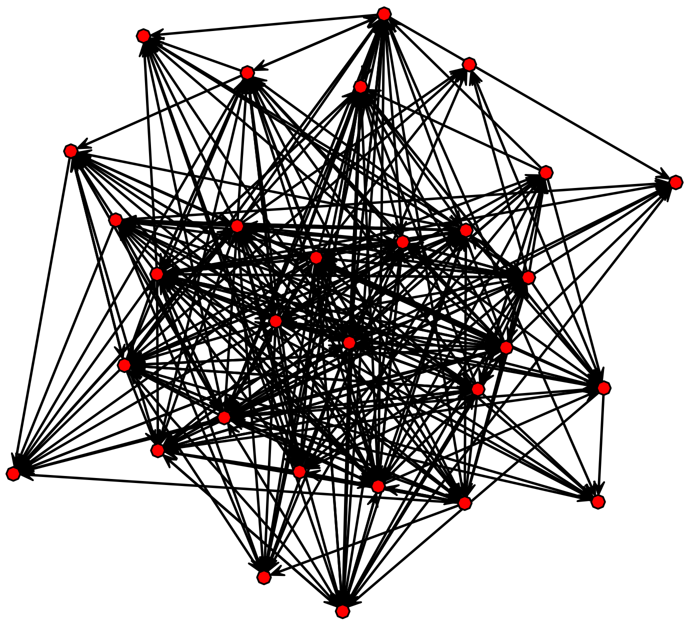
Consider testing:

Correlate advice x with friendship y?





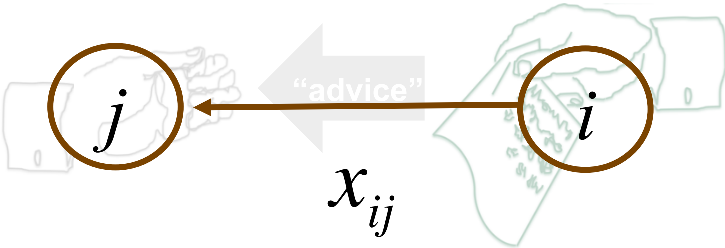


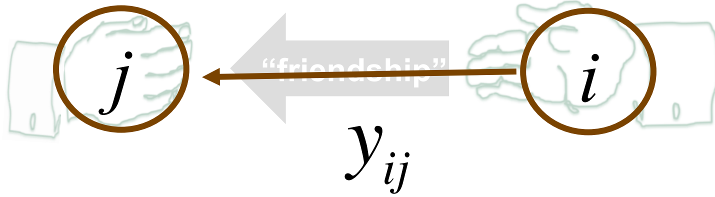


Can't we simply do t-tests?

Consider testing:

Correlate advice x with friendship y?



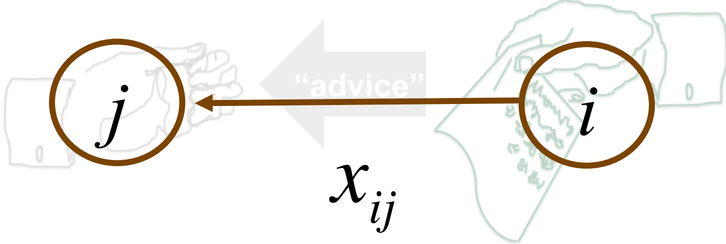


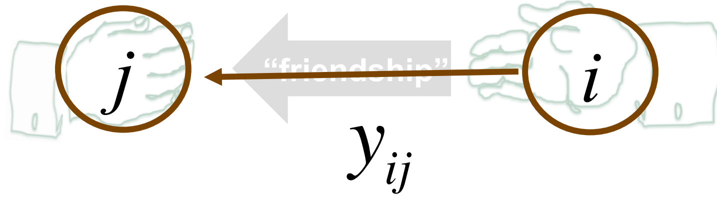
$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \qquad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & y_{jk} & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ y_{li} & y_{lj} & y_{lk} & - \end{pmatrix}$$

Can’t we simply do t-tests?

Consider testing:

Correlate advice x with friendship y?





$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \qquad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & y_{jk} & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ y_{li} & y_{lj} & y_{lk} & - \end{pmatrix}$$

Can’t we simply do t-tests?

Consider testing:

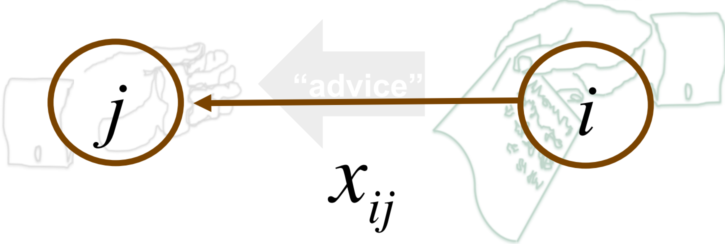
Correlate advice x with friendship y?

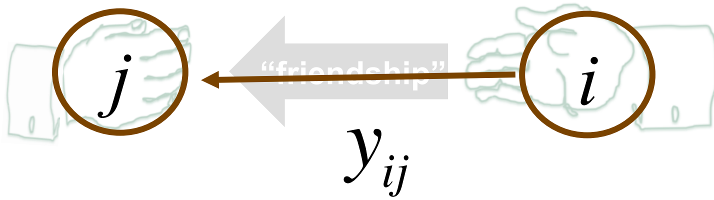
$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \qquad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & y_{jk} & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ y_{li} & y_{lj} & y_{lk} & - \end{pmatrix}$$

Can’t we simply do t-tests?

Consider testing:

Correlate advice x with friendship y?



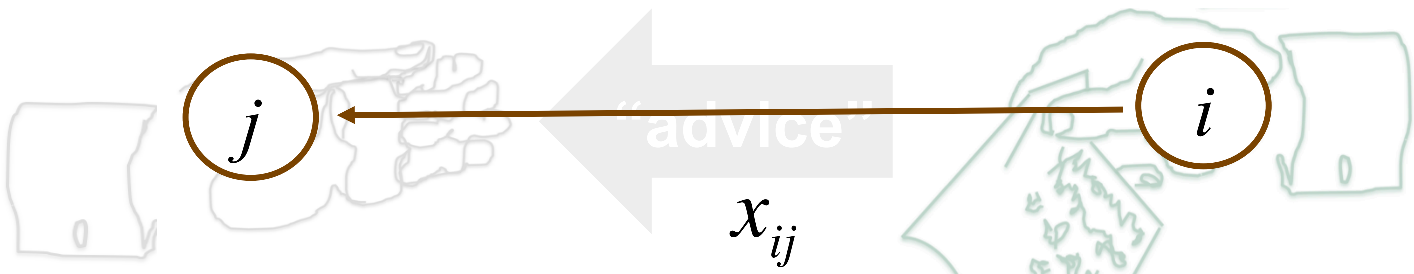


$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & x_{jk} & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \qquad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & y_{jk} & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ y_{li} & y_{lj} & y_{lk} & - \end{pmatrix}$$

Can't we simply do t-tests?

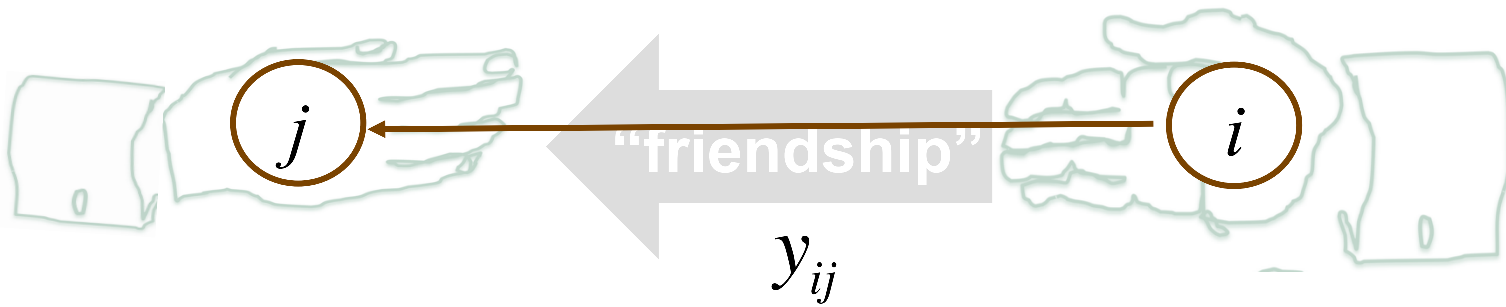
Consider testing:

Correlate advice x with friendship y?



Here we get r = 0.21

Large?



Can't we simply do t-tests?

Using standard statistical techniques

Is r = 0.21 **big?**

Standard* statistical approach:

Reject H_0 (no correlation) if

$$t = \left| \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \right|$$

is greater than 2

Here t = 6.44 (df 868)

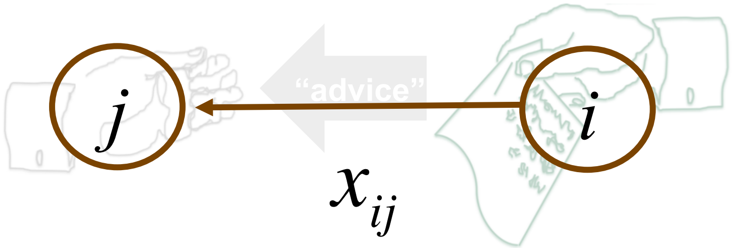
2-sided p-value: 2×10^{-10}

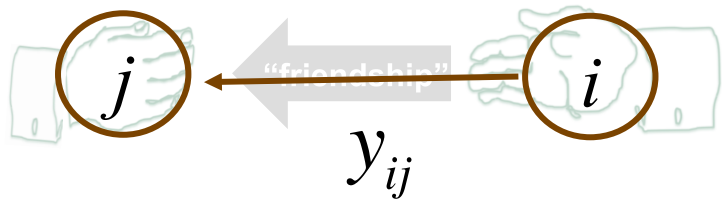
*though careless

Can't we simply do t-tests?

Does this – p-value of 2x10^{-10} – mean that

advice x and friendship y? are **truly associated**?

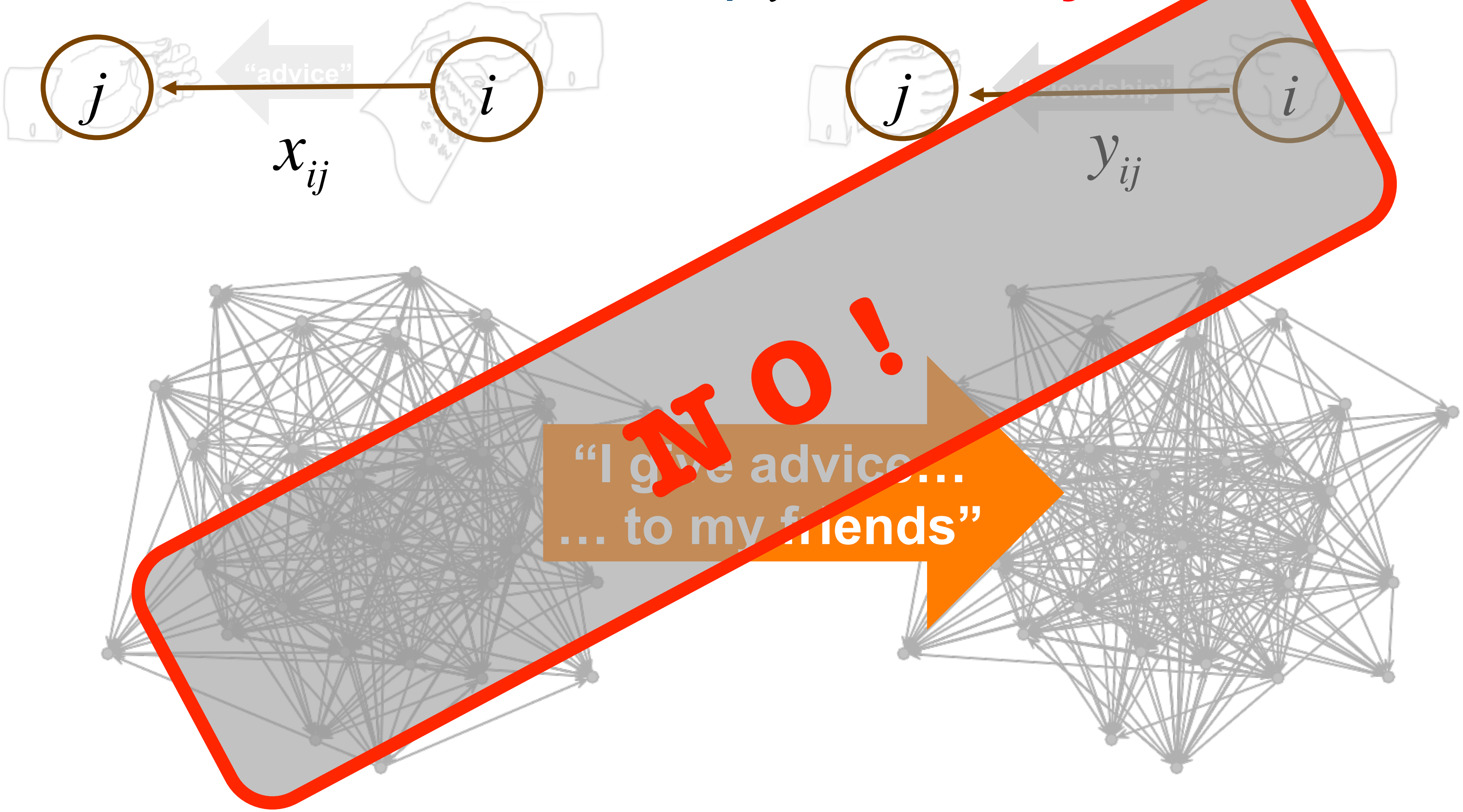




Can’t we simply do t-tests?

Does this – p-value of 2x10^{-10} – mean that

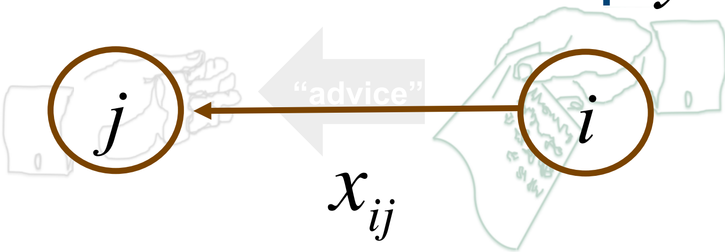
advice x and friendship y? are **truly associated**?



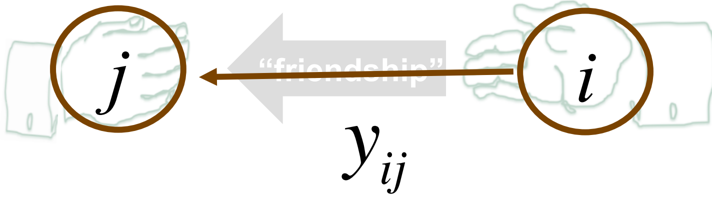
Can't we simply do t-tests?

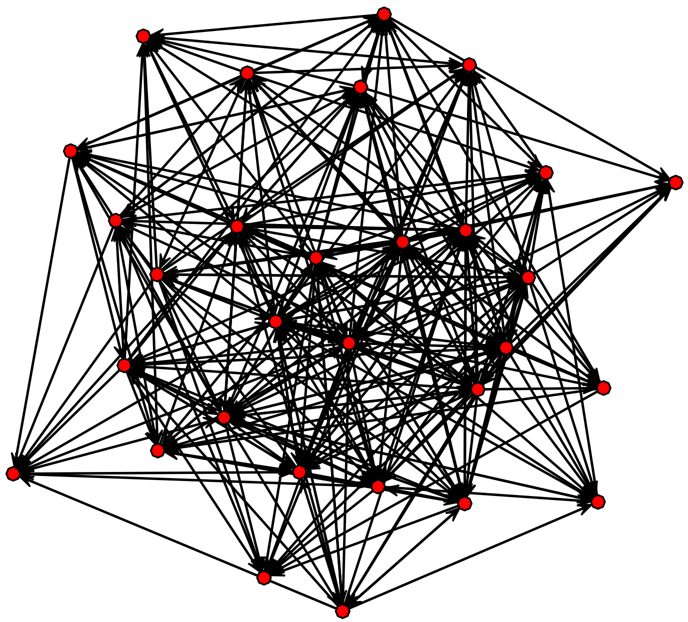
Here I generated

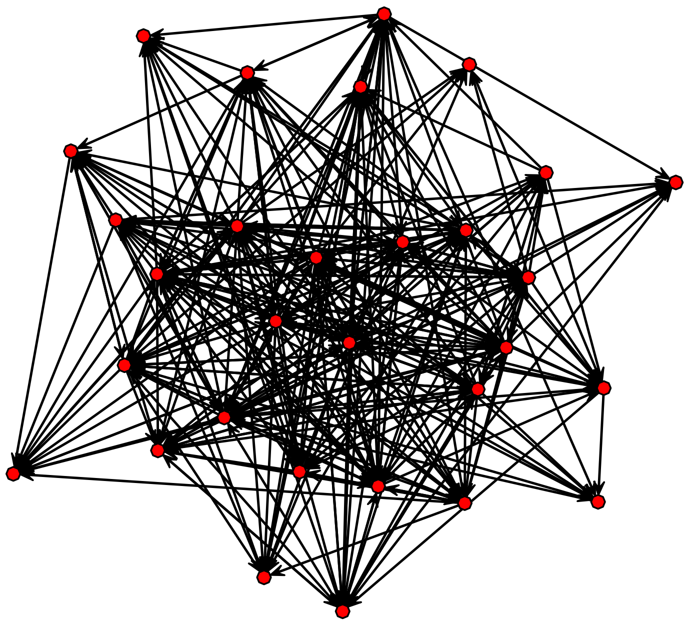
friendship y **independently** of advice x



$\neq$







Can't we simply do t-tests?

Friendship and advice ties are independent but

There may be dependence on actors

Some people:

Can't we simply do t-tests?

Friendship and advice ties are independent but

There may be dependence on actors

Some people:

Can't we simply do t-tests?

Friendship and advice ties are independent but

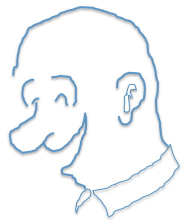
There may be dependence on actors

other people:

Can’t we simply do t-tests?

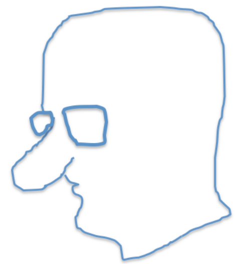
Sends ties to 0% others

$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \quad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ y_{li} & y_{lj} & y_{lk} & - \end{pmatrix}$$



MANCHESTER 1824 The University of Manchester

Can't we simply do t-tests?

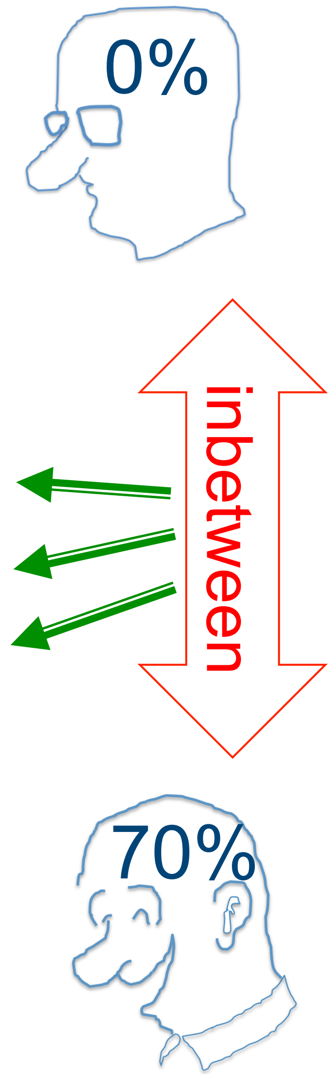


$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ \boxed{x_{li}} & \boxed{x_{lj}} & \boxed{x_{lk}} & - \end{pmatrix} \quad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ \boxed{y_{li}} & \boxed{y_{lj}} & \boxed{y_{lk}} & - \end{pmatrix}$$

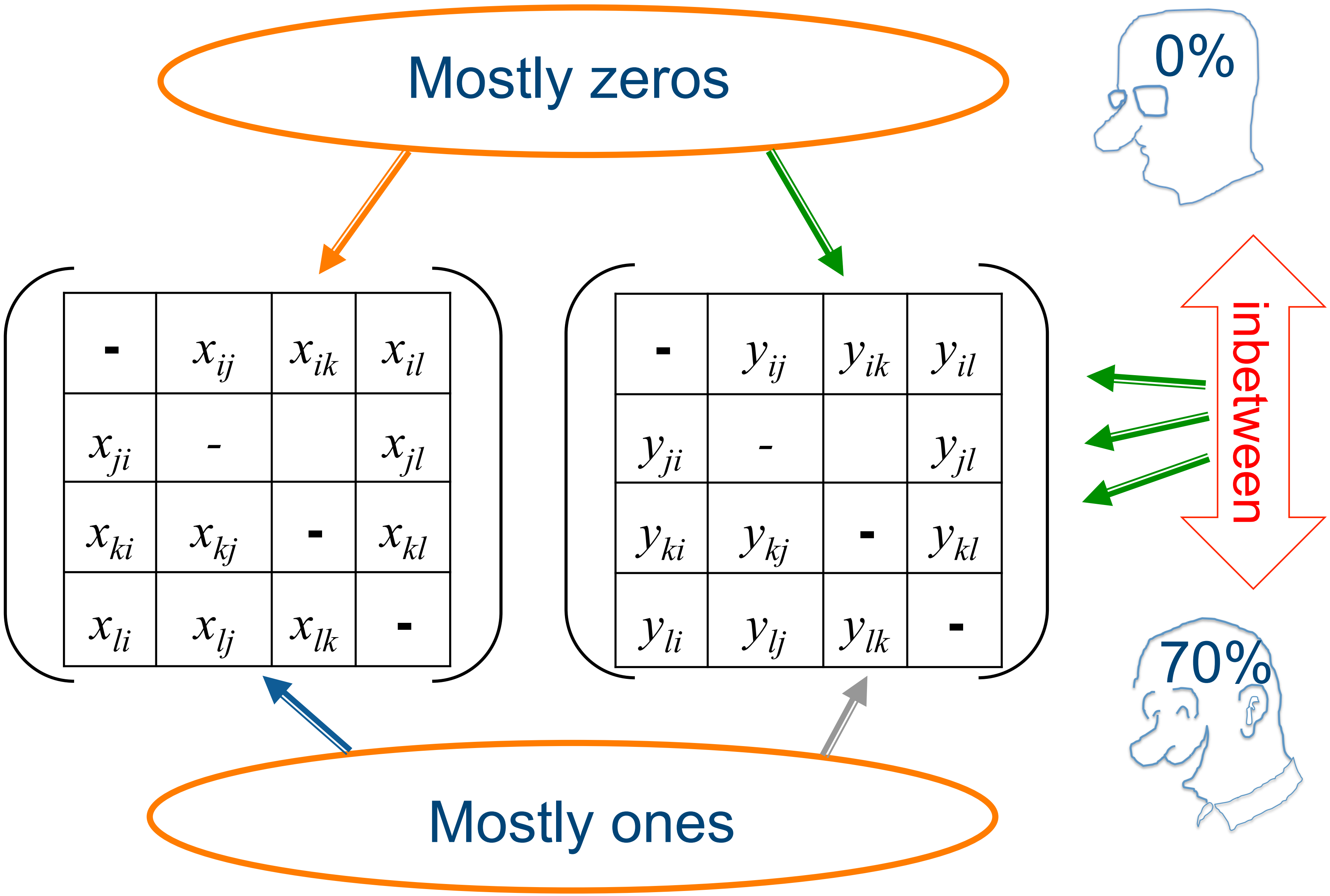
Sends ties to 70% others

Can’t we simply do t-tests?

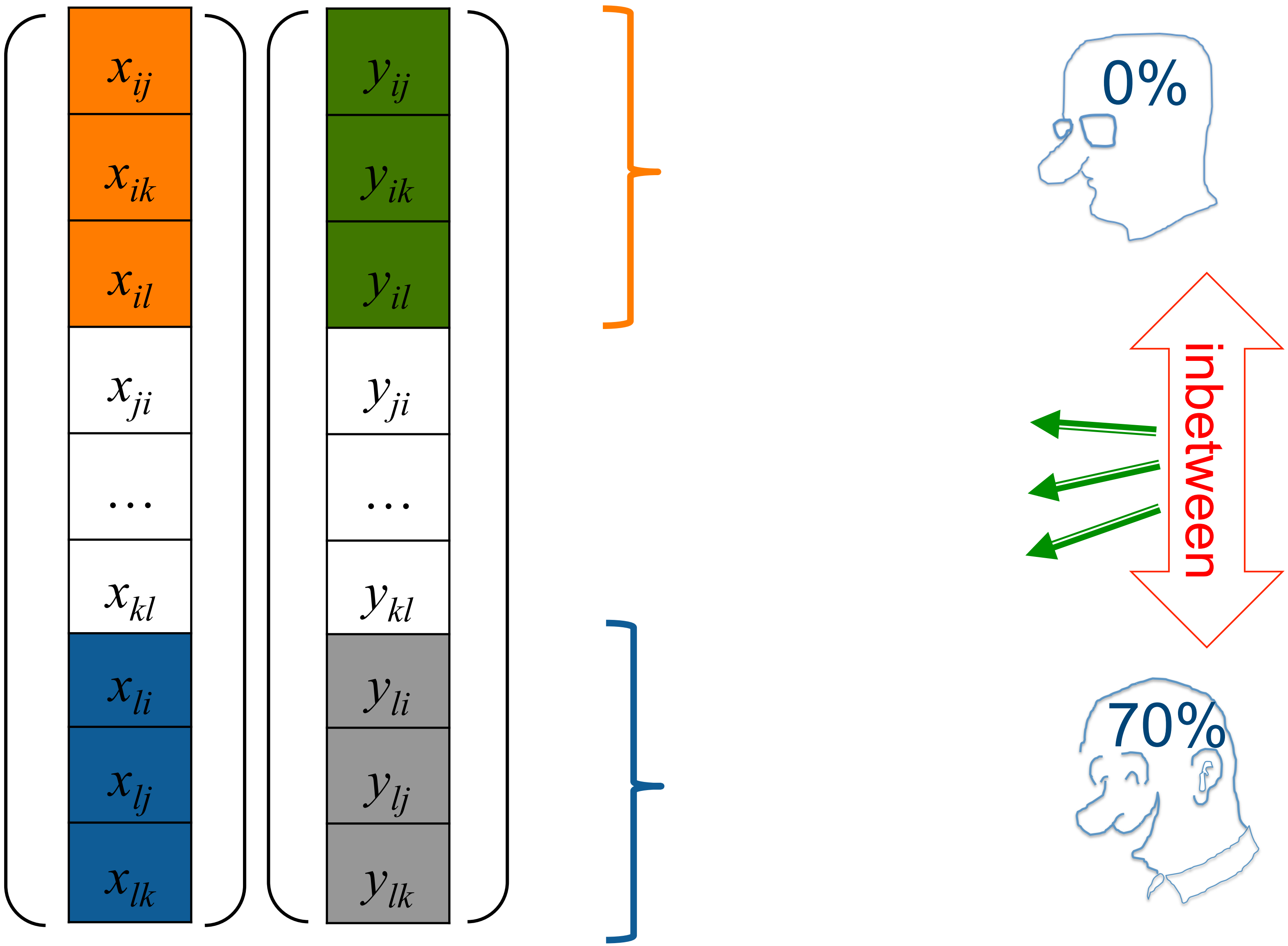
$$\begin{pmatrix} - & x_{ij} & x_{ik} & x_{il} \\ x_{ji} & - & & x_{jl} \\ x_{ki} & x_{kj} & - & x_{kl} \\ x_{li} & x_{lj} & x_{lk} & - \end{pmatrix} \quad \begin{pmatrix} - & y_{ij} & y_{ik} & y_{il} \\ y_{ji} & - & & y_{jl} \\ y_{ki} & y_{kj} & - & y_{kl} \\ y_{li} & y_{lj} & y_{lk} & - \end{pmatrix}$$



Can’t we simply do t-tests?



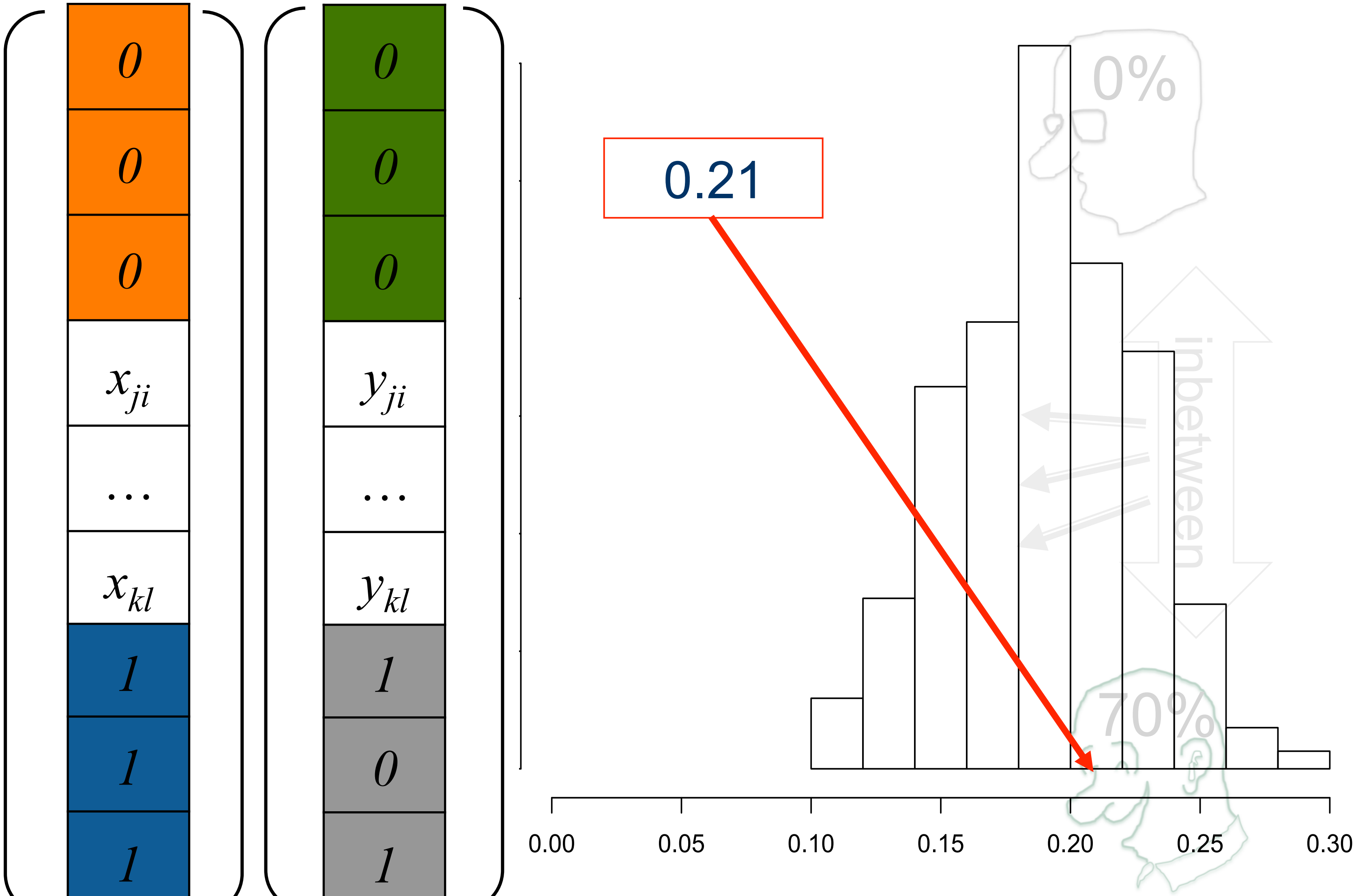
Can't we simply do t-tests?



Can't we simply do t-tests?

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ x_{ji} \\ \dots \\ x_{kl} \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ y_{ji} \\ \dots \\ y_{kl} \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Can't we simply do t-tests?

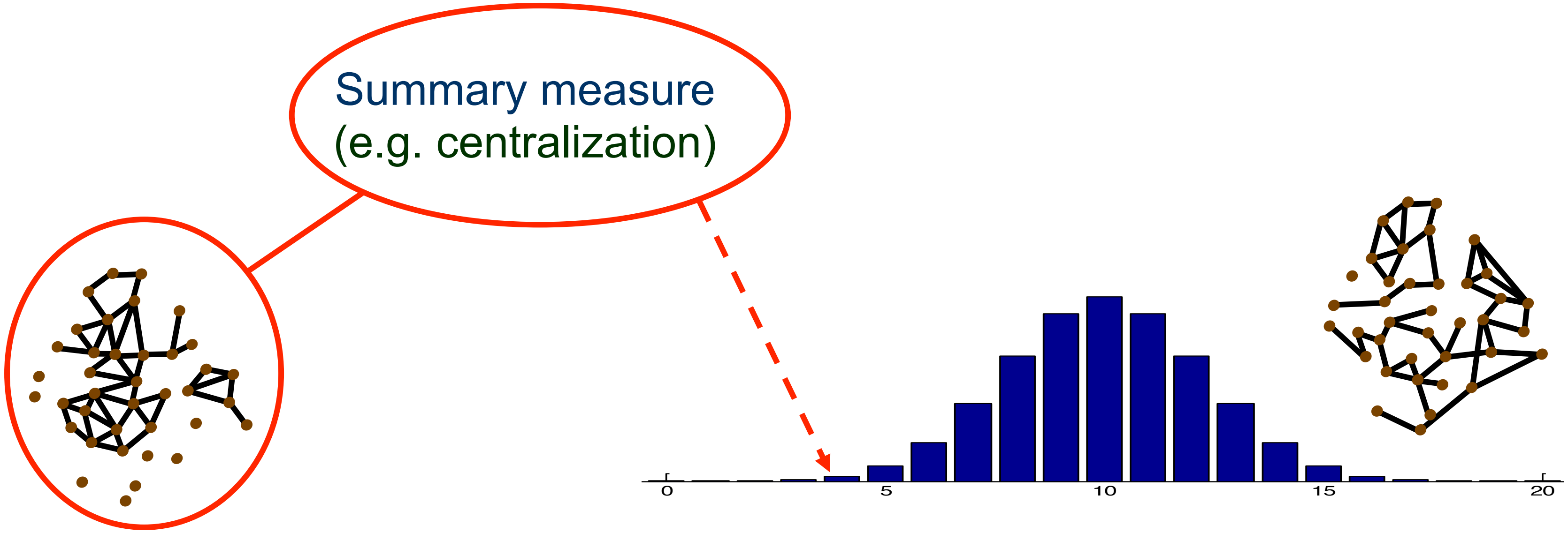


correlations assuming no association

History: non-parametric approaches

From late 1930s

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



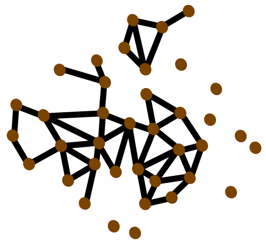
Observed network

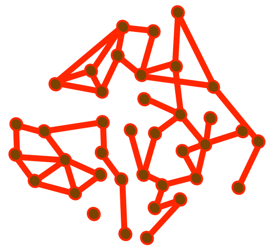
Distribution of measure under null distribution

Non-parametric: 2 relations

Conformity of 2 sociometric measures (Katz and Powell, 1953)

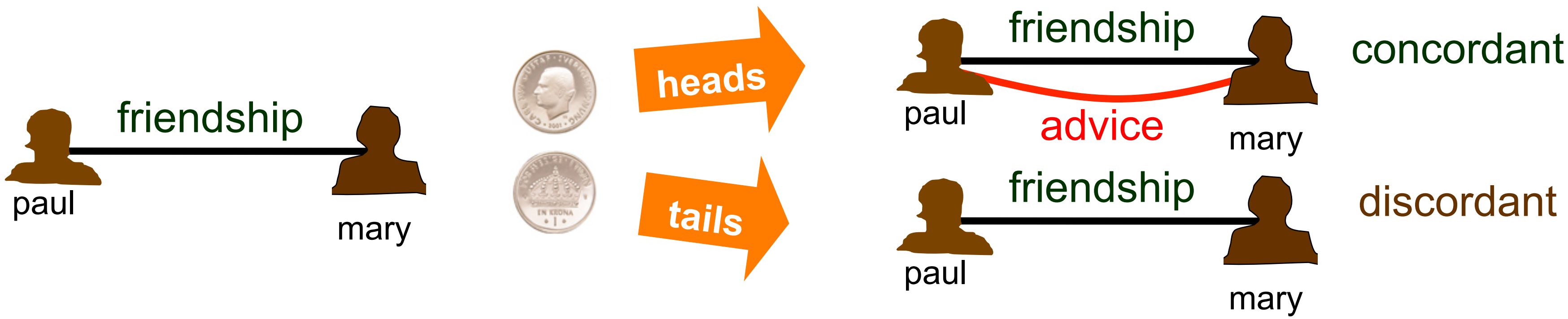
A: friendship network

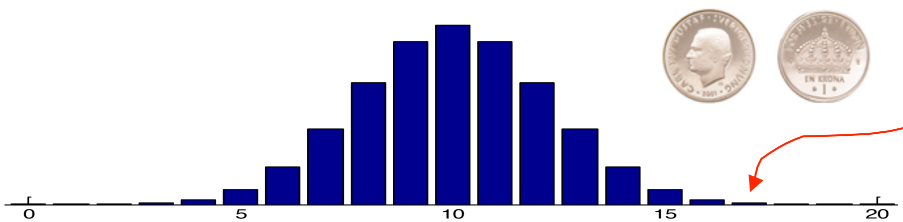




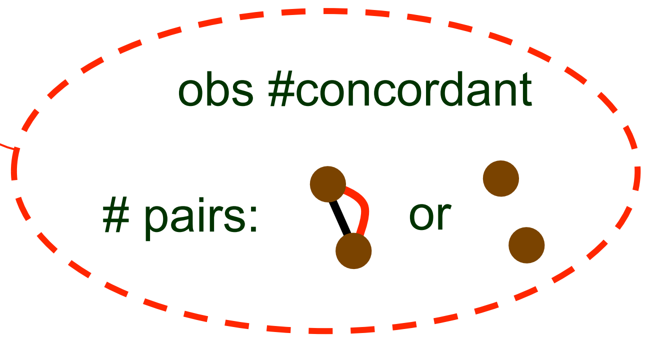
B: advice network

If **no** association between A and B, for each pair:





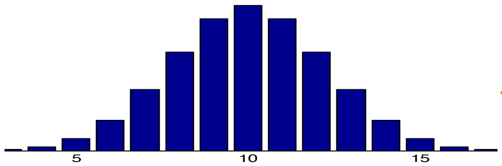
Distribution of #concordant under null distribution



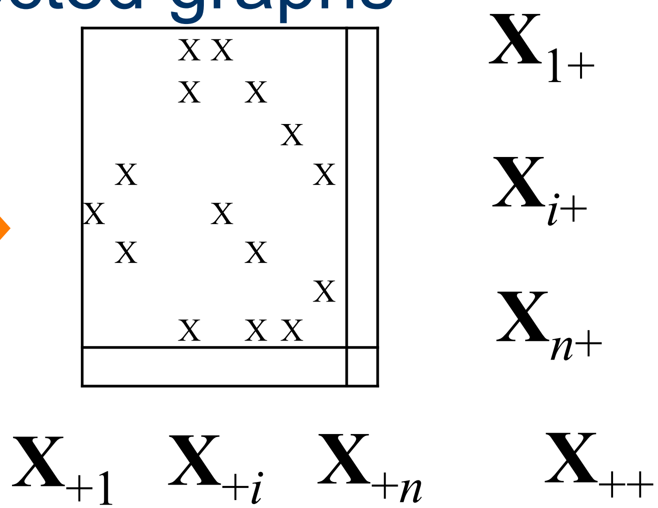
MANCHESTER 1824 The University of Manchester

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs

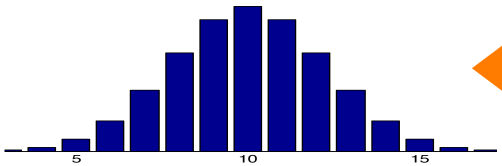


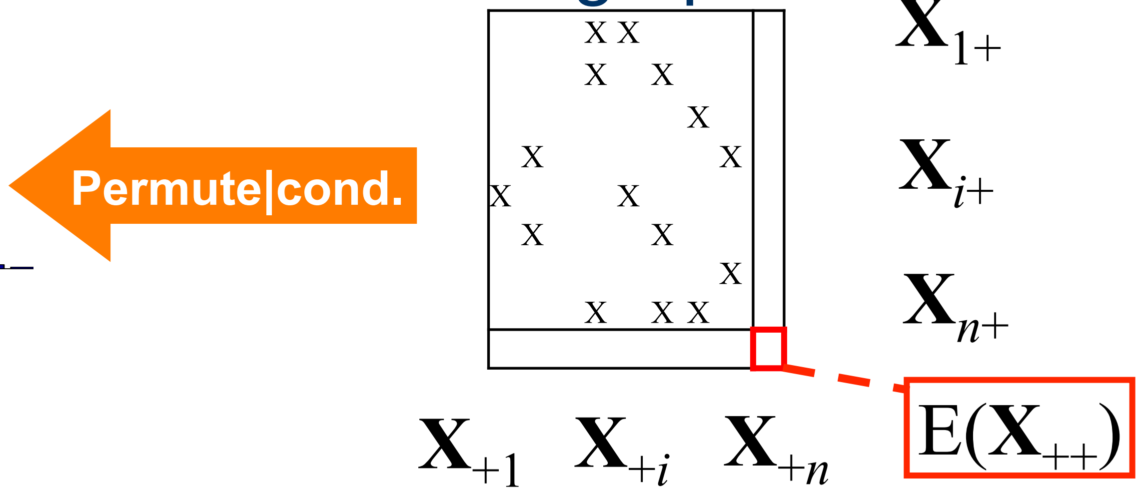
Compare?



Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



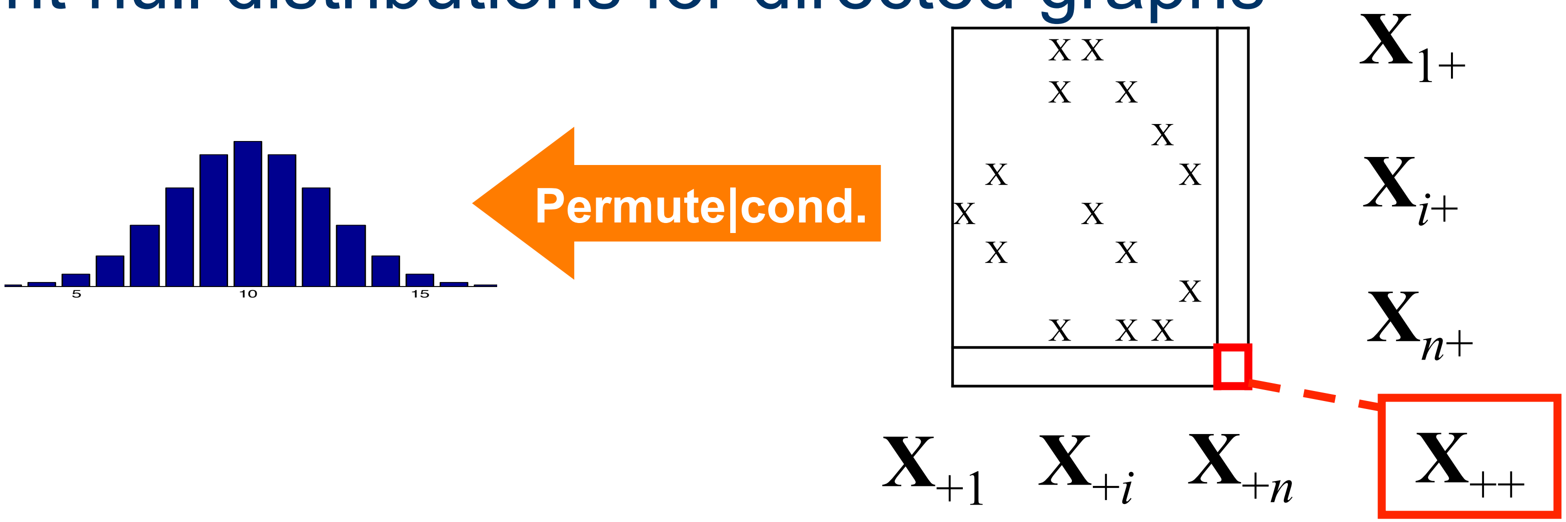


Condition on expected density $\mathscr{U}|E(\mathbf{X}_{++})$: Bernoulli

MANCHESTER 1824
The University of Manchester

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



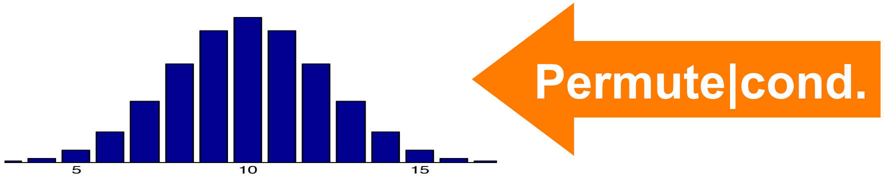
Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

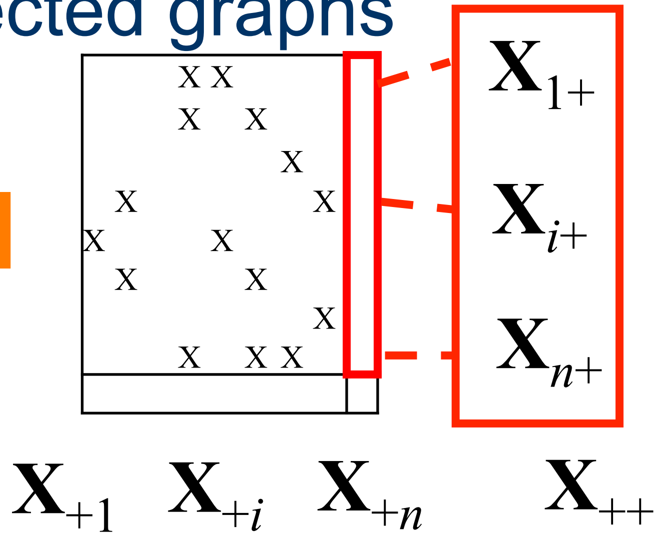
Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

MANCHESTER 1824 The University of Manchester

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





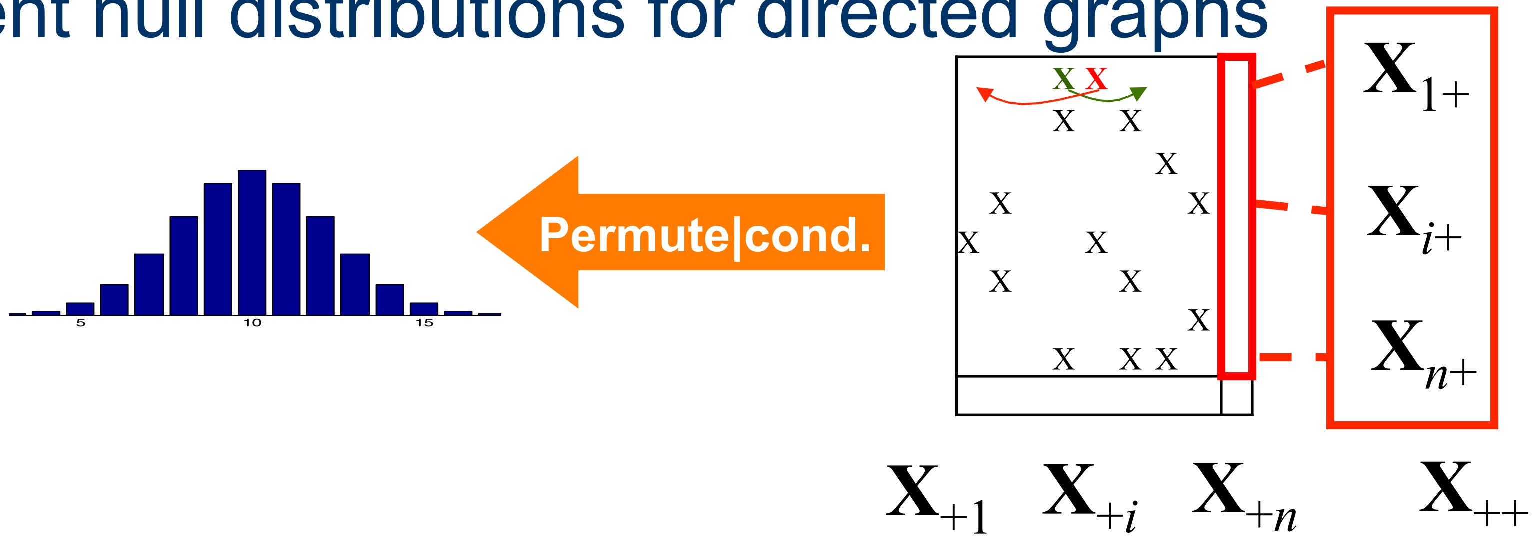
Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet +}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



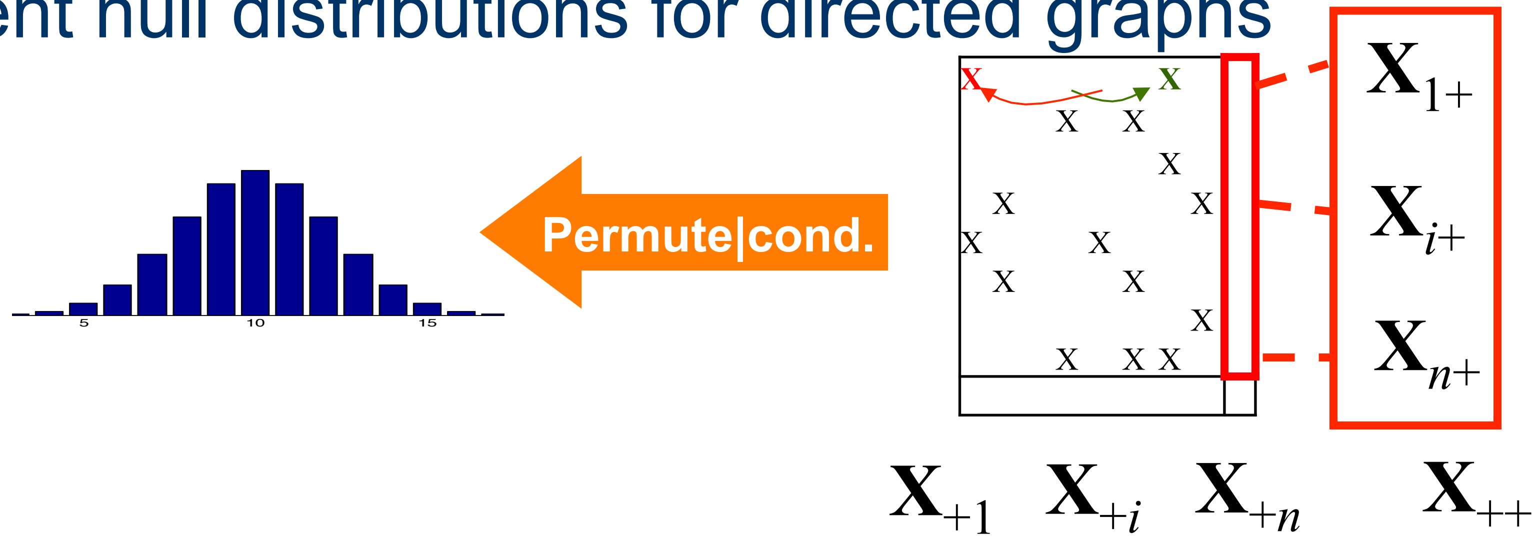
Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



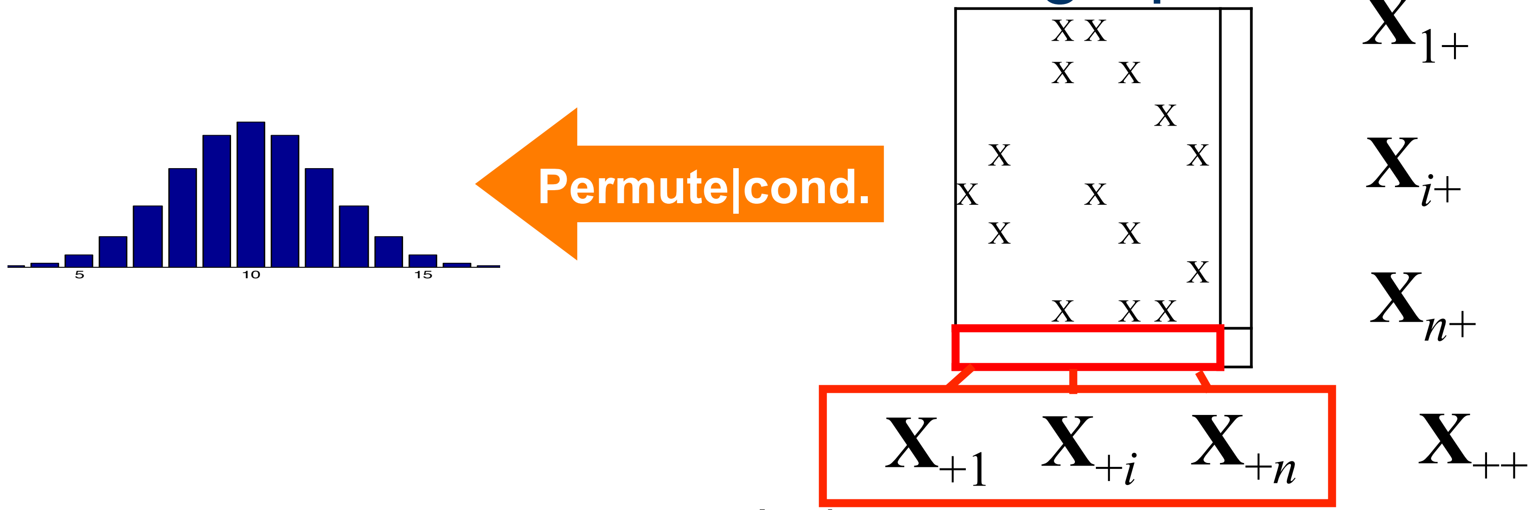
Condition on expected density $\mathcal{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathcal{U}|\mathbf{X}_{++}$

Condition on activity/out-degrees: $\mathcal{U}|\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

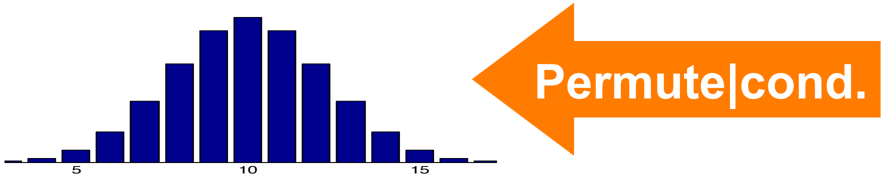
Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

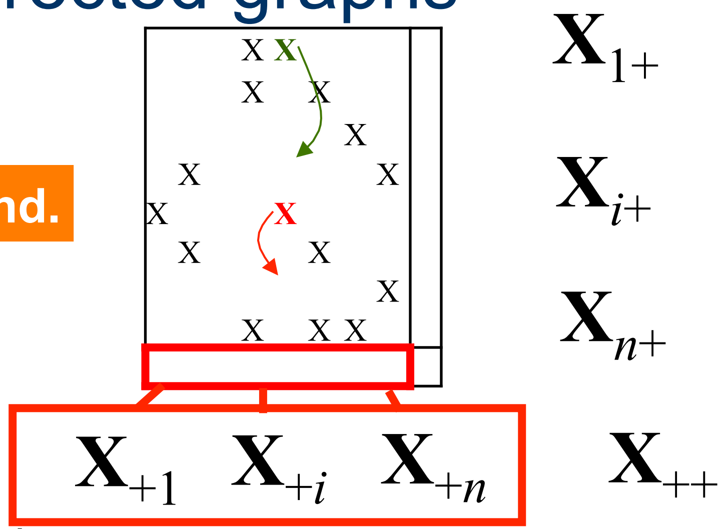
Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

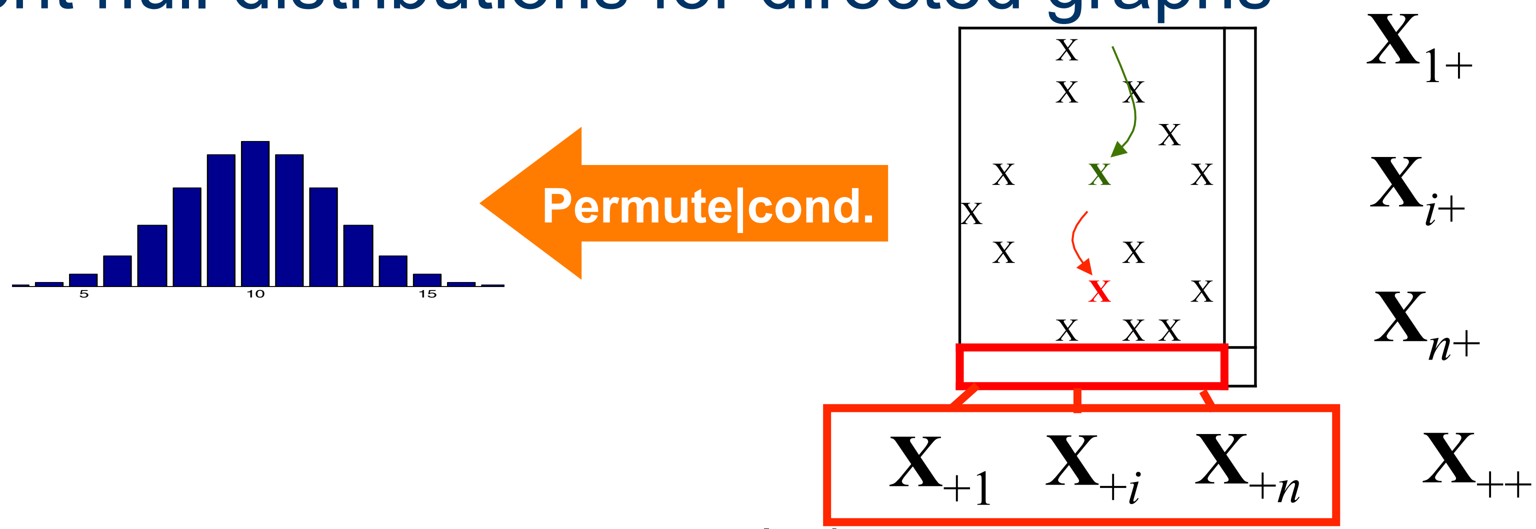
Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet +}$

Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

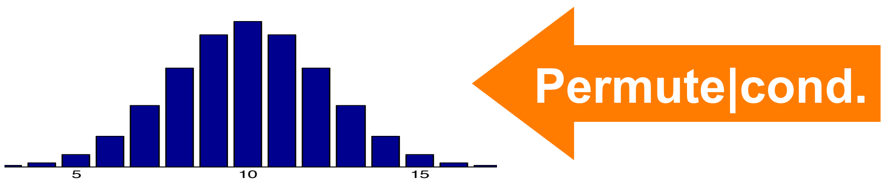
Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

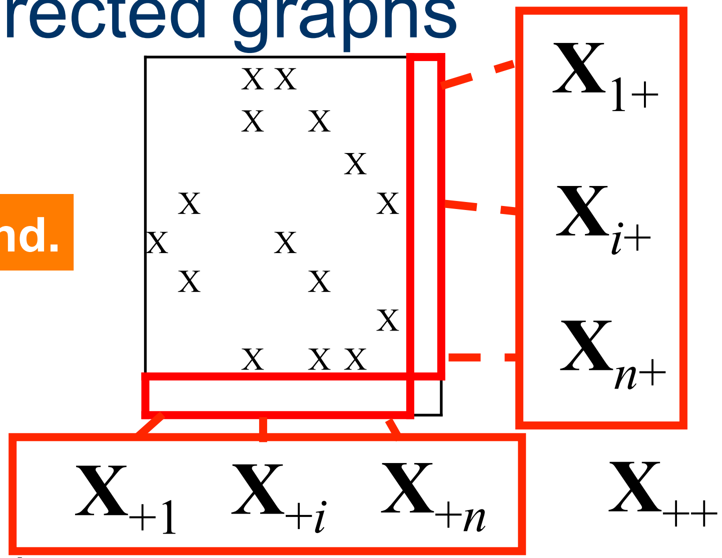
Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





Condition on expected density $\mathscr{U}|E(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

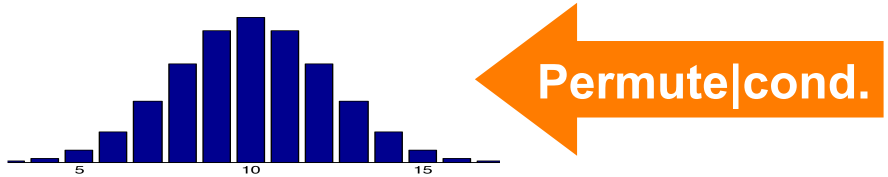
Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

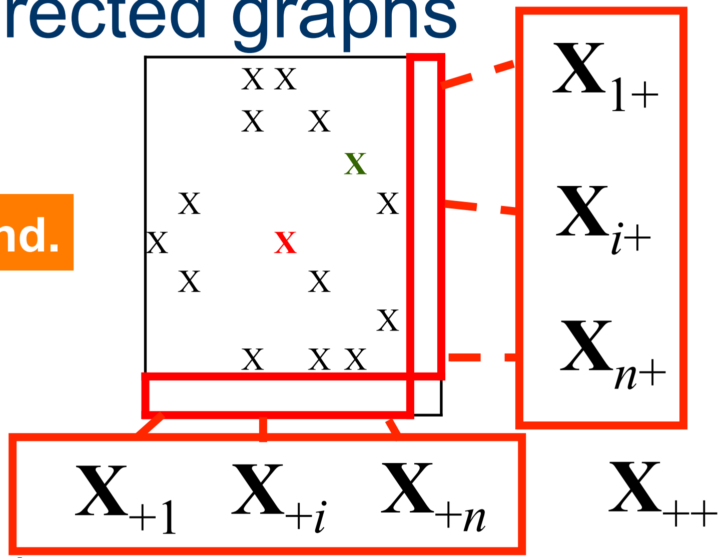
Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Condition on both in-degrees and out-degrees : $\mathscr{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

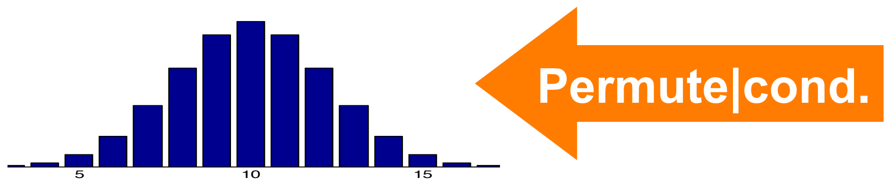
Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

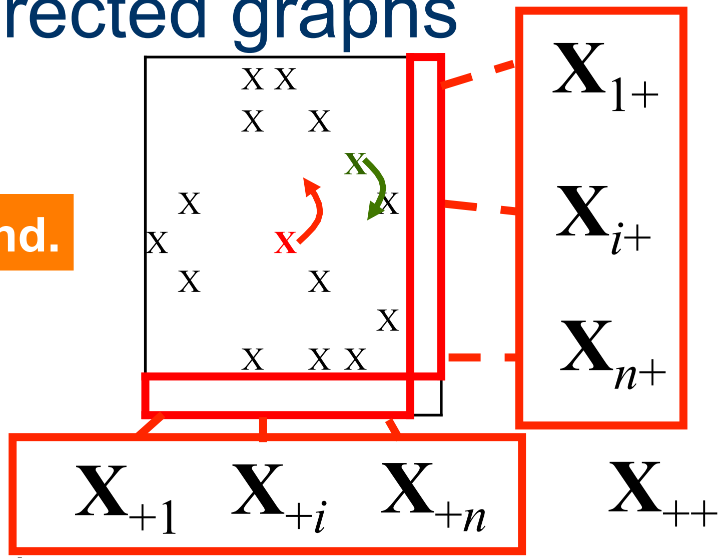
Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Condition on both in-degrees and out-degrees : $\mathscr{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





Condition on expected density $\mathscr{U}|E(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

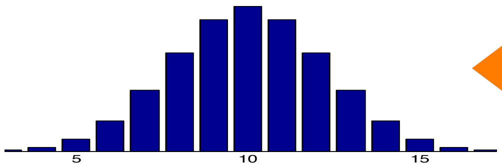
Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

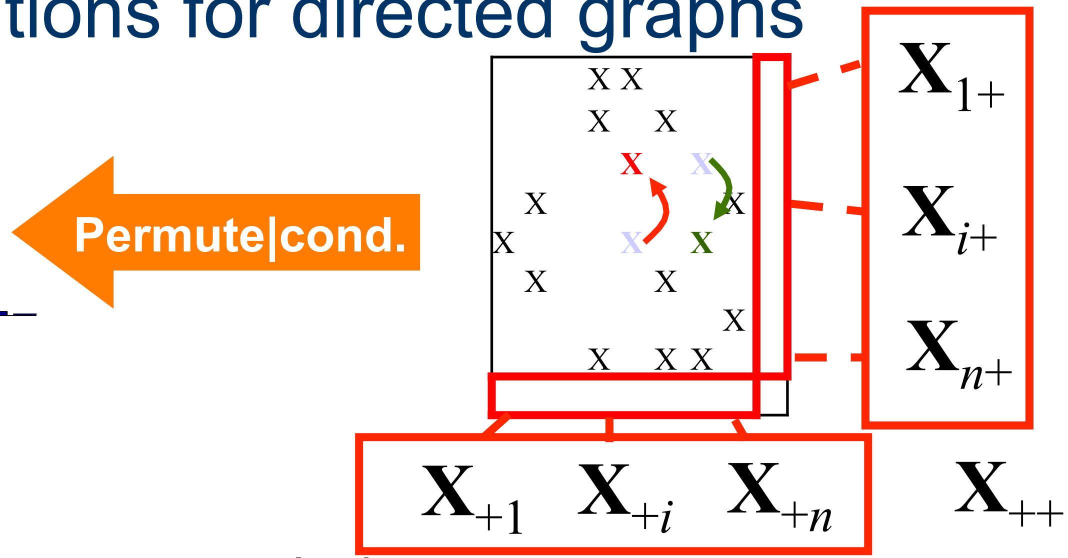
Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Condition on both in-degrees and out-degrees : $\mathscr{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





Condition on expected density $\mathcal{U}|E(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathcal{U}|\mathbf{X}_{++}$

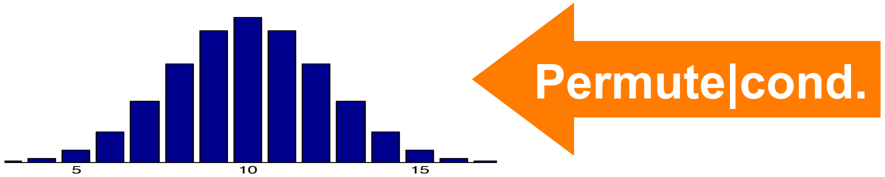
Condition on activity/out-degrees: $\mathcal{U}|\mathbf{X}_{\bullet+}$

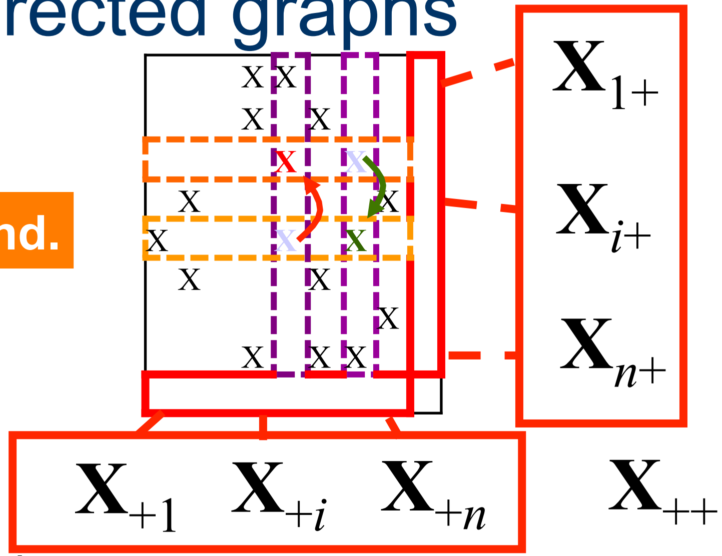
Condition on popularity/in-degrees: $\mathcal{U}|\mathbf{X}_{+\bullet}$

Condition on both in-degrees and out-degrees : $\mathcal{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs





Condition on expected density $\mathcal{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathcal{U}|\mathbf{X}_{++}$

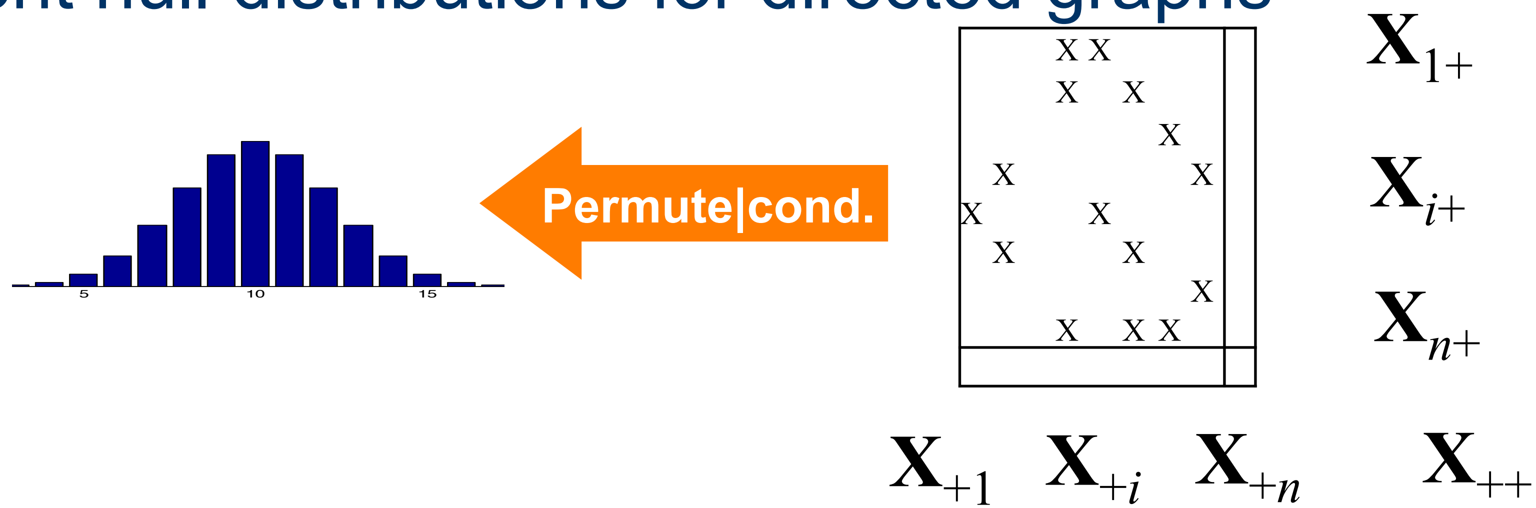
Condition on activity/out-degrees: $\mathcal{U}|\mathbf{X}_{\bullet+}$

Condition on popularity/in-degrees: $\mathcal{U}|\mathbf{X}_{+\bullet}$

Condition on both in-degrees and out-degrees : $\mathcal{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

Non-parametric: conditional uniform null distributions

Different null distributions for directed graphs



Condition on expected density $\mathscr{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathscr{U}|\mathbf{X}_{++}$

Condition on activity/out-degrees: $\mathscr{U}|\mathbf{X}_{\bullet+}$

Condition on popularity/in-degrees: $\mathscr{U}|\mathbf{X}_{+\bullet}$

Condition on both in-degrees and out-degrees : $\mathscr{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

For a systematic statistical approach to successive conditioning see Pattison et al., 2000

Non-parametric: conditional uniform null distributions

Condition on expected density $\mathcal{U}|\mathrm{E}(\mathbf{X}_{++})$: Bernoulli

Condition on density: $\mathcal{U}|\mathbf{X}_{++}$

Condition on activity/out-degrees: $\mathcal{U}|\mathbf{X}_{\bullet+}$

Condition on popularity/in-degrees: $\mathcal{U}|\mathbf{X}_{+\bullet}$

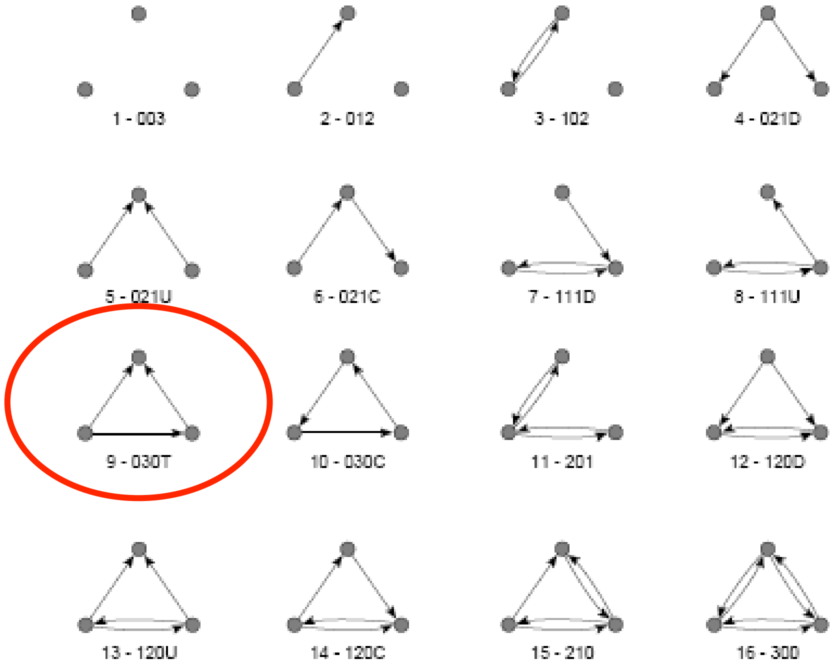
Condition on both in-degrees and out-degrees : $\mathcal{U}|\mathbf{X}_{+\bullet},\mathbf{X}_{\bullet+}$

Try and identify these distributions in ‘sna’:

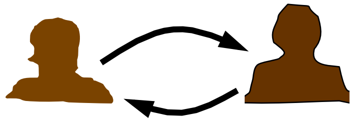
```
library(help=sna)
# e.g.: rgnm
```

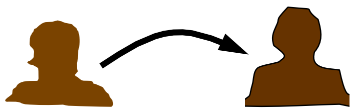
Investigating the triad census conditional on the dyad census (Holland & Leinhardt 1970)

Different directed triangles



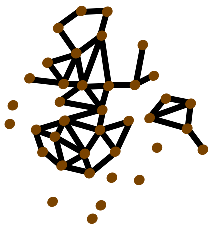
Types of dyads:

M (mutual): 

A (asymetric): 

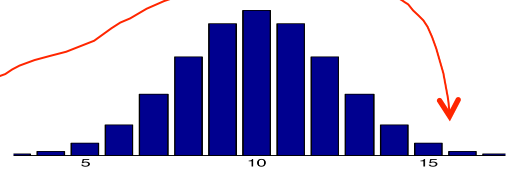
N (null): 

$\mathcal{U}|MAN$:  uniform graphs with same MAN as observed



Observed network

obs #030T

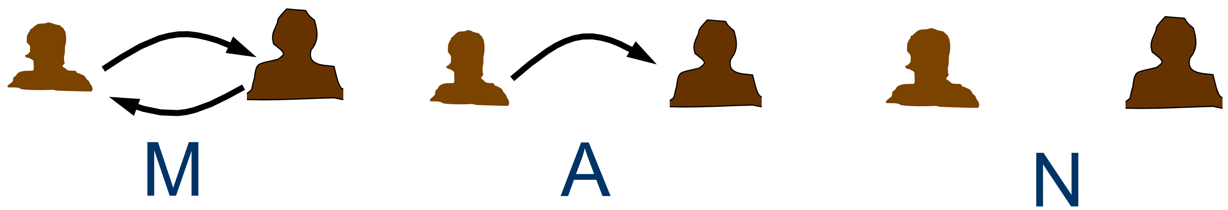


Distribution of #030T given observed MAN

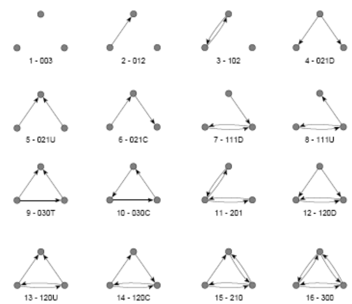
Investigating the triad census conditional on the dyad census

Interpretation

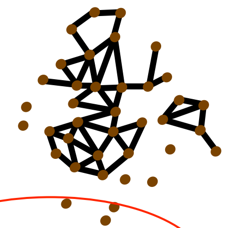
Given that we've accounted for different types of reciprocation

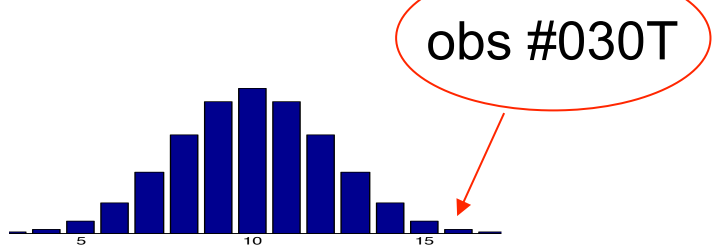


What triads occur more (less) freq. than chance?



Alt.: What triads occur more (less) freq. than what is explained by density and reciprocation alone?





Distribution of #030T given observed MAN

Triad census in R

Load the data set **coleman**

that comes with the package "sna"

```
?coleman
data(coleman) # loads data set
colenet <- as.network(coleman[1,,]) # create
network obj
colenet # check properties
plot(colenet) # plot
dyad.census(colenet)
ObsTriad <- triad.census(colenet)
```

Triad census in R

Generate a null-distribution of `NumReplics` graphs with the same MAN as `colenet`

```
NumReplics <- 500
g<-rguman(NumReplics,
73,mut=62,asym=119,null=2447,method = "exact")

TriadRes <- matrix(c(0),NumReplics,16)
for (i in 1:NumReplics)
{
      TriadRes[i,] <- triad.census(g[i,,])
}
```

Calculate the triad census for each simulated graph

Triad census in R

Generate a null-distribution of `NumReplics` graphs

with the same MAN as `colenet`

plot the simulated triad census against the observed

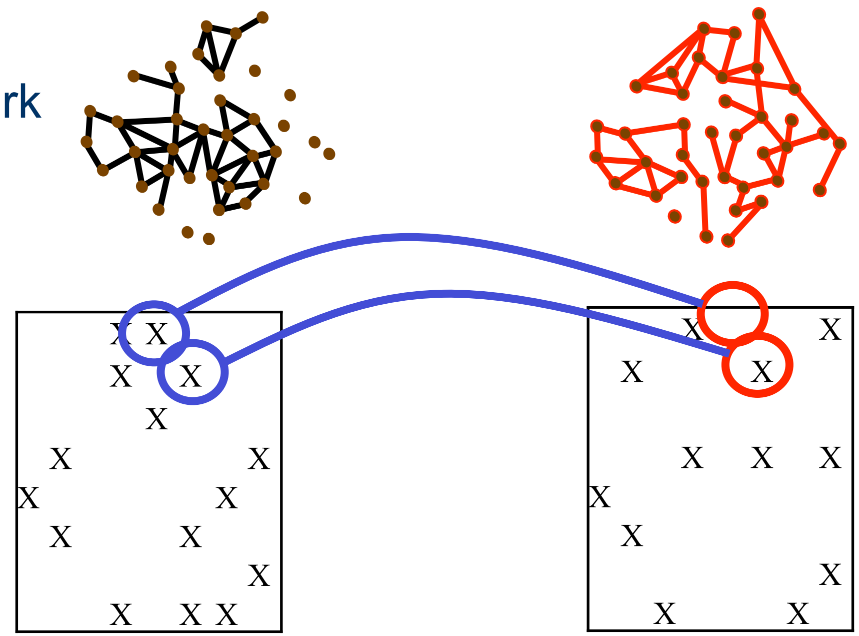
```
par( mfrow = c( 4, 4 ) )
 for (k in 1:16)
 {
      hist(TriadRes[,k],xlim =
c(min(ObsTriad[k],TriadRes[,k]),max(ObsTriad[k]
,TriadRes[,k] ) ) ,xlab=dimnames(ObsTriad)[[2]]
[k],main="")
      lines(ObsTriad[k],0,type="o", col="red")

 }
```

Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)

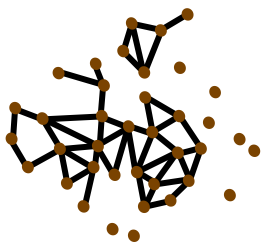
A: friendship network

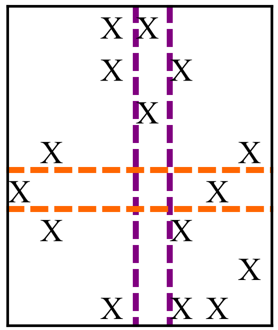
B: advice network



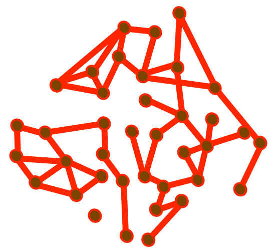
Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)

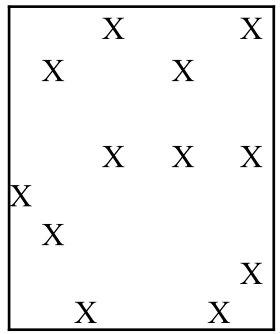
A: friendship network





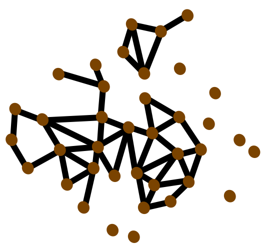
B: advice network



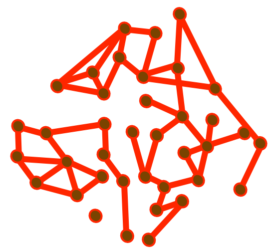


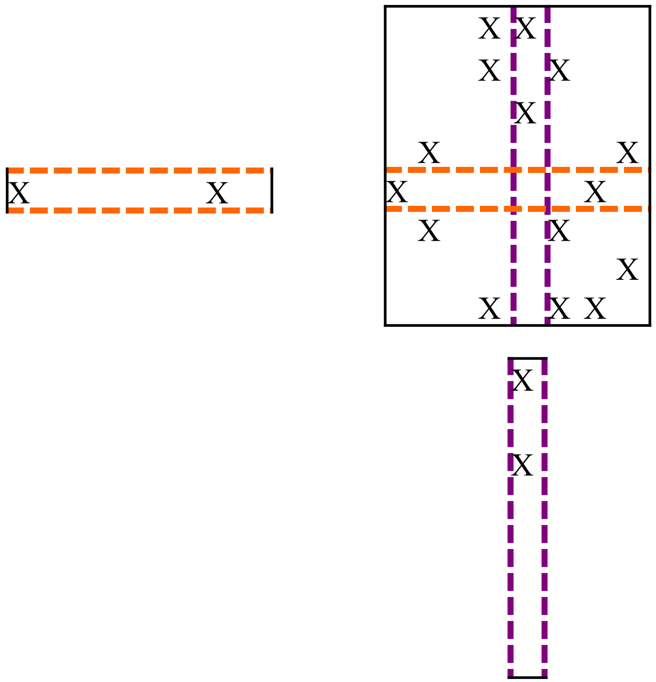
Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)

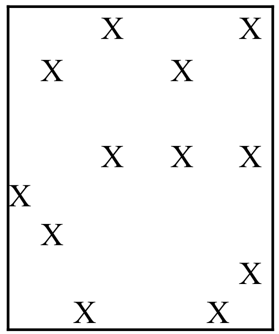
A: friendship network



B: advice network

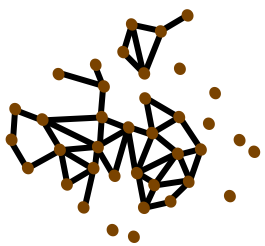


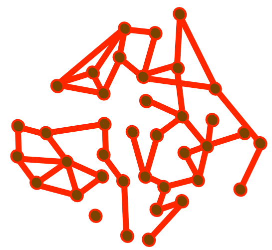




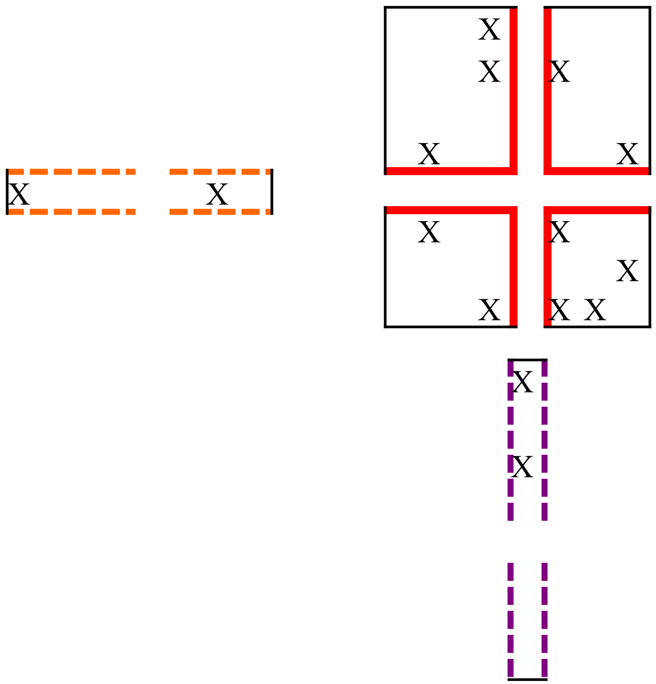
Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)

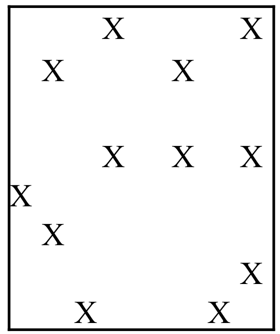
A: friendship network



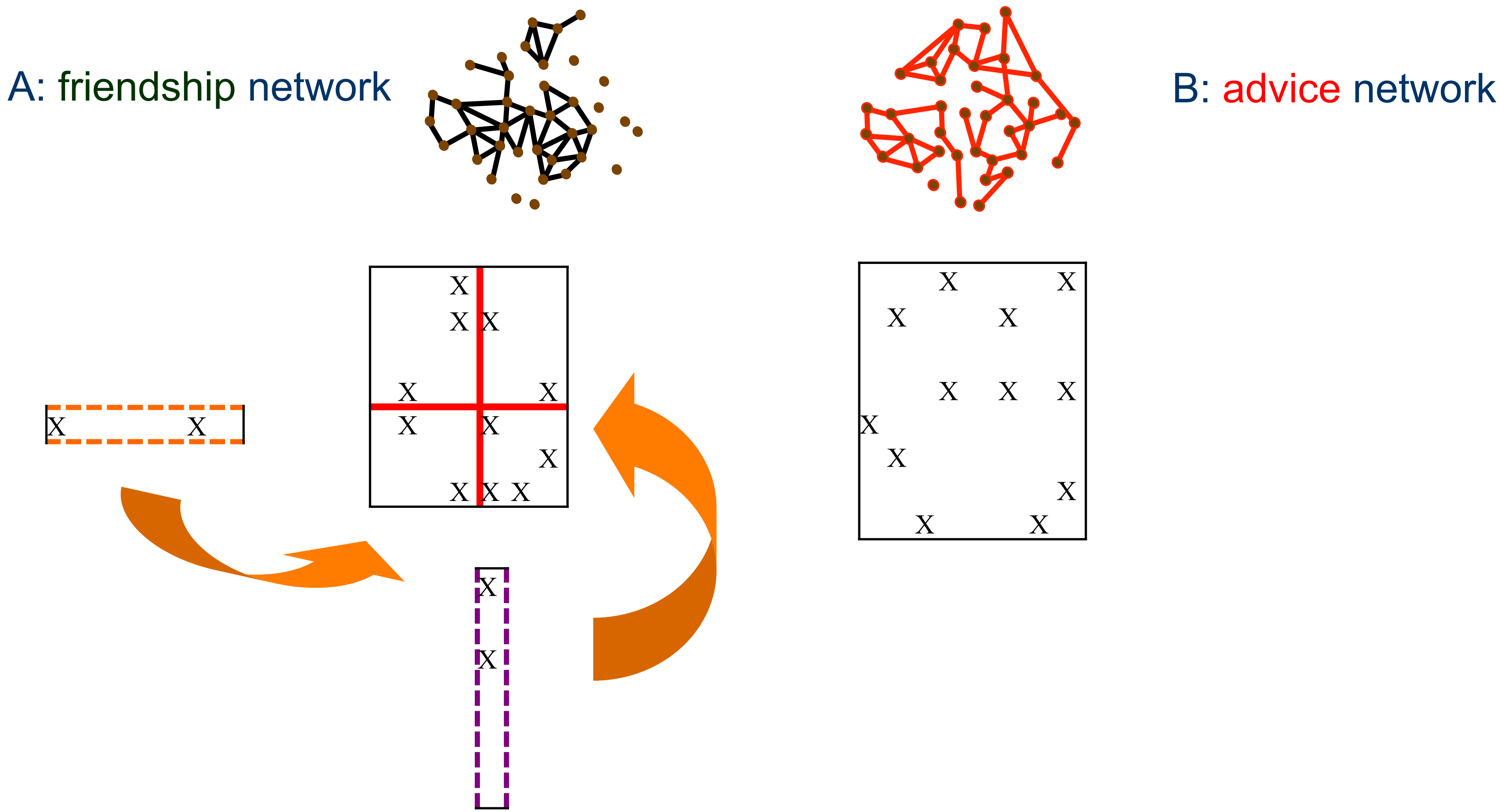


B: advice network



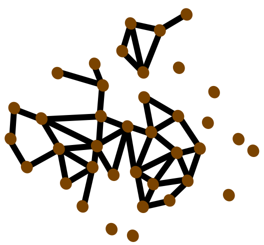


Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)

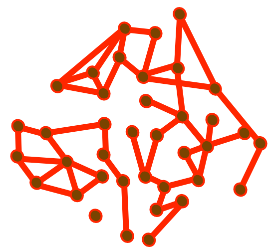


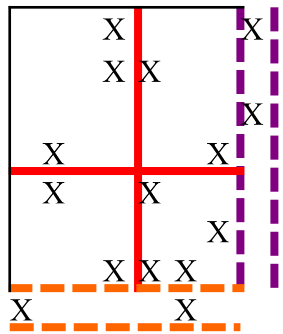
Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)

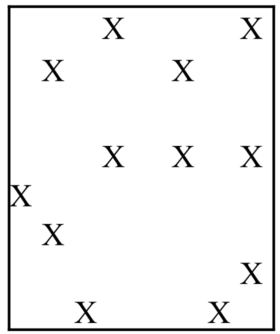
A: friendship network



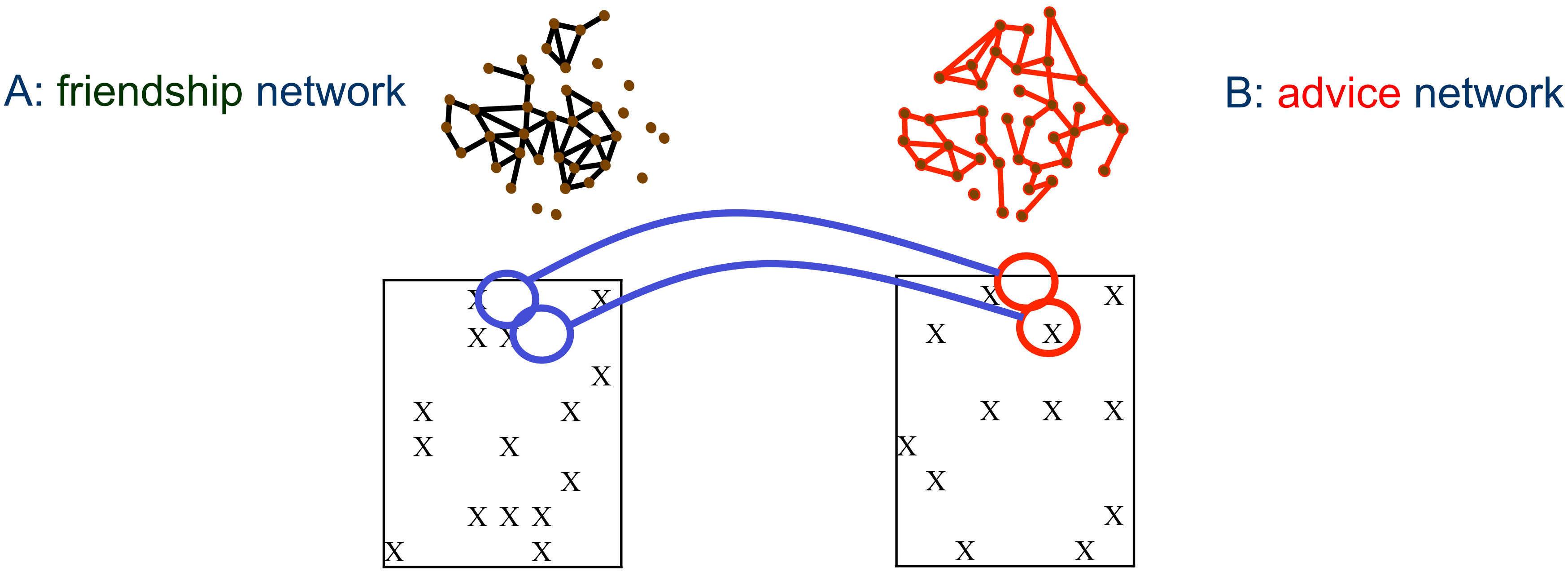
B: advice network







Quadratic Assignment Procedure (QAP) (Krackhardt, 1987)



How “unusual” is the observed number of concordant pairs compared to the permutation distribution?

QAP in R

Load the data set **coleman**

that comes with the package “sna”

```
?coleman
data(coleman) # loads data set
q.12<-qaptest(coleman,gcor,g1=1,g2=2)# qap test
summary(q.12)# summary of test
plot(q.12)# plot of null distribution
```

MANCHESTER 1824
The University of Manchester

Drawbacks of non model based statistical analysis

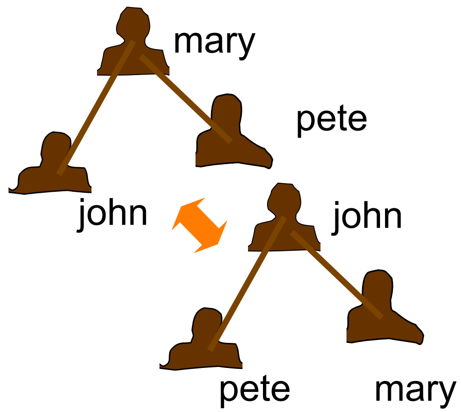
Weak (uninteresting) null hypotheses – what is it we are rejecting?

Test: Testing centralization using conditioning on density: $\mathscr{U}|\mathbf{X}_{++}$

Interpretation: network more centralised than expected by chance, but **also**, network not generated by randomly distributing edges

Test: Testing association between relations using QAP

Interpretation: relations are not unrelated, but **also**, ties are more concordant than if identities of vertices did not matter (sic)



Drawbacks of non model based statistical analysis

We have no model for what we are interested in – are "significant" effects artifacts of other effects ?

Test: Testing structural effects using $\mathcal{U}|MAN$

Limit in interpretation: what if we are interested in both reciprocity and triangulation?

Models for networks

- Models allow us to model features of the data that we are interested in
- If we are able to fit a model we (may) have adequately described the data (c.p. only holds true for non-parametric analysis when null hypothesis not rejected)
- Common critique:
  (a) only one observation
  (b) not inferring to population
  (c) where does “chance” come from?
- “chance” = “uncertainty”; possible process rather than sample (c.p. time series analysis)

MANCHESTER 1824
The University of Manchester

Models for networks

- Stochastic block models (e.g. Nowicki and Snijders, 2001)
- Latent class/ clustering models (e.g. Schweinberger, and Snijders, 2003; Handcock et al., 2007)
- Regressing variables on networks and covariates
  - the influence model (Robins et al., 2001)
  - the network effects and network autocorrelation models (Marsden and Friedkin, 1994)
- Models for longitudinal social network data (e.g. Snijders et al., 2007)