LISS2108

Statistical Inference for Social Networks Analysis

Session 3. Quadratic Assignment Procedure and Intro to ERGMs

Santiago Quintero

santiago.quintero_suarez@kcl.ac.uk

March 20th 2025

KING'S College LONDON

LISS DTP

Session outline

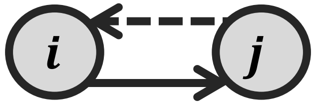
1. Review: Permutation-based tests and autoregressive models
2. Coding walkthrough
3. Inference at the dyad level
   3.1. Quadratic Assignment Procedure
      - QAP coding walkthrough

   3.2. Introduction to Exponential Random Graph Models (ERGMs)

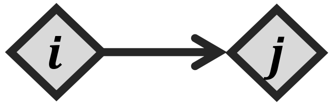
1. Review

Challenges of statistical inference in SNA

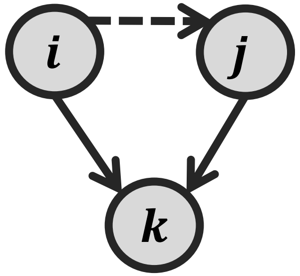
1. Usually, network analysis are made on one network (N = 1)
2. Data collection for network research is not random!
3. Observations in networks are not independent!
    - Networks exhibit many level of dyadic and triadic dependence
    - This might violate key assumption for identification (i.e., iid erros in classical regression framework)
    - We should be able to separate, identify, and control for them!



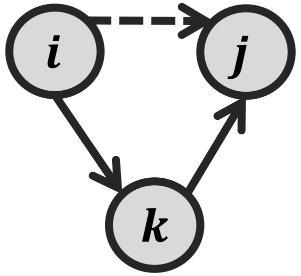
Reciprocity



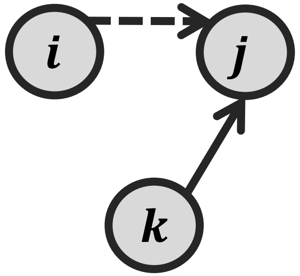
Homophily



Structural equivalence



Transitivity



Degree differentials

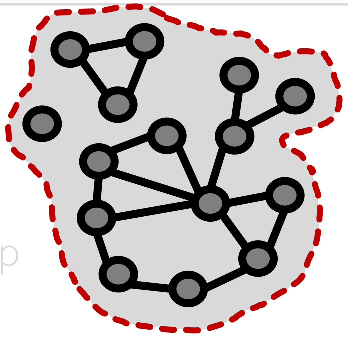
Modelling at the group level

Antecedents of networks
(Networks as *Dependent variable*)

<u>Group characteristics</u>
- Group size
- Composition
- Formal structure

Group level

Density or centralisation
Number of cliques
Level of "core-peripheriness"
Level of average homophily in a group



Consequences of networks
(Networks as *Inependent variable*)

<u>Group outcomes</u>
- Performance
- Group culture
- Intervention effect

Would classical statistical analyses (e.g., regressions) work here?

If we have a large enough (quasi)random sample of groups, and can demonstrate that "individuals" are not members of multiples groups: yes!

And what if we don't?

Modelling at the group level

Permutation (or randomisation) test

Procedure:

- Compute the actual correlation between the two variables
- Randomly shuffle (permute) one of the variables while keeping the network variable fixed
- Recalculate the correlation for each permutation (e.g., 1000 or 10,000 times)
- Compare the observed correlation to the distribution of permuted correlations

Significance testing: the p-value is *the proportion of permuted correlations that are as extreme as or more extreme than the observed correlation.*

$$p = \frac{\sum_{b=1}^{B} I\left(\left|r^{(b)}\right| \geq |r_{obs}|\right)}{B}$$

where $I(\cdot)$ is the indicator function that counts the number of permutations, B, where the permuted correlation, $r^{(b)}$, is as extreme as or more extreme than the observed correlation, r_{obs}.

- If p is small (e.g., $p < 0{,}05$), we reject the null hypothesis that the observed correlation is due to chance.
- Otherwise, we fail to reject the null, suggesting no strong evidence of a relationship.

(Borgatti et al, 2020)

Modelling at the node level

<u>Nodal (individual) characteristics</u>
- Age
- Values
- Reputation

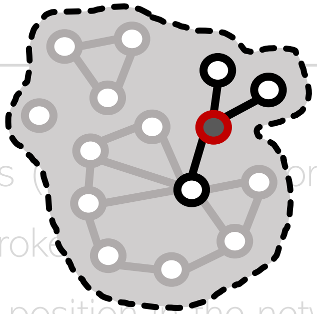
Node level

Degree, betweenness, closeness centrality

Level at which a node is similar to its alters (…ons)

How much a node fills a structural hole (broke…

How do node characteristics determine its position in the network



<u>Nodal (individual) outcomes</u>
- Performance
- Beliefs
- Well-being

Would classical statistical analyses (e.g., regressions) work here?

Can we use our permutation-based method to derive consistent test statistics?

How to control for network autocorrelation (i.e., the outcome of other observations affecting each other)?

Modelling at the node level

(Network/spatial) autoregressive models

We define our regression equation as:

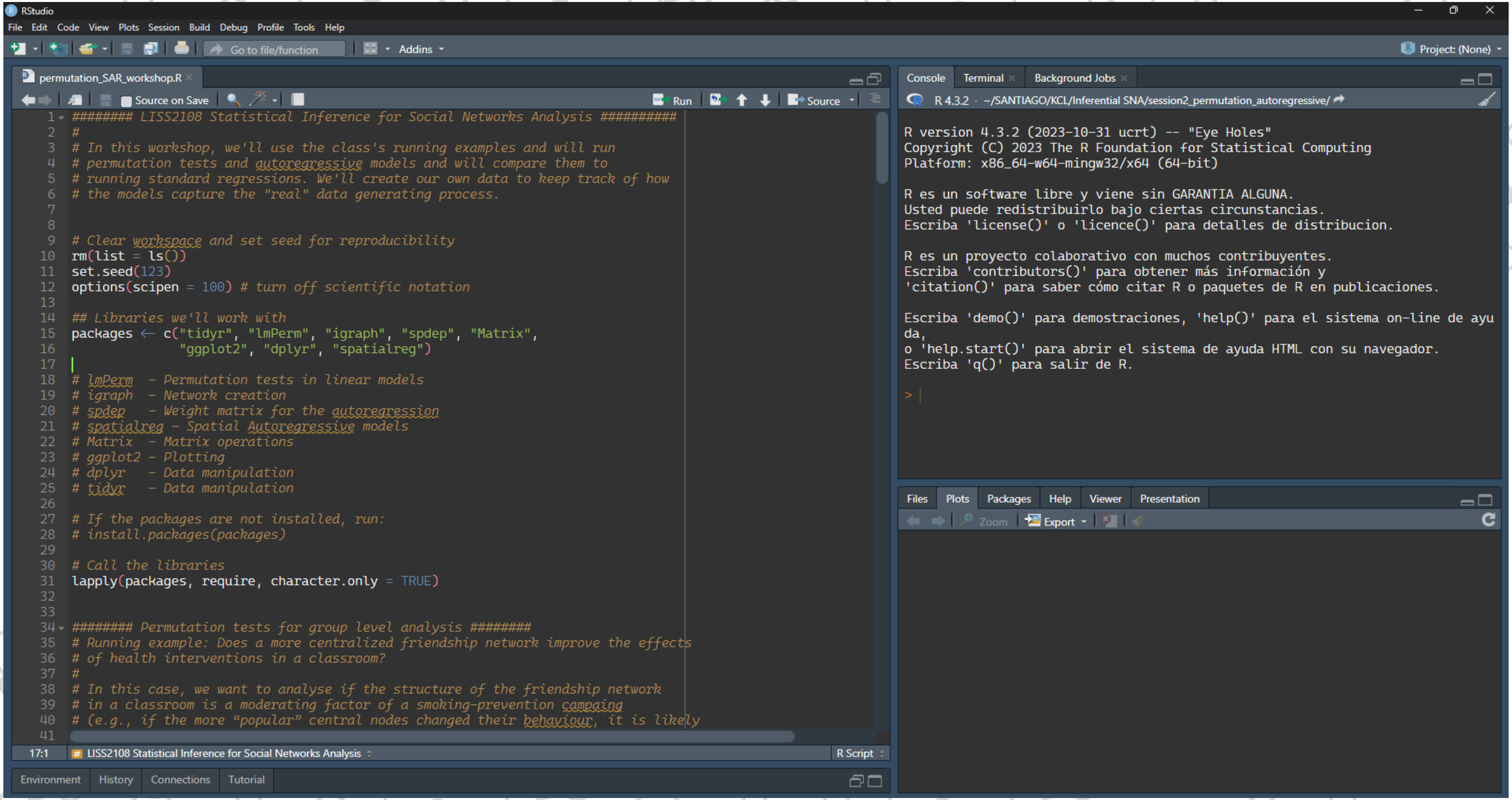
$$Y = \beta_0 + \boxed{\rho W Y} + \beta_1 X_1 + \beta_2 X_2 + \cdots + \varepsilon$$

Where the term $\rho W Y$ captures the other node's effect on each observation's main outcome Y.

- W represents a weights matrix that captures the strength of the relationship between the nodes
- Y is the matrix with the outcomes for the rest of the observations.
- ρ captures the size of the effect

(Griffith, 2009; Ver Hoef et al, 2018)

2. Coding walkthrough


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
permutation_SAR_workshop.R
Source on Save
Run
Source
######## LISS2108 Statistical Inference for Social Networks Analysis ##########
#
# In this workshop, we'll use the class's running examples and will run
# permutation tests and autoregressive models and will compare them to
# running standard regressions. We'll create our own data to keep track of how
# the models capture the "real" data generating process.
# Clear workspace and set seed for reproducibility
rm(list = ls())
set.seed(123)
options(scipen = 100) # turn off scientific notation
Libraries we'll work with
packages <- c("tidyr", "lmPerm", "igraph", "spdep", "Matrix",
"ggplot2", "dplyr", "spatialreg")
# lmPerm - Permutation tests in linear models
# igraph - Network creation
# spdep - Weight matrix for the autoregression
# spatialreg - Spatial Autoregressive models
# Matrix - Matrix operations
# ggplot2 - Plotting
# dplyr - Data manipulation
# tidyr - Data manipulation
# If the packages are not installed, run:
# install.packages(packages)
# Call the libraries
lapply(packages, require, character.only = TRUE)
######## Permutation tests for group level analysis ########
Running example: Does a more centralized friendship network improve the effects
of health interventions in a classroom?
#
In this case, we want to analyse if the structure of the friendship network
in a classroom is a moderating factor of a smoking-prevention campaing
(e.g., if the more "popular" central nodes changed their behaviour, it is likely
17:1
LISS2108 Statistical Inference for Social Networks Analysis
R Script
Environment History Connections Tutorial
Console Terminal Background Jobs
R 4.3.2 · ~/SANTIAGO/KCL/Inferential SNA/session2_permutation_autoregressive/
R version 4.3.2 (2023-10-31 ucrt) -- "Eye Holes"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.
R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.
Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.
>
Files Plots Packages Help Viewer Presentation
Zoom
Export

3. Inference at the dyad level

Modelling at the dyad level

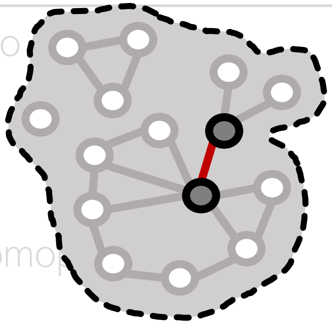
Dyad level

Presence or absence of a tie between two

Distance between two nodes

Number of common connections

Whether connected nodes are similar (homo



Dyadic characteristics

- Similarity (homophily)
- Peer-effects
- Surrounding

Dyadic outcomes

- Transfer

Examples:

- Who do people seek advice from?
- Why do firms do more business with some firms that with others?
- Why and which legislators work together to propose a bill?
- Are academics more likely to collaborate if they have a co-author in common?

Modelling at the dyad level

What are the challenges for inference in these cases?

Examples:

- Who do people seek advice from?
- Why do firms do more business with some firms that with others?
- Why and which legislators work together to propose a bill?
- Are academics more likely to collaborate if they have a co-author in common?

Modelling at the dyad level

What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Modelling at the dyad level

What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Friendship

	Ana	Tom	Jan	Lucy
Ana	0	1	0	0
Tom	1	0	1	1
Jan	0	1	0	1
Lucy	0	1	1	0

Advice-seeking

	Ana	Tom	Jan	Lucy
Ana	0	1	0	0
Tom	1	0	1	3
Jan	0	1	0	4
Lucy	0	3	4	0

Modelling at the dyad level

What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Friendship

	Ana	Tom	Jan	Lucy
Ana	0	1	0	0
Tom	1	0	1	1
Jan	0	1	0	1
Lucy	0	1	1	0

Advice-seeking

	Ana	Tom	Jan	Lucy
Ana	0	1	0	0
Tom	1	0	1	3
Jan	0	1	0	4
Lucy	0	3	4	0

Dyads	Friendship	Advice-seeking
Ana-Tom	1	1
Ana-Jan	0	0
Ana-Lucy	0	0
Tom-Jan	1	1
Tom-Lucy	1	3
Lucy-Jan	1	4

Modelling at the dyad level

What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Observations are not independent!

Dyads	Friendship	Advice-seeking
Ana-Tom	1	1
Ana-Jan	0	0
Ana-Lucy	0	0
Tom-Jan	1	1
Tom-Lucy	1	3
Lucy-Jan	1	4

Modelling at the dyad level

What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Observations are not independent!

- There is $n-1$ observations that involve node i.
  If, e.g., Lucy is very outgoing, it will affect the values of every other dyad involving Lucy.

Dyads	Friendship	Advice-seeking
Ana-Tom	1	1
Ana-Jan	0	0
Ana-Lucy	0	0
Tom-Jan	1	1
Tom-Lucy	1	3
Lucy-Jan	1	4

Modelling at the dyad level

What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Observations are not independent!

- There is $n-1$ observations that involve node i.
  If, e.g., Lucy is very outgoing, it will affect the values of every other dyad involving Lucy.
- There are other forms **dyadic and triadic dependence**.
  E.g., As Tom and Lucy seek advice from each other, and Lucy and Jan do as well, it is more likely for Tom and Jan seek advice from each other than from Ana.

Dyads	Friendship	Advice-seeking
Ana-Tom	1	1
Ana-Jan	0	0
Ana-Lucy	0	0
Tom-Jan	1	1
Tom-Lucy	1	3
Lucy-Jan	1	4

Modelling at the dyad level

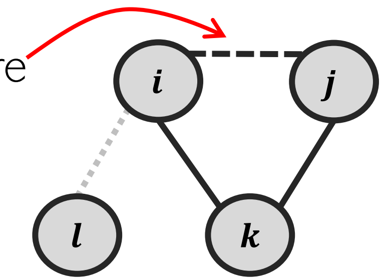
What are the challenges for inference in these cases?

Let's assume we have an edgelist containing information about two networks for the same group of people, one on friendships and one on advice-seeking:

Observations are not independent!

- There is $n-1$ observations that involve node i.
  If, e.g., Lucy is very outgoing, it will affect the values of every other dyad involving Lucy.
- There are other forms **dyadic and triadic dependence**.
  E.g., As Tom and Lucy seek advice from each other, and Lucy and Jan do as well, it is more likely for Tom and Jan seek advice from each other than from Ana.

This is known as triadic closure



Dyads	Friendship	Advice-seeking
Ana-Tom	1	1
Ana-Jan	0	0
Ana-Lucy	0	0
Tom-Jan	1	1
Tom-Lucy	1	3
Lucy-Jan	1	4

Modelling at the dyad level

Can't we simply run any correlation or regression analysis using the permutation tests to account for autocorrelation?

Modelling at the dyad level

Can't we simply run any correlation or regression analysis using the permutation tests to account for autocorrelation?

Well, we can. But it won't capture the structural dependencies of the network (i.e., the dyadic and triadic effects)

3.1. Quadratic Assignment Procedure

Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

The basic idea is the same as with permutation tests:

- We calculate the correlation statistic
- Then permute the data a good number of times and check if the observed correlation is rare enough—compared to the distribution of randomly generated correlations—as to not happen by chance.
- However, the way the permutations are done is special!

(Borgatti et al, 2020; Cranmer et al., 2017)

Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

For QAP, the rows and their corresponding columns in the (predictor) matrix are randomly permuted, while names and their attributes are held in the original order.

The main idea is that *we shuffle a node with which a particular label and set of attributes is associated while maintaining the structure of the network.*

(Borgatti et al, 2020)

Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

For QAP, the rows and their corresponding columns in the (predictor) matrix are randomly permuted, while names and their attributes are held in the original order.

The main idea is that *we shuffle a node with which a particular label and set of attributes is associated while maintaining the structure of the network.*

	Ana	Tom	Jan	Lucy
Ana	0	1	0	0
Tom	1	0	1	3
Jan	0	1	0	4
Lucy	0	3	4	0

Original advice-seeking network

→

	~~Lucy~~ Ana	~~Jan~~ Tom	~~Ana~~ Jan	~~Tom~~ Lucy
~~Lucy~~ Ana	0	4	0	3
~~Jan~~ Tom	4	0	0	1
~~Ana~~ Jan	0	0	0	1
~~Tom~~ Lucy	3	1	1	0

Permuted advice-seeking network

Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

For QAP, the rows and their corresponding columns in the (predictor) matrix are randomly permuted, while names and their attributes are held in the original order.

The main idea is that *we shuffle a node with which a particular label and set of attributes is associated while maintaining the structure of the network.*

	Ana	Tom	Jan	Lucy
Ana	0	1	0	0
Tom	1	0	1	3
Jan	0	1	0	4
Lucy	0	3	4	0

Original advice-seeking network

	~~Lucy~~ Ana	~~Jan~~ Tom	~~Ana~~ Jan	~~Tom~~ Lucy
~~Lucy~~ Ana	0	4	0	3
~~Jan~~ Tom	4	0	0	1
~~Ana~~ Jan	0	0	0	1
~~Tom~~ Lucy	3	1	1	0

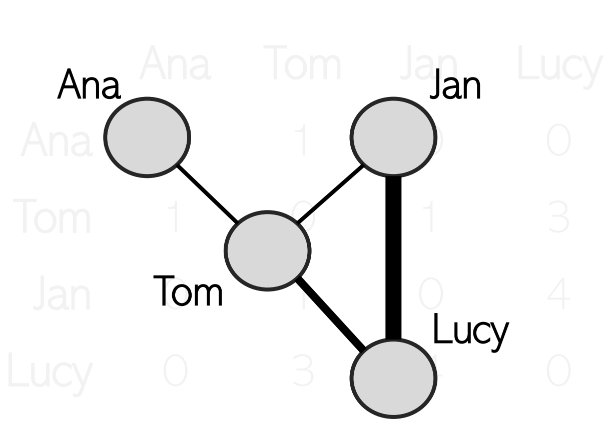
Permuted advice-seeking network

Modelling at the dyad level: QAP

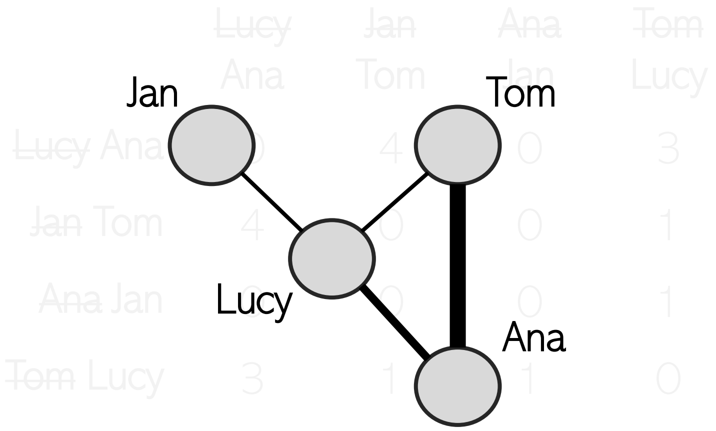
Quadratic Assignment Procedure (QAP)

For QAP, the rows and their corresponding columns in the (predictor) matrix are randomly permuted, while names and their attributes are held in the original order.

The main idea is that *we shuffle a node with which a particular label and set of attributes is associated while maintaining the structure of the network.*



Original advice-seeking network



Permuted advice-seeking network

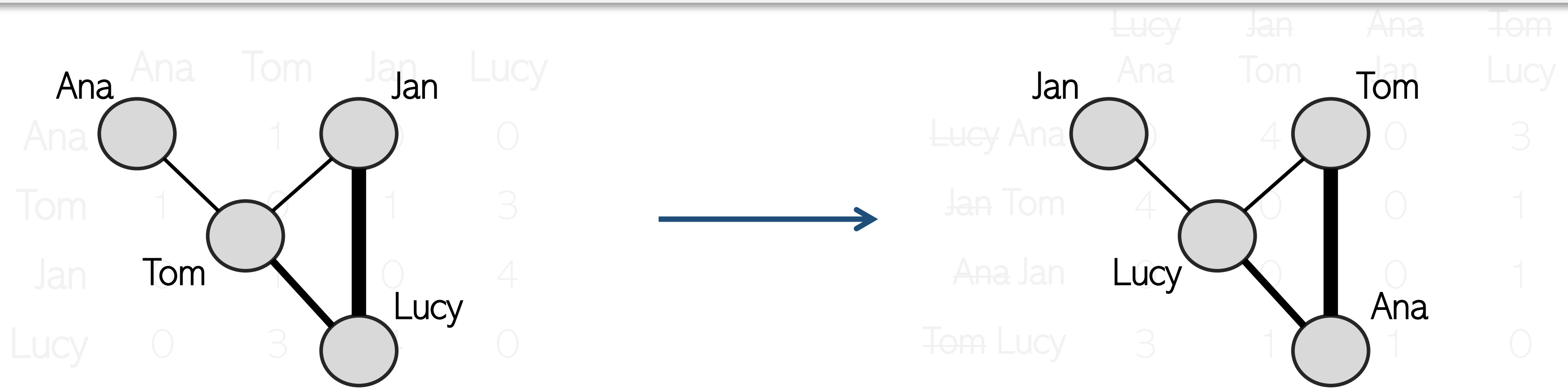
Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

For QAP, the rows and their corresponding columns in the (predictor) matrix are randomly permuted, while names and their attributes are held in the original order.

The main idea is that *we shuffle a node with which a particular label and set of attributes is associated while maintaining the structure of the network.*

This way, **we implicitly account for all structural dependencies automatically**, even if we are unaware of them



Original advice-seeking network

Permuted advice-seeking network

Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

QAP can be used for bivariate correlational analysis.

But there are also extensions that fit a regression framework

Modelling at the dyad level: MRQAP

Quadratic Assignment Procedure (QAP)

QAP can be used for bivariate correlational analysis.

But there are also extensions that fit a regression framework

- Multiple Regression Quadratic Assignment Procedure (MRQAP)
  - When the predicted network is weighted (i.e., the ties have values) it is essentially the same as an OLS but the *p*-value is calculated using QAP permutations.
  - When the predicted network is binary, it is essentially a *linear probability model*

Modelling at the dyad level: LRQAP

Quadratic Assignment Procedure (QAP)

QAP can be used for bivariate correlational analysis.

But there are also extensions that fit a regression framework

- Multiple Regression Quadratic Assignment Procedure (MRQAP)
  - When the predicted network is weighted (i.e., the ties have values) it is essentially the same as an OLS but the *p*-value is calculated using QAP permutations.
  - When the predicted network is binary, it is essentially a *linear probability model*

- Logistic Regression Quadratic Assignment Procedure (LRQAP)
  - Logit model using QAP permutations to calculate statistical significance

Modelling at the dyad level: QAP

Quadratic Assignment Procedure (QAP)

QAP can be used for bivariate correlational analysis.

But there are also extensions that fit a regression framework

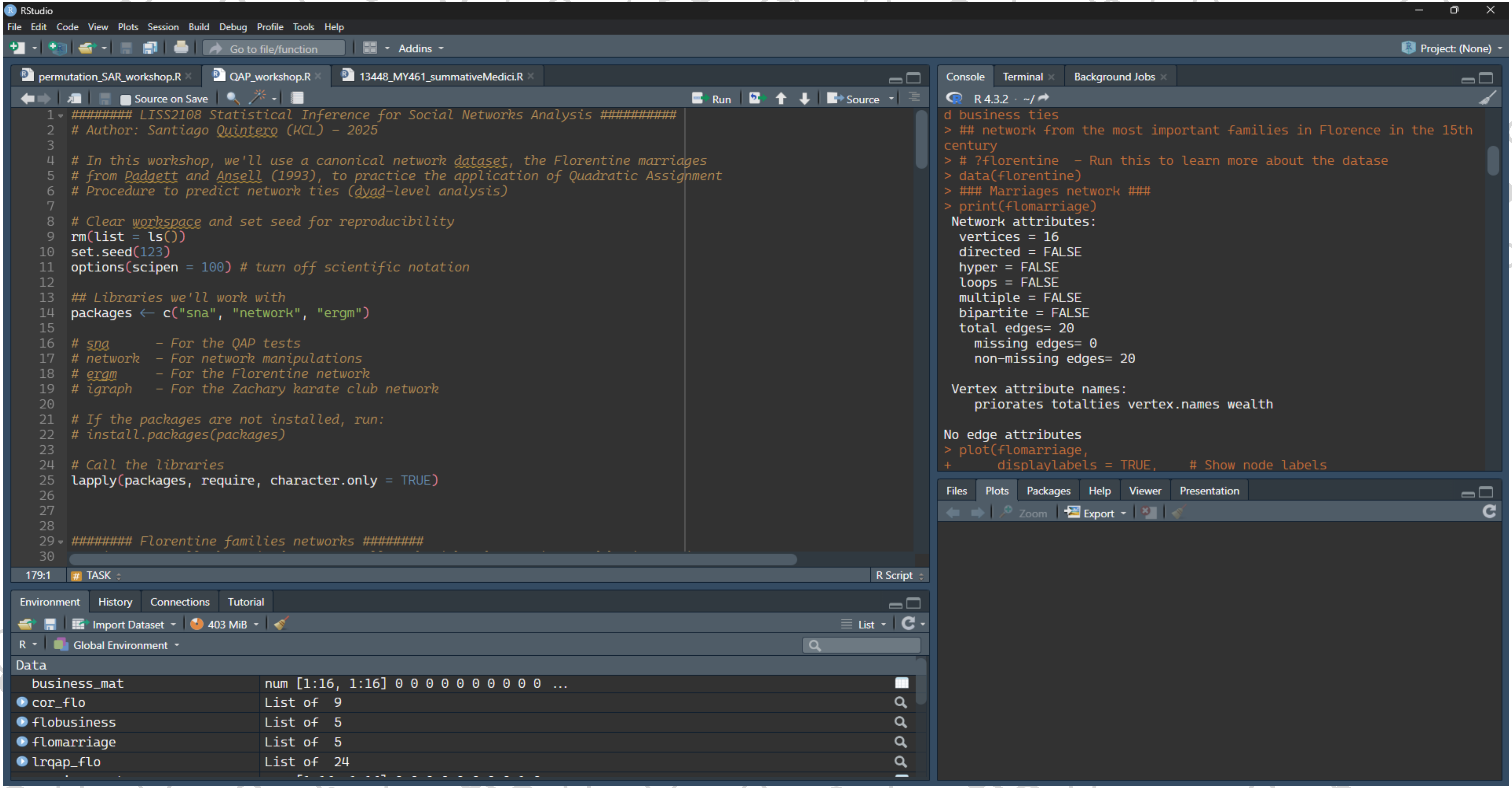
- Multiple Regression Quadratic Assignment Procedure (MRQAP)
    - When the predicted network is weighted (i.e., the ties have values) it is essentially the same as an OLS but the *p*-value is calculated using QAP permutations.
    - When the predicted network is binary, it is essentially a *linear probability model*

- Logistic Regression Quadratic Assignment Procedure (LRQAP)
    - Logit model using QAP permutations to calculate statistical significance

There are several methods for the specific QAP permutation, the most common being the *double semi-partialing method*, where we perform the permutations using the residual matrix from the original regression and we track the *t*-statistic rather than the regression coefficients (Dekker et al., 2007).

(Borgatti et al, 2020)

3.1.2. QAP coding walkthrough


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
permutation_SAR_workshop.R
QAP_workshop.R
13448_MY461_summativeMedici.R
Source on Save
Run
Source
######## LISS2108 Statistical Inference for Social Networks Analysis ##########
# Author: Santiago Quintero (KCL) - 2025
# In this workshop, we'll use a canonical network dataset, the Florentine marriages
# from Padgett and Ansell (1993), to practice the application of Quadratic Assignment
# Procedure to predict network ties (dyad-level analysis)
# Clear workspace and set seed for reproducibility
rm(list = ls())
set.seed(123)
options(scipen = 100) # turn off scientific notation
Libraries we'll work with
packages <- c("sna", "network", "ergm")
# sna   - For the QAP tests
# network - For network manipulations
# ergm   - For the Florentine network
# igraph   - For the Zachary karate club network
# If the packages are not installed, run:
# install.packages(packages)
# Call the libraries
lapply(packages, require, character.only = TRUE)
######## Florentine families networks ########
179:1
TASK
R Script
Console
Terminal
Background Jobs
R 4.3.2 · ~/
d business ties
> ## network from the most important families in Florence in the 15th century
> # ?florentine - Run this to learn more about the datase
> data(florentine)
> ### Marriages network ###
> print(flomarriage)
Network attributes:
vertices = 16
directed = FALSE
hyper = FALSE
loops = FALSE
multiple = FALSE
bipartite = FALSE
total edges= 20
missing edges= 0
non-missing edges= 20
Vertex attribute names:
priorates totalties vertex.names wealth
No edge attributes
> plot(flomarriage,
+   displaylabels = TRUE,   # Show node labels
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Environment
History
Connections
Tutorial
Import Dataset
403 MiB
List
R
Global Environment
Data
business_mat   num [1:16, 1:16] 0 0 0 0 0 0 0 0 0 0 ...
cor_flo   List of 9
flobusiness   List of 5
flomarriage   List of 5
lrqap_flo   List of 24

3.2. Introduction to ERGMs

Limitations of QAP models

While powerful, the QAP approach has a fundamental drawback: we cannot directly observe the network structure.

Network structures are not only statistical dependencies to account for. They are **meaningful social processes** we want to analyse.

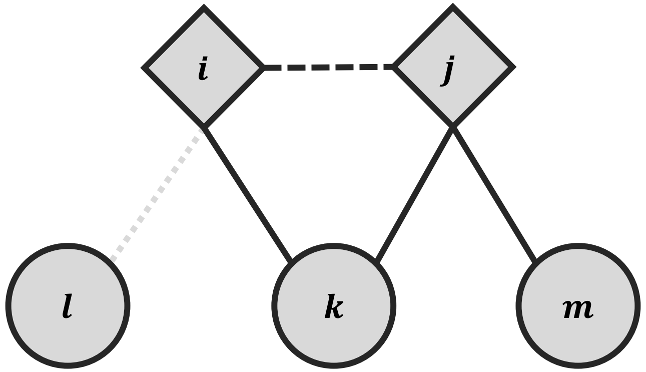
(Cranmer et al., 2017)

Limitations of QAP models

While powerful, the QAP approach has a fundamental drawback: we cannot directly observe the network structure.

Network structures are not only statistical dependencies to account for. They are **meaningful social processes** we want to analyse.

Take a case like the one we used to talk about triadic dependencies:

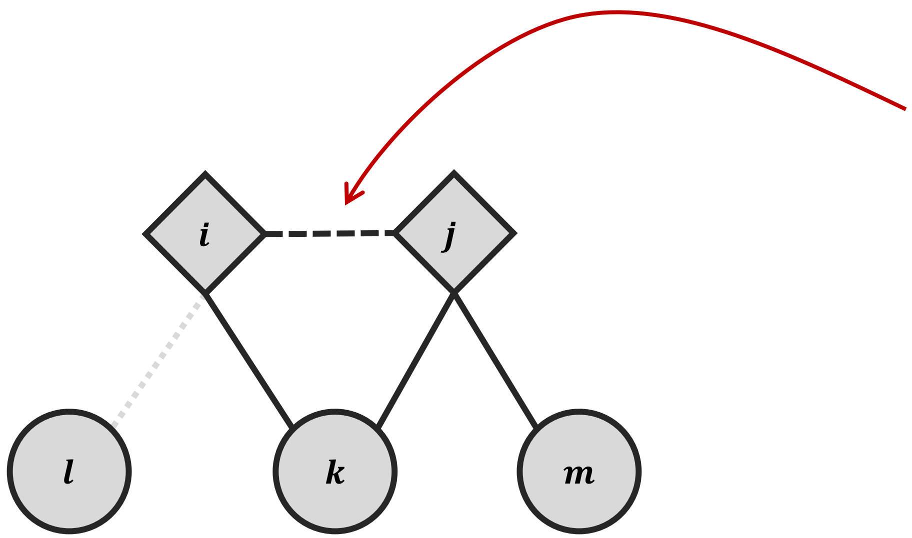


Limitations of QAP models

While powerful, the QAP approach has a fundamental drawback: we cannot directly observe the network structure.

Network structures are not only statistical dependencies to account for. They are **meaningful social processes** we want to directly analyse.

Take a case like the one we used to talk about triadic dependencies:



Did this tie emerge because of:

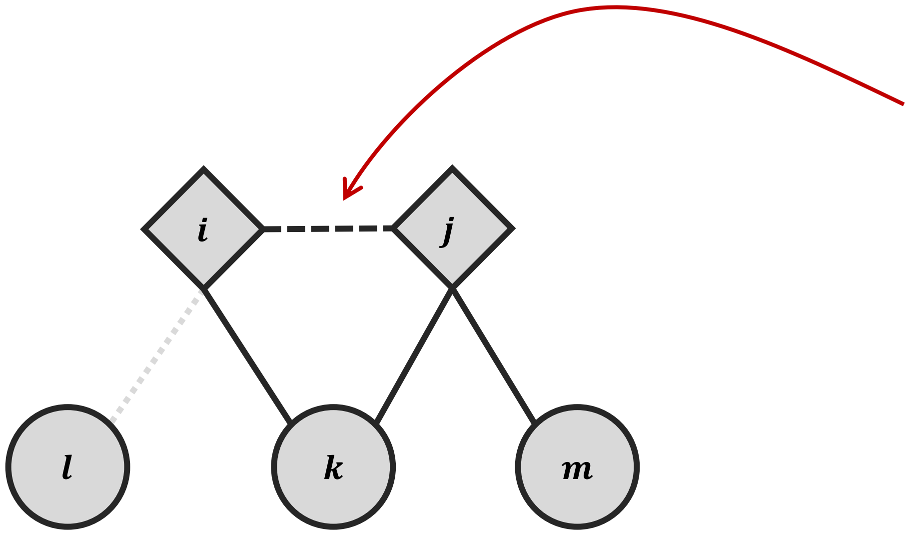
- A friend of a friend is my friend? (triadic closure)
- The rich get richer? (centralisation/popularity)
- Birds of a feather flock together? (homophily)

Limitations of QAP models

While powerful, the QAP approach has a fundamental drawback: we cannot directly observe the network structure.

Network structures are not only statistical dependencies to account for. They are **meaningful social processes** we want to directly analyse.

Take a case like the one we used to talk about triadic dependencies:



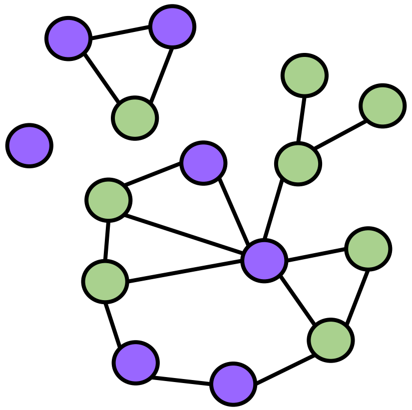
Did this tie emerge because of:

- A friend of a friend is my friend? (triadic closure)
- The rich get richer? (centralisation/popularity)
- Birds of a feather flock together? (homophily)

Can we model this factors explicitly, at the same time?

A (somewhat) intuitive introduction to ERGMS

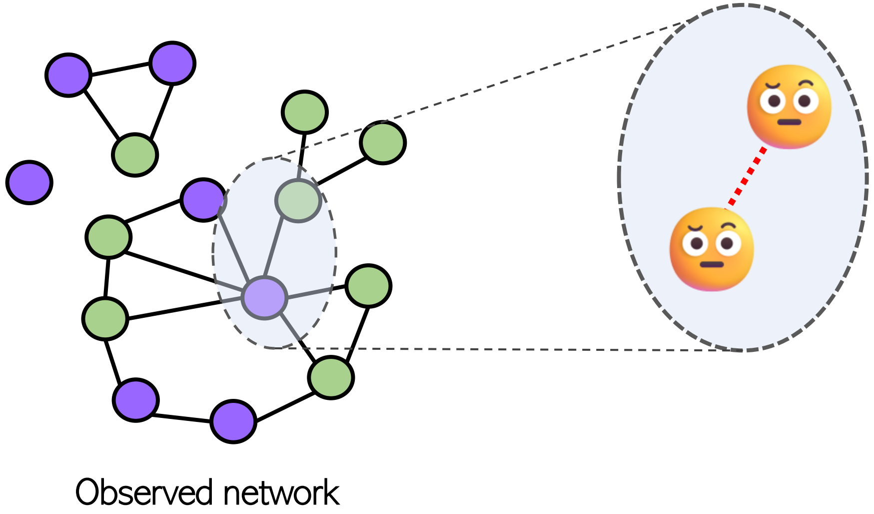
Consider an observed network as an aggregation of *local decision making-processes*—i.e., the sum of *individual* ***tie-based decisions*** *among pairs of nodes*



Observed network

A (somewhat) intuitive introduction to ERGMS

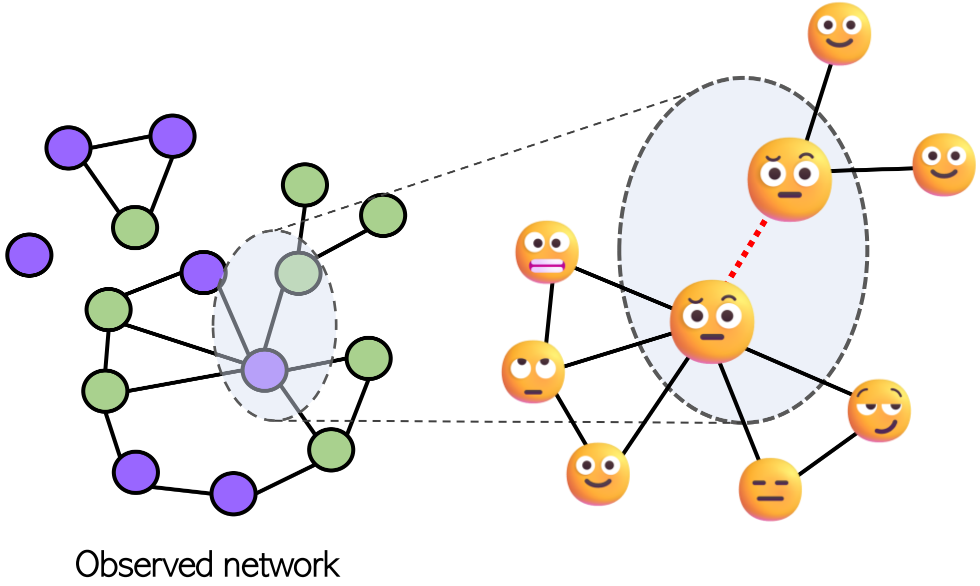
Consider an observed network as an aggregation of *local decision making-processes*—i.e., the sum of *individual* ***tie-based decisions*** *among pairs of nodes*



(Agneessens, 2023; Robins et al. 2007; Harris, 2014)

A (somewhat) intuitive introduction to ERGMS

Consider an observed network as an aggregation of *local decision making-processes*—i.e., the sum of *individual* ***tie-based decisions*** *among pairs of nodes, taking into account a* ***local environment.***

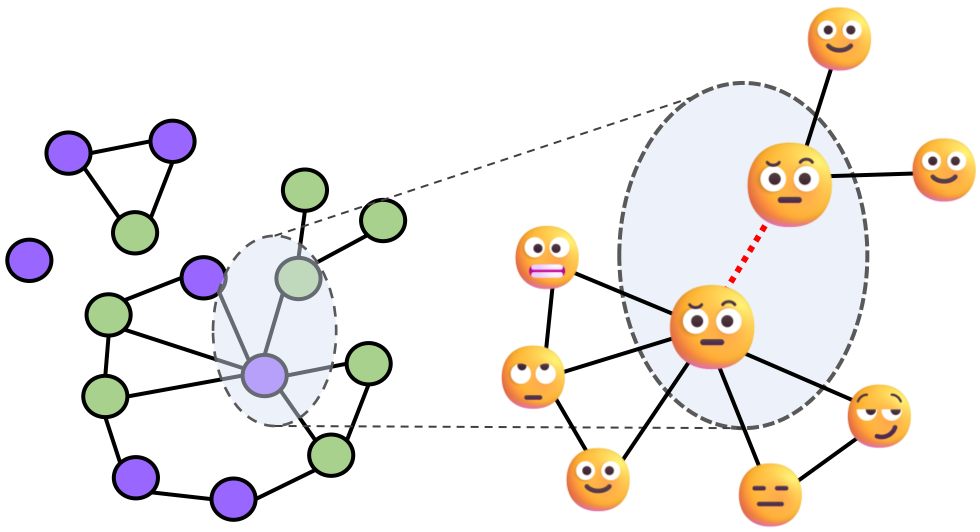


- This tie-based process is driven by an **aggregation** of **"local" social forces** that, we assume, led to the **current structure of the observed network.**

(Agneessens, 2023; Robins et al. 2007; Harris, 2014)

A (somewhat) intuitive introduction to ERGMS

Consider an observed network as an aggregation of *local decision making-processes*—i.e., the sum of *individual* ***tie-based decisions*** *among pairs of nodes, taking into account a* ***local environment.***



Observed network

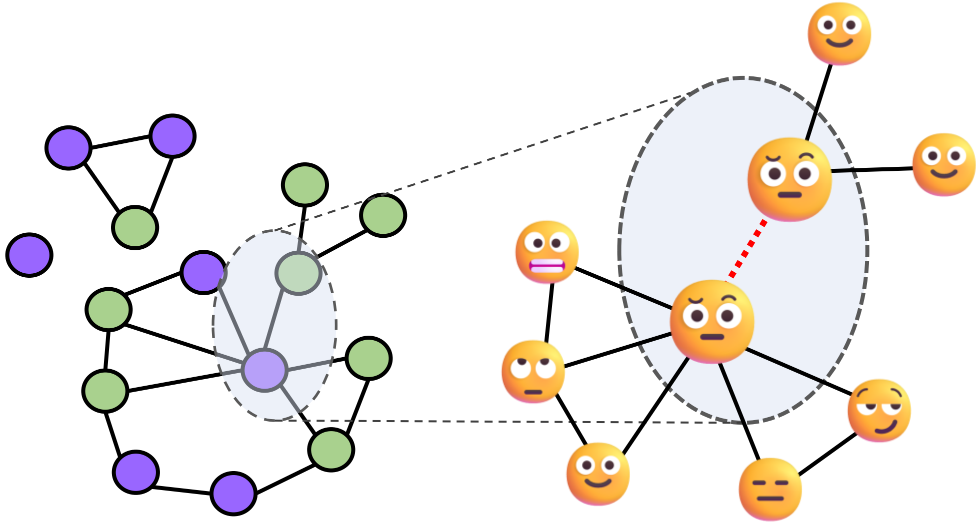
- This tie-based process is driven by an **aggregation** of **"local" social forces** that, we assume, led to the **current structure of the observed network.**

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

(Agneessens, 2023; Robins et al. 2007; Harris, 2014)

A (somewhat) intuitive introduction to ERGMS

Consider an observed network as an aggregation of *local decision making-processes*—i.e., the sum of *individual **tie-based decisions** among pairs of nodes, taking into account a **local environment**.*



Observed network

- This tie-based process is driven by an **aggregation** of **"local" social forces** that, we assume, led to the **current structure of the observed network**.

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

Y = Random variable for the state of the network
y = Actual network
$c(\theta)$ = Normalising constant to ensure probabilities $\{0,1\}$ (partition function)
θ_k = Coefficients for network statistics (e.g., number of edges, triangles, degree distribution, etc.)
$z_k(y)$ = Network statistics k

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

A (somewhat) intuitive introduction to ERGMS

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

However, we don't simply care about predicting the whole network, but this general formulation implies that:

$$\text{logit}\left(P\left(Y_{ij} = 1 \middle| n \text{ actors}, Y_{ij}^c\right)\right) = \sum_{k=1}^{K} \theta_k \delta_{z_k(y)}$$

And if we rearrange to have the actual conditional probability of a tie between i and j in the left-hand side:

$$P\left(Y_{ij} = 1 \middle| n\ actors, Y_{ij}^c\right) = \text{logistic}(\theta_1 \delta_{z_1(y)} + \theta_2 \delta_{z_2(y)} + \theta_2 \delta_{z_3(y)} + \cdots)$$

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

A (somewhat) intuitive introduction to ERGMS

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

However, we don't simply care about predicting the whole network, but this general formulation implies that:

$$\text{logit}\left(P\left(Y_{ij} = 1 \middle| n \text{ actors}, Y_{ij}^c\right)\right) = \sum_{k=1}^{K} \theta_k \delta_{z_k(y)}$$

The "change statistic", $\boldsymbol{\delta}$, refers to the change in the network statistic when Y_{ij} goes from 0 to 1

The probability of a tie between $\boldsymbol{i}$ and $\boldsymbol{j}$ …

… conditional on the rest of the network (Y_{ij}^c means all dyads other than Y_{ij})

We sum the product of all network statistics and their coefficients, $\boldsymbol{\theta\delta}$

And if we rearrange to have the actual conditional probability of a tie between i and j in the left-hand side:

$$P\left(Y_{ij} = 1 \middle| n\ actors, Y_{ij}^c\right) = \text{logistic}(\theta_1 \delta_{z_1(y)} + \theta_2 \delta_{z_2(y)} + \theta_2 \delta_{z_3(y)} + \cdots)$$

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

A (somewhat) intuitive introduction to ERGMS

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

- In practical terms, we use our model of the network to explore the many processes that could have generated it.
- We then generate thousands of simulations of the observed network to se how closely they approximate it.
- If some formation processes don't seem to generate networks that look like the observed one, they probably don't matter.

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

A (somewhat) intuitive introduction to ERGMS

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

- In practical terms, we use our model of the network to explore the many processes that could have generated it.
- We then generate thousands of simulations of the observed network to se how closely they approximate it.
- If some formation processes don't seem to generate networks that look like the observed one, they probably don't matter.

The problem:

- There are too many possible networks to be simulated.

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

A (somewhat) intuitive introduction to ERGMS

$$P(Y = y) = \frac{\exp\{\sum_{k=1}^{K} \theta_k z_k(y)\}}{c(\theta)}$$

The problem:

- There are too many possible networks to be simulated.
- The normalising constant requires summing all possible networks of the size of the observed one
- For a network with N nodes, the number of possible undirected graphs is $2^{\frac{N(N-1)}{2}}$ (each edge can be present or absent).
- This grows exponentially, making it **computationally intractable**!
- $c(\theta)$ has no closed-form solution and cannot be computed using Maximum Likelihood Estimation (as other GLM)

So, what do we do now?

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

A (somewhat) intuitive introduction to ERGMS

$$P(Y=y)=\frac{\exp\{\sum_{k=1}^{K}\theta_k z_k(y)\}}{c(\theta)}$$

The problem:

- There are too many possible networks to be simulated.
- The normalising constant requires summing all possible networks of the size of the observed one
- For a network with N nodes, the number of possible undirected graphs is $2^{\frac{N(N-1)}{2}}$ (each edge can be present or absent).
- This grows exponentially, making it **computationally intractable**!
- $c(\theta)$ has no closed-form solution and cannot be computed using Maximum Likelihood Estimation (as other GLM)

So, what do we do now?

Pseudo-MLE
MCMC!

(Agneessens, 2023; Robins et al. 2007; Harris, 2014; Power, 2001)

Bibliography

- Agneessens, F. (2023). *Statistical Analysis for Cross-Sectional and Longitudinal Social Network Analysis*. Essex Summer School in Social Science Data Analysis. University of Essex
- Borgatti, S., Everett, M.G., Johnson, J.C. & Agneessens, F. (2022). *Analyzing Social Networks Using R*. SAGE Publications
- Cranmer, S. J., Leifeld, P., McClurg, S. D., & Rolfe, M. (2017). Navigating the Range of Statistical Tools for Inferential Network Analysis. *American Journal of Political Science, 61(1*), 237–251
- Dekker, D., Krackhardt, D. & Snijders, T.A.B. (2007). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika, 72*, 563-81
- Harris, J. K. (2014). *An introduction to exponential random graph modeling*. Thousand Oaks: SAGE
- Padgett, J.F. & Ansell, C.K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology, 98*(6), 1259-1319
- Power, E. (2021). *MY461. Social Network Analysis*. Department of Methodology, London School of Economics and Political Science
- Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 173–191