SOCIAL NETWORK ANALYSIS for DATA SCIENTISTS

today's menu: LECTURE: ERGM II (LECTURE Week 4)

Your lecturer: Claudia

Playdate: September, 24th, 2025

Disclaimer: The LLB conundrum

SNA4DS

Lectures + Labs + Books = Good preparation

I'm not covering the book in class

The lab puts in practice what we do in class (including the tricks to run the models)

Menu' for today

- Recap of the previous episode
- How to fit the models in R
- Classification of the types of effects
- Finding your way in the term/effect Jungle
- Fitting ERGMs with exogenous terms

ERGM for a Dutch speaker

Phonetic pronunciation for Dutch speakers

"eurcoom"

or

"eurghum"

Recap of the previous episode

Lecture ERGM I

- ERGMs are models for causal inference - if you do not have Hypotheses to test they make no sense
- Inferential network analysis test Hypotheses about why an observed network is the way it is.
- Two types of effects:
    - Endogenous: structural - terms that predict the probability of observing a certain network structure
    - Exogenous: as in GLM - variables external to the network that predict the occurrence of ties
- ERGM with edges term (Erdos Renyi model)
- ERGM with edges + sender + receiver + mutual terms (P1 model)
- ERGM concepts and math

How to fit the models in R

Data set: Florentine Families

16 Families - each node is a family that lived in Florence in the 15th century

Acciaiuoli - Albizzi - Barbadori - Bischeri - Castellani - Ginori - Guadagni - Lamberteschi - Medici - Pazzi - Peruzzi - Pucci - Ridolfi - Salviati - Strozzi - Tornabuoni

Data collected by historian John Padgett.

Two networks: 1) business 2) marriage

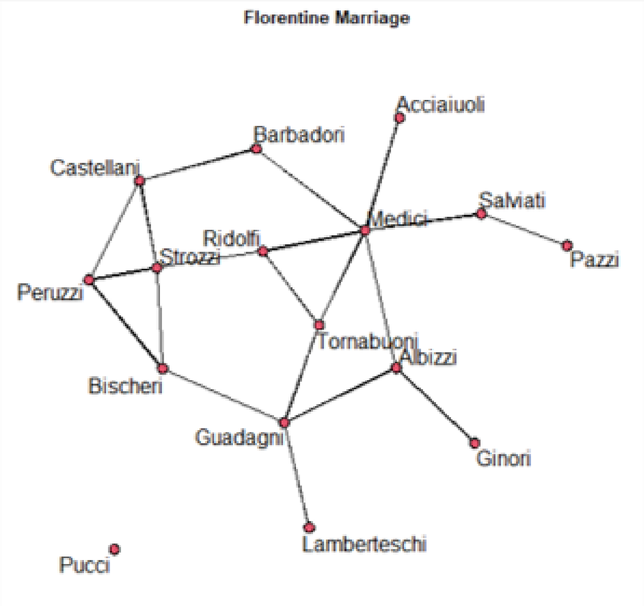
Attributes:

- Wealth (v/n): how reach
- Number Priorates (v/n): How many times a member of their family was Prior of Florence between 1282 and 1344
- Weight (e/n): number of times two families do business/marry to each other

ERGM with edges term (Erdos Renyi)

Florentine Marriage

- 16 nodes - 20 edges - UNDIRECTED
- We observe a marriage pattern (certain network structure). Is it random?
- RQ: Do these Florentine families get married in a random way? Is love blind?
- H1: Love is not blind in Renaissance Florence



RQ: Is love blind in Renaissance Florence?

H: Love is not blind in Renaissance Florence

```
flomodel.01 <- ergm::ergm(flomarriage ~ edges)
summary(flomodel.01)
```

```
## Call:
## ergm::ergm(formula = flomarriage ~ edges)
##
## Maximum Likelihood Results:
##
##       Estimate Std. Error MCMC % z value Pr(>|z|)
## edges  -1.6094     0.2449      0  -6.571   <1e-04 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 166.4  on 120  degrees of freedom
##  Residual Deviance: 108.1  on 119  degrees of freedom
##
## AIC: 110.1  BIC: 112.9  (Smaller is better. MC Std. Err. = 0)
```

Love has very good sight (The p-value is significant)

negative coefficient in this case means that the network is sparse.

ERGM with the edge term (ER)

SNA4DS

Very simple model

We can use it to see whether the edges are random or if they exist for a reason

You will insert the `edge` term in every model -always important

We use it as the intercept of the model - still it has a meaning

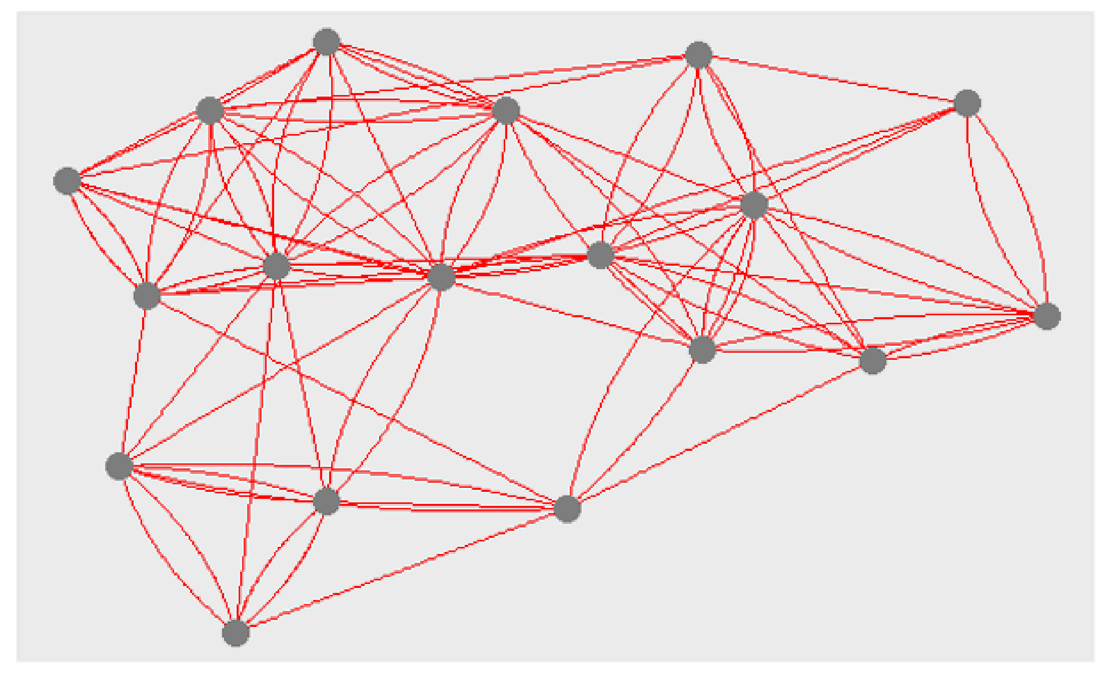
This will be the first term in every model you run

ERGM with 4 endogenous parameters (PI)

SNA4DS

Holland and Leinhardt analyzed Sampson's monks dataset

- Ethnographic study of community structure in a New England monastery by Samuel F. Sampson 1968.
- Social relationships among a group of men (novices) who were preparing to join a monastic order.
- Simplified data set - 18 nodes - 88 edges - DIRECTED



Holland and Leinhardt analyzed Sampson's monks dataset (P1 model)

RQ: Are the social relationships between these monks random or driven by specific social dynamics?

H1: Monks are not friends with everyone (friendship between monks is not random)

looking at

- Probability of forming edges (ER model: edges are not random)
- probability of a monk wanting to be friends with another monk (sender)
- Probability that the other monks want to be friends with the first one (receiver)
- Probability that the two monks mutually want to be each other's friends (mutual)

Hypothesis testing on Sampson data with P1 formulation

- NB: multiple terms for one hypothesis
- Look at p-values to see whether we can discard the null Hypothesis
- Look at the coefficient to appraise the intensity of the effect
- Relate results of sender and receiver to the focal node only (!!!)

```
p1 <- ergm::ergm(sampson ~ edges + sender + receiver + mutual)

summary(p1)
```

```
## Call:
## ergm::ergm(formula = sampson ~ edges + sender + receiver + mutual)
##
## Monte Carlo Maximum Likelihood Results:
##
##            Estimate Std. Error MCMC % z value Pr(>|z|)
## edges      -2.92437    0.90438      0  -3.234  0.00122 **
## sender2    -0.54652    1.12866      0  -0.484  0.62823
## sender3     0.25538    1.06199      0   0.240  0.80997
## sender4    -0.35179    1.05470      0  -0.334  0.73872
## sender5    -0.22805    1.10388      0  -0.207  0.83633
## sender6     0.02975    1.13023      0   0.026  0.97900
## sender7    -0.50708    1.14970      0  -0.441  0.65917
```

Individuals VS network structure

Sender and Receiver provide you with results about each node other than the focal one

They are an exception!

most terms inform us about the network overall

statistical models are mostly about collective behavior

rather than individual behavior

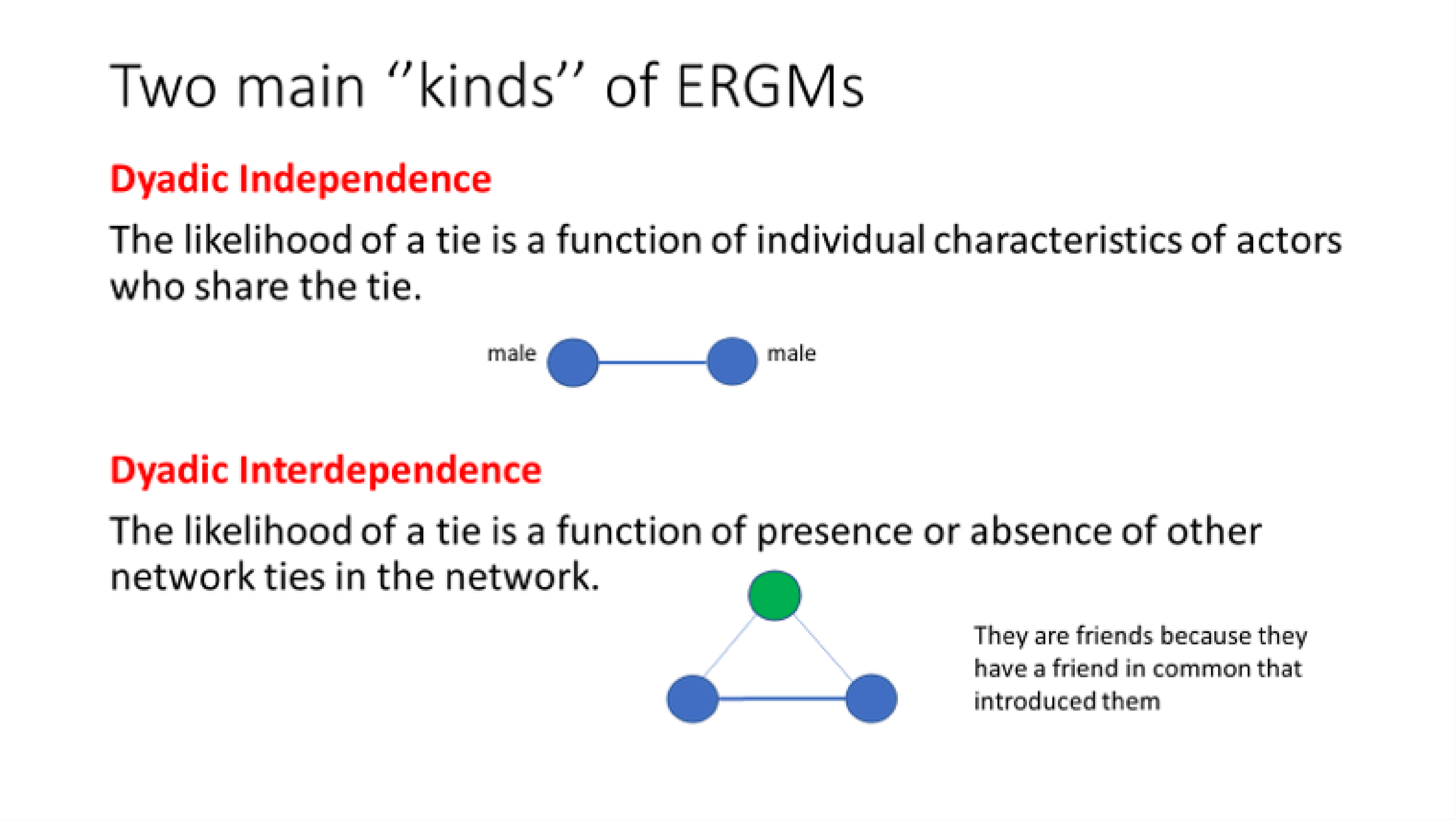
Classification of the types of terms/effects

Classification of the types of terms/effects

classif. 1

Classification of the types of terms/effects

classif. 2



Estimation VS Simulation

SNA4DS

Independence

- Mathematically tractable
- Solved by Maximum Likelihood Estimation MLE
- (or other, it depends)

Interdependence

- Mathematically INtractable
- Solved with approximation via simulation using Markov Chains Monte Carlo MCMC
- (or other, it depends)

Example

TERM Mutual

- endogenous
- dyadic interdependent (or dependent)
- Mathematically INtractable: Solved with approximation via simulation using Markov Chains Monte Carlo MCMC

TERMS edges/sender/receiver

- endogenous
- dyadic independent
- Mathematically tractable: Solved by Maximum Likelihood Estimation MLE

Other classifications

- Unipartite - Bipartite
- Directed - Undirected

  D: istar - ostar

  UND: star
- Quadratic - Markovian

  triangles

  GWESP (geometrically weighted edgewise shared partners)
- Binary - Weighted
- ...

Finding your way in the Term/Effect Jungle

A term/effect Jungle

Many Effects (or terms)

- Impossible to cover them all. Let's
  - Categorize them
  - Search for them
  - Understand the most common
  - Learn how to use them

- In order to fit the best possible dress to predict the network data you have

Use the help file!

Categorization https://cran.r-project.org/web/packages/ergm/vignettes/ergm-term-crossRef.html

Term name	binary	dyad-independent	frequently-used	directed	undirected	quantitative nodal attribute	valued	categorical nodal attribute	curved	triad-related	bipartite	quantitative nodalattribute	deprecated	non-negative
absdiff	✓	✓	✓	✓	✓	✓								
absdiff		✓		✓	✓	✓	✓							
absdiffcat	✓	✓		✓	✓			✓						
absdiffcat		✓		✓	✓		✓	✓						
altkstar	✓				✓			✓	✓					
asymmetric	✓	✓		✓						✓				
atleast		✓		✓	✓		✓							
atmost		✓		✓	✓		✓							
b1concurrent	✓				✓			✓			✓			
b1cov	✓	✓	✓		✓						✓	✓		
b1cov		✓	✓		✓	✓	✓				✓			
b1degrange	✓				✓						✓			
b1degree	✓		✓		✓			✓			✓			
b1dsp	✓				✓						✓			
b1factor	✓	✓	✓		✓			✓			✓			
b1factor		✓	✓		✓		✓	✓			✓			
b1mindegree	✓				✓						✓			
b1nodematch	✓	✓	✓		✓			✓			✓			

https://cran.r-project.org/web/packages/ergm/vignettes/ergm-term-crossRef.html

Use the help file!

```
ergm::search.ergmTerms(keyword, net, categories, name)
```

Arguments

search: optional character search term to search for in the text of the term descriptions. Only matching terms will be returned. Matching is case insensitive.

net: a network object that the term would be applied to, used as template to determine directedness, bipartite, etc

keywords: optional character vector of keyword tags to use to restrict the results (i.e. 'curved', 'triad-related')

name: optional character name of a specific term to return

packages: optional character vector indicating the subset of packages in which to search

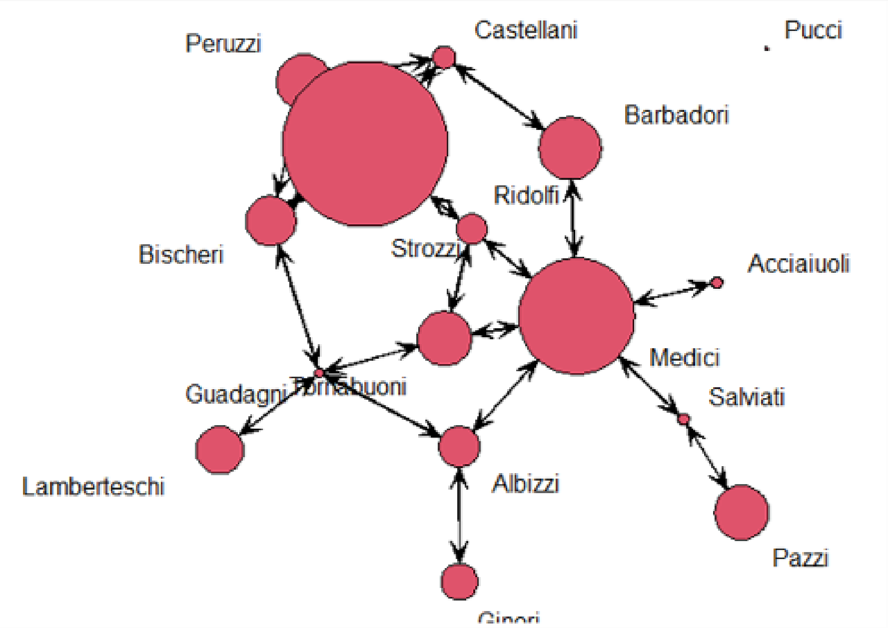
Fitting ERGMs with exogenous terms

Fitting an ergm with exo terms

So far, we only explored endogenous terms. Let's explore exogenous

- something that comes from the outside of the network structure

Wealth of the Florentine families.



Exogenous covariates make the ERG model equivalent to logistic regressions. no simulation (MCMC)

One Exogenous Dyadic independent term

Q: Are rich people more likely to get married?

H1: The richer a family is, the more likely its members are to marry into other families

TERM: `nodecov()` Probability of a tie given the receiver having the similar attribute.

```
flo.wealth <- ergm::ergm(flomarriage ~ edges + nodecov("Wealth"))
```

Results

- `edges` is (like) the intercept
- `nodecov("Wealth")` is like any other numeric variable in a Logistic regression

```
summary(flo.wealth)
```

```
## Call:
## ergm::ergm(formula = flomarriage ~ edges + nodecov("Wealth"))
##
## Maximum Likelihood Results:
##
##                  Estimate Std. Error MCMC % z value Pr(>|z|)
## edges           -2.594929   0.536056      0  -4.841   <1e-04 ***
## nodecov.Wealth   0.010546   0.004674      0   2.256   0.0241 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 166.4  on 120  degrees of freedom
##  Residual Deviance: 103.1  on 118  degrees of freedom
##
## AIC: 107.1  BIC: 112.7  (Smaller is better. MC Std. Err. = 0)
```

Other option

Likelihood to marry in general or to marry whom?

RQ: Do families of equal wealth seek marital alliances with one another?

H1: Families of equal wealth tend to form marital alliances with one another.

TERM: `absdiff` - Probability of a tie given the similarity between sender and receiver.

```
flo.wealth1 <- ergm::ergm(flomarriage ~ edges  + absdiff("Wealth"))
```

Results second hypothesis

```
flo.wealth1 <- ergm::ergm(flomarriage ~ edges + absdiff("Wealth"))

summary(flo.wealth1)
```

```
## Call:
## ergm::ergm(formula = flomarriage ~ edges + absdiff("Wealth"))
##
## Maximum Likelihood Results:
##
##                 Estimate Std. Error MCMC % z value Pr(>|z|)
## edges          -2.302042   0.401906      0  -5.728   <1e-04 ***
## absdiff.Wealth  0.015519   0.006157      0   2.521   0.0117 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 166.4  on 120  degrees of freedom
##  Residual Deviance: 102.0  on 118  degrees of freedom
##
## AIC: 106  BIC: 111.5  (Smaller is better. MC Std. Err. = 0)
```

Final Model and Results

Can we reject the two null Hypotheses?

```
flo.wealth2 <- ergm::ergm(flomarriage ~ edges + nodecov("Wealth") + absdiff("Wealth"))

summary(flo.wealth2)
```

```
## Call:
## ergm::ergm(formula = flomarriage ~ edges + nodecov("Wealth") +
##     absdiff("Wealth"))
##
## Maximum Likelihood Results:
##
##                 Estimate Std. Error MCMC % z value Pr(>|z|)
## edges          -2.527091   0.535994      0  -4.715   <1e-04 ***
## nodecov.Wealth  0.004506   0.006791      0   0.664    0.507
## absdiff.Wealth  0.011143   0.008950      0   1.245    0.213
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      Null Deviance: 166.4  on 120  degrees of freedom
##  Residual Deviance: 101.5  on 117  degrees of freedom
##
## AIC: 107.5  BIC: 115.9  (Smaller is better. MC Std. Err. = 0)
```

Model Comparison

```
texreg::screenreg(list(flo.wealth,flo.wealth1, flo.wealth2))
```

```
##
## =====================================================
##                  Model 1      Model 2      Model 3
## -----------------------------------------------------
## edges             -2.59 ***    -2.30 ***    -2.53 ***
##                   (0.54)       (0.40)       (0.54)
## nodecov.Wealth     0.01 *                    0.00
##                   (0.00)                    (0.01)
## absdiff.Wealth                  0.02 *       0.01
##                                (0.01)       (0.01)
## -----------------------------------------------------
## AIC              107.11       105.95       107.52
## BIC              112.68       111.53       115.88
## Log Likelihood   -51.55       -50.98       -50.76
## =====================================================
## *** p < 0.001; ** p < 0.01; * p < 0.05
```

Conclusions?

Is love in the air? Or is money in the bank?

Reading Results -significance

ERGMs results are interpreted the same way as logistic regressions' results.

Is it significant?

1) IF beta > 0 and stat. signif. --> X increases P(1) and decreases P(0)

2) IF beta < 0 and stat. signif. --> X decreases P(1) and increases P(0)

3) IF beta NOT stat. signif. --> X does not affect P(1) and P(0)

Reading Results -intensity

If the effect is significant, move on checking the intensity.

Two options:

- odds ratios exp(coef)
- probability I'll give you the formula later :)

How to use ERGMs

Multiple hypotheses to test

Combine endogenous and exogenous terms (always)

DO NOT FIT ERGM with Exogenous variables only (!!!) - in case it was not clear

Check the combination of your terms

Make hypotheses as sharp as possible (in this example either nodecov or absdiff)

See you next week!