Everything you ever wanted to know about network statistics*

*BUT WERE AFRAID TO ASK

Momin M. Malik, Ph.D.

SUNBELT 2019, 18 JUNE 2019, MONTRÉAL

Disclaimer

Title is referencing the book, NOT the movie (which I've never seen, as I boycott its creator).

Network statistics in one sentence:

Everything is terrible, and nothing works.

Sorry.

"Too long, didn't read" version:

- Everything is terrible, and nothing works. Sorry.
- Logistic regression on the edges isn't a bad first pass
- Use whatever model is accepted by your community
- Or…
  - Give up on empirical analysis and do simulation modeling (but I do not endorse this)
  - Give up on modeling and do qualitative analysis

Outline

- Preliminaries
- Example
- Why statistics?
- Probability in statistics
- The likelihood principle
- Centralities are not statistical
- Dependencies
- Models to *control* for network structure
- Models *of* network structure
- Models of networks in time
- Graphical models

Preliminaries

Motivation, rules, learning objectives

Motivation: My own confusion

- There were many things I didn't understand about network statistics
- I pestered statisticians, and also learned statistics from the ground up (statistics PhD coursework at Carnegie Mellon)
- I found explanations are out there... but are useless unless you already understand them
- The *expert blind spot* impedes communication
- I want to help people who are where I was, who don't have the time or inclination to do half a statistics PhD

Let's introduce ourselves!

Name, pronouns

Affiliation

Background

Most/one frustrating thing about network statistics

Rules: Ask questions!

- I am succumbing to the expert blind spot, so **I need you to help me** collect a list of all the confusions people have
- ASK QUESTIONS!! Constantly, throughout. This is the real benefit of being here.
- If you have a “stupid” question, I guarantee somebody else has it! Let’s address it!

Topics and learning objectives

We can't cover everything in 3h. But we can:

- Cover fundamental concepts and their contingencies, so you know what statistical statements even mean
- Give "glossary of models," with their purposes and drawbacks, so you can decide where to learn more
- Cover which communities consider what kind of modeling legitimate, and why, to help you choose
- Show you what the modeling process looks like, to demystify it

Some questions you've already asked

- Power law and small world networks?
- Including latent variables, and how do they differ in treating dependencies
- Calculating eigenvector centrality?
- Using igraph?
- Challenges, traps, and ways to do good stats?
- Disconnected components?
- Criticisms of ERGMs, and alternatives?
- Interpreting for application?

Questions before I start?

Things we'll cover?

More things you'd *like* to cover?

A worked example

(A bit like an) idea contributed by Chloe Bracegirdle

Manipulate in R

- Two groups in a social network
- Data: attributes, attitudes, and ties
- Do attitudes correlate with by-group ties?
- Don’t have attitudes, so just look at attributes
- Use “Lazega lawyers” network, and:

Download data

https://www.stats.ox.ac.uk/~snijders/siena/LazegaLawyers.zip

Copy this into an R script

```
names(nodes) <- c("seniority",
                  "status",
                  "sex",
                  "office",
                  "tenure",
                  "age",
                  "practice",
                  "lawschool")

nodes$status <- nodes$status %>%
  factor(labels=c("partner",
                  "associate"))
nodes$sex <- nodes$sex %>%
  factor(labels=c("male",
                  "female"))
nodes$office <- nodes$office %>%
  factor(labels=c("Boston",
                  "Hartford",
                  "Providence"))
nodes$practice <- nodes$practice %>%
  factor(labels=c("litigation",
                  "corporate"))
nodes$lawschool <- nodes$lawschool %>%
  factor(labels=c("Harvard/Yale",
                  "UConn",
                  "Other"))
```

Notice: 3 modes of engagement

Math: $$P_\theta(\mathbf{A}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left\{ \theta_0 L(\mathbf{A}) + \sum_{k=1}^{n-1} \theta_k S_k(\mathbf{A}) + \theta_\tau T(\mathbf{A}) \right\}$$

Code:

```r
n <- 100
df <- data.frame(from = rep(1:n, each = n),
                 to = rep(1:n, times = n))
df <- df[df$from!=df$to,]
# nrow(df) == 2*choose(n, 2)
df$edge <- rbinom(n = nrow(df), size = 1, prob = 0.3)

library(igraph)
library(magrittr)
df[df$edge==1,c("from","to")] %>% graph.data.frame %>% plot
```

Concepts: Likelihood

(Worked example was done here)
Back to the presentation!

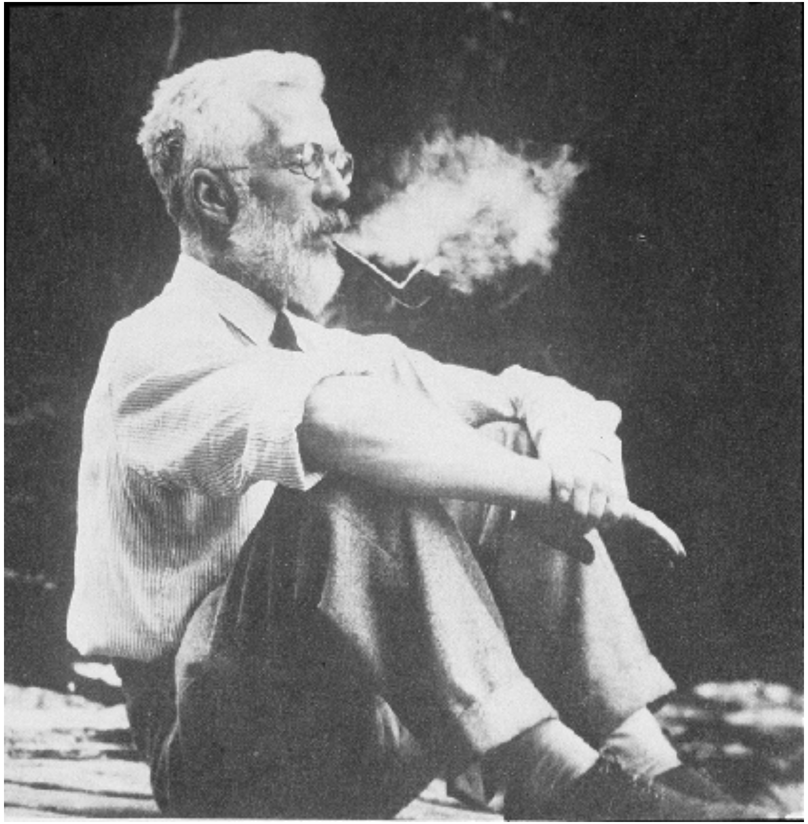
Why statistics?

The fundamentals you (maybe never knew you) missed

SUNBELT 2019 JUNE 18-23 MONTRÉAL

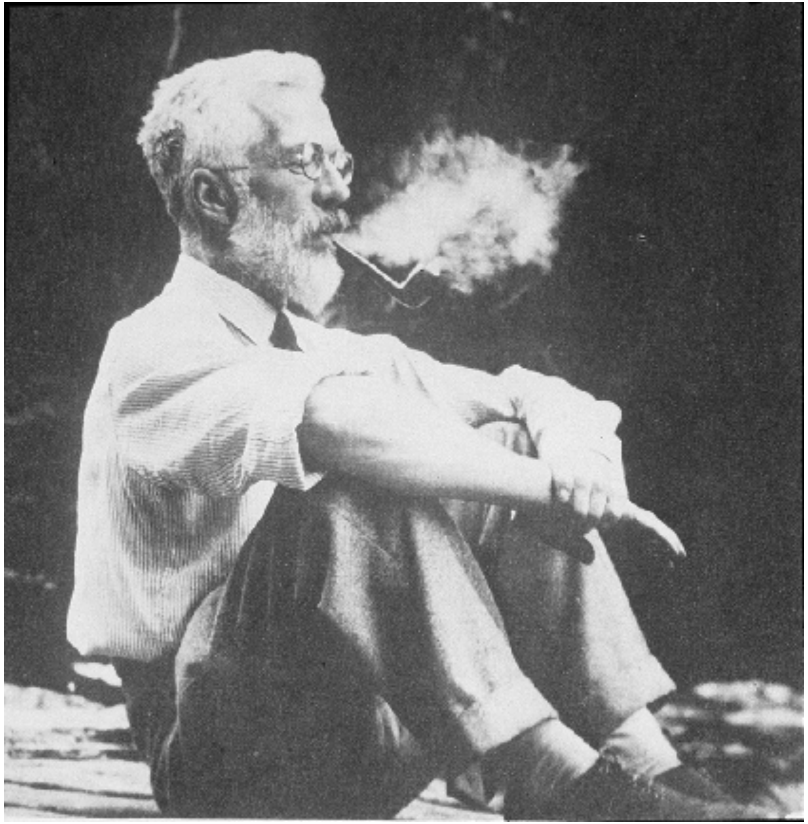
Purpose of stats: "the reduction of data"

"briefly, and in its most concrete form, ***the object of statistical methods is the reduction of data****."* (Fisher, 1922)



Purpose of stats: "the reduction of data"

*"A quantity of data, which usually by its **mere bulk** is **incapable of entering the mind**, is to be replaced by **relatively few quantities** which shall **adequately represent the whole**, or… as much as possible… of the **relevant information** contained in the original data."* (Ibid.)

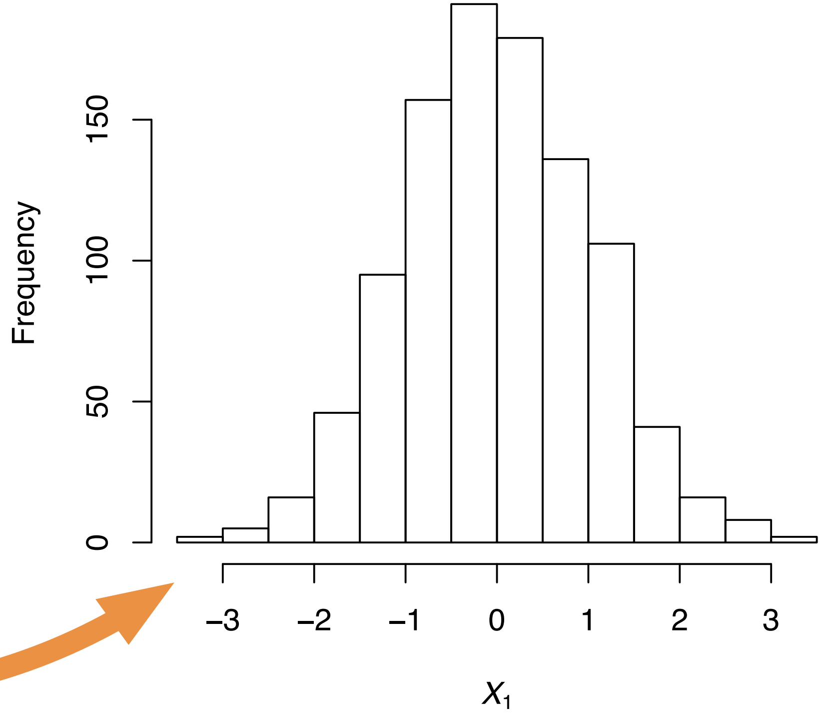


Requires a philosophical commitment

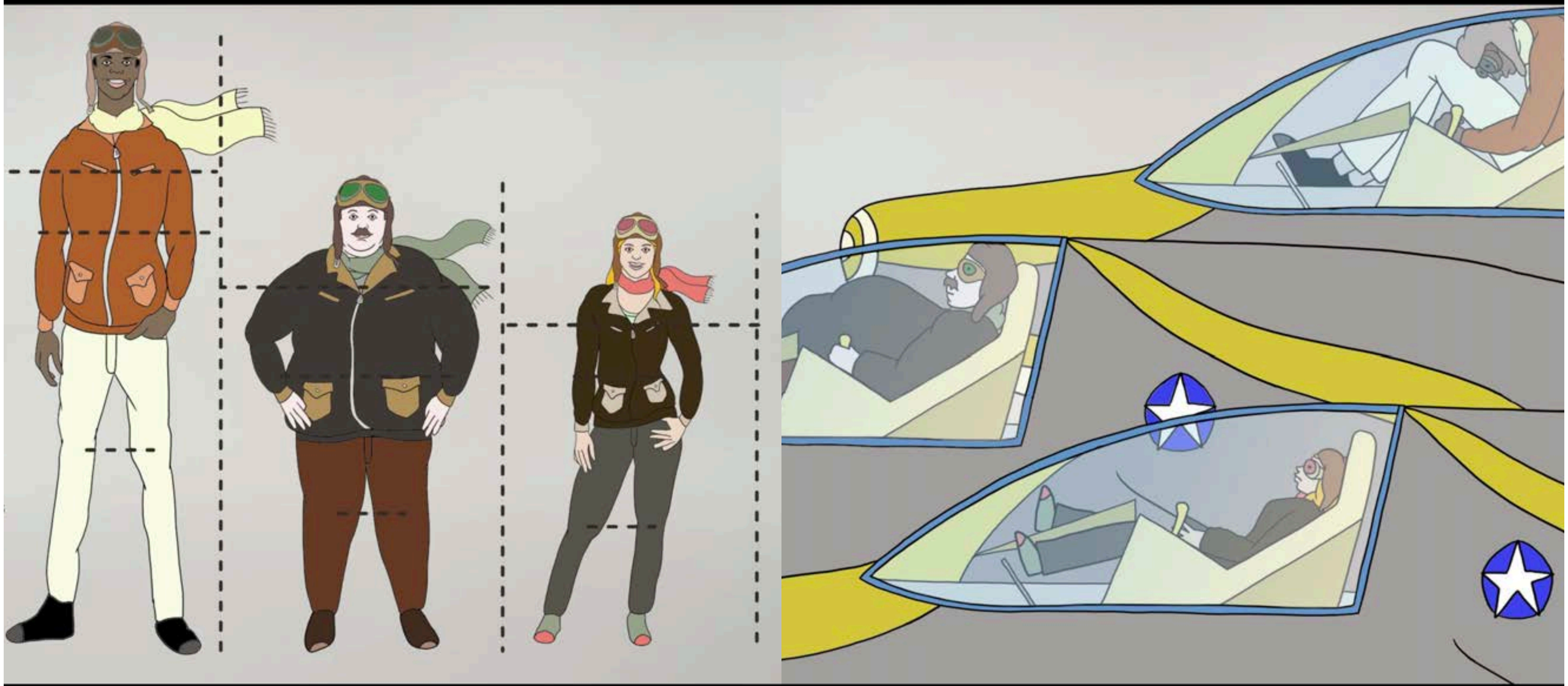
- A fundamental philosophical commitment of statistics (and machine learning):
  - *There are distinct entities in the world that are comparable.*
- When/How are they comparable?
  - Entities can share a *central tendency.* (This is the "relevant information" to which data are "reduced")
- If not a single entity achieves the central tendency (i.e., nobody is exactly average height), how can all entities nonetheless share it?
  - There is *variability* around the central tendency.

Statistics forces the world into a specific form

$$
\begin{array}{cccc}
X_1 & X_2 & \cdots & X_d \\
\hline
x_{11} & x_{12} & \cdots & x_{1d} \\
x_{21} & x_{22} & \cdots & x_{2d} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nd}
\end{array}
$$



This is not inherently true/meaningful...



Source: Todd Rose. Illustration: Future for Learning.

...and lacks sociological justification.

"**...it is striking how absolutely these assumptions** [fixed entities with properties] **contradict those of the major theoretical traditions of sociology**. Symbolic interactionism rejects the assumption of fixed entities and makes the meaning of a given occurrence depend on its location—within an interaction, within an actor's biography, within a sequence of events. Both the Marxian and Weberian traditions deny explicitly that a given property of a social actor has one and only one set of causal implications." (Abbott, 1988)

So: why statistics?

- If we commit to a philosophical belief that there is variability in the world:
  - Entities/processes have underlying similarities without being identical
  - We want to not be fooled by variability, thinking that we have found patterns in data when there are none
- Statistics is a way to analyze data and systematically manage variability (using probability as a model: we'll return to this)
- **Institutional/professional pressure, that if there isn't statistics, it isn't "science" and won't get published**

Alternatives?

- Nothing but statistics both reduces data *and* accounts for variability
  - Metrics (especially centrality) reduce data but don't account for variability
  - Simulation modeling and "mathematical models" (small-world, power law) account for variability but don't reduce data
    - Only qualitatively compares simulations outputs to data
    - (Simulation modeling uses statistics, and statistics uses simulation, but they are two different modeling logics)
- Qualitative research
  - I personally believe this the most, but it "doesn't scale"
- Giving up. After this presentation, you may be tempted...

Questions so far?

What is the purpose of statistics?

Why would we use statistics?

Why might we *not* use statistics?

SUNBELT 2019 JUNE 18-23 MONTRÉAL

Covering *probability* in statistics

Will seem abstract, but underlies everything

Why we need to cover this:

"There are two reasons for not using tests of statistical significance for the coefficients in these models. First, **the data for each community come from a theoretical population instead of a sample, which means that all coefficients are necessarily significant**. We might, however, elect to apply such tests as a guide to important relations or as a guard against findings due to random measurement error (see Stinchcombe 1968, p. 23n.), were it not for extremely complicated problems involved in the determination of the appropriate number of degrees of freedom for such tests." (Laumann et al., 1977)

This is deeply confused!

- One statistician: "I can't even begin to understand the levels of confusion that led to this statement."
  - Coefficients are *never* "necessarily significant."
  - "guard against findings due to random measurement error" misunderstands nature of uncertainty
  - Determining effective sample size (ESS) is neither necessary nor sufficient for correct inferences

Probability in statistics

How does statistics "reduce data" to the "relevant information" of the central tendency? With the mathematical abstraction of *probability*.

Statistics uses probability for two things

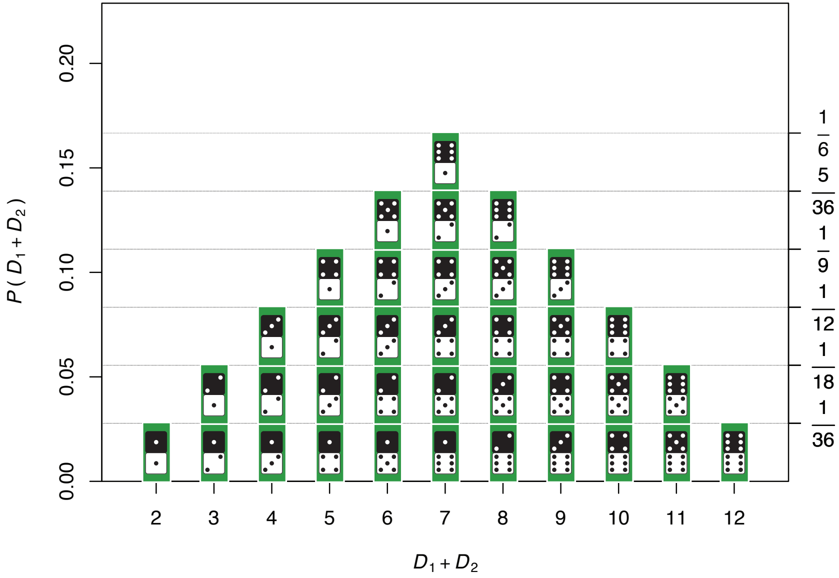
"Probability is used [in statistics] in two distinct, although interrelated, ways in statistics, **phenomenologically to describe haphazard variability arising in the real world** and epistemologically to represent uncertainty of knowledge." (Cox, 1990)

SUNBELT 2019 JUNE 18-23 MONTRÉAL

(Use of probability is not obvious!)

"It is remarkable that a science which began with the **consideration of games of chance** should have become the most important object of human knowledge." (Pierre Simon Laplace, *Théorie analytique des probabilités*, 1812)

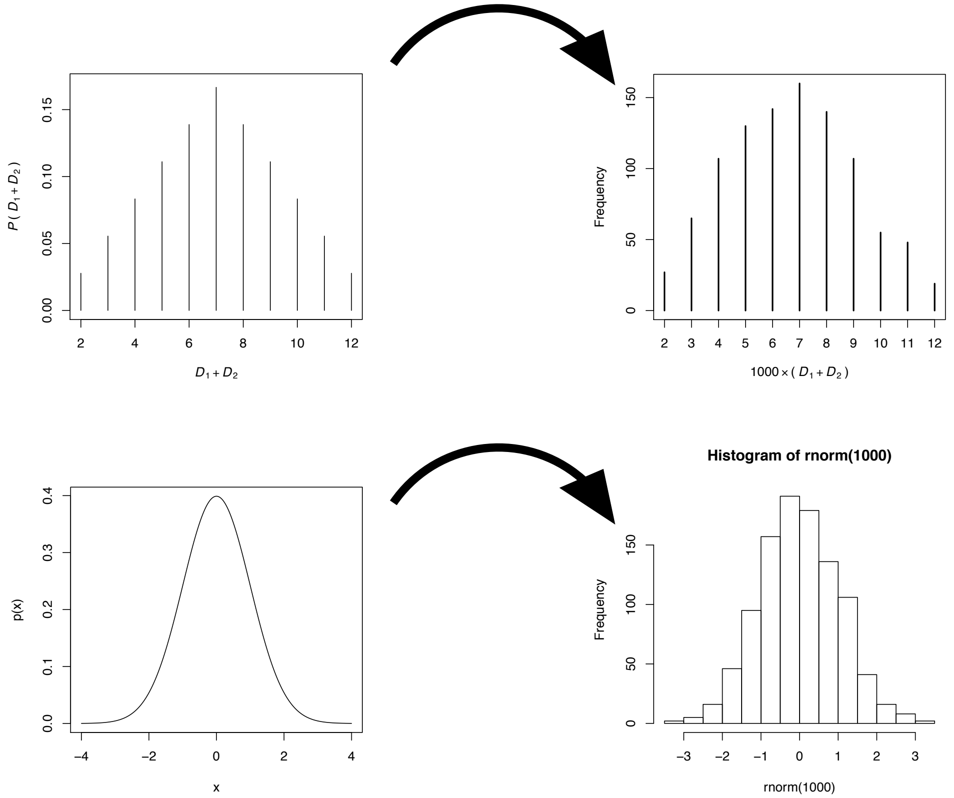
1. Probability represents *variability*



- We conceive of *variability* in data
- We use *probability distributions* to represent this variability.
- Idea that randomness has a "shape"! Random but still a *central tendency* we can reduce to
- (Note: like a histogram)

1. Probability represents *variability*

Data are "draws" from a theoretical distribution, with the *empirical* distribution coming to resemble the underlying one



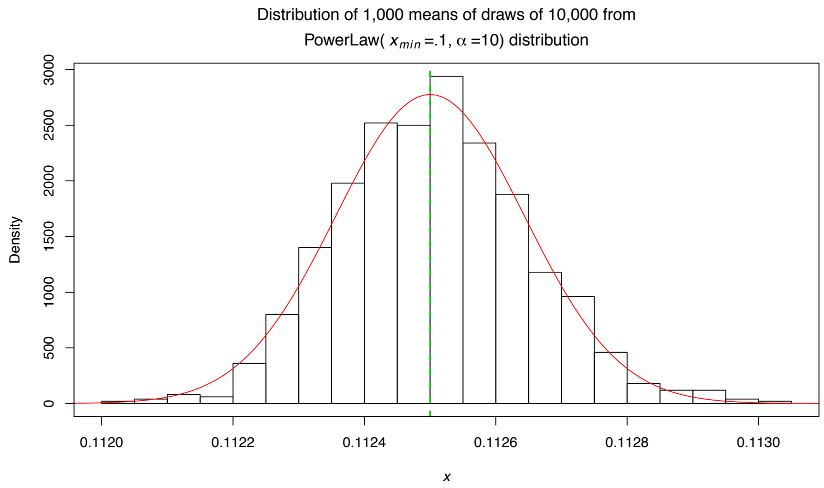
SUNBELT 2019 JUNE 18-23 MONTRÉAL

2. Probability represents *uncertainty*

- Variability implies our estimates of the central tendency are uncertain
- Probability represents this too, through *distributions* of the *estimates*
- There's a strange idea of the distribution of the central tendencies of several distributions
- (After all, why wouldn't you just combine multiple datasets?)

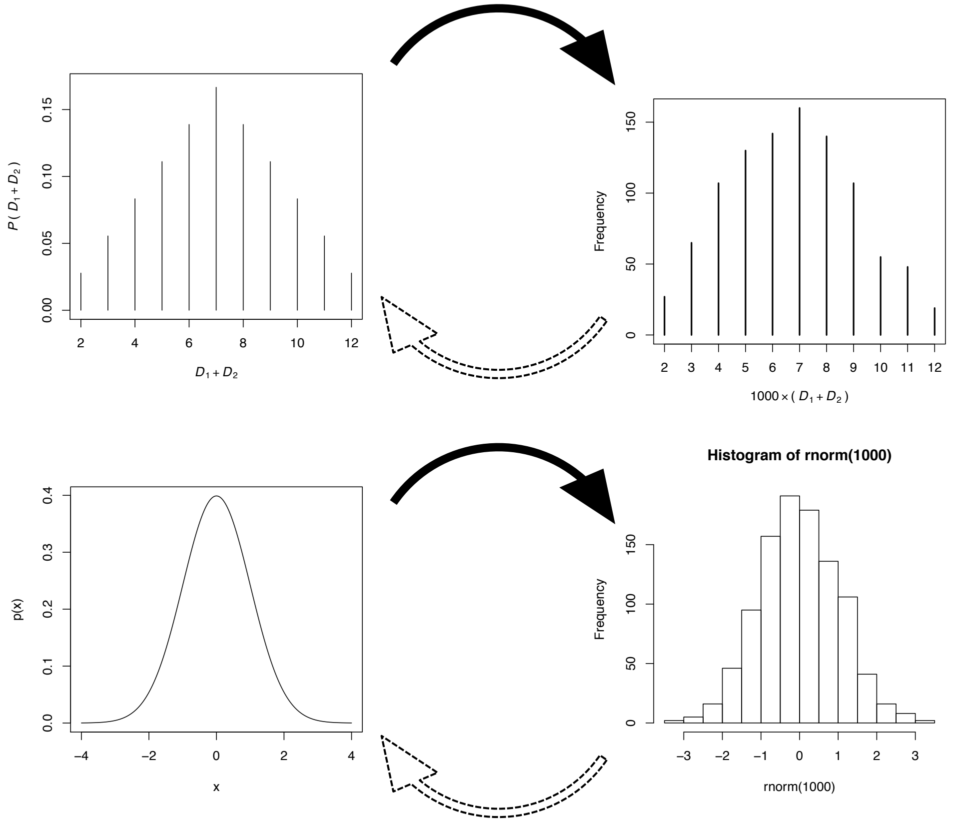
(Central limit theorem links both uses)

The Central Limit Theorem is about a *distribution of distributions*: the *uncertainty* of distributions of *central tendencies* is normally distributed.



"Statistical inference"

We work backwards from the observed distribution to say something about the underlying distribution



(Full ontology is complicated)

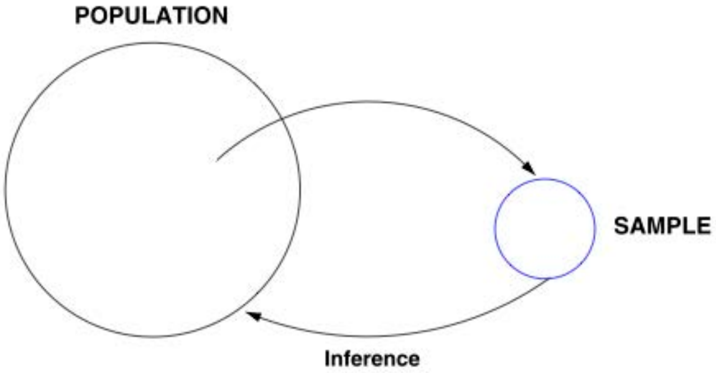


FIG. 3. *The big picture of statistical inference according to the standard conception. Here, a random sample is pictured as a sample from a finite population.*

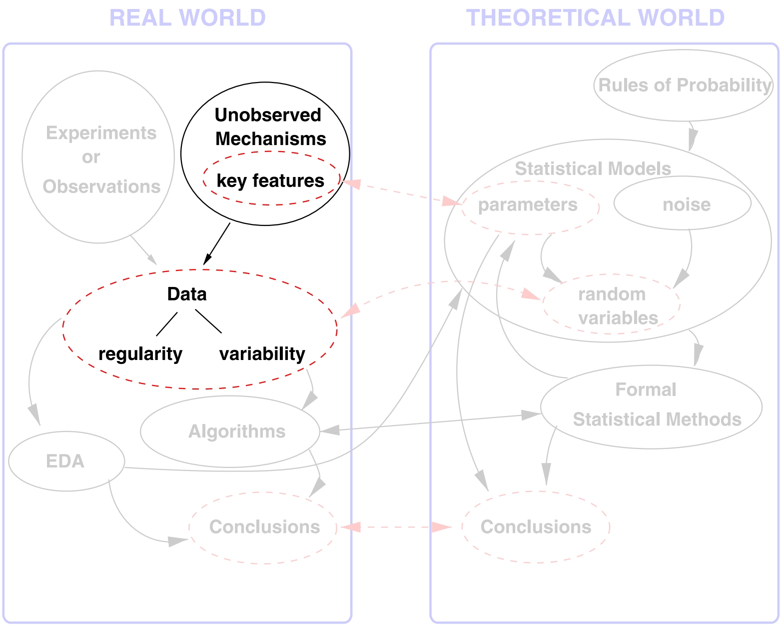


FIG. 4. *A more elaborate big picture, reflecting in greater detail the process of statistical inference. As in Figure 1, there is a hypothetical link between data and statistical model but here the data are connected more specifically to their representation as random variables.*

Kass (2011)

So: statistics is metaphysical!

- There is no such thing as an the "underlying distribution," or the *data-generating process*.
- (Jerzy Neyman and Egon Pearson tried to take the metaphysics out, but that's what led to the confusion in Laumann et al.)
- So "explanation" is founded on appealing to something that doesn't exist
- But that's the framework within which statistics exists — and again, for handling variability and reducing data, there's nothing else

Questions so far?

Use and meaning of probability in statistics?

Metaphysics of statistics?

The likelihood principle: Connecting data and probability

Working up to understanding this:

$$P_\theta(\mathbf{A}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left\{\theta_0 L(\mathbf{A}) + \sum_{k=1}^{n-1} \theta_k S_k(\mathbf{A}) + \theta_\tau T(\mathbf{A})\right\}$$

SUNBELT 2019 JUNE 18-23 MONTRÉAL

Probability of data

- We think of data as a *realization* of a *random variable* (think of a coin, tumbling in mid-air)
- We assume a distribution
- For a coin Y, this is a *Bernoulli* distribution,

$$\mathbb{P}(Y = 1) = p$$

$$\mathbb{P}(Y = 0) = 1 - p$$

- As a single equation:

$$\mathbb{P}(Y = y) = p^y(1 - p)^{1-y}$$

- (Implicit that y, the "support," can only be 0 or 1)

Interpreting probability

- What does it mean for an outcome to have, say, a $p = 0.2$ probability? Either it happens or it doesn't!
  - Frequentists: probability is long-term frequency, So data *collectively* reveal the "relevant information"(problem: incoherent for one-off events).
  - Bayesians: probability is "reasonable expectations" or "personal beliefs" (problem: what we really want are frequency guarantees)

Discrete versus continuous

- Discrete distributions make sense
- But *continuous* ones, like the normal distribution, require calculus
- The probability of a tree being π meters tall, for example, is zero. For any specific number, actually, it is zero
- (Probability of a specific height is like velocity; what does it mean to have a given speed and direction at a moment, frozen in time?)

Continuous distributions

- So instead, we take the probability of a range of values, which is nonzero
- We take the derivative of that function
- We end up with something that we can manipulate like a discrete probability distribution

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

Normal distribution

$$p(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

- The negative squared term inside the exponential makes a bell curve, and the term outside makes it integrate to 1.
- The “relevant information” is the mean (μ) and the variance (σ^2), although data points do not individually exhibit either.

Multiple data

- (*Bold, lowercase* letters are vectors; *bold, uppercase* letters are matrices.)
- We never observe just one data point. The probability of multiple events is called the *joint probability*.

$$p(y_1, y_2, \ldots, y_n) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\}$$

(Matrix notation 1/3)

- For n observations on d variables, we have n equations, each d terms on the right.

$$
\begin{aligned}
y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_d x_{1d} \\
y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_d x_{2d} \\
&\vdots \\
y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_d x_{nd}
\end{aligned}
$$

- A *summation* can write d terms more succinctly. And a *range* can collapse n terms.

$$
y_i = \sum_{j=1}^{d} \beta_j x_{ij}, \text{ for all } i \in \{1, \ldots, n\}
$$

(Matrix notation 2/3)

- Linear algebra is a way to not only write, but *manage* all of these operations simultaneously.
- Specifically, through *matrix notation*.

$$\left.\begin{aligned} y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_d x_{1d} \\ y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_d x_{2d} \\ &\vdots \\ y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_d x_{nd} \end{aligned}\right\} \longrightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta}$$

(Matrix notation 3/3)

- All of these are equivalent:

$$\mathbf{X}\beta = \begin{bmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_d \end{bmatrix} = \begin{bmatrix} — & \mathbf{x}_1{}^T & — \\ & \vdots & \\ — & \mathbf{x}_n{}^T & — \end{bmatrix} \begin{bmatrix} | \\ \boldsymbol{\beta} \\ | \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1{}^T \boldsymbol{\beta} \\ \vdots \\ \mathbf{x}_n{}^T \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{d} x_{1j}\beta_j \\ \vdots \\ \sum_{j=1}^{d} x_{nj}\beta_j \end{bmatrix}$$

- Matrix-by-vector multiplication represents *systems of equations*.
- We can also solve for these systems simul-taneously, e.g.:

$$\max_{\boldsymbol{\beta}} \left\{ \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \right\} \right\}$$
$$= \min_{\boldsymbol{\beta}} \left\{ (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \right\}$$
$$= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Probability of data

- If observations are *independent*, we can split the probability. But each y can still have its own distribution (its own mean and variance).

$$p(y_1, y_2, \ldots, y_n) = p_1(y_1) \times p_2(y_2) \times \cdots \times p_n(y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i{}^2}} \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma_i{}^2}\right\}$$

- If observations are also *identically distributed*, they have the same mean and variance.

$$p(y_1, y_2, \ldots, y_n) = p(y_1) \times p(y_2) \times \cdots \times p(y_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \mu)^2}{2\sigma^2}\right\}$$

The *likelihood principle*

- The *likelihood principle*: instead of looking at the probability of the data we observed, we say, "*given* the data we observe, what are the most likely *parameters*?"
- We do this by reinterpreting the probability as *likelihood:*

$$p(y; \mu, \sigma^2) \to \mathcal{L}(\mu, \sigma^2; y)$$

- The functional form stays the same, but now is a function of the *parameters*, not the variable

$$\mathcal{L}(\mu, \sigma^2; y) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

Maximum likelihood

- Assume independence and identical distribution, then the likelihood from multiple data is:

$$\mathcal{L}(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}$$

- Plug in the data points for the y_i's, and maximize over μ and σ^2 (easier to solve for squared term)
- Turns out an equivalent, and easier, problem is to maximize the *log* of the likelihood

$$\ell(\mu, \sigma^2) = \log(1) - \tfrac{1}{2}\log(2\pi\sigma^2) + \frac{-(y_1-\mu)^2 - \cdots - (y_n-\mu)^2}{2\sigma^2}$$

- Use calculus! Take derivative and set to zero.

Maximum likelihood

- *The likelihood principle is what connects probability and data*
- Whatever maximizes this likelihood is what makes the data "most likely." That's what we want to find
- We call the maximizer an "estimate" and notate it with a "hat"
- For a normal distribution, the maximum likelihood estimate is the sample mean!

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- Statistical theory looks at how, based on n, an estimator compares to the "underlying truth" (goes deep)

Network statistics also about likelihood

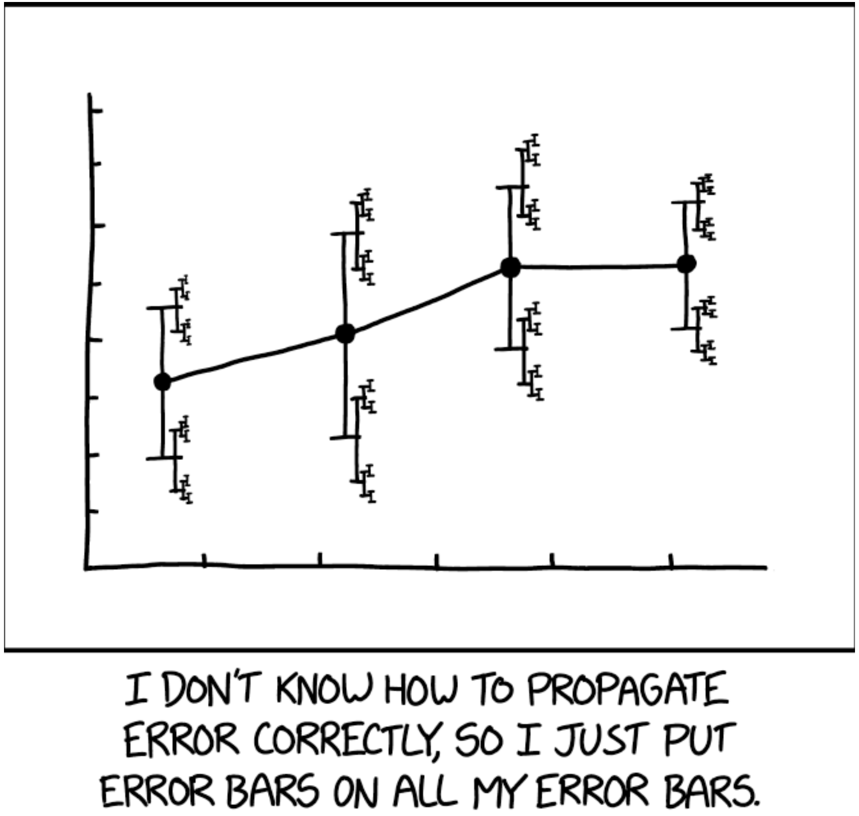
- Discrete outcomes: take a sum of attributes times parameters, then dividing by all possible outcomes added

$$P_\theta(\mathbf{A}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left\{\theta_0 L(\mathbf{A}) + \sum_{k=1}^{n-1} \theta_k S_k(\mathbf{A}) + \theta_\tau T(\mathbf{A})\right\}$$

- We express our *models* as *probability functions*. (To *estimate* the parameters, we interpret it as a likelihood.)
- (Note: this isn't something negative squared inside the exponent, so it's more like a Poisson model's probability distribution than a normal distribution, but works similarly)
- For networks, making sure things integrate to 1 becomes enormously difficult, and many tricks come into play
- But *principle* is the same

Note: Inference in statistics

- "Statistical inference" is the overall process
- Within statistics, *inference* is specifically: *quantifying the uncertainty of estimates* to make conclusions. How?
- The estimator is itself a random variable!
- *Estimate its variability*



https://xkcd.com/2110

Whew! Questions?

Always remember: this is not a natural or inevitable way of thinking.

Centralities are not statistical

Filling in some explanations that statisticians have so far failed to give

The difference in approaches

Social scientists:

- Regression as an omnibus framework
- "What corrections do I have to use to make regression work for my problem?"
- "*Why* don't the standard approaches work?"

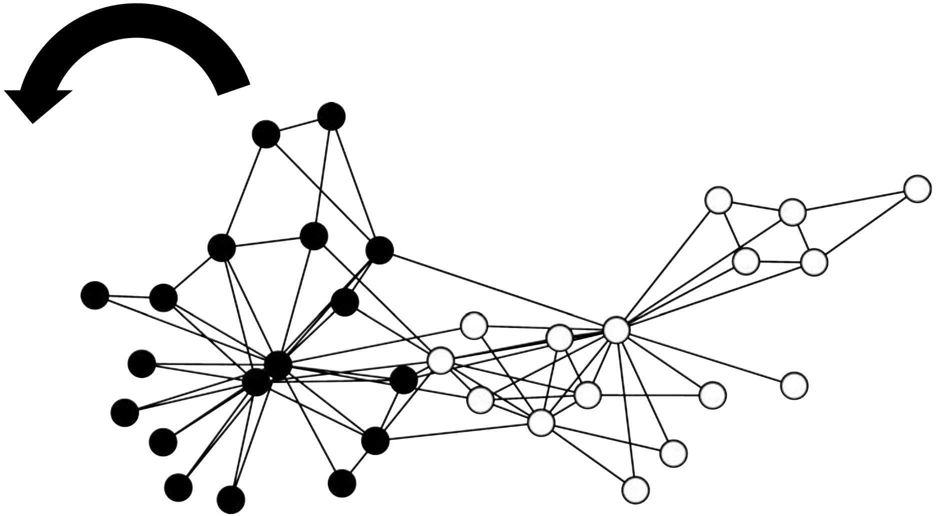
Statisticians:

- Think about the data-generating process
- Not so much that standard approaches "don't work", but don't give us what we want (inferences to DGP)
- If it's wrong/unhelpful, why bother exploring?

SUNBELT 2019 JUNE 18-23 MONTRÉAL

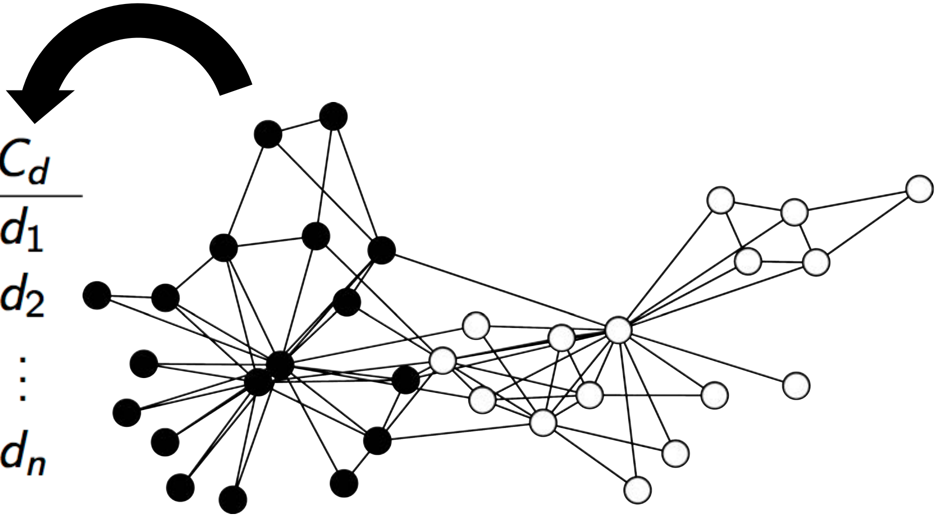
Social scientists: "Obvious" first pass

	Y	X_1	X_2	$\cdots$	X_k
v_1	y_1	x_{11}	x_{12}	$\cdots$	x_{1k}
v_2	y_2	x_{21}	x_{22}	$\cdots$	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
v_n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}



Social scientists: “Obvious” first pass

	Y	X_1	X_2	$\cdots$	X_k	C_d
v_1	y_1	x_{11}	x_{12}	$\cdots$	x_{1k}	d_1
v_2	y_2	x_{21}	x_{22}	$\cdots$	x_{2k}	d_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
v_n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nk}	d_n



The problem for statisticians

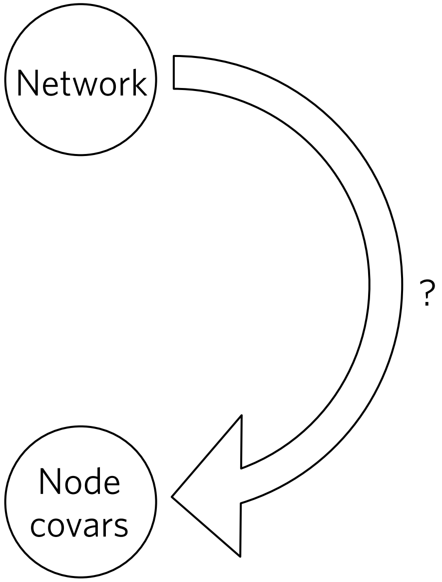
- *Ceteris paribus* ("holding all else constant") interpretation
- How do we change the (undirected) degree of one node (or some centrality like eigenvector, betweenness, closeness) and hold those of all other nodes constant?
- Deeper question: **what are we trying model?**

What are we trying to model?

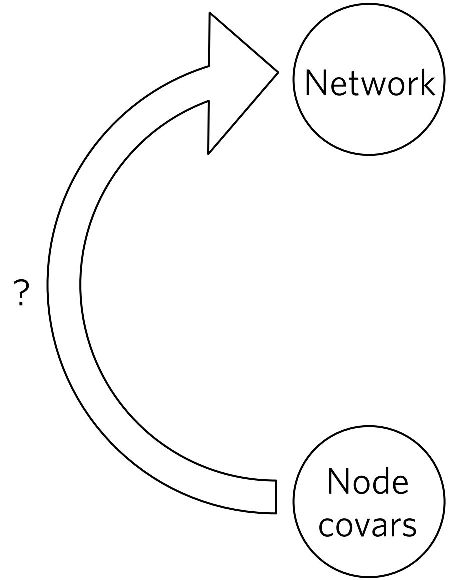
- Centralities are a very crass way of capturing network structure
  - Are a *by-product* of network structure/processes, not what produces them
- Even if we have a directed graph, e.g. an advice network,
  - Modeling in-degree centrality would be getting at who is sought out
  - But not *by whom*
  - Out-degree centrality would be getting at who seeks out advice
  - But not *from whom*
- Model the *process* also to manage dependencies

Network: explanatory or response?

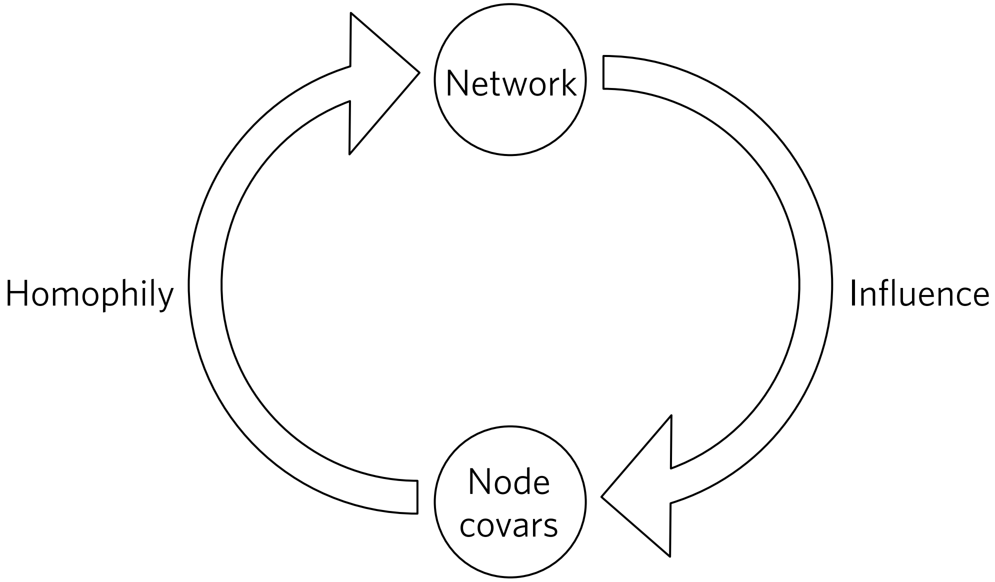
Network as cause? (as explanatory/IV?)



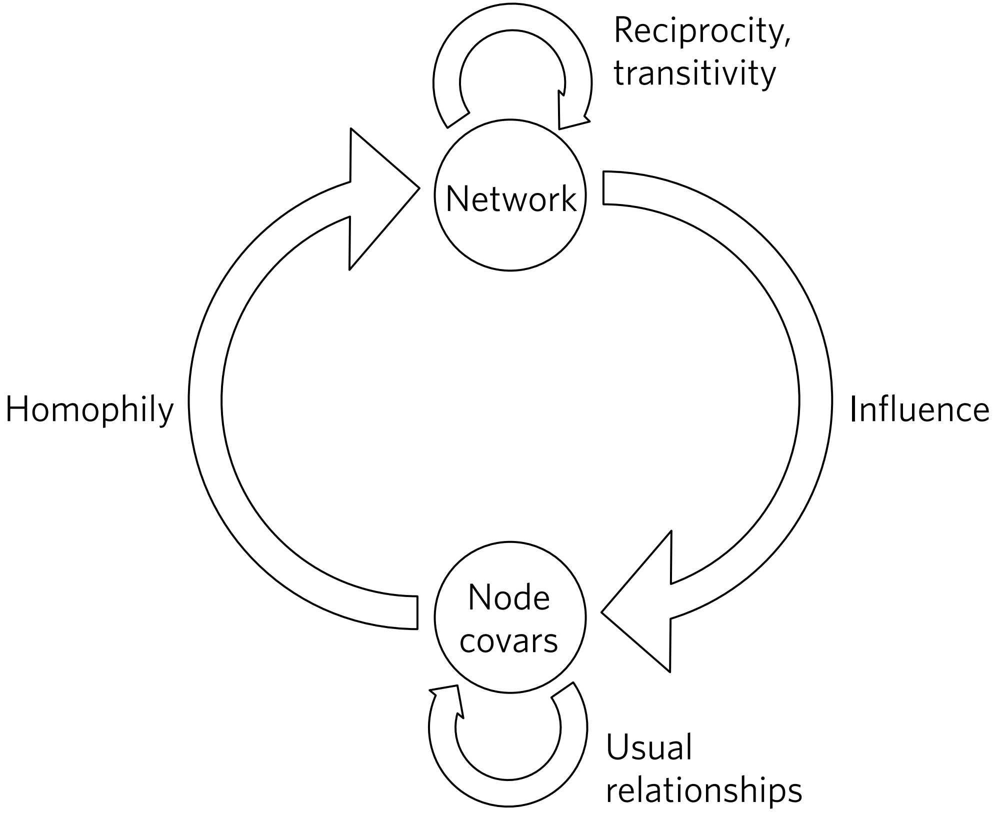
Network as effect? (as response/DV?)



The problem: both happen.



And they aren't the only things.



Technical vocabulary

- "Model misspecification"
  - The wrong functional form, and/or
  - The wrong variables
- Omitted variable bias (OVB)
  - For a powerful example, see Arceneaux, Gerber, & Green (2010). Omitted variable of "reachability by phone" is so powerful, nothing can control for it, and it makes all of our conclusions wrong
- Synonymous: "Non-iid data," "dependent data," "autocorrelation," "endogeneity," "pseudoreplication"

Dependencies

"Dependencies" is an overloaded term. Network ties are dependencies, but themselves *have* dependencies

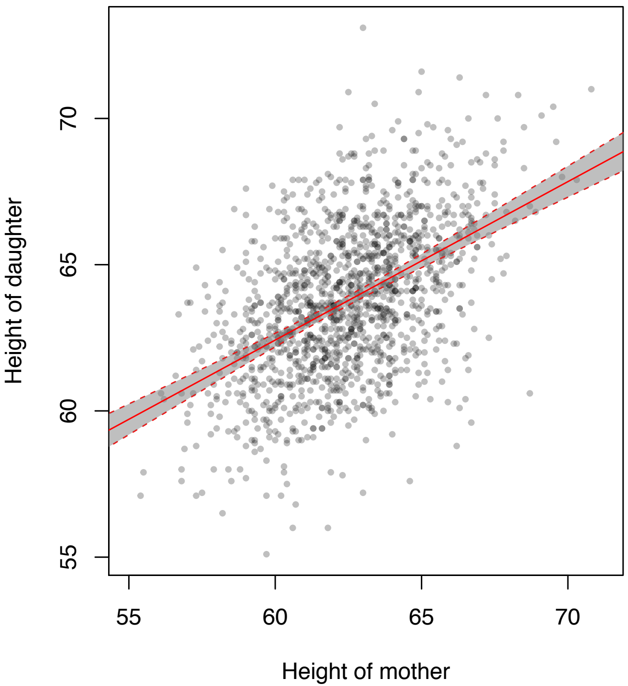
What do dependencies do?

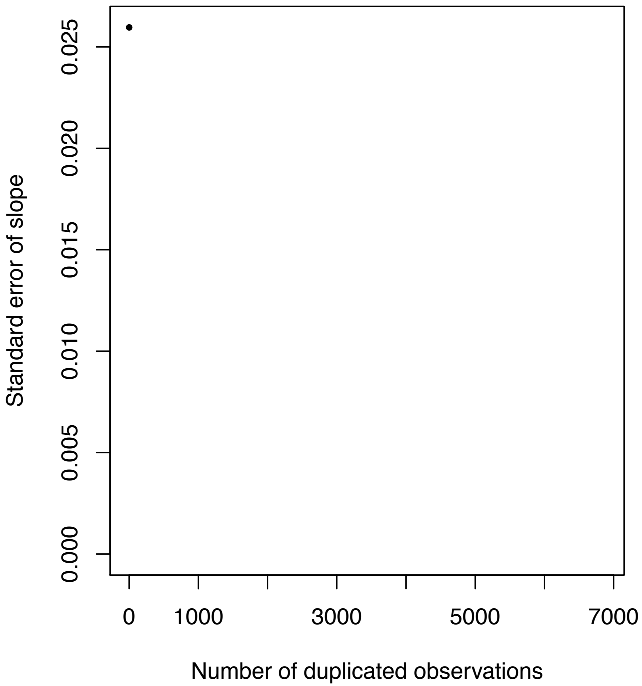
- From Wikipedia: "Asking two people in the same household whether they watch TV, for example, does not give you statistically independent answers. The sample size, *n*, for independent observations in this case is one, not two."
- **The simplest form of dependence between observations: duplicate observations**

Exploring using simulation

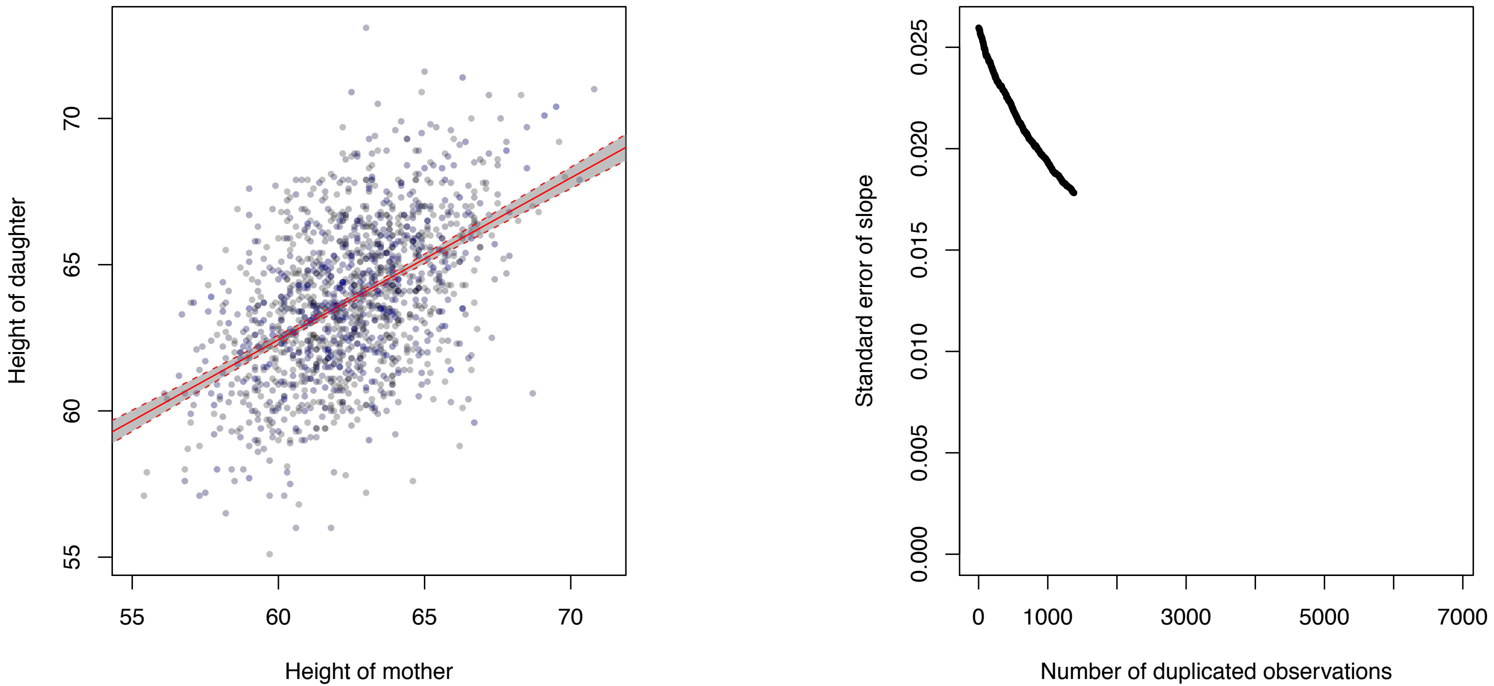
- Let's use Galton's height data
- Sample from the observations at random, and append a copy of that observation to the data set
- What happens to our fitted regression line?

99% CI, duplicates in blue

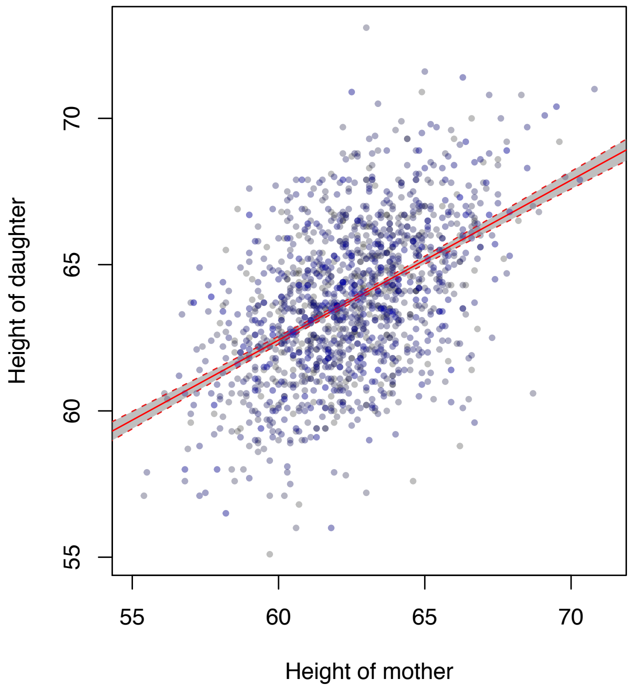


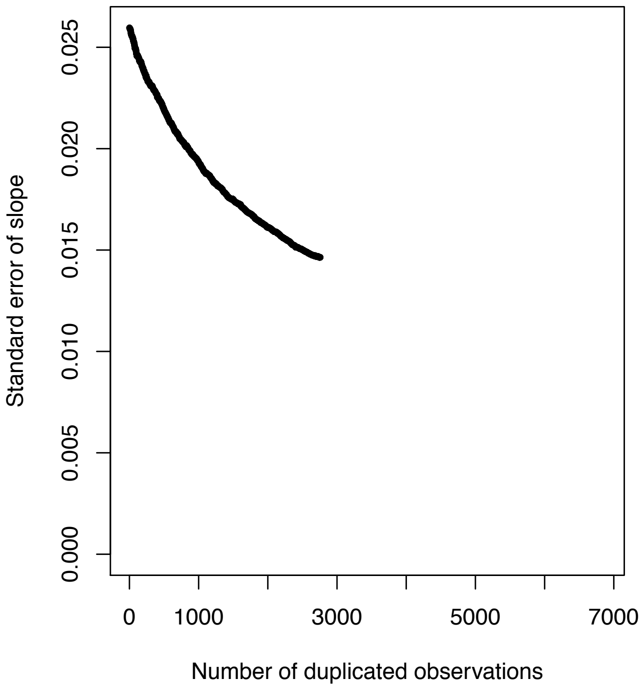


99% CI, duplicates in blue

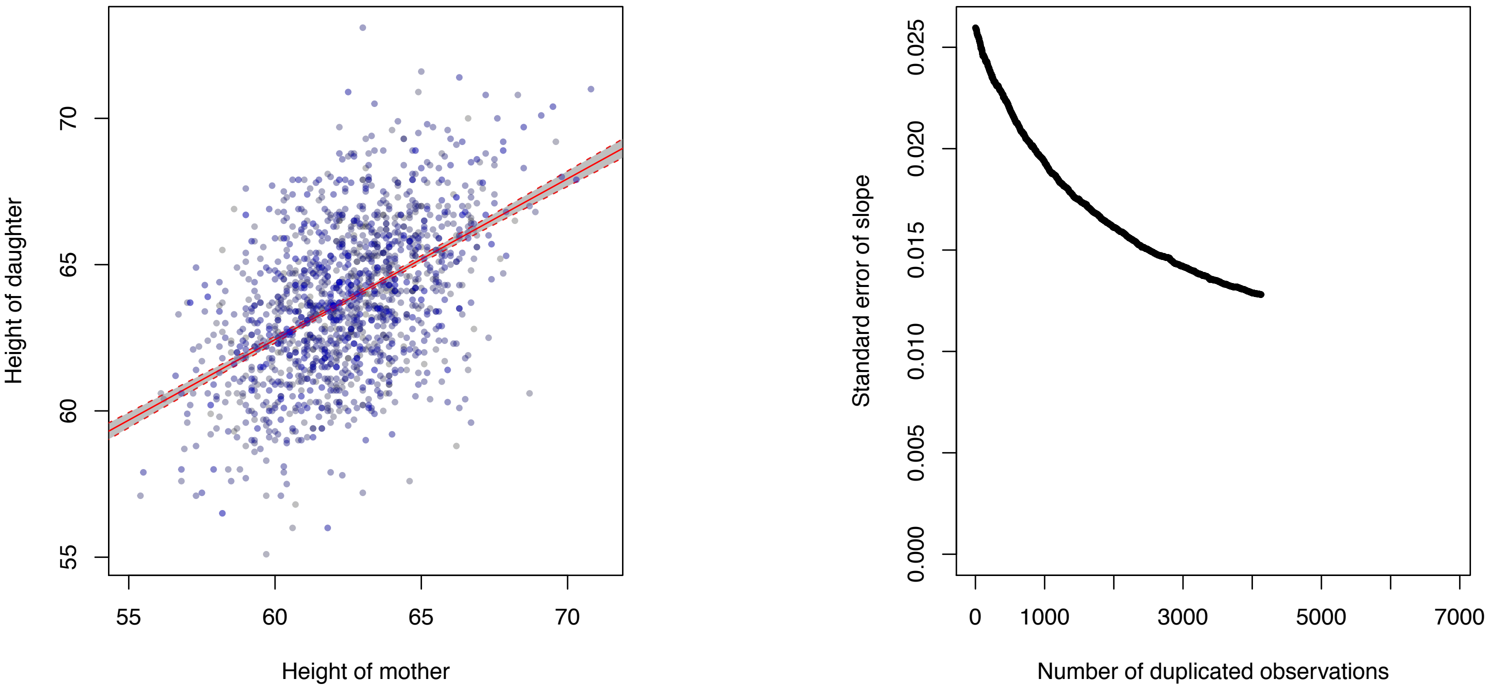


99% CI, duplicates in blue

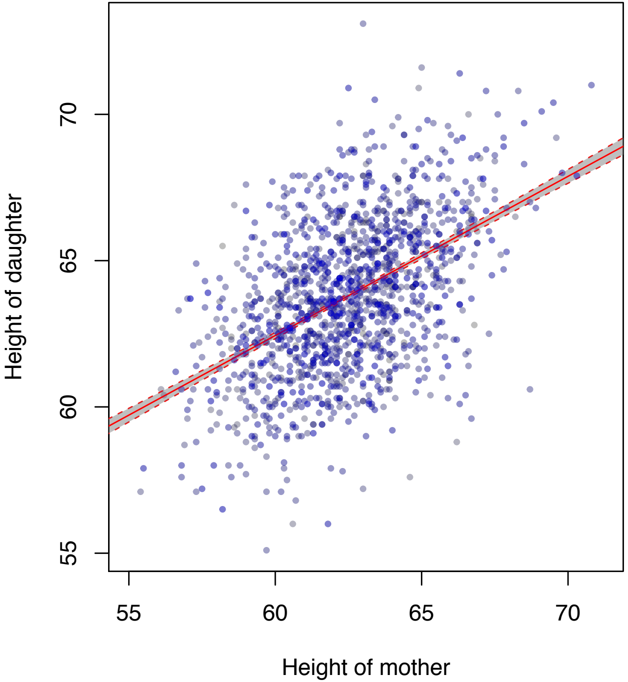


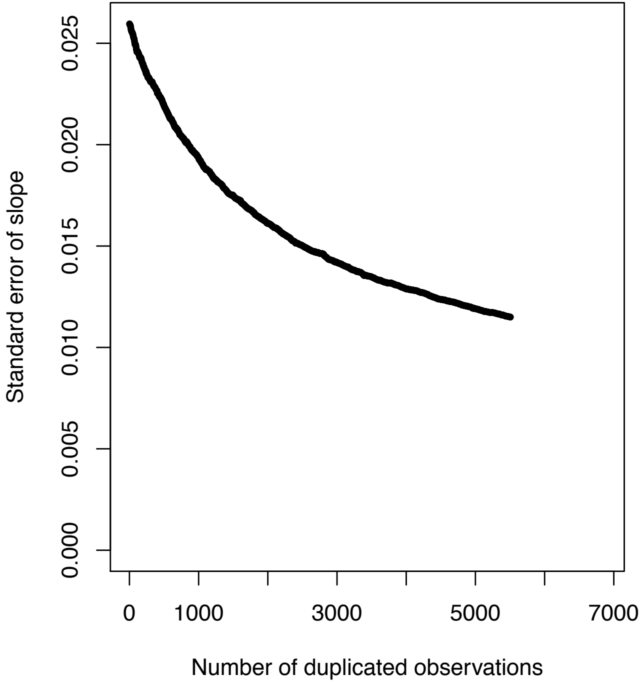


99% CI, duplicates in blue

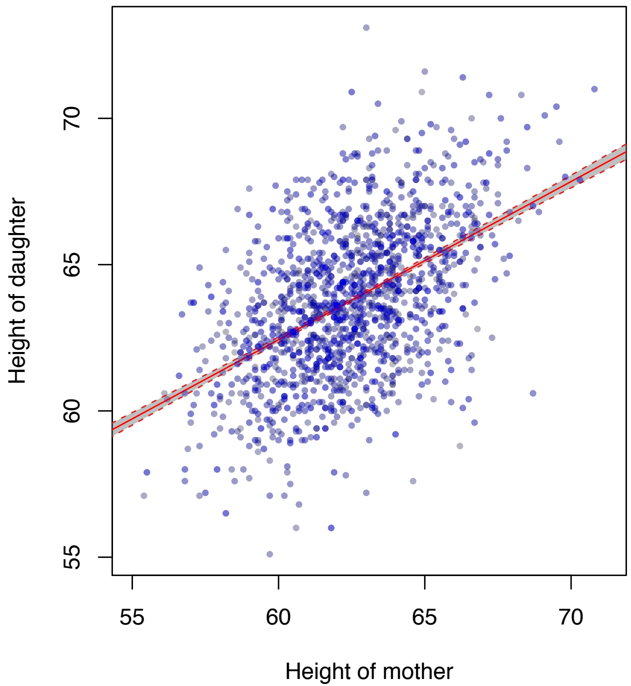


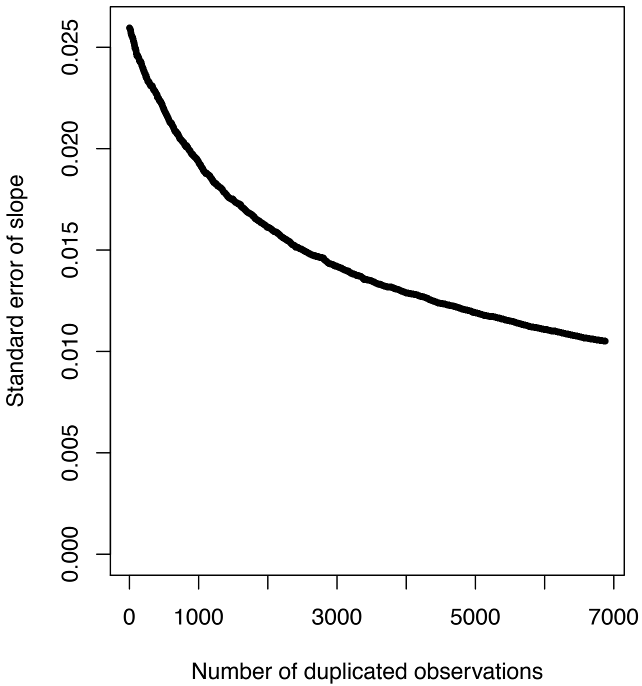
99% CI, duplicates in blue





99% CI, duplicates in blue

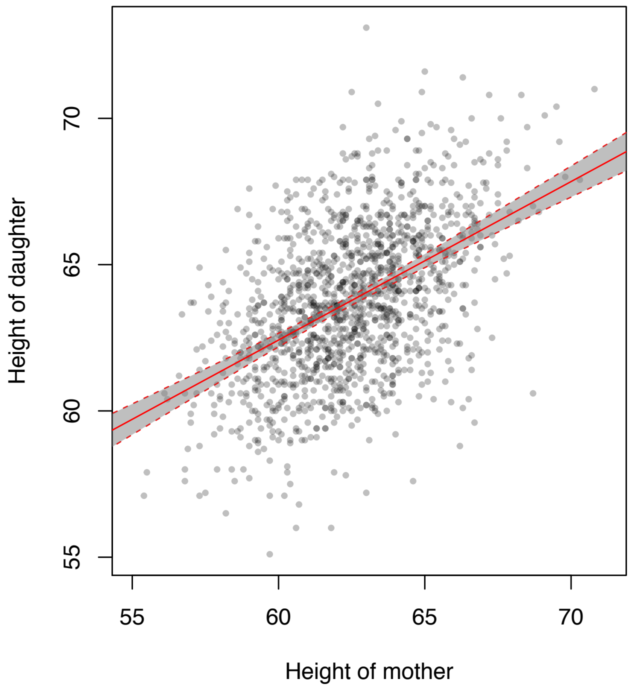


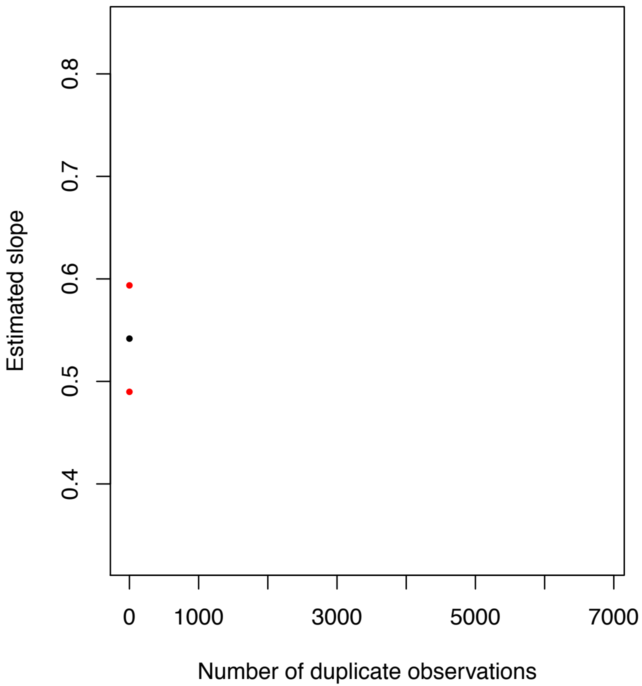


“Duplicated at random” is not so bad

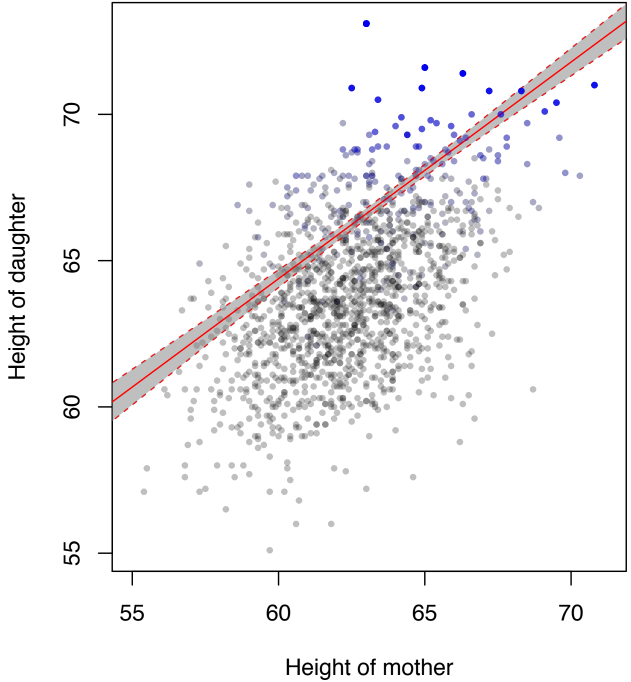
- Standard errors shrink (at a rate of $n^{-1/2}$), but no bias.
- If observations duplicated not at random, but instead proportionately to the dependent variable…

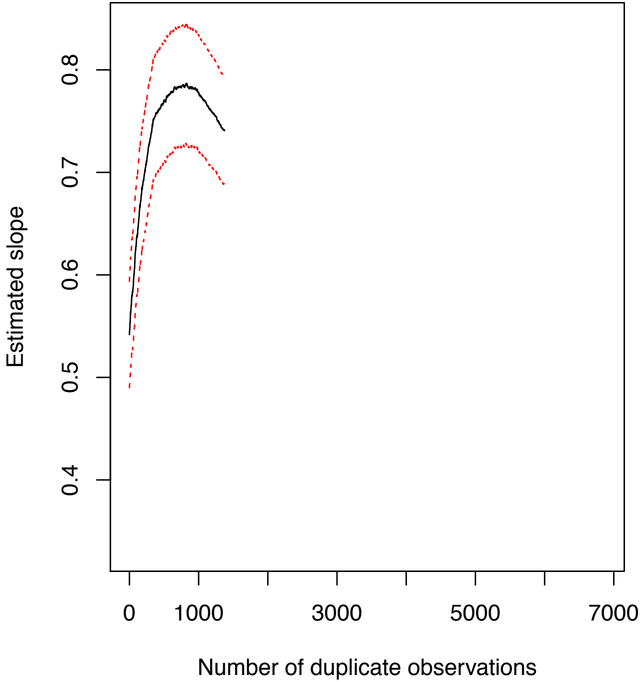
99% CI, duplicates in blue



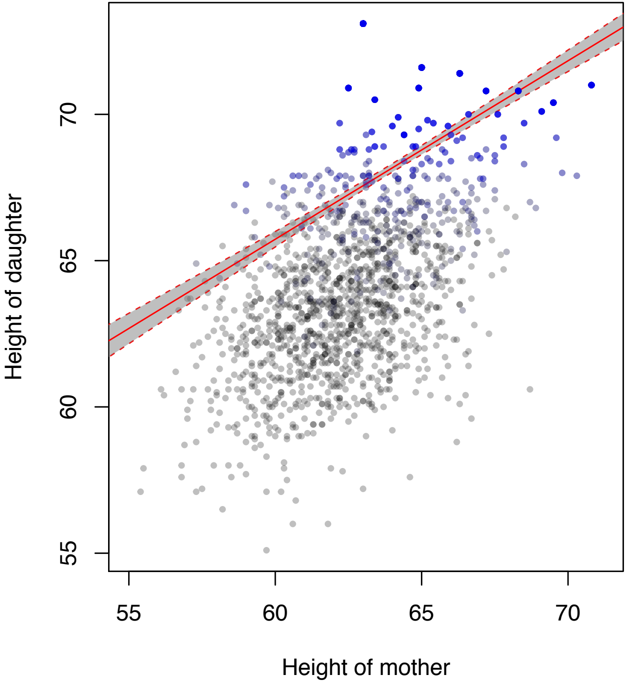


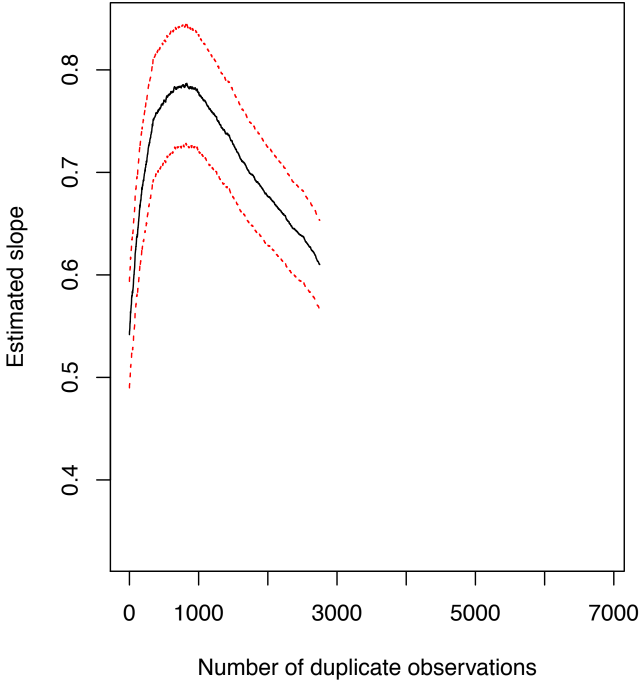
99% CI, duplicates in blue



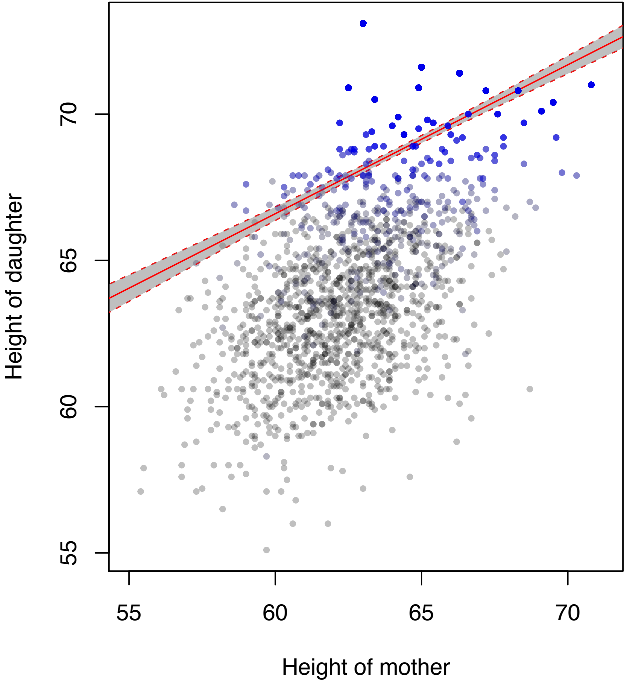


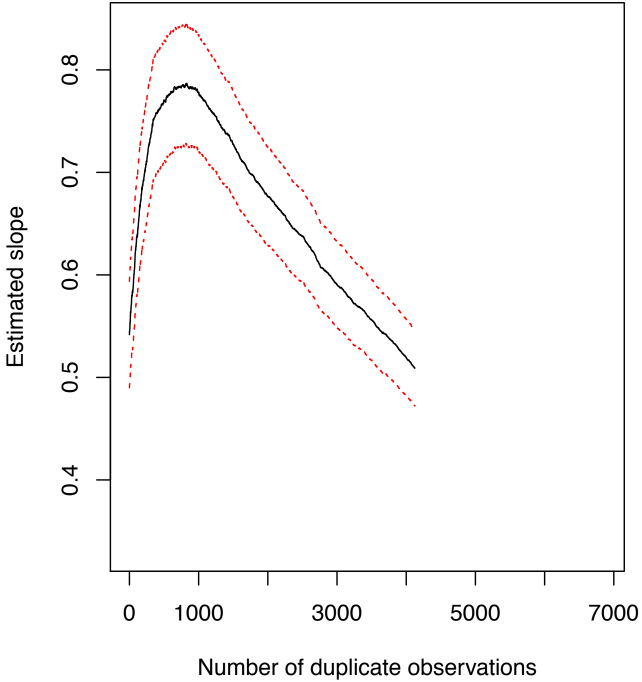
99% CI, duplicates in blue



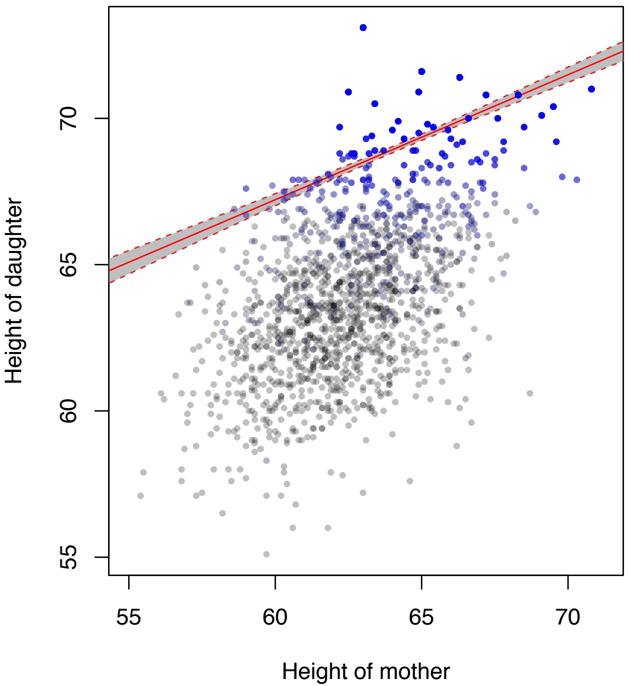


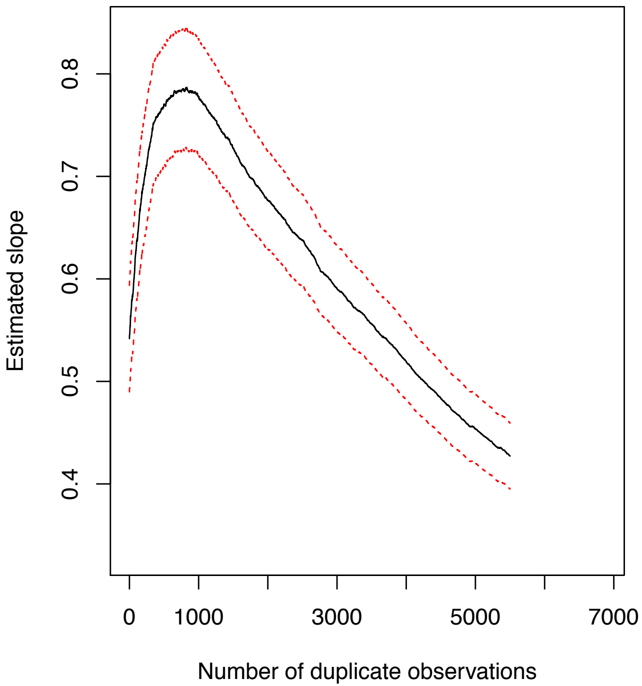
99% CI, duplicates in blue





99% CI, duplicates in blue





Preliminaries

Example

Why statistics?

Probability in statistics

The likelihood principle

Centralities are not statistical

Dependencies

Control for network structure

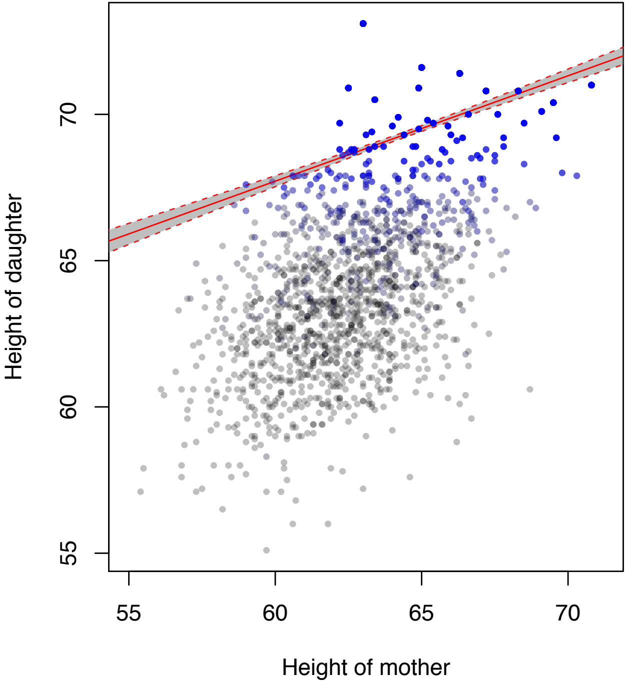
Model network structure

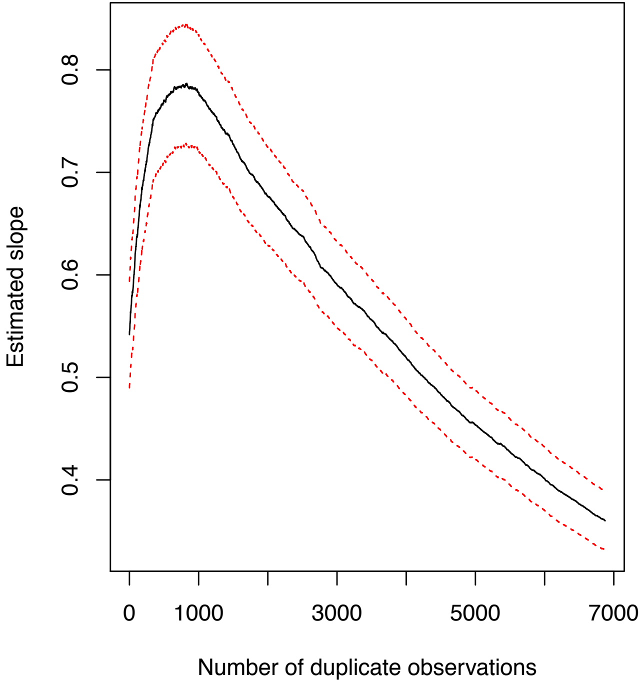
Networks in time

Graphical models

Final thoughts

99% CI, duplicates in blue





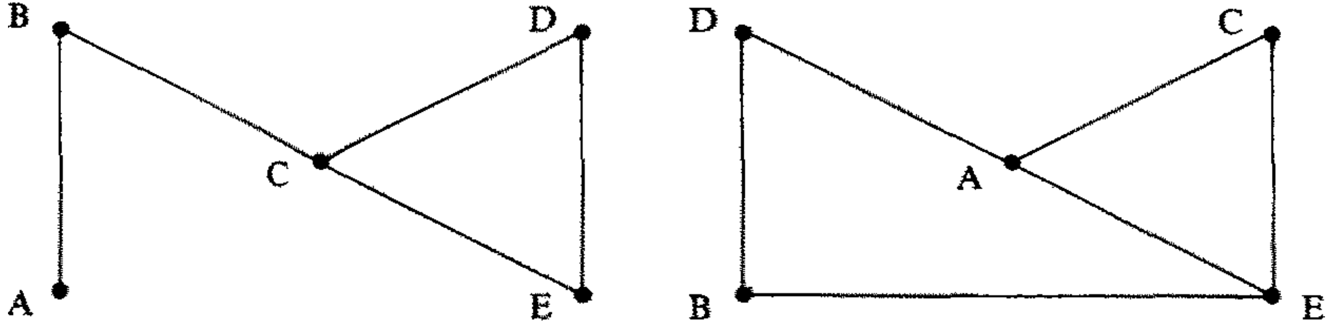
Not at random: Anything goes

- "Dependencies" can shrink standard errors *and* cause bias
- If the dependence is regular enough, we can try to model it directly…
- Time series does this: "temporal autocorrelation" is when an observation is dependent with "itself" at different times
- Network dependencies don't have the same regularity

Models to *control for* network structure

Caution: These seem attractive, but are seldom what we want

Quadratic Assignment Procedure (QAP)



(a) (b)

	A	B	C	D	E
A	x	1			
B	1	x	1		
C		1	x	1	1
D			1	x	1
E			1	1	x

	A	B	C	D	E
A	x		1	1	1
B		x		1	1
C	1		x		1
D	1	1		x	
E	1	1	1		x

Krackhardt (1987)

QAP: The good

- "Quadratic Assignment Procedure" is a *nonparametric permutation test*, same as the Mantel test in ecology (Krackhardt, 1987)
- Procedure: take the adjacency matrix **A** and another matrix **X** of attributes/similarities, turn both into vectors, find the correlation
- Permute the node labels of the graph (apply the same reordering of rows and columns of the adjacency matrix), take the new adjacency matrix **A**′, again turn into a vector, and calculate correlation again. Do many times (often 1000) to get a *null distribution*
- If **X** were be correlated with **A** "by chance," actual correlation should fall in the middle of this null distribution
- If correlation is at the tails of the null distribution, can reject a null of no association

QAP: The bad

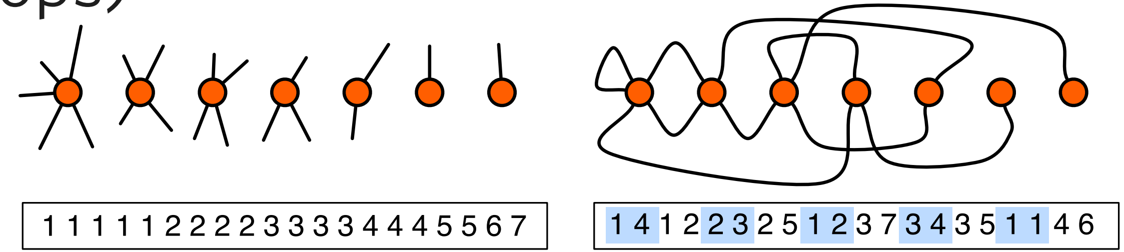
- Can extend to "Multiple Regression QAP" (Dekker, Snijders, & Krackhardt, 2007), same as "Mantel regression" (Ibid.)
- Problem: permutation tests are *tests*, not models
- When using such tests as models, you "get the standard errors from the null model": your *standard errors are a feature of the variability of permutations*, not the variability of your data **X** and **A**
- Further problem: can only *control* for network structure, not model it

Network autocorrelation

- A great frame for understanding dependencies (Dow et al., 1984)
- Analogous to temporal/spatial autocorrelation and time series models: fit a parameter for "lag"
$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \beta + \varepsilon$$
- Problem: can only fit a single parameter for all network autocorrelation
- Problem: is the adjacency matrix the "right" weights matrix? Maybe not! (Leenders, 2002)
- Further problem: again, control for network dependencies at best

Bootstrapping

- Whatever relationship you are interested in: measure on the observed graph, and compare to a null distribution
- Bootstrap: “resampling data”
- Can also sample from a Bernoulli random network
- Configuration model: “rewire” (allow multi-edges and self-loops)



Clauset (2013)

Models *of* network structure

Getting serious

Model the *edges!*

	Y	X_1	X_2	$\cdots$	X_d
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1d}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2d}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nd}

index	*from*	*to*	Y	W_1	W_2	W_3	$\cdots$
e_1	1	2	y_{12}	$\mathbf{1}(x_{11} = x_{21})$	$x_{12} - x_{22}$	x_{13}	$\cdots$
e_2	2	3	y_{23}	$\mathbf{1}(x_{11} = x_{31})$	$x_{12} - x_{32}$	x_{13}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
e_{n+1}	2	1	y_{21}	$\mathbf{1}(x_{21} = x_{11})$	$x_{22} - x_{12}$	x_{23}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$e_{2\binom{n}{2}}$	$n-1$	n	$y_{(n-1)n}$	$\mathbf{1}(x_{(n-1)1} = x_{n1})$	$x_{(n-1)2} - x_{n2}$	$x_{(n-1)3}$	$\cdots$

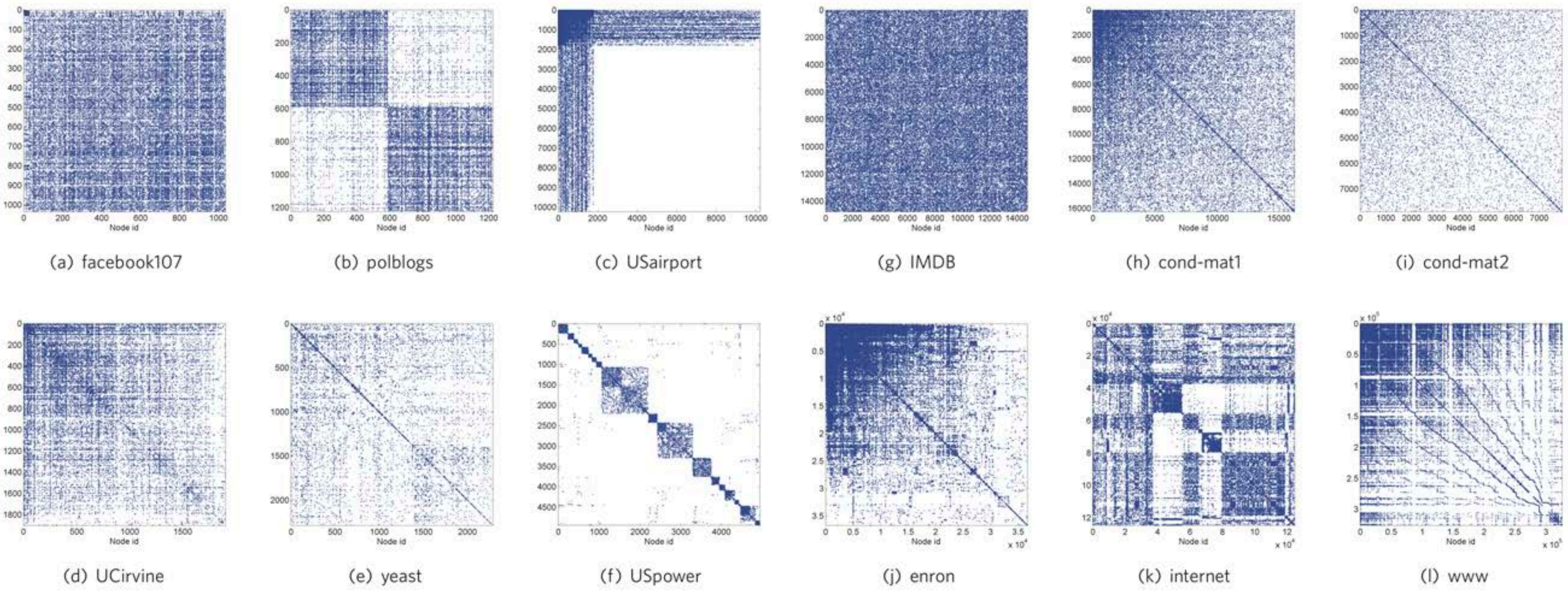
Model the *edges!*

- For maybe two years, I didn't realize that you actually transform your data set
- The edges are dependencies between observations
- Problem: the edges are dependent, too!
- Transitivity, reciprocity, Dunbar's number: these are "dependencies between dependencies"
- (Useful language: "dyad dependent" vs. "dyad independent")
- Not only are we not measuring important forces, but we assume them away (get OVB!)

Logistic regression

- As I showed in the demo, you can create a data set of 0s and 1s for the edges and edge attributes
- Put this into a logistic regression
- This is *misspecified*, but **it is not a bad first pass**
- Sometimes network processes aren't that strong
- Many models build on logistic regression anyway

Block Models: Can be compelling



Caron & Fox (2015)

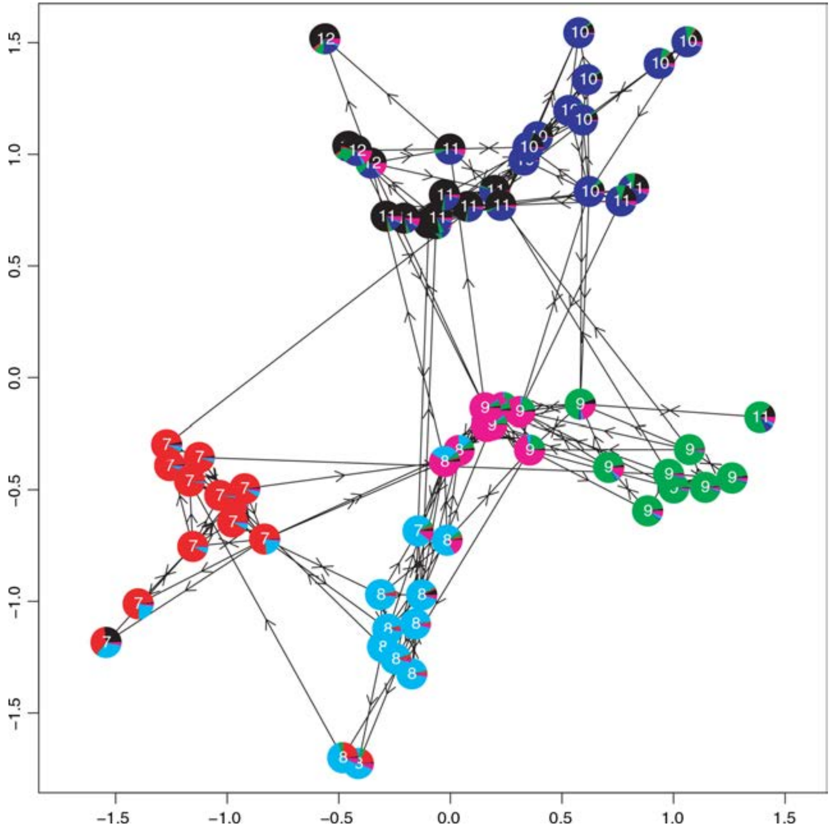
Stochastic Block Models

- A random graph model with "community structure:" separate parameters for within-group ties and out-of-group ties, otherwise everything is a Bernoulli random graph
- A foundational model for statistics, because it is *analytically* tractable
- But for social scientists: it can only model community dependencies, so its use cases are extremely limited
- And finding the "right" ordering and number of groups is hard (like in previous slide)

SUNBELT 2019 JUNE 18-23 MONTRÉAL

Latent Space Models

- Consider networks as existing in an extremely high-dimensional space, where the graph neighbors of a node are the ones it is geometrically closest to
- The dimensions of this space "soak up" all dependencies
- Pro: Unlike other models, this has good theoretical properties
- Con: Pretty much the only information is pictures like this: tells nothing about processes of interest, just gives a visual grouping



Handcock et al. (2006)

p_1, p_2

- Logistic regression on edges can't model dependencies between edges, like reciprocity
- Solution: multinomial regression (with a cross-term). Each pair is an observation, with values in $\{i \quad j, i \rightarrow j, i \leftarrow j, i \leftrightarrow j\}$
- Fixed effects for sending, receiving, and reciprocity
- This is the "p_1 model", recently redescribed as the "β model" or "sender-receiver model"
- p_2 model: random effects version of p_1

Exponential[-family] Random Graph Models (ERGMs)

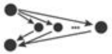
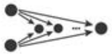
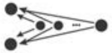
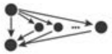
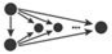
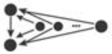
- The crown jewel of 30+ years of research, came out of p_2 model
- (Main version treats graphs as the response: graphs as explanatory are called "autologistic actor attribute models" [ALAAMs], and I never see)
- Logic: specify a set of *sufficient statistics*, calculated over whole network
- These can include terms for anything you can think of
- By construction, these are the sufficient statistics for a graph. Question is if there is any weighting of these statistics that can produced the observed graph

Terms in ERGMs

$$S_1(y) = \sum_{1 \le i < j \le n} y_{ij} \quad \text{number of edges}$$

$$S_k(y) = \sum_{1 \le i \le n} \binom{y_{i+}}{k} \quad \text{number of } k\text{-stars } (k \ge 2)$$

$$T(y) = \sum_{1 \le i < j < h \le n} y_{ij}\, y_{ih}\, y_{jh} \quad \text{number of triangles}$$

Network statistics	Description	Structural signature
Univariate parameters		
Dyadic parameters		
Reciprocity	Occurrence of mutual ties	
Degree parameters		
Mixed 2-star	Correlation of indegrees and outdegrees	
Alternating-in-star (A-in-S)	Network centralisaiton around indegree	
Isolate	Occurrence of actors with zero indegree and zero outdegree	
Sink	Occurrence of actors with an outdegree of zero and indegree of at least one	
Triangle parameters		
Multiple connectivity (A2P-T)	Multiple paths of indirect connectivity	
Shared out-ties (A2P-U)	Activity based structural equivalence: multiple sets of out-ties to the same third others	
Shared in-ties (A2P-D)	Popularity based structure equivalence: multiple sets of in-ties from the same third others	
Transitive closure (AT-T)	Transitive closure of multiple 2-paths	
Activity closure (AT-U)	Closure of multiple in-2-stars	
Popularity closure (AT-D)	Closure of multiple out-2-stars	

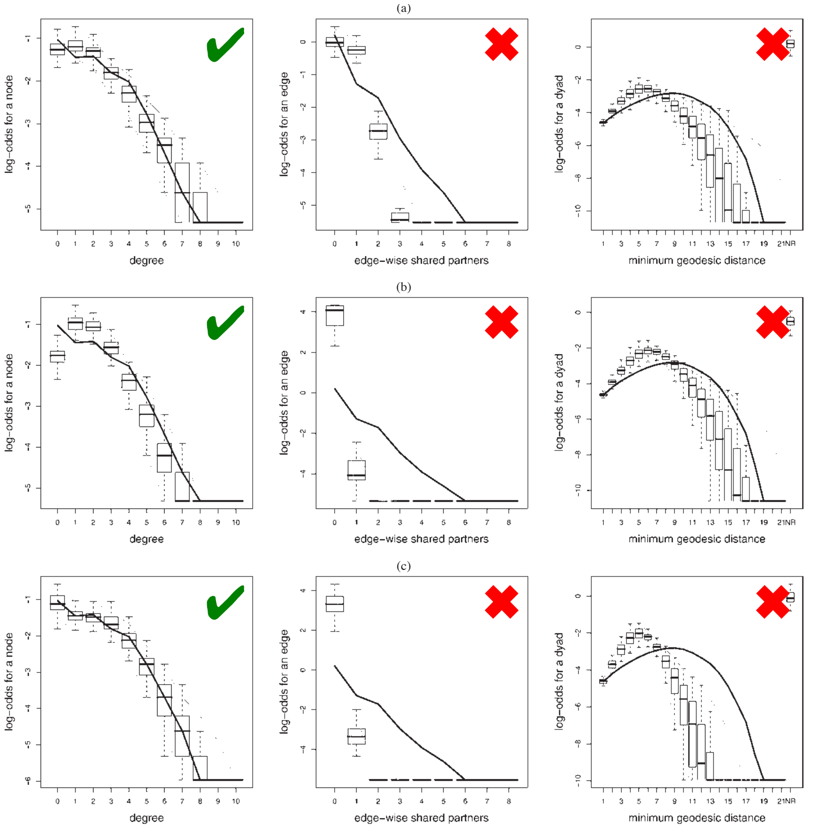
Snijders et al. (2006), Simpson (2015)

ERGMs: Procedure

- Take the observed graph, do counts of sufficient statistics, and initialize weights of terms (through logistic regression)
- Holding the rest of the graph constant, consider a single edge.
- How would removing this edge (if present) or adding it (if absent) change the count of sufficient statistics? Would a higher/lower count make the graph more likely based on current weights?
- Do this for some time to explore the parameter space (an MCMC procedure)
- Do the counts of the actual observed graph fit within the distribution?
- If yes, you're done
- If not, adjust the weights up/down, and start everything all over!
- At the end: if the terms put in were indeed the "correct" ones, these would be their weights

ERGMs: Goodness-of-fit testing

- Excellent goodness-of-fit (GOF) testing framework.
- See if the sufficient statistics that you put into the model can recover the distribution of statistics that were not among your sufficient statistics
- E.g., can density, reciprocity and transitivity alone as sufficient statistics recover the graph's degree distribution?
- Can test with anything (e.g., any subgraph/graph motif density), but should be theoretically important
- Gives a complete framework for finding a parsimonious explanation

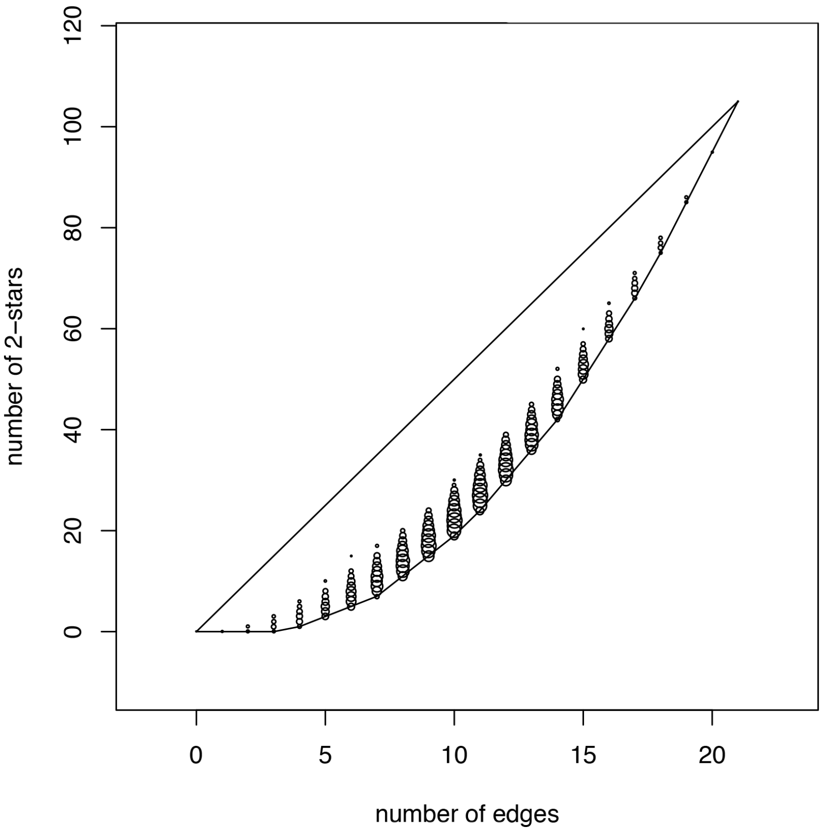


Hunter et al. (2008)

SUNBELT 2019 JUNE 18-23 MONTRÉAL

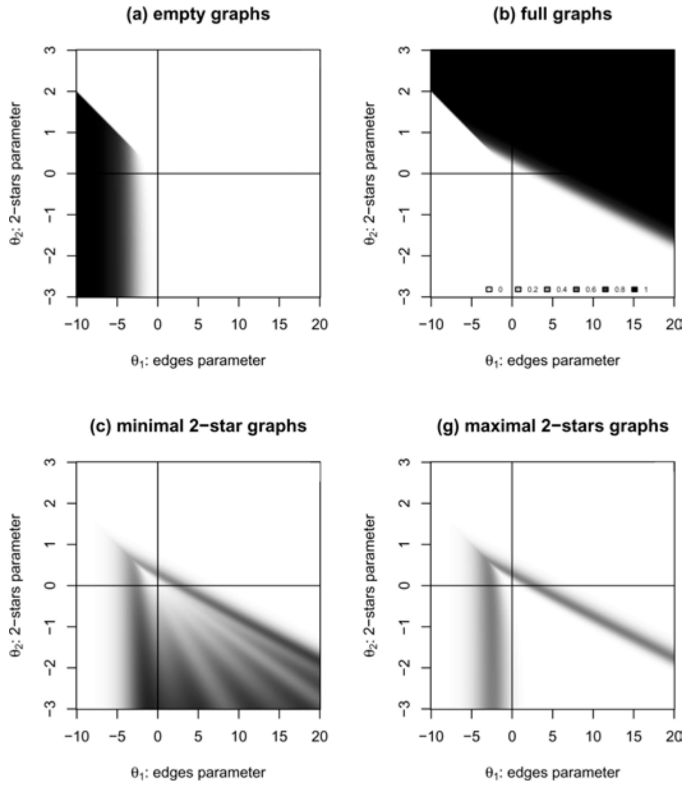
ERGMs: The bad news

- LOTS of problems.
- The space of graphs doesn't play nice with probabilities
- There are only a certain number of graphs of any given size, and only a certain number of graphs with a combination of sufficient statistics



Handcock (2003)

ERGMs: The bad news



Handcock (2003)

- Sometimes, under large portions of the parameter space, the most likely graph is either the complete graph or the empty graph: such specifications are are *degenerate*
- Because the space of graphs is so large, don't know if a model is degenerate or if our MCMC procedure is bad
- Model degeneracy (arguably) has nothing to do with the social phenomena of interest
- Better specifications are (arguably) technical, not sociological, entities: e.g., "geometrically weighted edgewise shared partners", GWESP
- Alternatively, maybe the math gives us insight into sociological processes (e.g., the effect of increasing friends in common is not linear; there's much more difference between 1 friend in common and 2 friends in common than with 20 and 30, and GWESP models this)

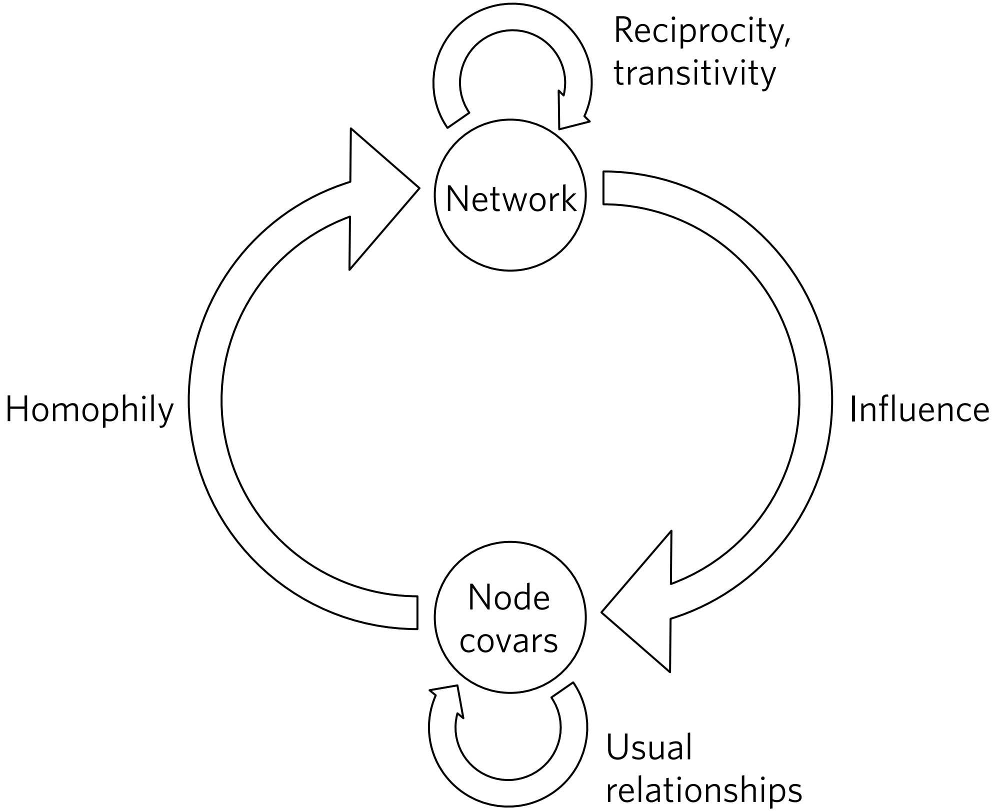
ERGMs: More bad news

- Another: ERGMs are not "projective" (Shalizi & Rinaldo, 2013)
- Practically: if you are missing one node, it could have ties to every other single node, which would completely change the estimates of all the network effects. Very fragile.
- But maybe this is an issue of research design, not arcane statistical theory far removed from practice (Schweinberger et al., 2017)

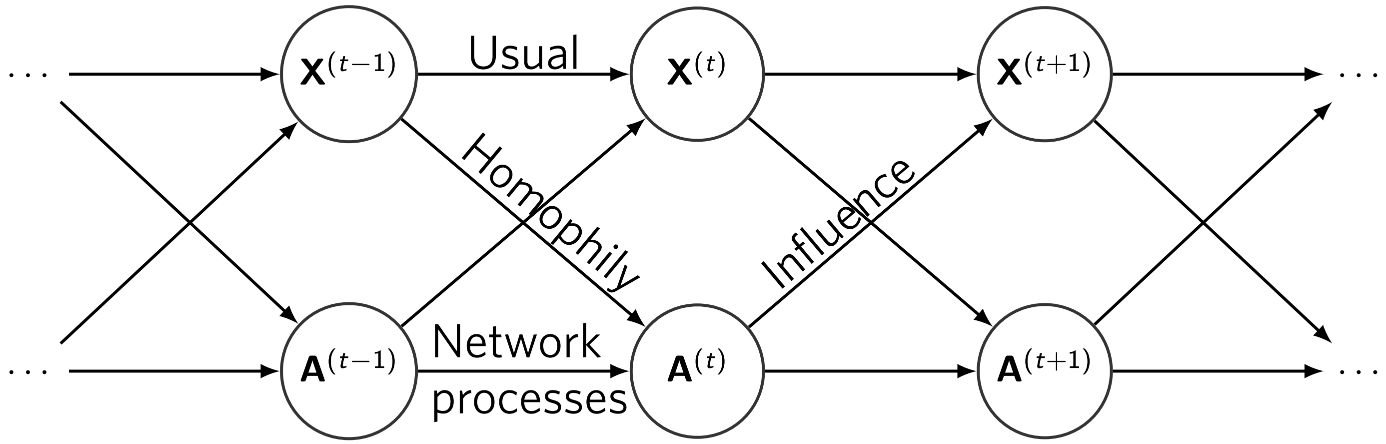
Models for networks in time

We can do a lot more if we have temporal information, either longitudinal (discrete time) or timestamped (continuous time)

Remember this?



Time (maybe) lets us sort it out



Stochastic Actor-Oriented Models (SAOMs)/ SIENA

- A different perspective: model actor decision-making ("utility")
- Main SAOM is SIENA (Simulation Investigation for Empirical Network Analysis)
- Create utility functions with ERGM-like terms (SIENA manual gives 100+ built-in terms)
- Uses something like an agent-based model to fit the terms
- Elegant, only model to get at co-evolution of behavior and networks, but layers upon layers of assumptions
- And in practice, SIENA can be very temperamental, it's hard to models to successfully run

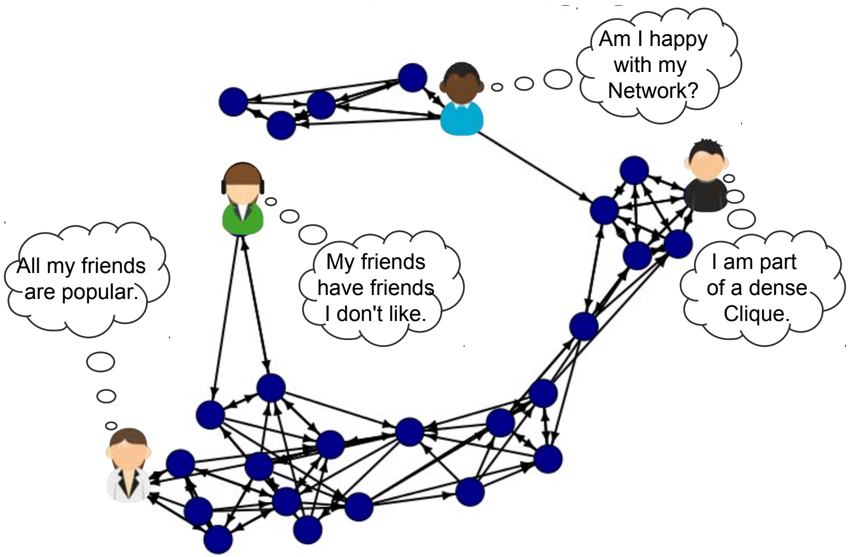


Image: Christoph Stadtfeld and Zsófi Boda, Introduction to SIENA – Part 1, SIENA Workshop Sunbelt 2016. Cite: Steglich et al. (2010).

Relational Event Models (REMs)

- Relational Event Models (Butts, 2008b) model continuous-time network data (network ties with time steps, e.g. emails or calls, each of which is called an “event”)
- It is similar to (and builds on) ERGMs and SIENA in the terms it uses to express processes like transitivity, reciprocity, etc. Like SIENA, it models actor decision-making (the likelihood function tries to capture actor “utility”)
- REMs normalize the probability of an observed event stream by possible alternative actions (e.g., all other possible sender-receiver pairs) at a the time of each event given all previous events until then
- A good, reasonable model, but has extremely low predictive performance

Scalability

- Yet another problem: none of the “good” models (LSMs, ERGMs, SIENA, REMs) scale past a few hundred nodes at best
  - They all require intensive computation (generally, MCMC procedures through a space of graphs or at least alternative edges)
- So, forget using any of these to model all of Facebook, or any other big dataset

Graphical models

Helpful conceptual tool for understanding dependencies

More background: Expectation, conditioning

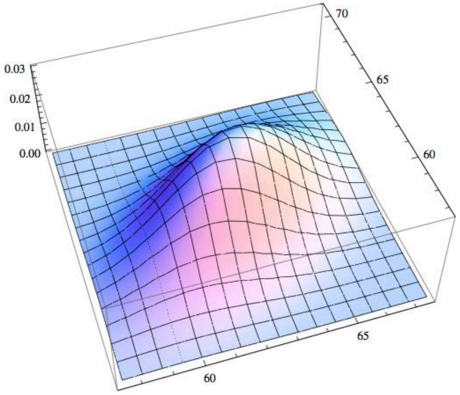
- Expectation: Take all possible outcomes, multiple each by its probability, and add up
- For a six-sided die:

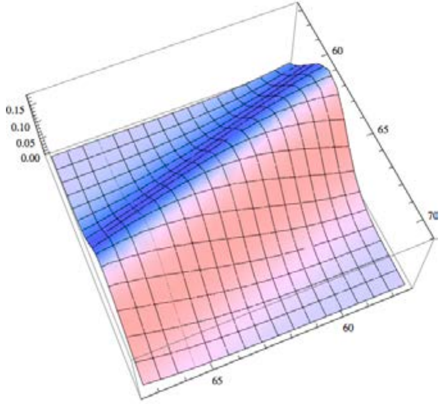
$$\left(1 \times \tfrac{1}{6}\right) + \left(2 \times \tfrac{1}{6}\right) + \cdots + \left(6 \times \tfrac{1}{6}\right) = 3.5$$

- *Conditional* probability of value of two dice, given that the first one is 6
- *Conditional expectation* of the value of two dice, given that the first one is 6

Regression as *conditional mean*

- Regression: conditional expectation of y given x
- Or, taking the joint distribution of y and x, then "fixing" (conditioning on) values of x
- Images to the right: top is a bivariate normal, bottom is distributions of y conditioned on values of x
- The *expectation* is the line along the "ridge" in the bottom image (only slices of x are probability, the overall surface is no longer a probability)





The data table

	Y	X_1	X_2	$\cdots$	X_d
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1d}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2d}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nd}

Full joint probability

- Full conditional probability includes *everything* in the data table together

$$p(Y|\mathbf{X}) = p(y_1, \ldots, y_n | x_{11}, \ldots, x_{1d}, x_{21}, \ldots, x_{2d}, \ldots\ldots, x_{n1}, \ldots, x_{nd})$$

- We never take this; we only separate out by *observations* by assuming observations are independent and identically distributed (iid)

What is independent? Observations?

- The "independent and identically distributed" assumption is that: all observations tell us about the same underlying phenomenon
- Let $\mathbf{x}_i$ be the vector of person i's covariates
- Then, in math:

$$p(Y|\mathbf{X}) = p(y_1, \ldots, y_n|\mathbf{x}_1, \ldots, \mathbf{x}_n) \overset{\text{iid}}{=} \prod_{i=1}^{n} p(y_i|\mathbf{x}_i)$$

What is independent? Or variables?

- In regression, we assume Y is dependent on each covariate individually. Covariates are mutually independent
- *Collinearity*, (linear) dependence between covariates, breaks this assumption
- There's a whole area of statistics and computer science that models these dependencies, descended from path diagrams

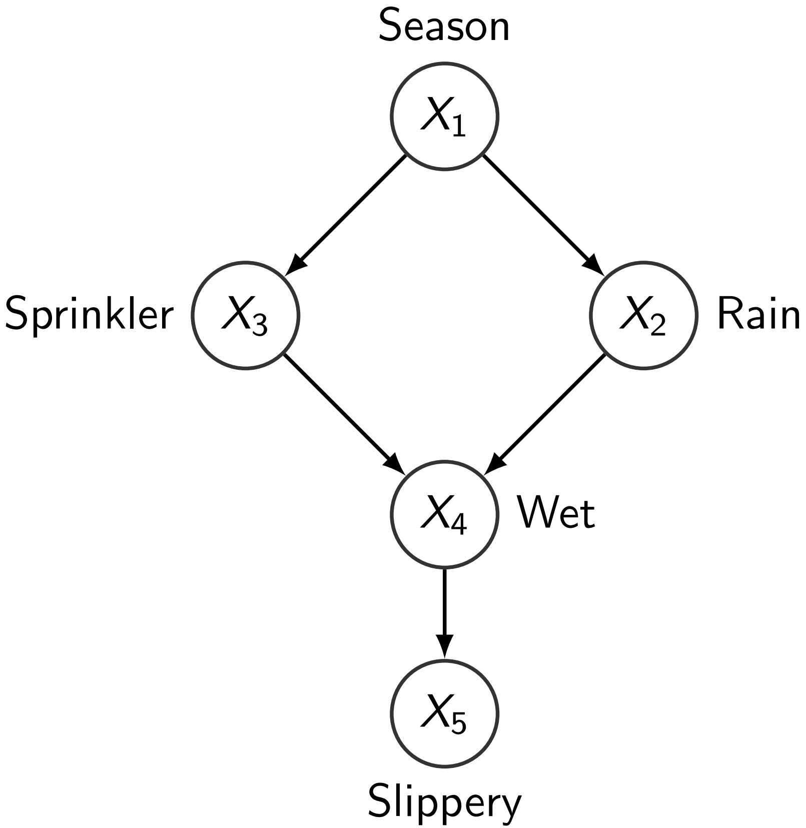
Graphical models

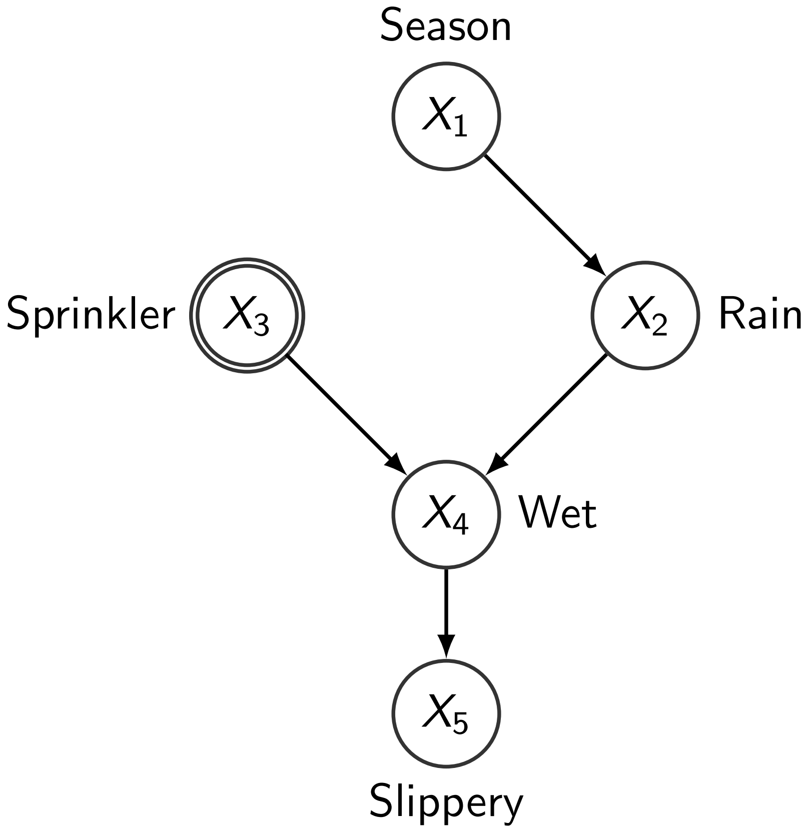
- Descended from path diagrams
- Represent dependencies between *variables*
- Regard covariates as random, not fixed, so have $p(X)$'s
- If all covariates are independent,

$$p(Y, \mathbf{X}) = p(Y, X_1, \ldots, X_d) \overset{\text{iid}}{=} p(Y | X_1, \ldots, X_d) \prod_{j=1}^{d} p(X_j)$$

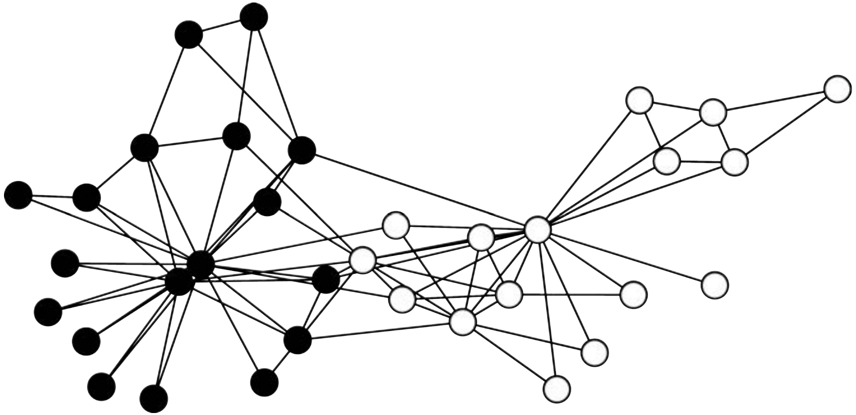
- If not independent, we have extra terms to worry about
- This also lets us represent causality and intervention

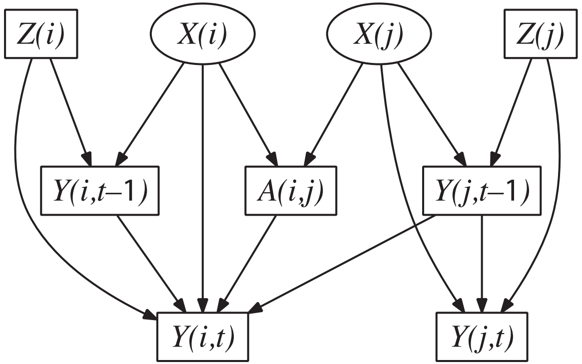
Causality, and intervention





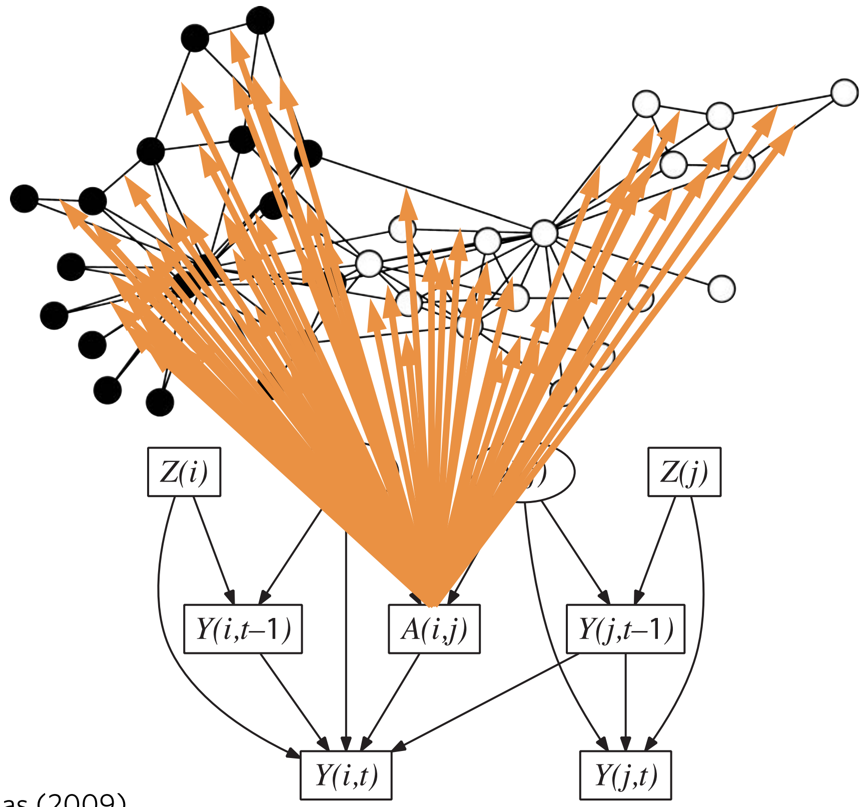
Networks: Dependencies between the *edges*





Graphical model: Shalizi & Thomas (2009)

Networks: Dependencies between the *edges*

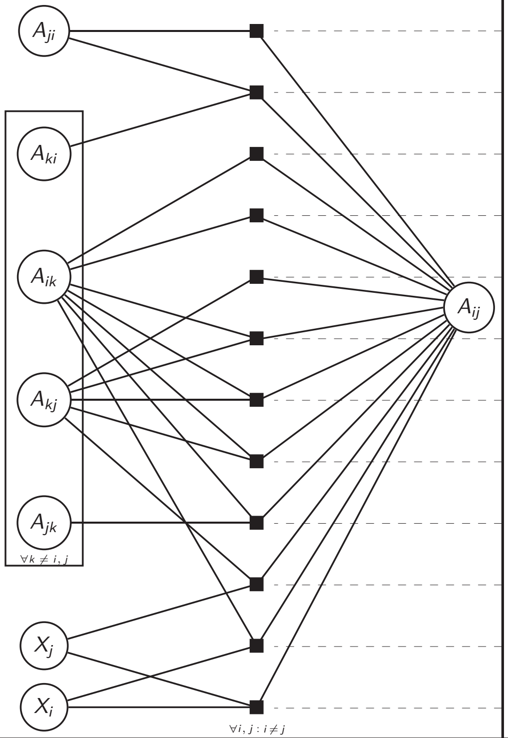


Graphical model: Shalizi & Thomas (2009)

ERGMs as a graphical model

Terms: Snijders et al. (2006). From (unpublished) joint work with Antonis Manousis, and Naji Shajarisales.

Factor graph



Parameter name	Network Motif	Parameterization	Matrix notation
mutual dyads		$\sum_{i<j} A_{ij}A_{ji}$	$\frac{1}{2}\,\text{tr}\left(\mathbf{AA}^T\right)$
in-two-stars		$\sum_{(i,j,k)} A_{ji}A_{ki}$	$\text{sum}\left(\mathbf{AA}^T\right) - \text{tr}\left(\mathbf{AA}^T\right)$
out-two-stars		$\sum_{(i,j,k)} A_{ij}A_{ik}$	$\text{sum}\left(\mathbf{A}^T\mathbf{A}\right) - \text{tr}\left(\mathbf{A}^T\mathbf{A}\right)$
geom. weighted out-degrees	—	$\sum_i \exp\left\{-\alpha \sum_k A_{ik}\right\}$	$\text{sum}\left(\exp\{-\alpha\,\text{rowsum}\left(\mathbf{A}\right)\}\right)$
geom. weighted in-degrees	—	$\sum_j \exp\left\{-\alpha \sum_k A_{kj}\right\}$	$\text{sum}\left(\exp\{-\alpha\,\text{colsum}\left(\mathbf{A}\right)\}\right)$
alternating transitive k-triplets		$\lambda \sum_{i,j} A_{ij}\left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\,\text{sum}\left(\mathbf{A}\odot\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\text{diag}(\mathbf{AA})}\right)\right)$
alternating indep. two-paths		$\lambda \sum_{i,j}\left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\,\text{sum}\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\text{diag}(\mathbf{AA})}\right)$
two-paths (mixed two-stars)		$\sum_{(i,k,j)} A_{ik}A_{kj}$	$\text{sum}\left(\mathbf{AA}\right) - \text{tr}\left(\mathbf{AA}\right)$
transitive triads		$\sum_{(i,j,k)} A_{ij}A_{jk}A_{ik}$	$\text{tr}\left(\mathbf{AAA}^T\right)$
activity effect		$\sum_i X_i \sum_j A_{ij}$	$\text{sum}\left(\mathbf{X}\odot\text{rowsum}\left(\mathbf{A}\right)\right)$
popularity effect		$\sum_j X_j \sum_i A_{ij}$	$\text{sum}\left(\mathbf{X}\odot\text{colsum}\left(\mathbf{A}\right)\right)$
similarity effect		$\sum_{i,j} A_{ij}\left(1-\frac{\lvert X_i-X_j\rvert}{\max_{k,l}\lvert X_k-X_l\rvert}\right)$	$\text{sum}\left(\mathbf{A}\odot\mathbf{S}\right)$

Final thoughts

Now you know how everything is terrible, and nothing works.

So, what do we do?

- ERGMs, SAOMs, or REMs… if you believe them
- Other models, if you really have no transitivity
- Make better models?
  - Include transitivity!
- Clever study design?
- Or…
  - Give up on explanation and do only prediction
  - Give up on empirical analysis and do simulation modeling
  - Give up on modeling and do qualitative analysis

Come across a fancy (new) network model and wondering if it's the answer?

- (Don't worry, it's not.)
- My heuristic: "[how] does it model transitivity?"
- If it doesn't, I'm not interested
  - I care about network processes, for which transitivity (which happens between node triplets) is exemplary
- E.g., "degree-corrected stochastic block model"? Nope. "Kronecker graphs"? Nope. The "influence model"? Nope.
- AMEN? Sort of. Doesn't model transitivity itself, only the clustering that results.
- Caveat: if you are doing prediction, not explanation (Shmueli, 2010; Breiman, 2001), the data-generating process is irrelevant and you should use whatever can perform well

The eternal caveat

- "All models are wrong…"
- "…but some are useful." –George Box
- Networks are hard to measure
- All network data is highly uncertain
  - Perfect and complete trace data (e.g., online social media) doesn't give us what's important
  - Getting at what's important (e.g., through surveys and interviews) gives us imperfect and incomplete data
- Networks are an abstraction. They may not be the "right" abstraction.

References (1 of 4)

Abbott, A. (1988). Transcending general linear reality. *Sociological Theory, 6*(2), 169–186. https://dx.doi.org/10.2307/202114.

Arceneaux, K., Gerber, A. S., & Green, D. P. (2010). A cautionary note on the use of matching to estimate causal effects: An empirical example comparing matching estimates to an experimental benchmark. *Sociological Methods & Research*, *39*(2), 256–282. https://dx.doi.org/10.1177/0049124110378098.

Besag, J. (2001). Markov Chain Monte Carlo for statistical inference. Working Paper No. 9, Center for Statistics and the Social Sciences, University of Washington. https://www.csss.washington.edu/files/working-papers/2003/wp39.pdf.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231. https://dx.doi.org/10.1214/ss/1009213726.

Butts, C. T. (2008a). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology*, *11*, 13–41. https://dx.doi.org/10.1111/j.1467-839X.2007.00241.x.

Butts, C. T. (2008b). A relational event model for social action. *Sociological Methodology*, *38*, 155–200. https://dx.doi.org/10.1111/j.1467-9531.2008.00203.x.

Caron, F., & Fox, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society Statistical Methodology Series B*, *79*(5), 1295–1366. https://dx.doi.org/10.1111/rssb.12233.

Clauset, A. (2013). Lecture 11: The configuration model. CSCI 5352: Network analysis and modeling. University of Colorado Boulder. http://tuvalu.santafe.edu/~aaronc/courses/5352/fall2013/csci5352_2013_L11.pdf.

Cox, D. R. (1990). Role of models in statistical analysis. *Statistical Science, 5*(2), 169–174. https://dx.doi.org/10.1214/ss/1177012165.

References (2 of 4)

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. https://dx.doi.org/10.1111/ajps.12263.

Dekker, D., Krackhardt, D., & Sniggers, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika*, *72*(4), 563–581. https://dx.doi.org/10.1007/S11336-007-9016-1.

Dow, M. M., Burton, M. L., White, D. R., and Reitz, K. P. (1984). Galton's problem as network autocorrelation. *American Ethnologist*, *11*(4), 754–770. https://dx.doi.org/10.1525/ae.1984.11.4.02a00080.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222*, 309–368. https://dx.doi.org/10.1098/rsta.1922.0009.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, *2*(2), 129–233. https://arxiv.org/abs/0912.5410.

Handcock, M. S. (2003). Degeneracy in statistical models of social networks. Working Paper No. 39, Center for Statistics and the Social Sciences, University of Washington. https://www.csss.washington.edu/Papers/2003/wp39.pdf.

Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society A*, *170*(Part 2), 301–354. https://dx.doi.org/0.1111/j.1467-985X.2007.00471.x.

Hanneman, R. A. and Riddle, M. (2005). Chapter 18: Some statistical tools. In *Introduction to social network methods*. University of California, Riverside. http://faculty.ucr.edu/~hanneman/nettext/C18_Statistics.html.

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008) Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248–258. https://dx.doi.org/10.1198/016214507000000446.

References (3 of 4)

Kass, R. E. (2011). Statistical inference: The big picture. *Statistical Science*, *26*(1), 1–9. https://dx.doi.org/10.1214/10-STS337.

Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical analysis of network data with R.* Springer-Verlag. https://dx.doi.org/10.1007/978-1-4939-0983-4.

Krackhardt, D. (1987). QAP partialling as a test of spuriousness. *Social Networks*, *9*, 171–186. https://dx.doi.org/10.1016/0378-8733(87)90012-8.

Laumann, E. O., Marsden, P.V., & Galaskiewicz, J. (1977). Community-elite influence structures: Extension of a network approach. *American Journal of Sociology*, *83*(3), 594–631. http://www.jstor.org/stable/2778146.

Leenders, R. T. A. J. (2002). Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, *24*, 21–47. https://dx.doi.org/10.1016/S0378-8733(01)00049-1.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://dx.doi.org/10.1257/jep.31.2.87.

O'Malley, A. J., & Marsden, P. V. (2008). The analysis of social networks. *Health Services and Outcomes Research Methodology*, *8*, 222–269. https://dx.doi.org/10.1007/s10742-008-0041-z.

Pattison, P. and Robins, G. (2008). Chapter 18: Probabilistic network analysis. In T. Rudas (Ed.), *Handbook of probability: Theory and applications* (pp. 291–313). SAGE. https://www.sagepub.com/sites/default/files/upm-binaries/18551_Chapter18.pdf.

Robins, G. (2015). Drawing conclusions: Inference, generalization, causality and other weighty matters. In *Doing social network research: Network-based research design for social scientists* (pp. 211–230). SAGE.

References (4 of 4)

Robins, G., Lewis, J. M., & Wang, P. (2012). Statistical network analysis for analyzing policy networks. *The Policy Studies Journal*, *40*(3), 375–401. https://dx.doi.org/10.1111/j.1541-0072.2012.00458.x.

Rose, T. (2016). *The end of average: How we succeed in a world that values sameness*. New York: HarperOne. See animated video at https://vimeo.com/237632676 and excerpt at https://www.thestar.com/news/insight/2016/01/16/when-us-air-force-discovered-the-flaw-of-averages.html.

Schweinberger, M., Krivitsky, P. N., & Butts, C. T. (2017). A note on the role of projectivity in likelihood-based inference for random graph models. *arXiv*:1707.00211v1. https://arxiv.org/abs/1707.00211.

Shalizi, C. R., & Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *The Annals of Statistics*, *41*(2), 508–535. https://dx.doi.org/10.1214/12-AOS1044.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, *40*(2), 211–239. https://dx.doi.org/10.1177/0049124111404820.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://dx.doi.org/10.1214/10-STS330.

Simpson, C. R. (2015). Multiplexity and strategic alliances: The relational embeddedness of coalitions in social movement organisational fields. *Social Networks*, *42*, 42-59. https://dx.doi.org/10.1016/j.socnet.2015.02.007.

Steglich, C., Snijders, T. A. B., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, *40*(1), 329–393. https://dx.doi.org/10.1111/j.1467-9531.2010.01225.x.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, *37*(1):131–153. https://dx.doi.org/10.1146/annurev.soc.012809.102709.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for Exponential Random Graph Models. *Sociological Methodology*, *36*, 99–153. https://www.jstor.org/stable/25046693.

Acknowledgements

(In chronological order, and partial)

- Tom Snijders
- Cosma Shalizi
- Andrew Thomas
- Brian Junker
- Cohen Simpson
- David Krackhardt
- Hemank Lamba
- Alex Loewi
- Philipp Burckhardt
- Dave Choi
- Valerie Ventura
- Seth Flaxman
- The stat-network group at CMU
- Brandy Aven
- Jeremy Koster
- Kate Anderson
- Nynke Niezink
- The participants of this 2019 Sunbelt workshop!

All errors, overgeneralizations, and unfair characterizations of models are my own.