THE GRADUATE INSTITUTE GENEVA

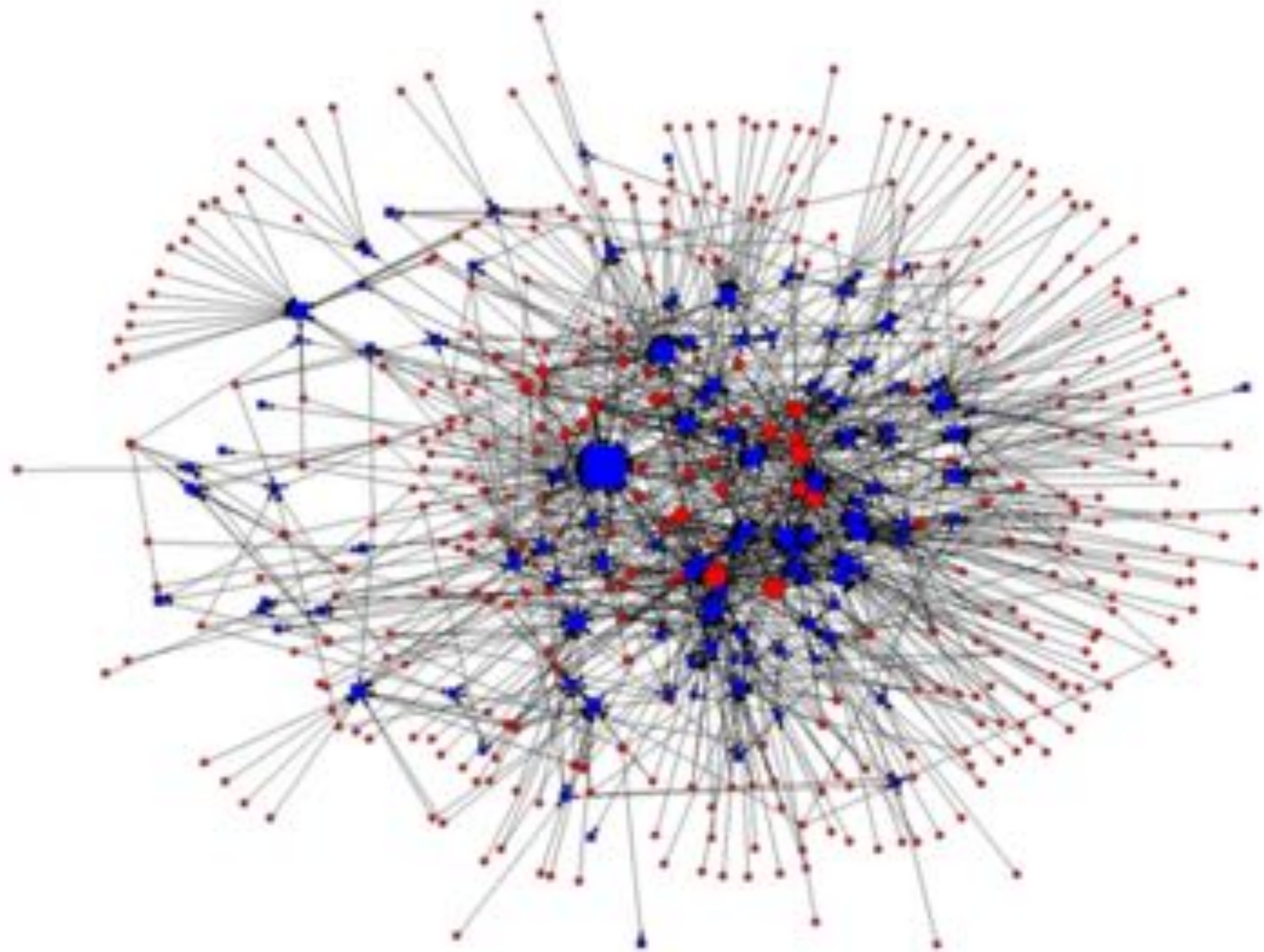
INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

ERGM

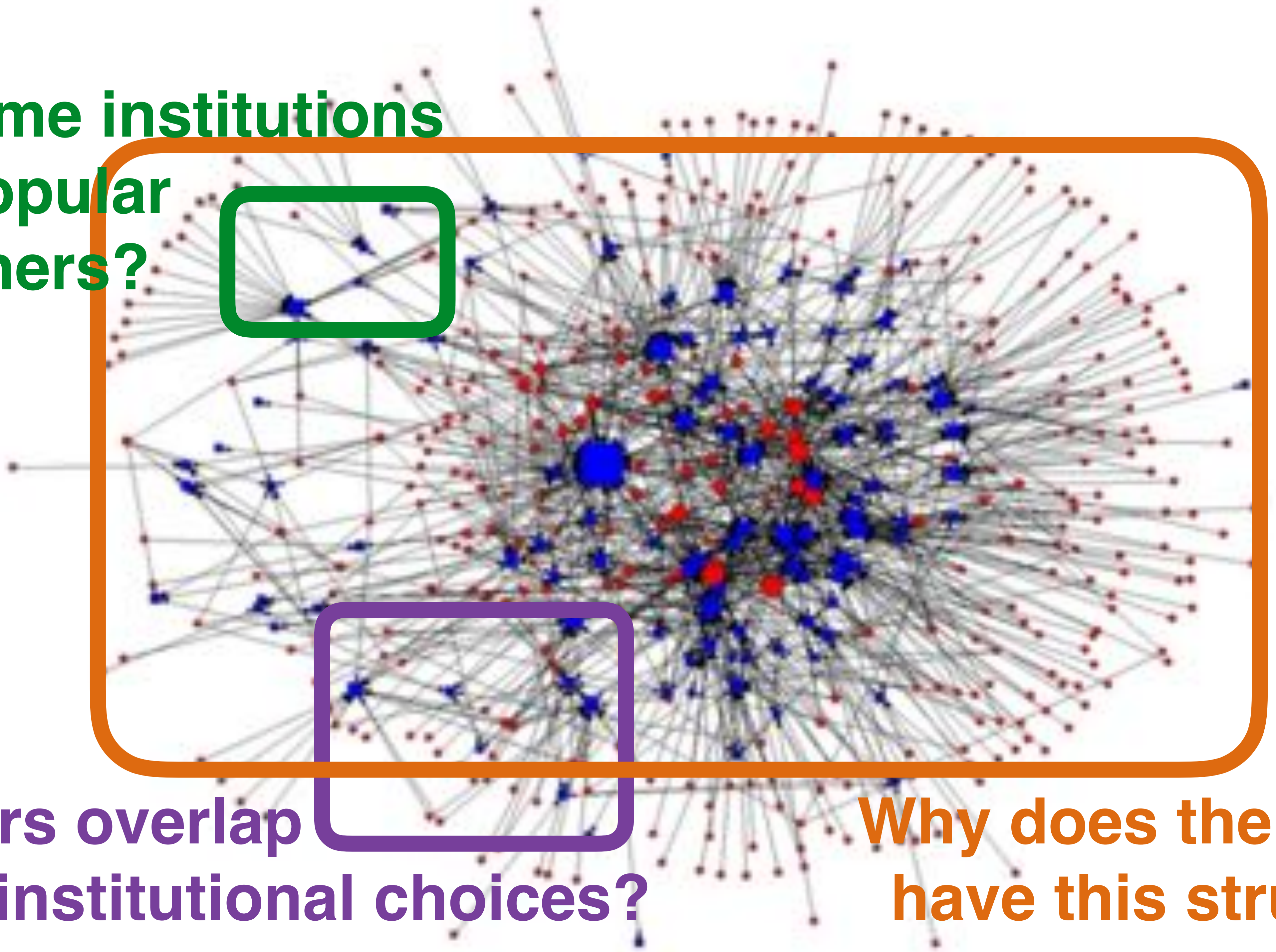
Social Networks Theories and Methods

James Hollway

Bay area network of actors and water management institutions



Questions?



Why some institutions more popular than others?

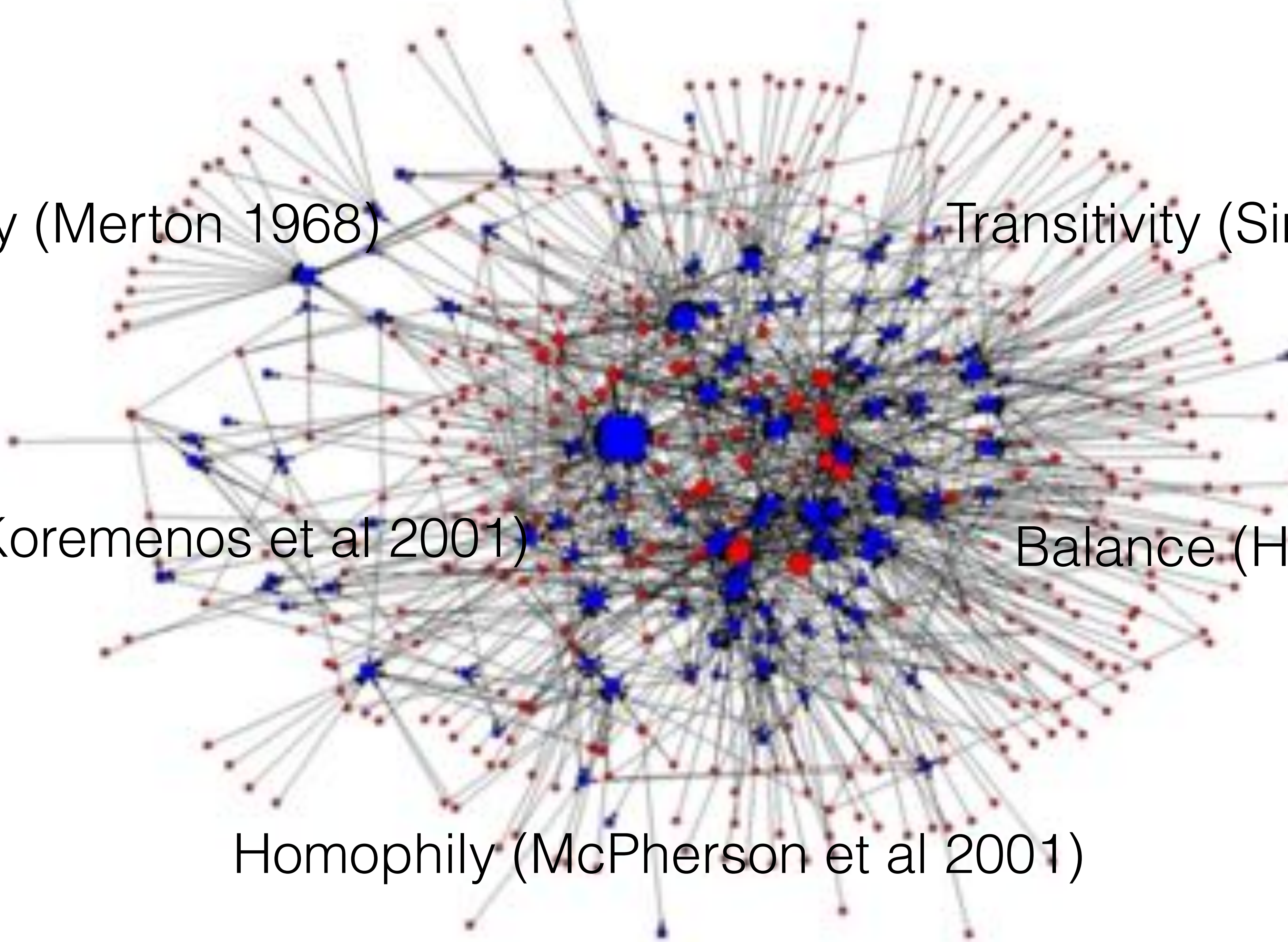
Do actors overlap in their institutional choices?

Why does the network have this structure?

Answers?

Random (Erdős and Rényi 1960)

Propinquity (Festinger et al 1950)



Popularity (Merton 1968)

Transitivity (Simmel 1902)

Design (Koremenos et al 2001)

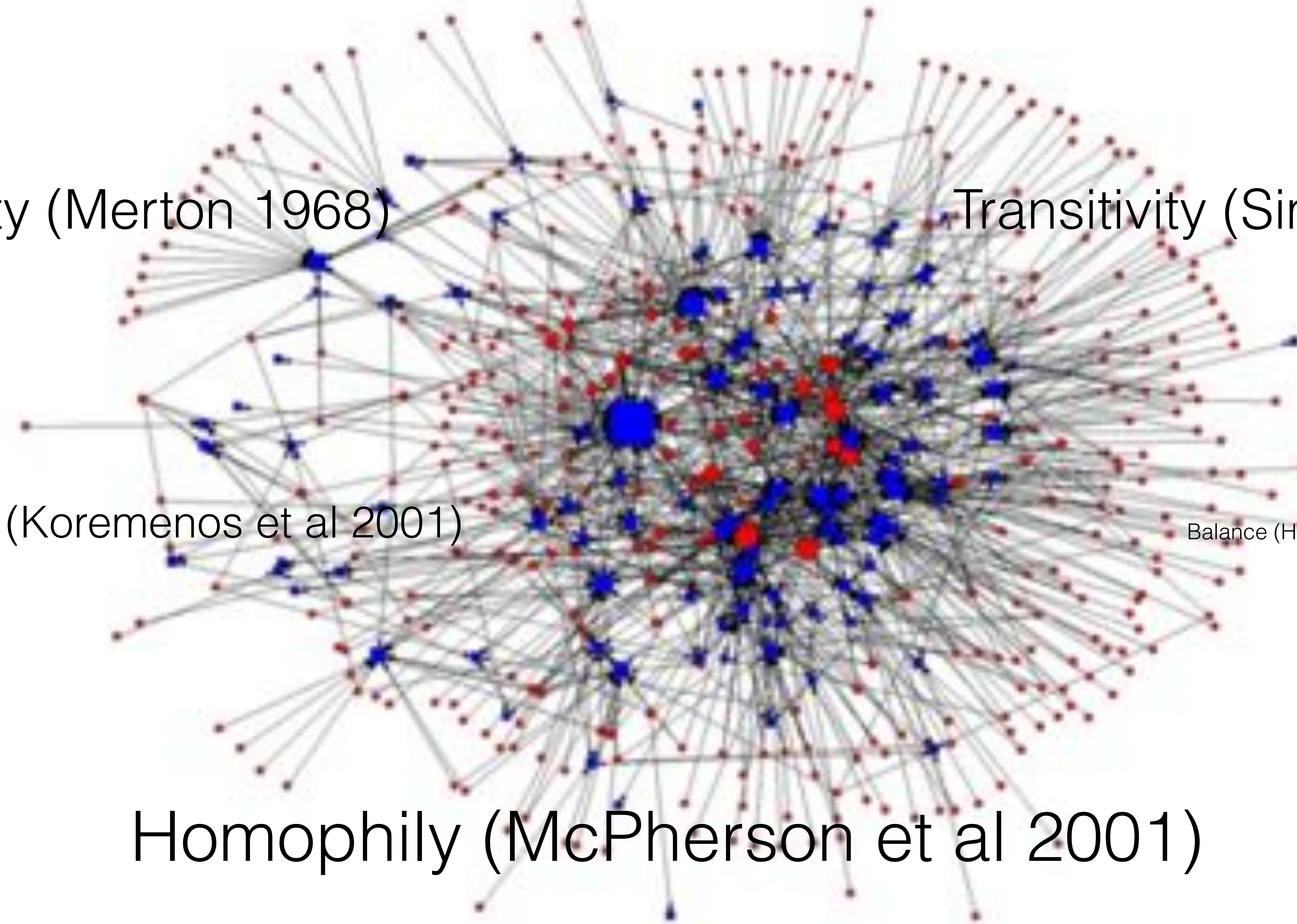
Balance (Heider 1958)

Homophily (McPherson et al 2001)

Could be a bit of all of these, maybe some more than others?

Random (Erdős and Rényi 1960) Propinquity (Festinger et al 1950)

Popularity (Merton 1968) Transitivity (Simmel 1902)



Design (Koremenos et al 2001) Balance (Heider 1958)

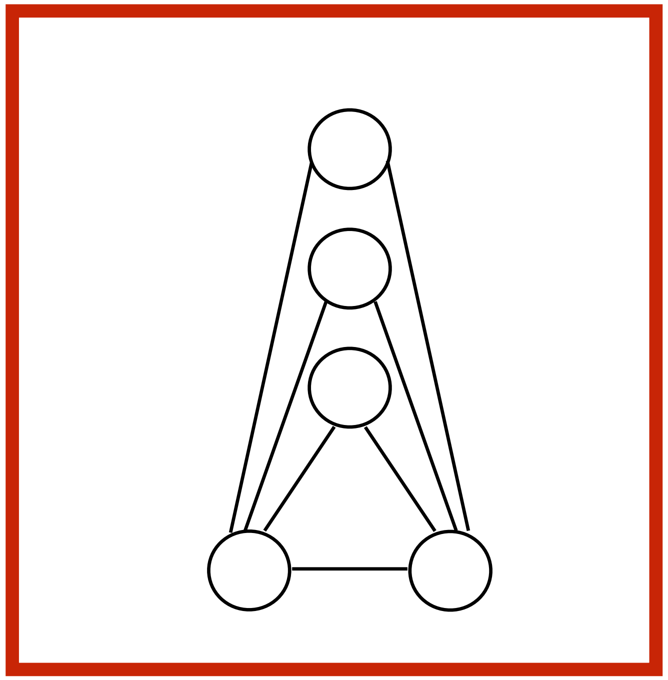
Homophily (McPherson et al 2001)

Strategy…

- 1. Get out all plausible **ingredients** (Effects)
- 2. Use special **oven** (ERGM)
- 3. Test look and **taste** (Diagnostics)

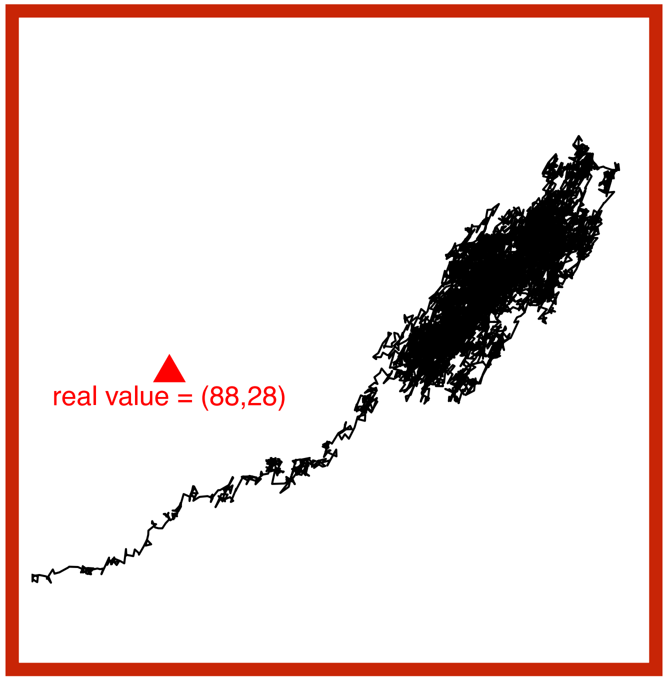
ERGM

Ingredients



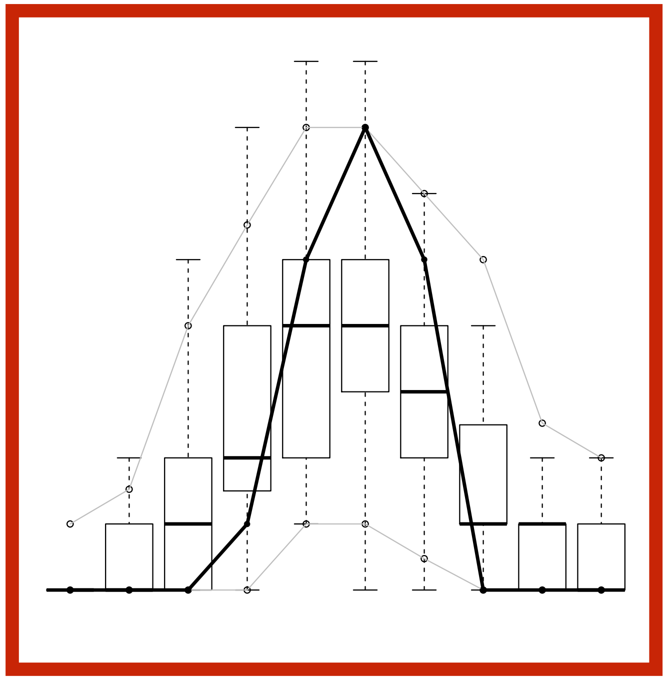
Explore various effects available

Oven



Understand model intuition and estimation

Taste Test



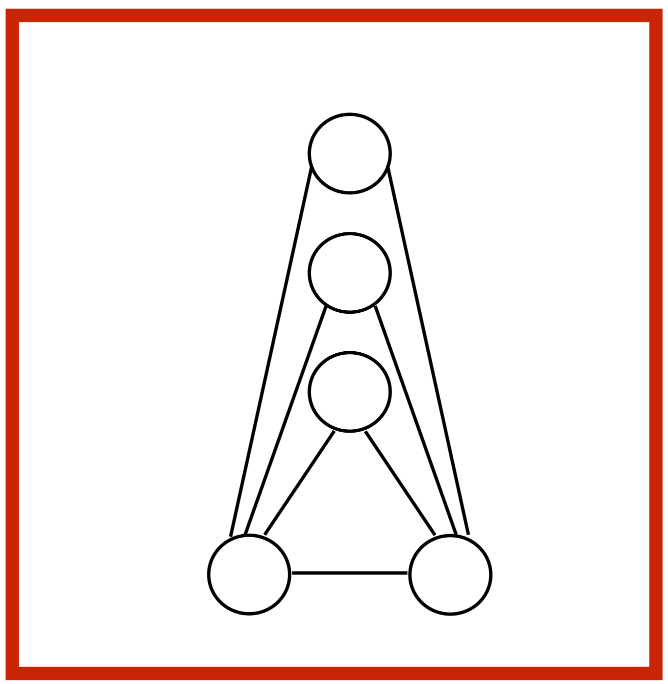
Recognise when a model converges and fits

ERGM

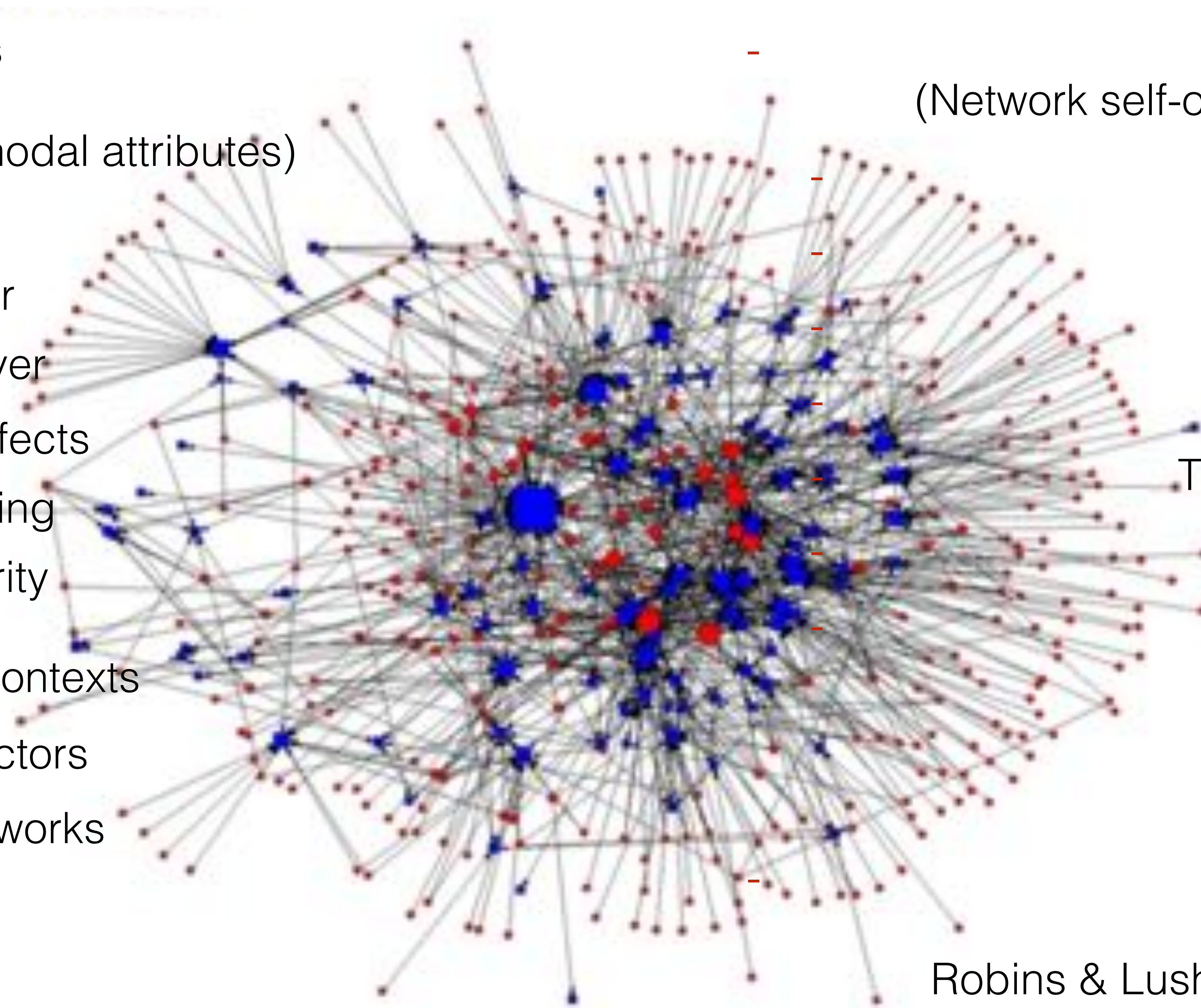
Effects

Model

Diagnostics



How & Why Ties Form?



- Randomness
- Covariates (nodal attributes)
  - Monadic
    - Sender
    - Receiver
  - Dyadic effects
    - Matching
    - Similarity
- Exogenous contexts
  - Spatial factors
  - Other networks

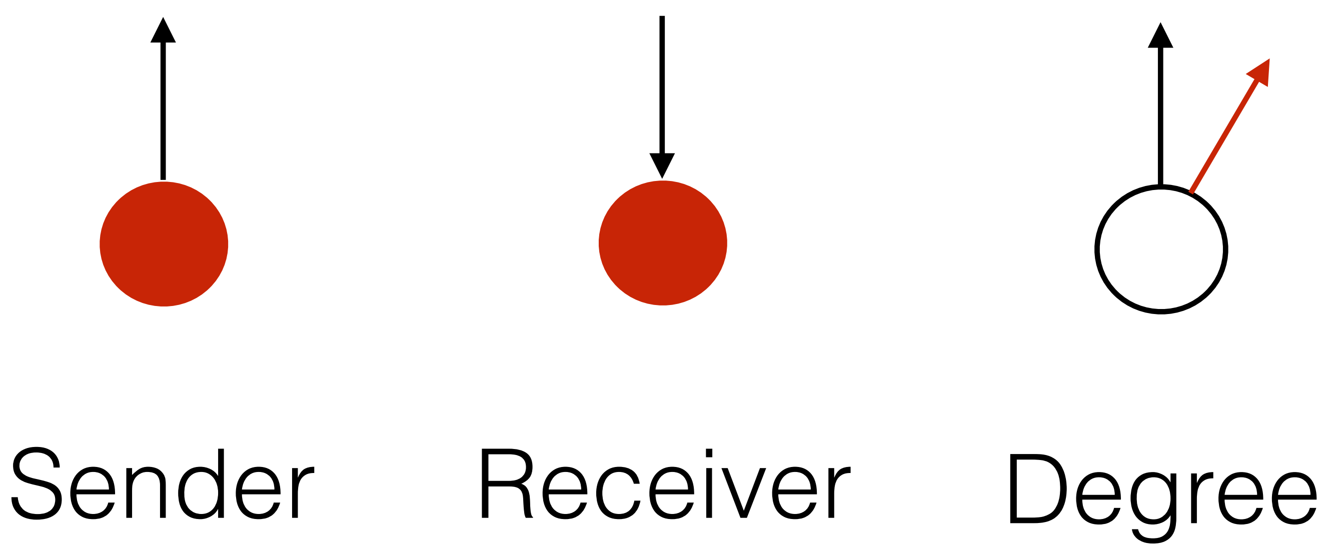
- Structural (Network self-organization)
  - Activity
  - Popularity
  - Reciprocity
  - Transitivity
  - Three-Cycles
  - Four-Cycles
  - Brokerage

Robins & Lusher 2012: 23

Levels of Dependence

- Independence
    - Logistic regression
    - Density, attributes

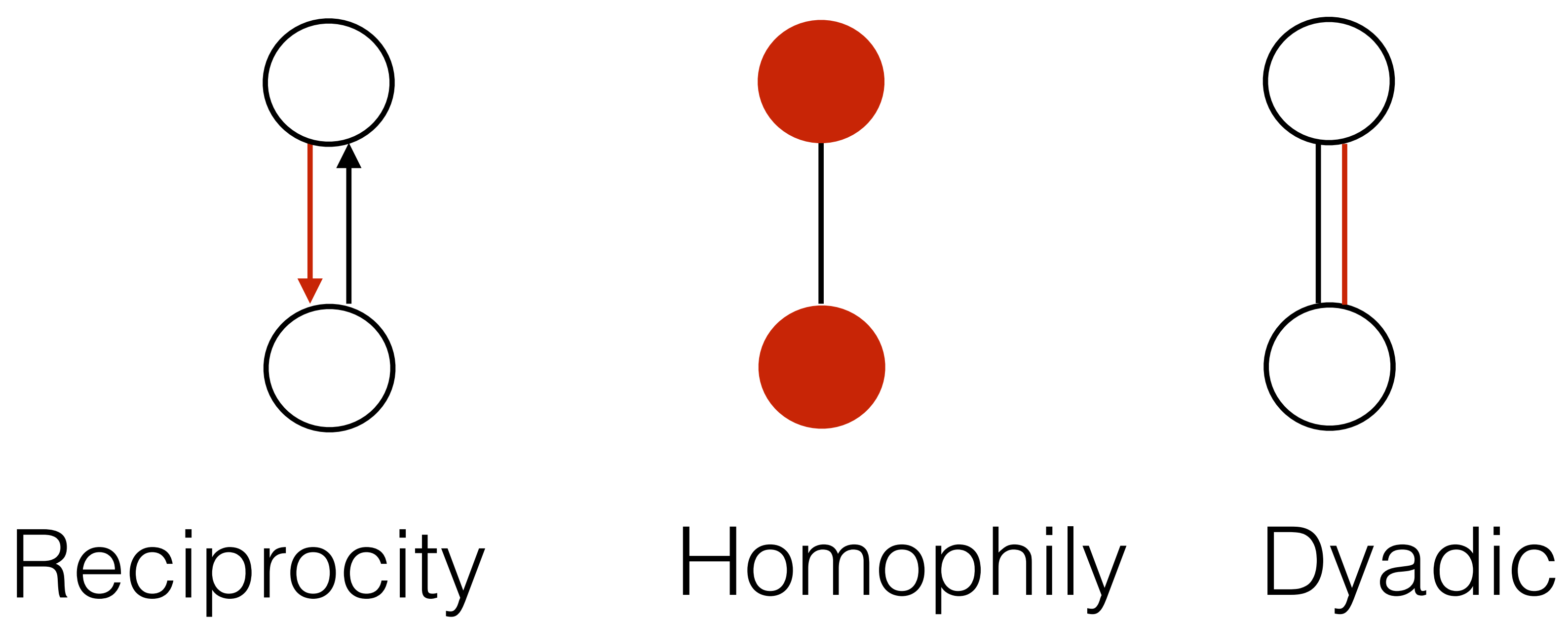
If we believe that particular attributes
are responsible for ties,
then include counts of



Levels of Dependence

- Independence
  - Logistic regression
  - Density, attributes
- Dyad-independence
  - p1 (Holland & Leinhardt 1981; Fienberg & Wasserman 1979; 1981)
  - Reciprocity, homophily

If we believe that reciprocity or homophily are responsible for ties, then include counts of

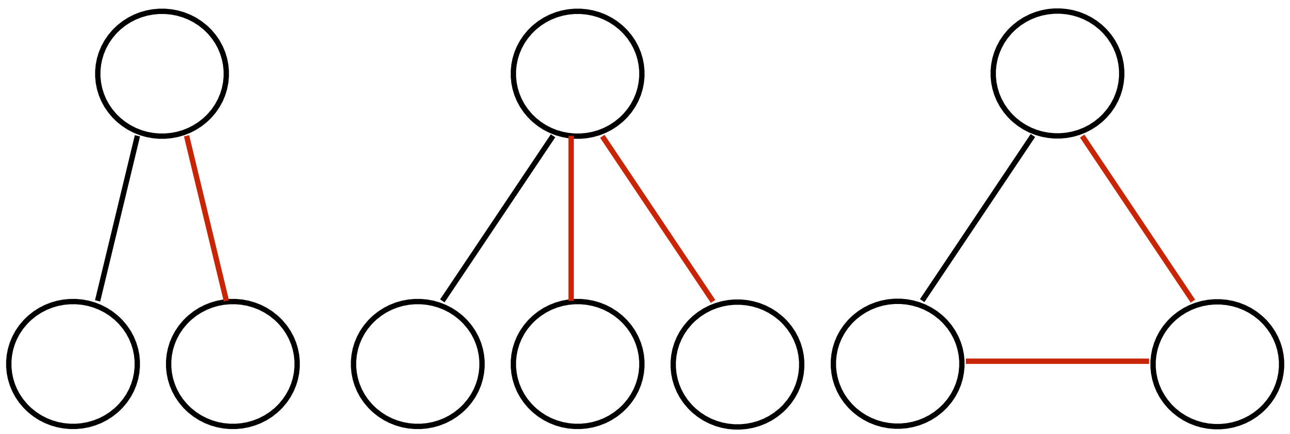


Levels of Dependence

- Independence
  - Logistic regression
  - Density, attributes
- Dyad-independence
  - p1
  - Reciprocity, homophily
- Markov-dependence
  - p*/ERGM (Frank & Strauss 1986)
  - Transitivity, popularity (Pattison & Wasserman 19999; Robins, Pattison & Wasserman 1999; Wasserman & Pattison 1996)

If we believe that popularity or transitivity are responsible for ties, then include counts of

may depend on one step removed…



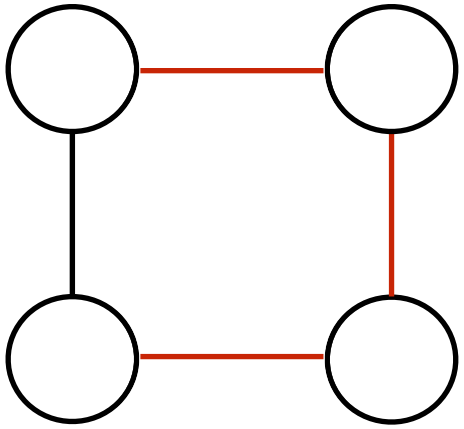
k-Stars

Triangles

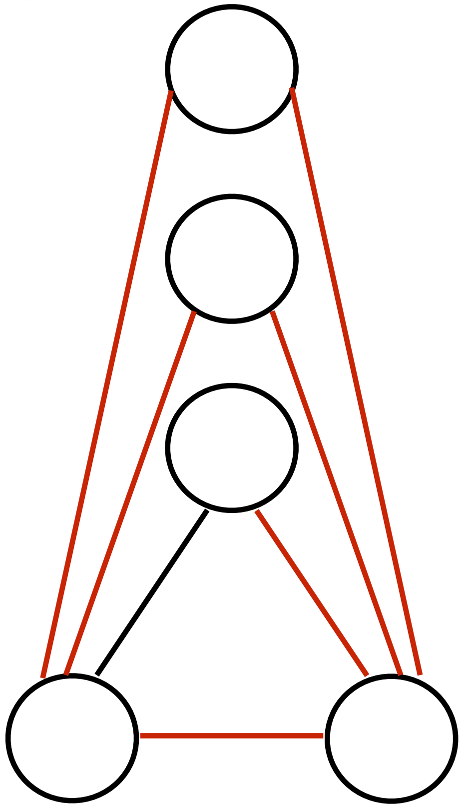
Levels of Dependence

- Independence
  - Logistic regression
  - Density, attributes
- Dyad-independence
  - p1
  - Reciprocity, homophily
- Markov-dependence
  - p*/ERGM
  - Transitivity, popularity
- Social circuit dependence
  - New specifications (Snijders et al 2006; Hunter & Handcock 2006)
  - Geometrically weighted edgewise shared partners (GWESP), four-cycles

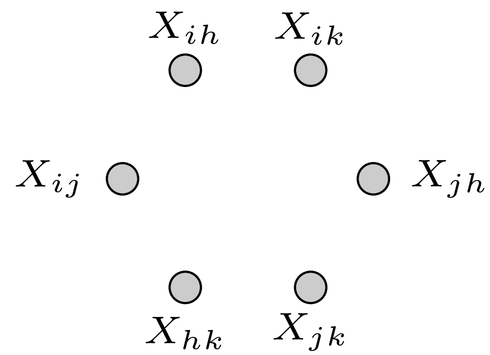
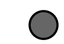
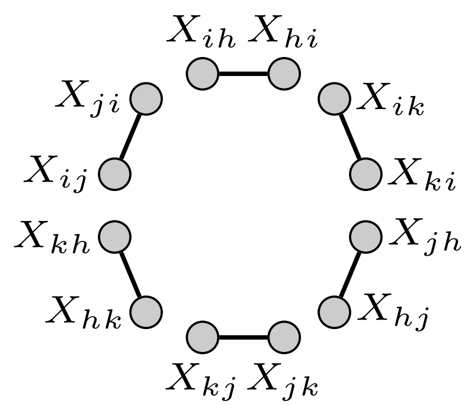
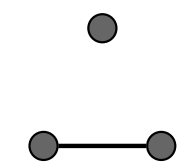
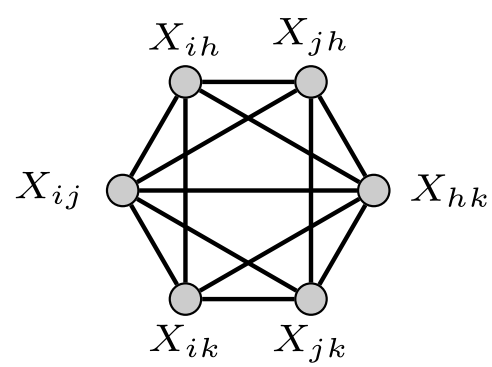
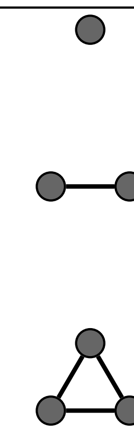
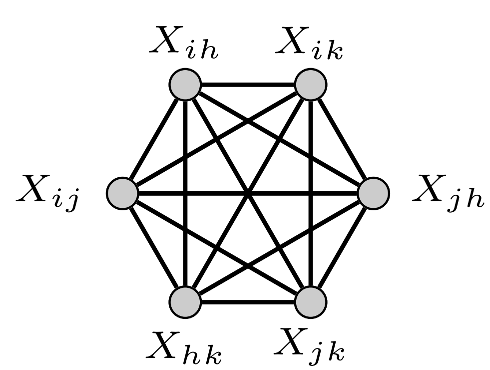
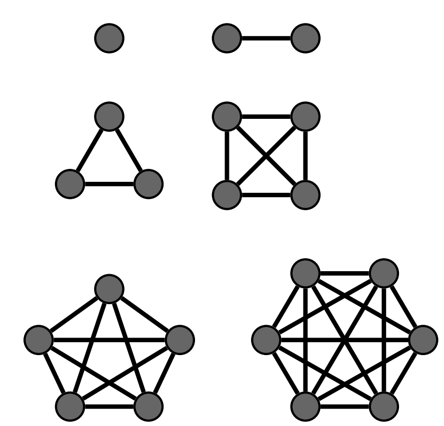
If we believe that ties are coordinated or that clustering aggregates, then include counts of

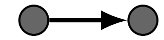
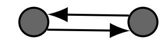
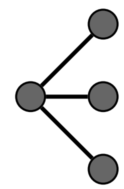
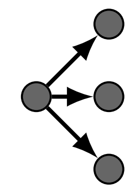
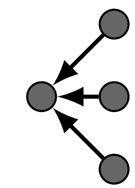
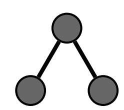
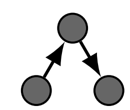
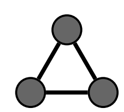
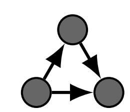
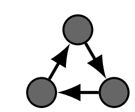
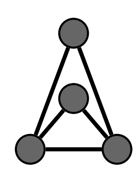
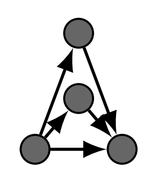
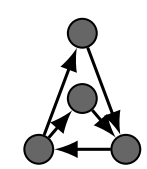
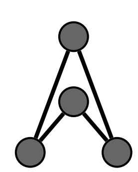
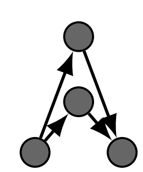


Four Cycles



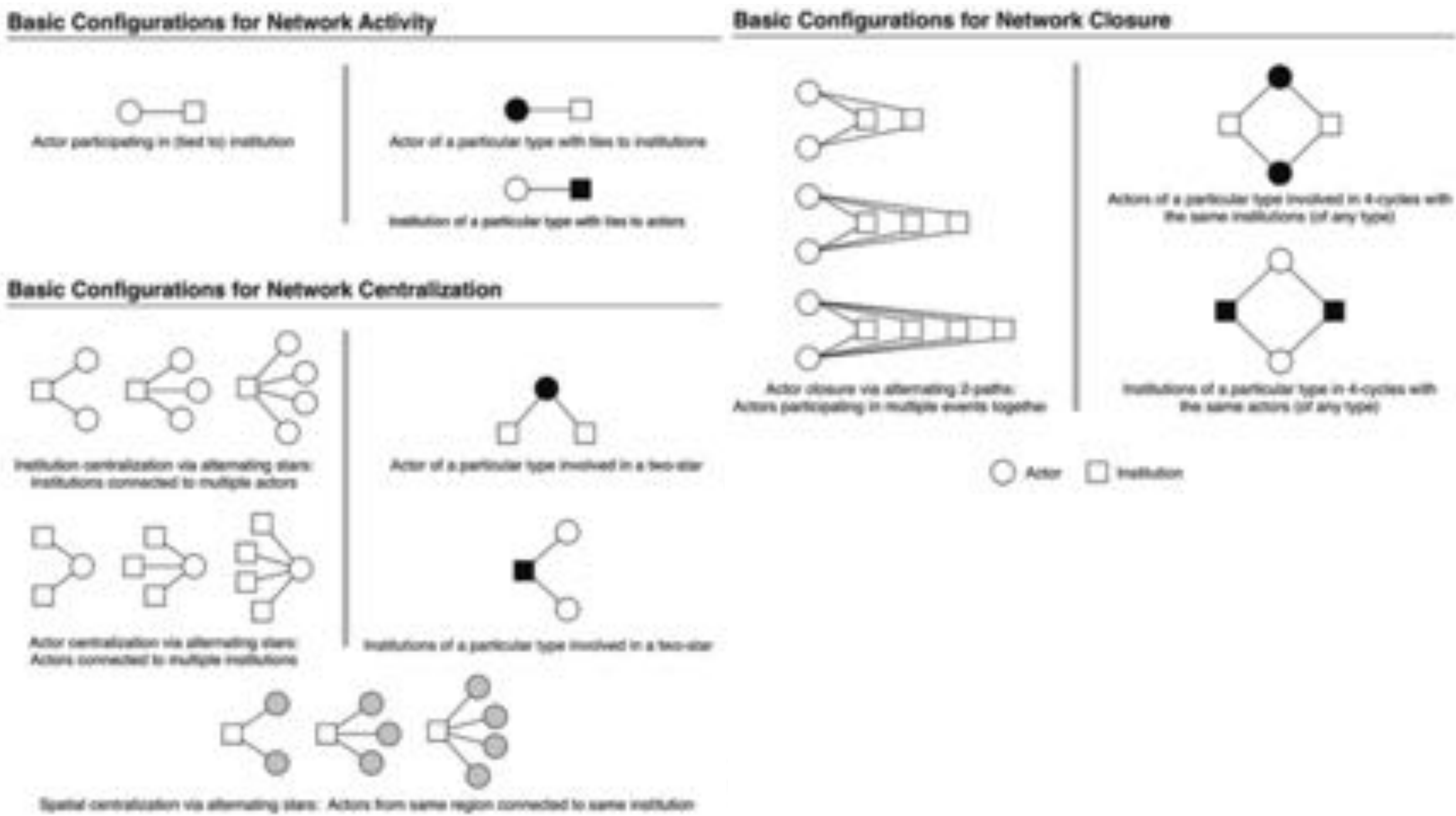
GWESP

Dependence assumption and corresponding model	Dependence graph D	Cliques of D
Independence $X_{ij} \perp\!\!\!\perp X_{hk}, \ \forall\ i, j, h, k \in \mathcal{N}$ Bernoulli random graph models		
Dyadic Dependence $X_{ij} \not\perp\!\!\!\perp X_{hk}, \ \forall\ \{i, j\} = \{h, k\}$ Dyadic dependence models		
Markov Dependence $X_{ij} \not\perp\!\!\!\perp X_{hk}\ if\ \{i, j\} \cap \{h, k\} \neq \emptyset$ Markov graphs		
Partial conditional dependence E.g., social circuit dependence $X_{ij} \not\perp\!\!\!\perp X_{hk}\ if\ X_{ih} = X_{jk} = 1\ or$ $X_{ik} = X_{jh} = 1$ Exponential random graph models		

Network property	Statistics		
	Undirected	Directed	
Density	 Edges	 Arcs	
Reciprocity		 Mutual dyads	
Degree distribution	 Stars	 Out-stars	 In-stars
Connectivity	 Two-paths	 Two-paths	
Closure	 Triangles	 Transitive triads	 3-cycles
Clustering of triangles	 Alternating-k-triangles	 Alternating-transitive-triangles	 Alternating-3-cycles
Clustering of 2-paths	 Alternating-k-paths	 Alternating-k-twopaths	

Amati, Viviana, Alessandro Lomi, and Antonietta Mira.
2018. "Social Network Modeling."
Annual Review of Statistics and Its Application 5 (1): 343–69.

Relevant statistics in the water management case



Notes on ingredients

- Start with most basic effects (e.g. density)

- Add effects from increasing levels of dependence (e.g. Markov, social circuit)

- Always include more fundamental forms from within more complex configurations (e.g. monadic before homophily, degree before closure)

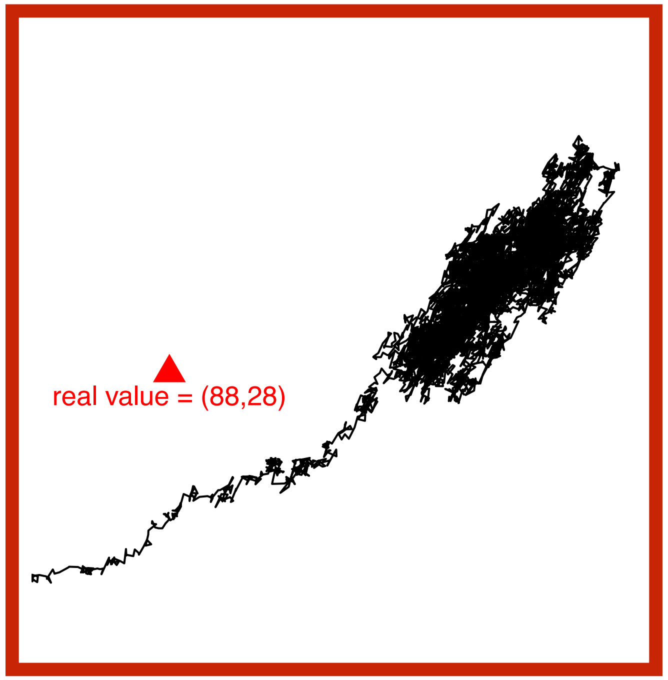
- Often useful to contrast structural-only models with with covariates-added models

ERGM

Effects

Model

Diagnostics

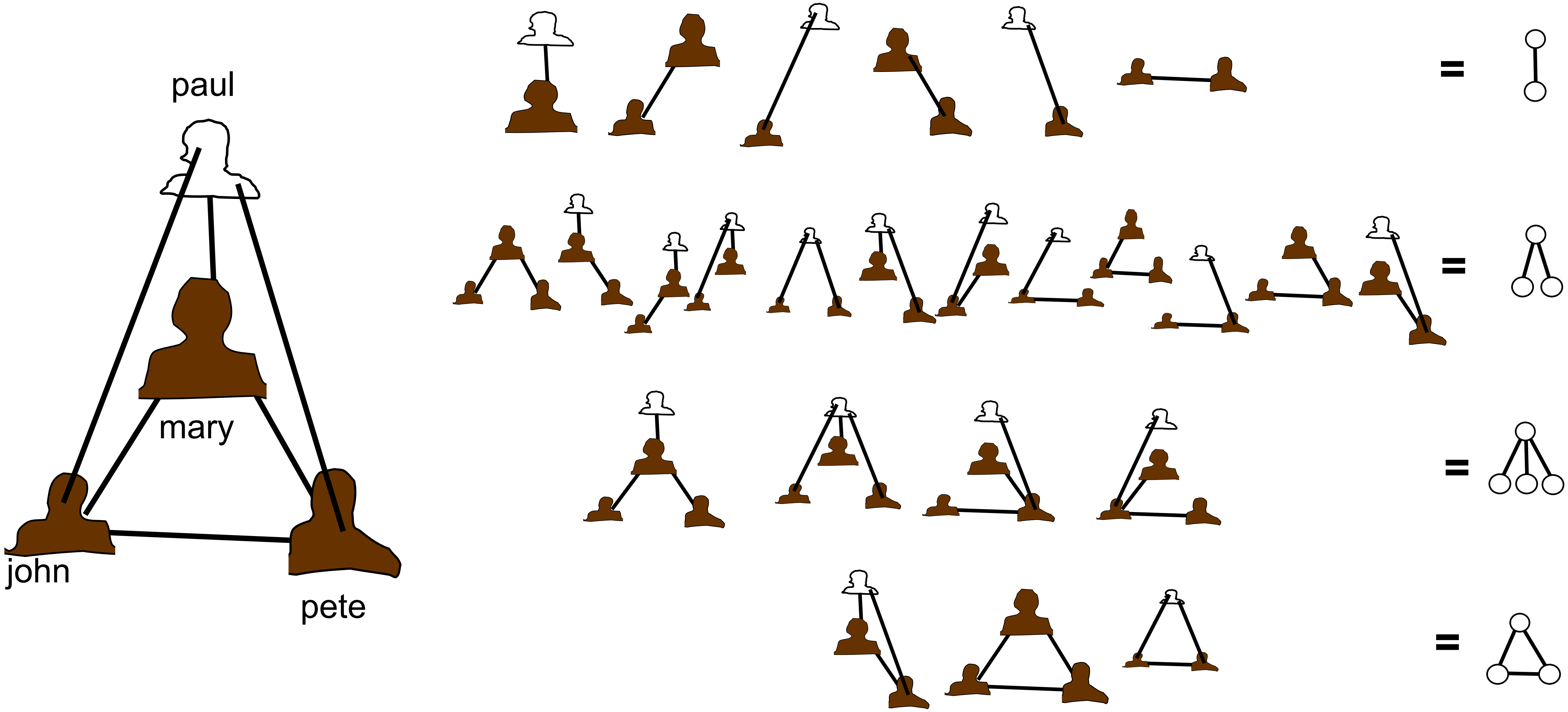


Nota bene..

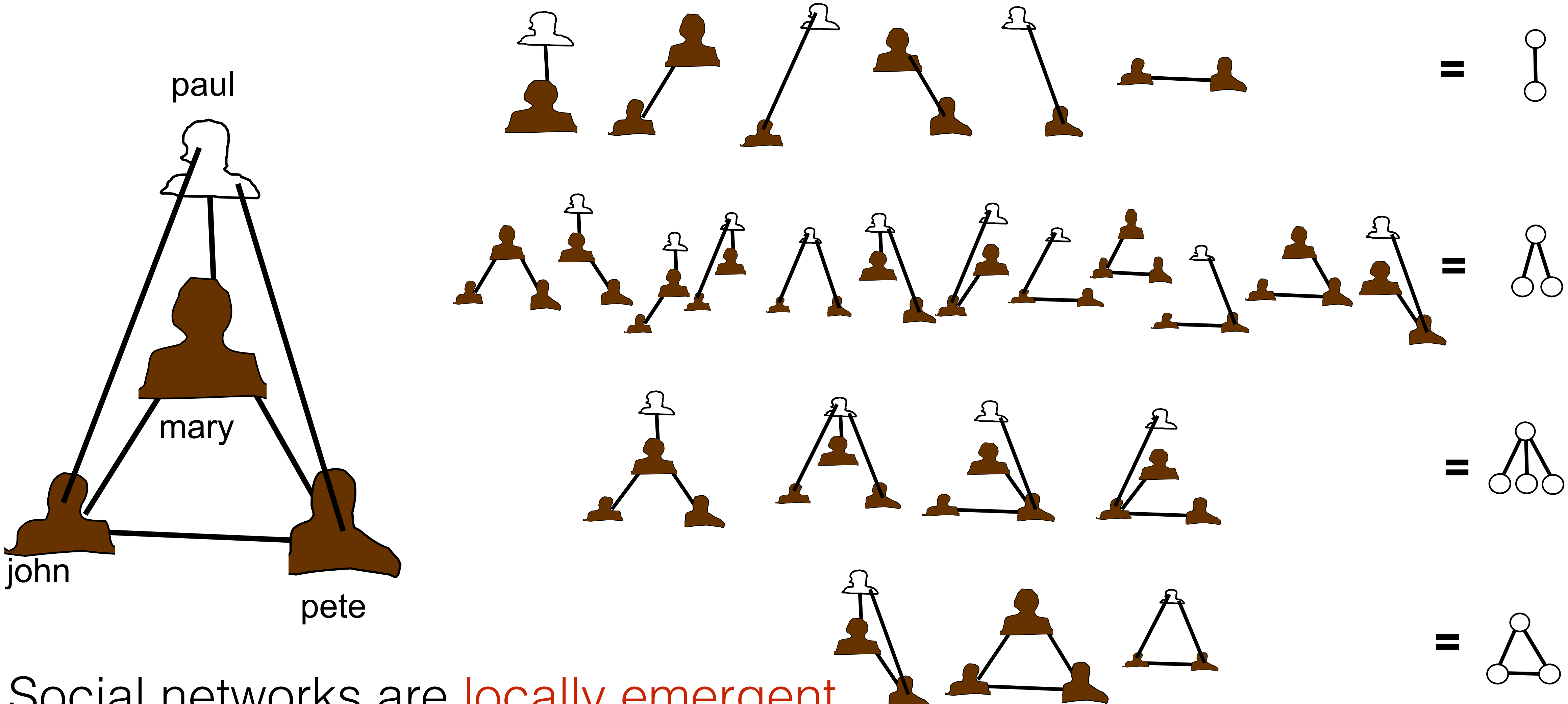
- We'll cover a lot of ground here
  - Some vocabulary may be unfamiliar
- Don't worry if you don't understand everything
  - Focus on getting the big picture
  - Statnet puts a lot of this behind the curtain, so you often don't have to deal with this
  - But the details matter when a model is problematic
- So: don't be afraid to ask questions! (today, in practical, in office hours, on Moodle, during consultation sessions)

What are ERGMs?

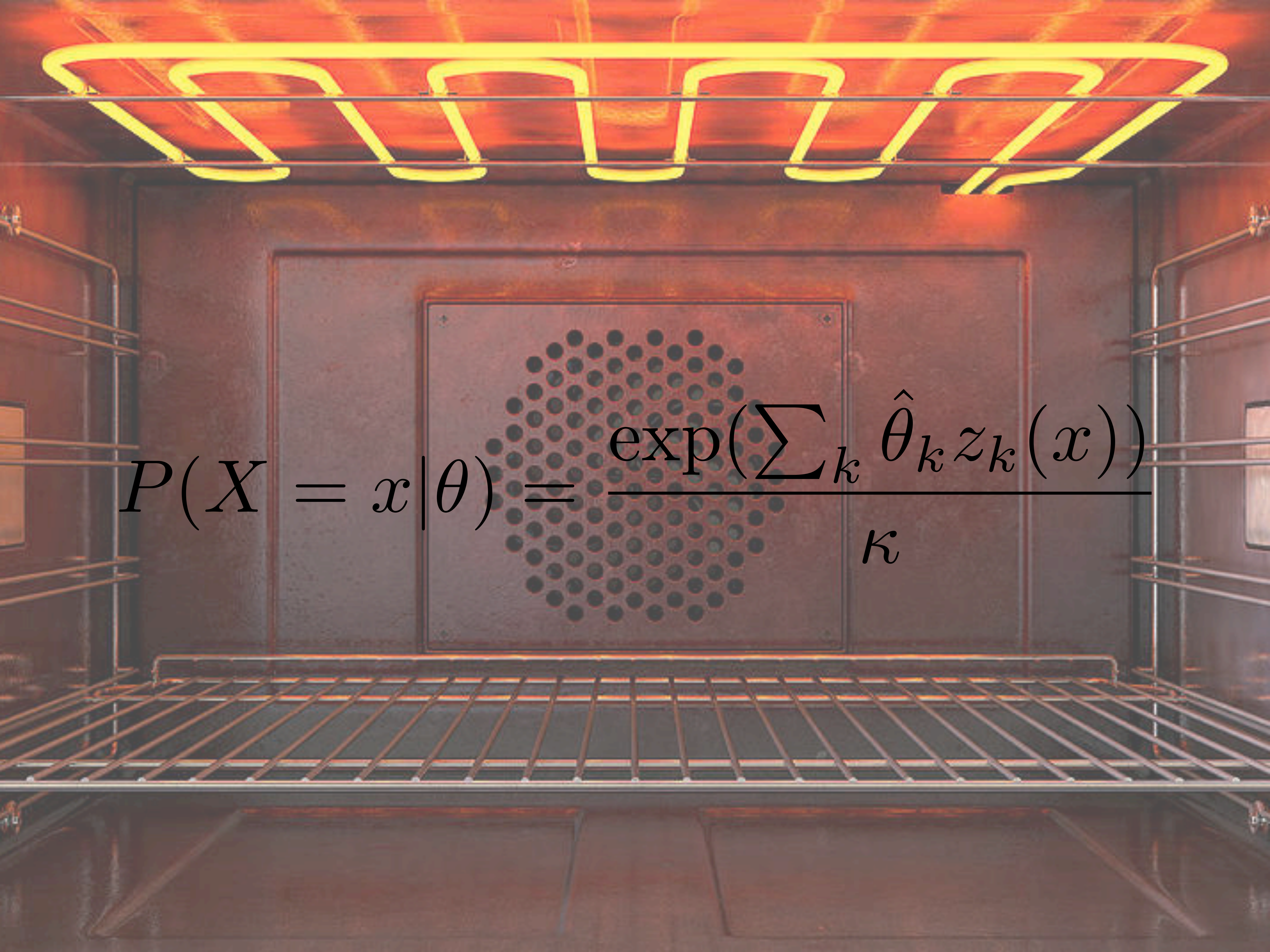
- ERGMs (pronounced *örgums* - this is <u>important</u>)
  - "are statistical models for network structure, permitting inferences about how network ties are patterned" (Robins & Lusher 2012)
- Since the random graphs in our model form an exponential family, we call the model an *exponential (family) random graph model* (ERGM, EFRGM would be *too* cumbersome!)



- Aim to explain observed network ties or structure as function of “ingredients” you put in to it
    - these ingredients can be exogenous (monadic and dyadic covariates), or
    - endogenous (structural effects, like activity/popularity or transitivity)
- Once you have a model of how much of each “ingredient” to put in, we can use this model to:
    - predict ties (e.g. how likely is it that Paul and Ringo have a tie?)
    - simulate networks (e.g. how would the cake look if I added more butter?)



- Social networks are locally emergent
- Local configurations homogenous
- Network configurations that appear more often than by chance and over attribute explanations evince structural processes
- Multiple processes can operate simultaneously
- Social networks are structured, yet stochastic



$$P(X = x|\theta) = \frac{\exp(\sum_k \hat{\theta}_k z_k(x))}{\kappa}$$

The Secret Sauce

$$P(X = x | \theta) = \frac{\exp(\sum_k \hat{\theta}_k z_k(x))}{\kappa}$$

- Probability of network *x* is given by
  - a sum of network statistics (*z*)
    - expresses counts of network configurations (e.g. counts of reciprocal, transitive, or homophilic subgraphs)
  - that is weighted (θ)
    - expresses the importance of each configuration
  - inside an exponential (*e*)
    - this is an exponential-*family* random graph model, so that probabilities [0,1]
  - and is normalised (κ)
    - over all possible graphs of the same size (x' in X)

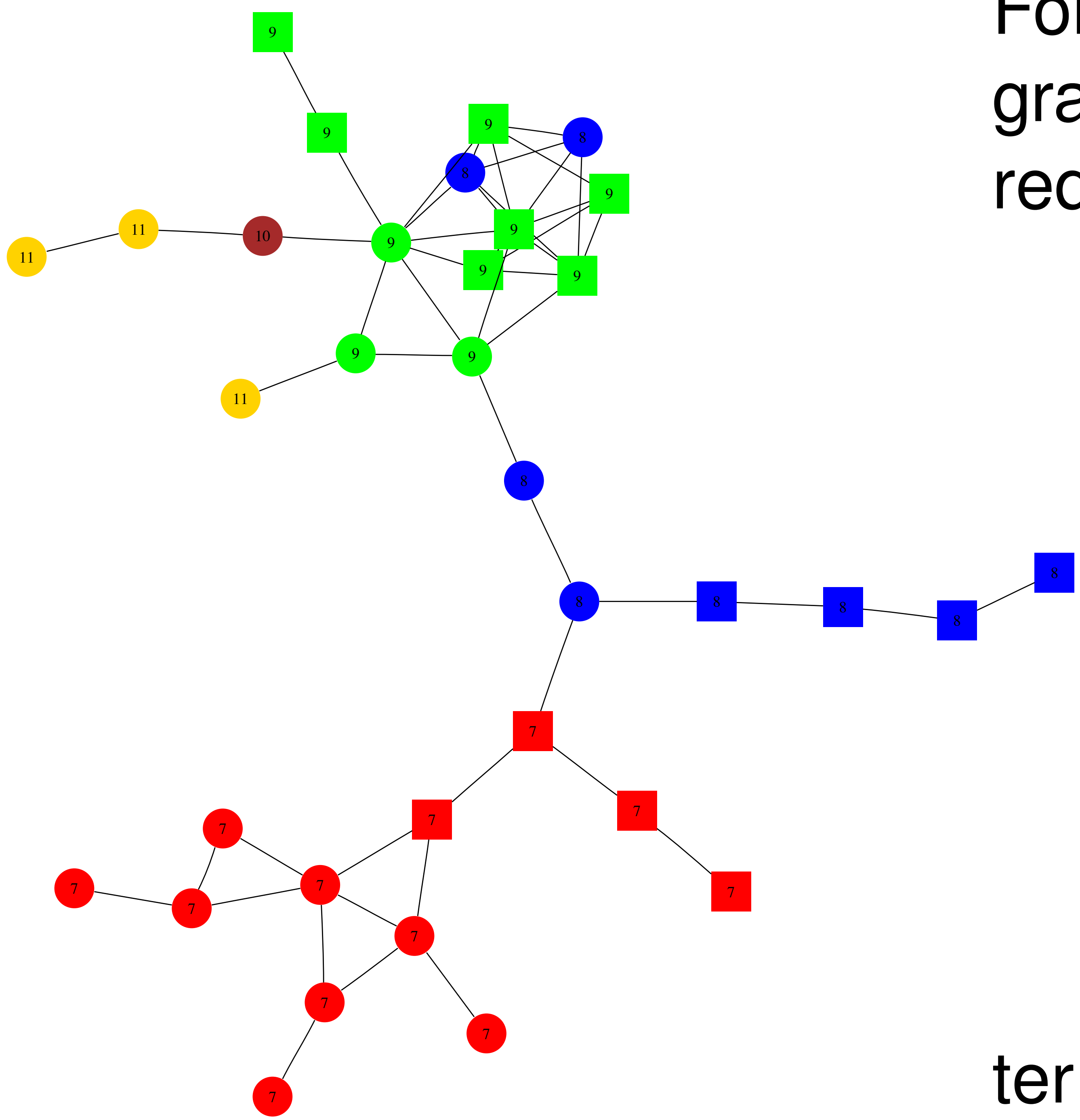
Problem: Oh κ!

- Ideally use maximum likelihood estimation, $L(\theta|x)$, directly, to find estimates of θ that make x most likely
- But remember κ?

$$\kappa = \sum_{x' \in X} \exp\left(\sum_k \theta_k z_k(x')\right)$$

  - Directed, binary network of n nodes has $2^{n(n-1)}$ states
  - Really, really large, making κ not computable except for very small graphs
- How large?…

How large?



For this undirected, 34-node graph, computing $c(\theta)$ directly requires summation of

*7,547,924,849,643,082,704,483,
109,161,976,537,781,833,842,
440,832,880,856,752,412,600,
491,248,324,784,297,704,172,
253,450,355,317,535,082,936,
750,061,527,689,799,541,169,
259,849,585,265,122,868,502,
865,392,087,298,790,653,952*

terms.

Moment mal

MLE

$$E_\theta(z(X)) = z(x_{obs})$$

- Inferential goal is to centre the distribution of statistics over those of the observed network
  - "fitting a model that gives maximal support to the data"

Moment equation

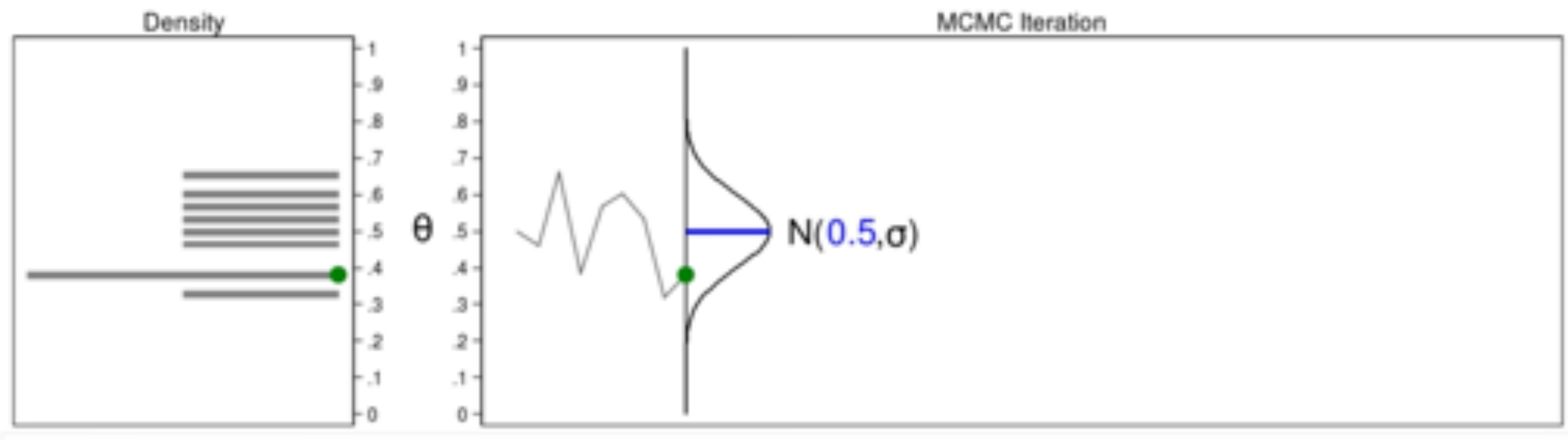
$$E_\theta(z(X)) - z(x_{obs}) = 0$$

- We define a distribution as centred when values of the statistics from the distribution are the same as those observed on average

Oκ, how do we get κ?

- Markov Chain Monte Carlo (MCMC)
  - Different variations available (Gibbs, Metropolis-Hastings)
- Main idea: Simulate a discrete-time Markov chain whose stationary distribution is the distribution we want to sample from

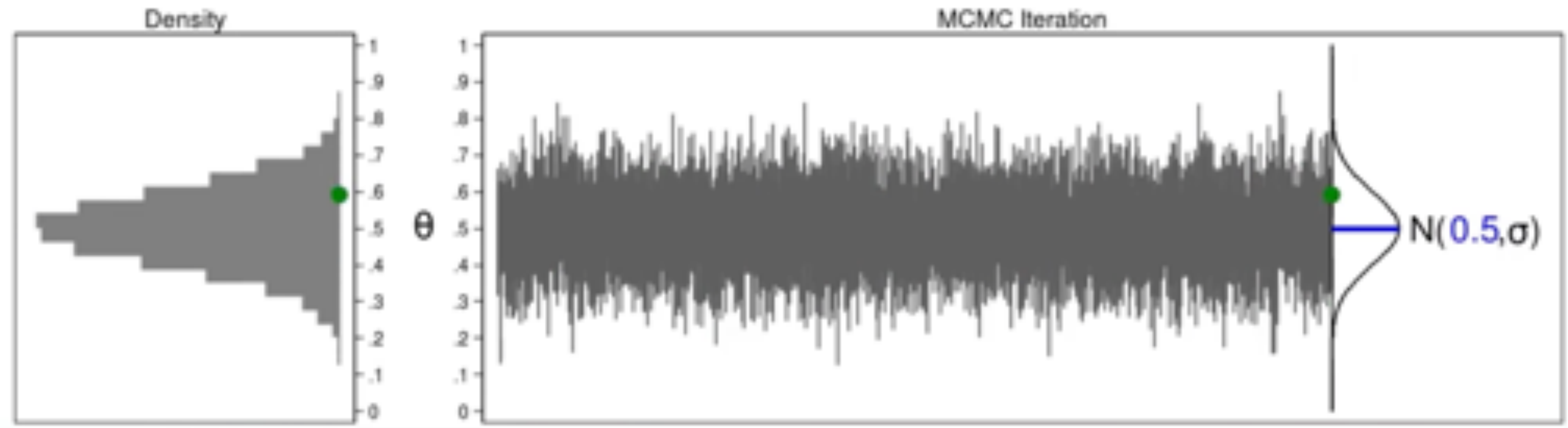
Markov Chain Monte Carlo (MCMC)



- refers to the part that relies on the generation of random numbers

- note that the distribution on the left resembles the distribution we are drawing from and that the proposal distribution does not move

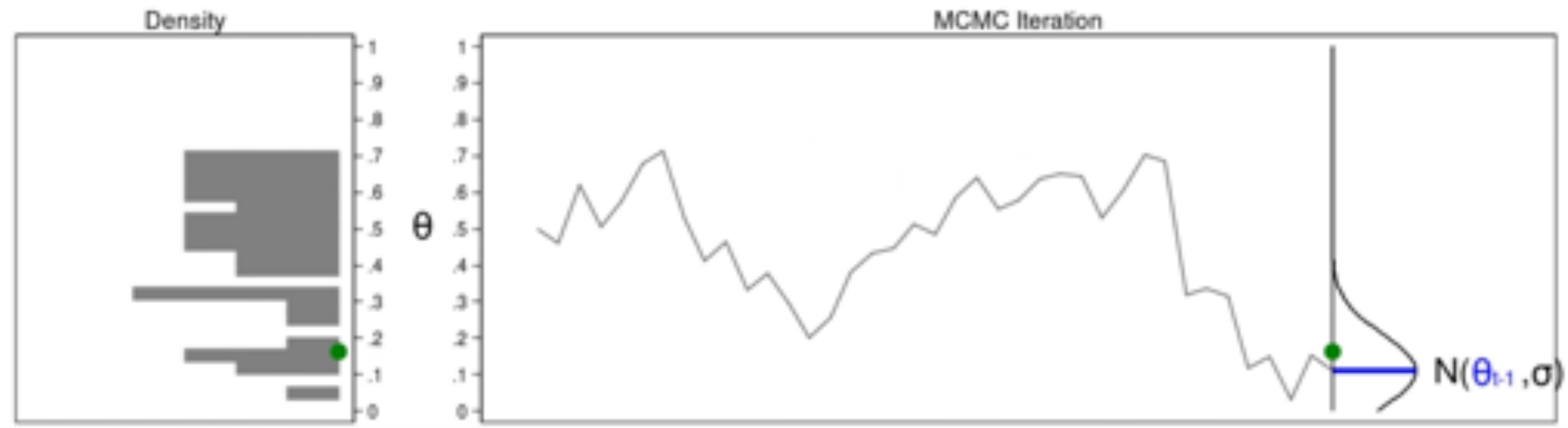
Markov Chain Monte Carlo (MCMC)



- refers to the part that relies on the generation of random numbers

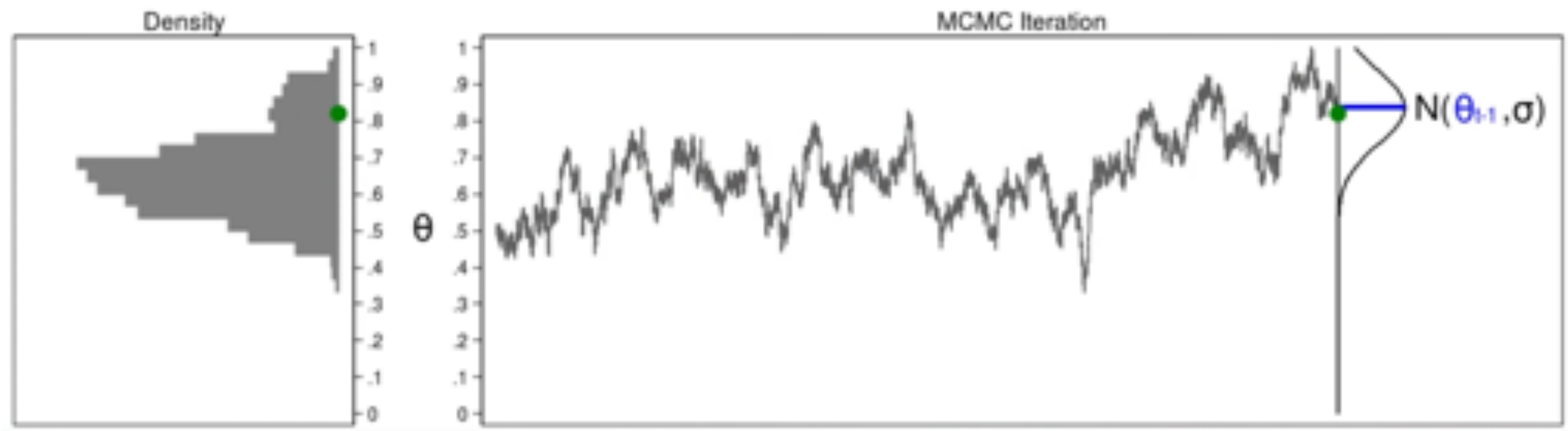
- note that the distribution on the left resembles the distribution we are drawing from and that the proposal distribution does not move

Markov Chain Monte Carlo (MCMC)



- is a sequence of numbers in which each number is dependent (only) on the previous number
- traceplot seems to wander like in a random walk

Markov Chain Monte Carlo (MCMC)



- is a sequence of numbers in which each number is dependent (only) on the previous number

- traceplot seems to wander like in a random walk

An underlying Markov chain

- The ERGM is also the stationary distribution of a Markov random walk with transition probabilities

$$p(x \to x^{i \leadsto j}; \theta) = \frac{1}{N(N-1)} \cdot \frac{\exp\left(\sum_k \theta_k z_k(x^{i \leadsto j})\right)}{\exp\left(\sum_k \theta_k z_k(x)\right) + \exp\left(\sum_k \theta_k z_k(x^{i \leadsto j})\right)}$$

- In theory, if we just let this random walk run long enough, it will approximate the stationary distribution and thus the ERGM for a given parameter θ

- In practice, this problem is again intractable

Sampling from the Markov chain to estimate $\hat{\theta}$

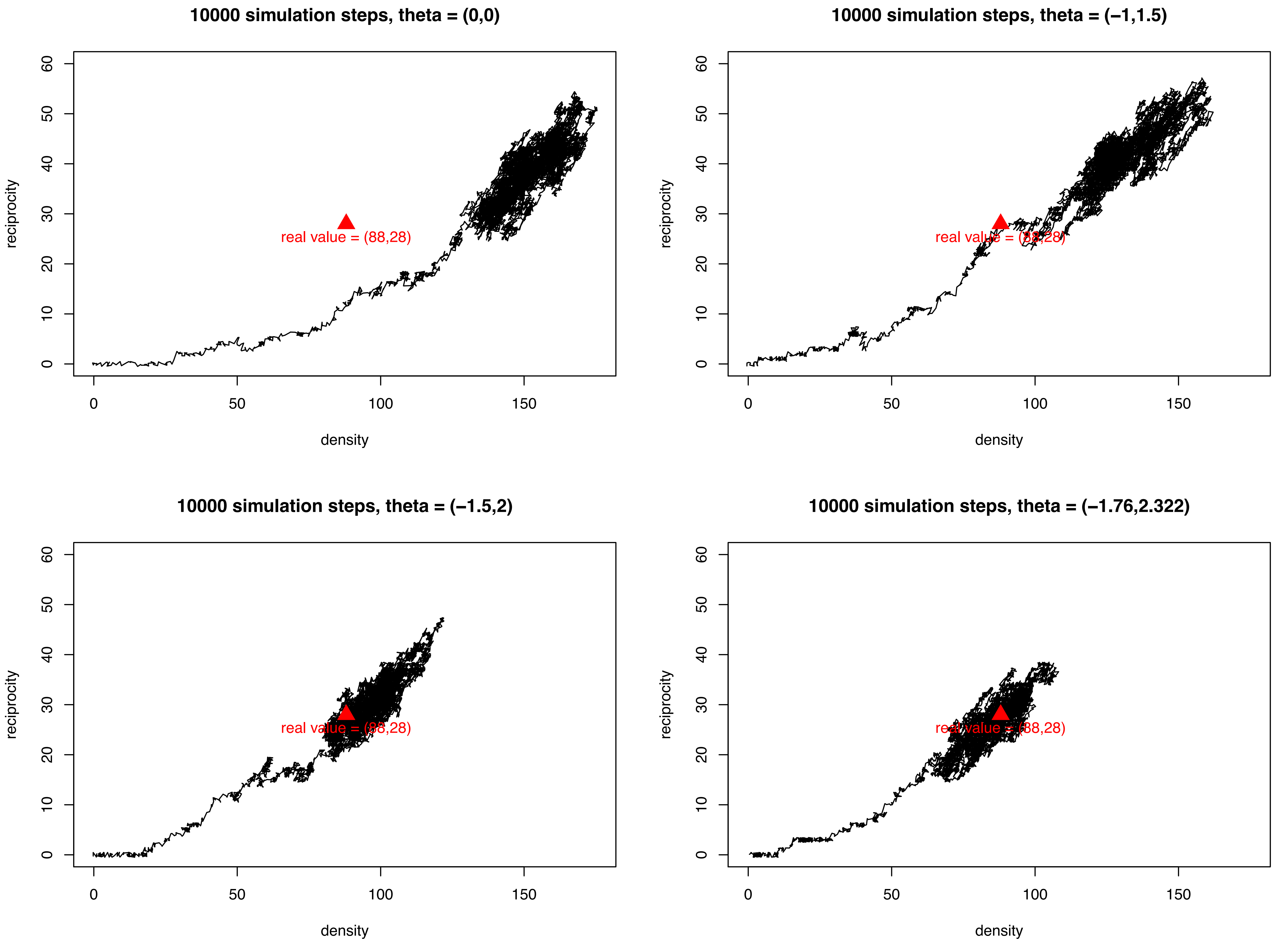
- However, we can use the Markov chain to simulate networks $x^{(1)}$, $x^{(2)}$, …, $x^{(M)}$ that are a good sample of the space of all networks

- We need to make sure that these simulated networks have a low autocorrelation and are representative for the sample space

- Calculate the sample equivalent of $E_{\hat{\theta}}(z(X))$

$$\bar{z}_\theta = \frac{1}{M}\left(z(x^{(1)}) + z(x^{(2)}) + \ldots + z(x^{(M)})\right)$$

- Check whether $\bar{z}_\theta - z(x_{obs}) = 0$
    - If yes, $\theta = \hat{\theta}$
    - If no, update $\hat{\theta}$



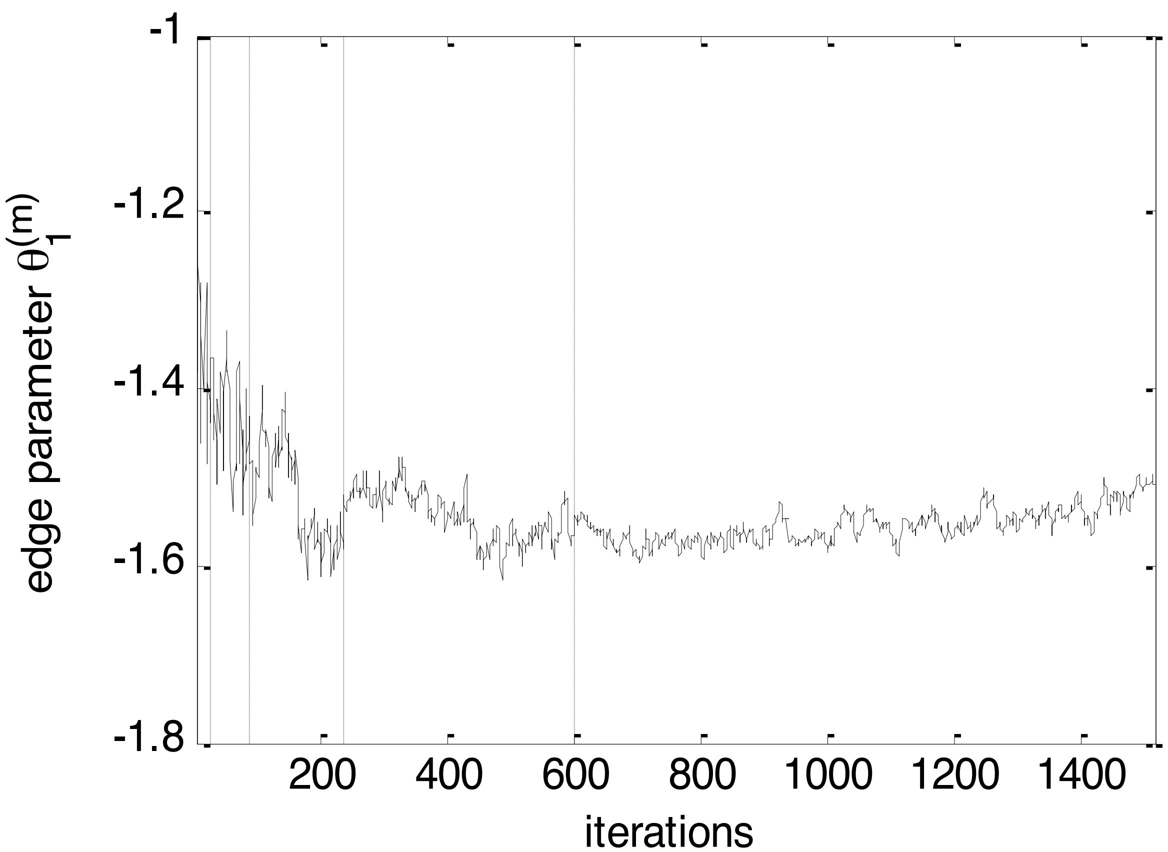
10000 simulation steps, theta = (0,0)
10000 simulation steps, theta = (−1,1.5)
10000 simulation steps, theta = (−1.5,2)
10000 simulation steps, theta = (−1.76,2.322)
real value = (88,28)
reciprocity
density

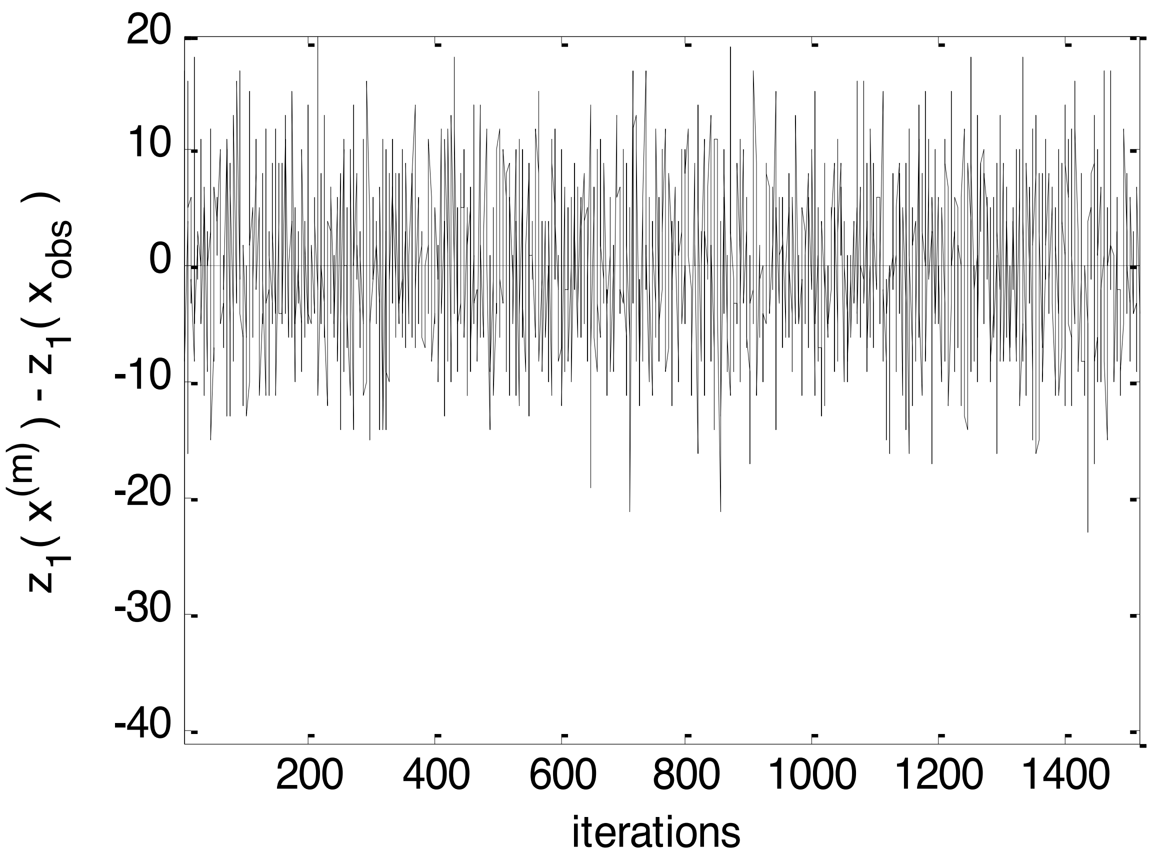

IN ORDER TO SAVE TIME, THE REMAINDER OF THIS MARRIAGE PROPOSAL WILL BE GENERATED USING MARKOV CHAINS.

Method I: Importance Sampling

- Handcock 2003, based on Geyer-Thompson 1992, see also
- Implemented in `statnet`
- Generates one sample of graphs at the beginning of the estimation that is supposedly representative of the complete model space
    - $\bar{z}_\theta$ is weighted to take into account that the sample is based on a Markov chain with a parameter starting value: $\tilde{\theta}$
    - To solve the method of moments equation, a chain of parameters $\tilde{\theta} = \theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(G)}$ is generated using Newton-Raphson updating
    - The algorithm can be restarted, with $\tilde{\theta} = \theta^{(G)}$ as a starting vector
    - It may be difficult (and time-consuming) to find a good starting value $\tilde{\theta}$

Method II: Stochastic Approximation





- Snijders 2002, based on Robbins-Monro 1951
- Implemented in `PNet`
- Three phases: initialization, estimation, convergence/standard errors

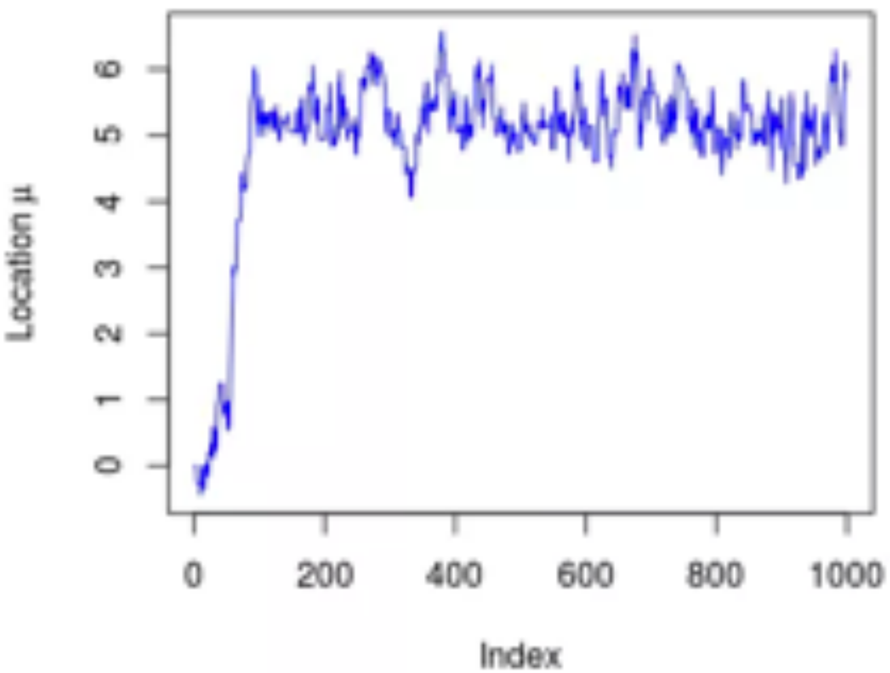
Convergence

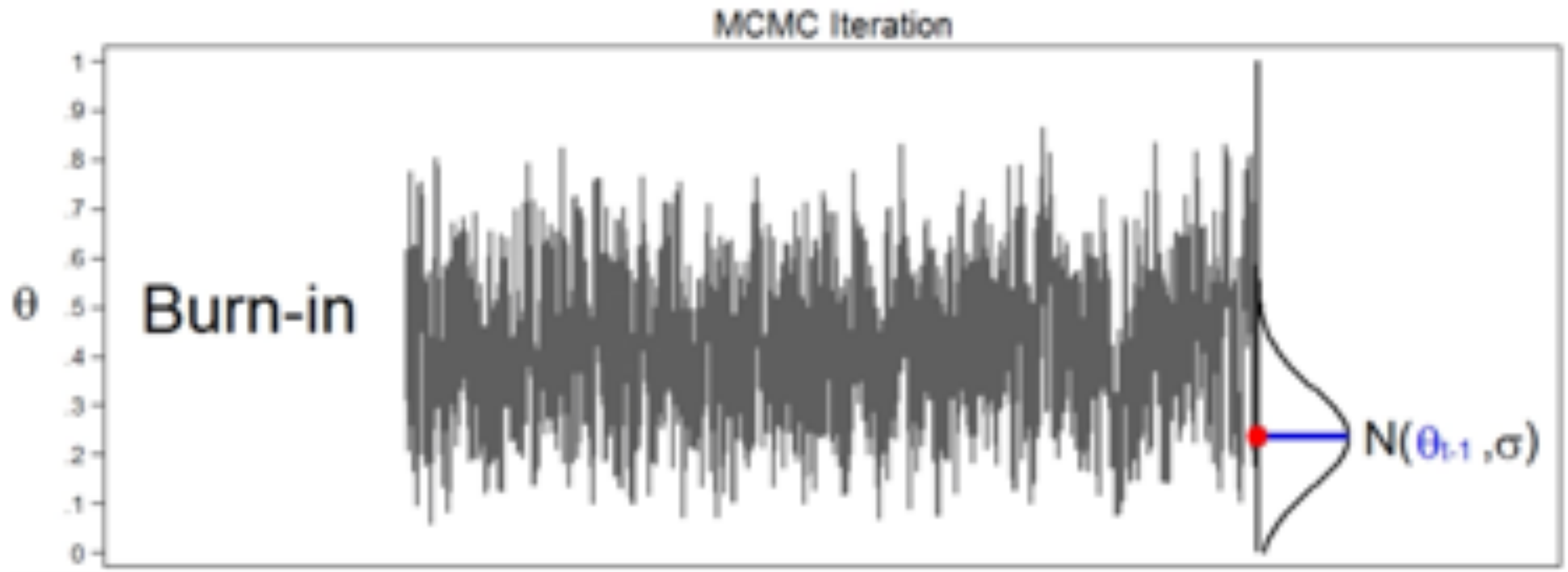
- The ERGM tries to produce a combination (vector) of parameter estimates that together generate simulated networks that don't differ (much) from the observed network on the salient statistics

- When it has settled on estimates (any updates are very small and tend to oscillate around a particular point estimate) we can say that the model has converged

So far, the empirical approach to Zeno's Paradox has been inconclusive.

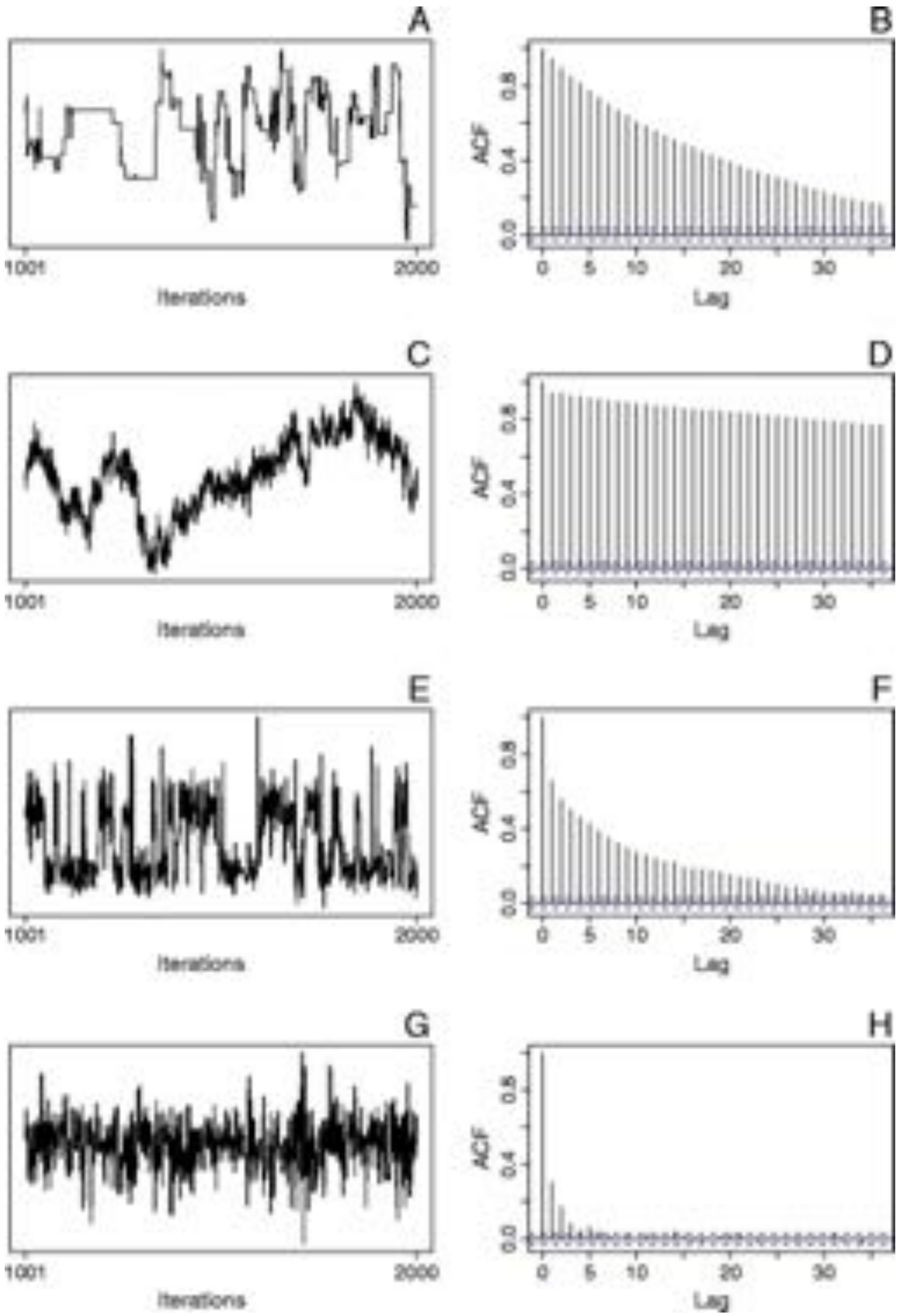
Two Main Issues



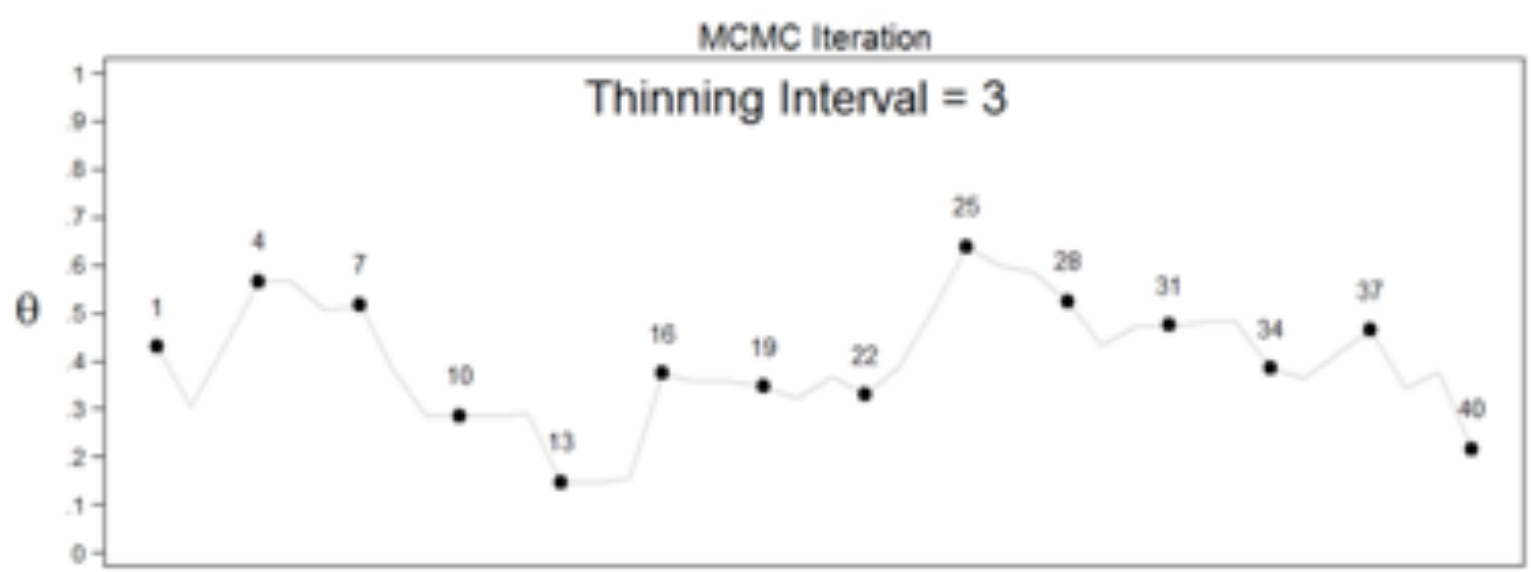


- 1. Dependence on starting values
    - Problem: Some starting values (e.g. 0) may be biased
    - Solution: Increase burnin period to discard first samples from Markov chain to give it time to stabilise or restart with new values

Two Main Issues



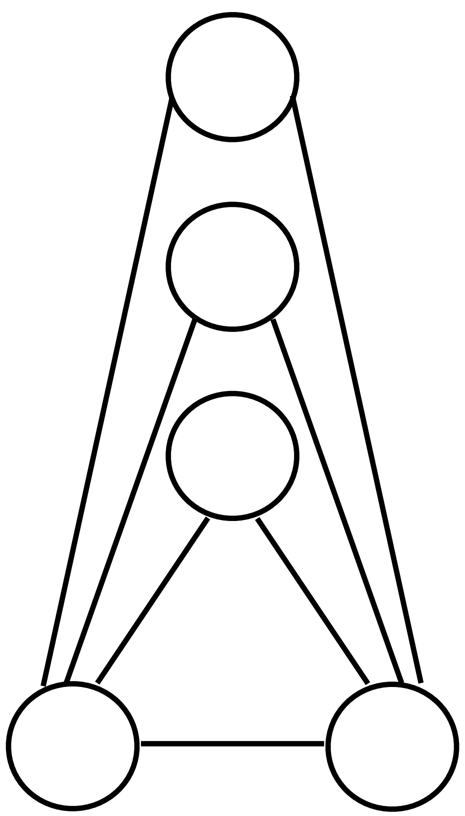
- 2. Autocorrelation due to Markov chain
    - Problem: Some normal, but should drop down and waver around 0 quite quickly or is considered excessive (not mixing well)
    - Solution: Increase thinning or change model



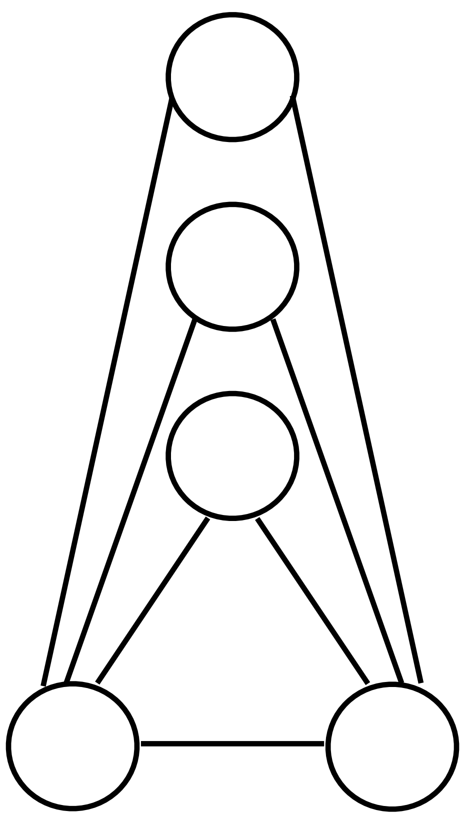



GWESP WTF?

- A major problem with ERGMs is that sometimes the amount of the ingredients shouldn't be scaled linearly:
  - Having a friend in common obviously makes our friendship more likely
  - But should each additional friend contribute the same? Three friends = thrice as likely? Four friends = …*fource* as likely? Same "info" in each?
- Moreover, if some configurations, like triangles, scaled linearly, then simulated networks would end up *degenerate*: impossibly dense, sparse, etc.



GWESP WTF?



- Maybe better to discount additional friends (see Snijders et al 2006; Hunter 2007)
  - *Alternating* k-stars and triangles effectively alternate the contributions of successive ties positively and negatively
  - *Geometrically-weighted* degrees and edgewise-shared partners discounts additional contributions by α
- Basically the same:
  - $\alpha = 0$, then GWESP statistic = number of edges in at least one triangle
  - $\alpha \to \infty$, then GWESP statistic -> 3x number of triangles
  - so as $\alpha \to 0$, subsequent ties/partners discounted more
- The lower α, model less likely to be degenerate, so start by fixing α low, say 0.25 or so (possible to estimate together with the coefficient, but slooooow)

A results table

	Naïve Actor Model	Political Capacity Model	Strategic Decision Model	Strategic Geography Model
General Parameters				
Density	-3.88 (0.03)*	-3.75 (0.07)*	-7.01 (0.35)*	-5.77(0.36)*
Centralization (actors)	–	–	0.61 (0.11)*	-0.21(0.11)
Centralization (institutions)	–	–	1.36 (0.18)*	0.56(0.18)*
Closure (actors)	–	–	-0.19(0.05)*	-0.06(0.04)
Geographic Centralization	–	–	–	1.57(0.05)*
Actor Type Activity Parameters (Local Government is Excluded Category)				
Federal Government	–	0.45 (0.15)*	0.43 (0.16)*	1.82(0.18)*
State Government	–	0.19 (0.14)	0.16 (0.13)	1.35(0.16)*
Water Special District	–	0.13 (0.09)	0.12 (0.09)	0.42(0.10)*
Environmental Special District	–	0.29 (0.17)	0.26 (0.17)	0.46(0.19)*
Environmental Group	–	-0.18 (0.10)	-0.16 (0.09)	-0.01(0.10)
Industry Group	–	-0.59 (0.26)*	-0.50 (0.23)*	0.05(0.29)
Education/Consulting	–	-0.40 (0.18)*	-0.32 (0.17)	-0.06(0.19)
Actor Coalition	–	-0.03 (0.34)	-0.03 (0.33)	0.44(0.38)
Other Activity	–	0.07 (0.48)	0.11 (0.43)	1.33(0.54)*
Institution Type Activity Parameters (Collaborative Partnership is Excluded Category)				
Interest Group Association Activity	–	-0.22 (0.10)*	-0.09 (0.09)	-0.04(0.06)
Advisory Committee Activity	–	-0.16 (0.12)	-0.10 (0.11)	-0.03(0.06)
Regulatory Process Activity	–	-0.78 (0.16)*	-0.61(0.15)*	-0.36(0.12)*
Actor as Venue Activity	–	-0.70 (0.19)*	-0.47 (0.16)*	-0.26(0.13)*
Joint Powers Authority Activity	–	0.16 (0.16)	0.15 (0.15)	0.06(0.10)
Mahalanobis distance as an indicator of model fit (smaller values indicate greater fit)	46,208	15,541	4,173	638

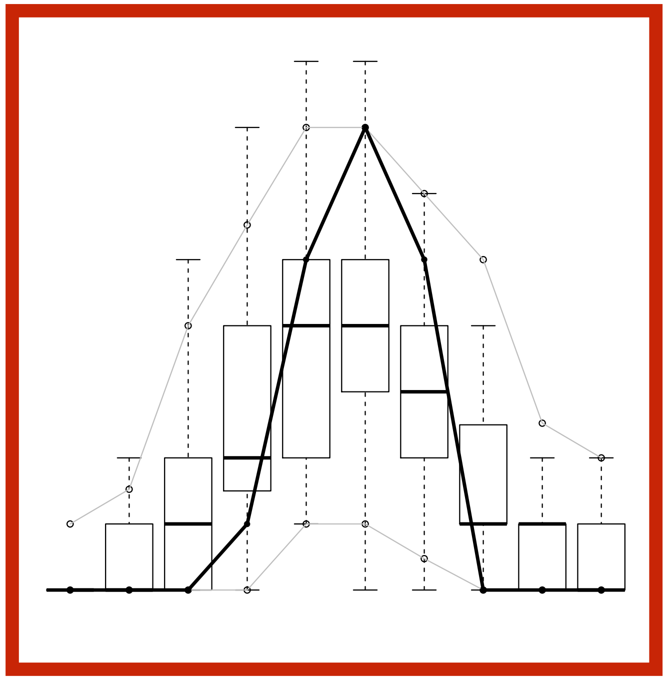
Note: Cell entries are ERGM parameter estimates with standard errors in parentheses. All models are estimated with "exogenous hubs," with fixed degree distributions for nodes with more than 20 edges. *Reject null hypothesis of parameter = 0, $p < 0.05$.

ERGM

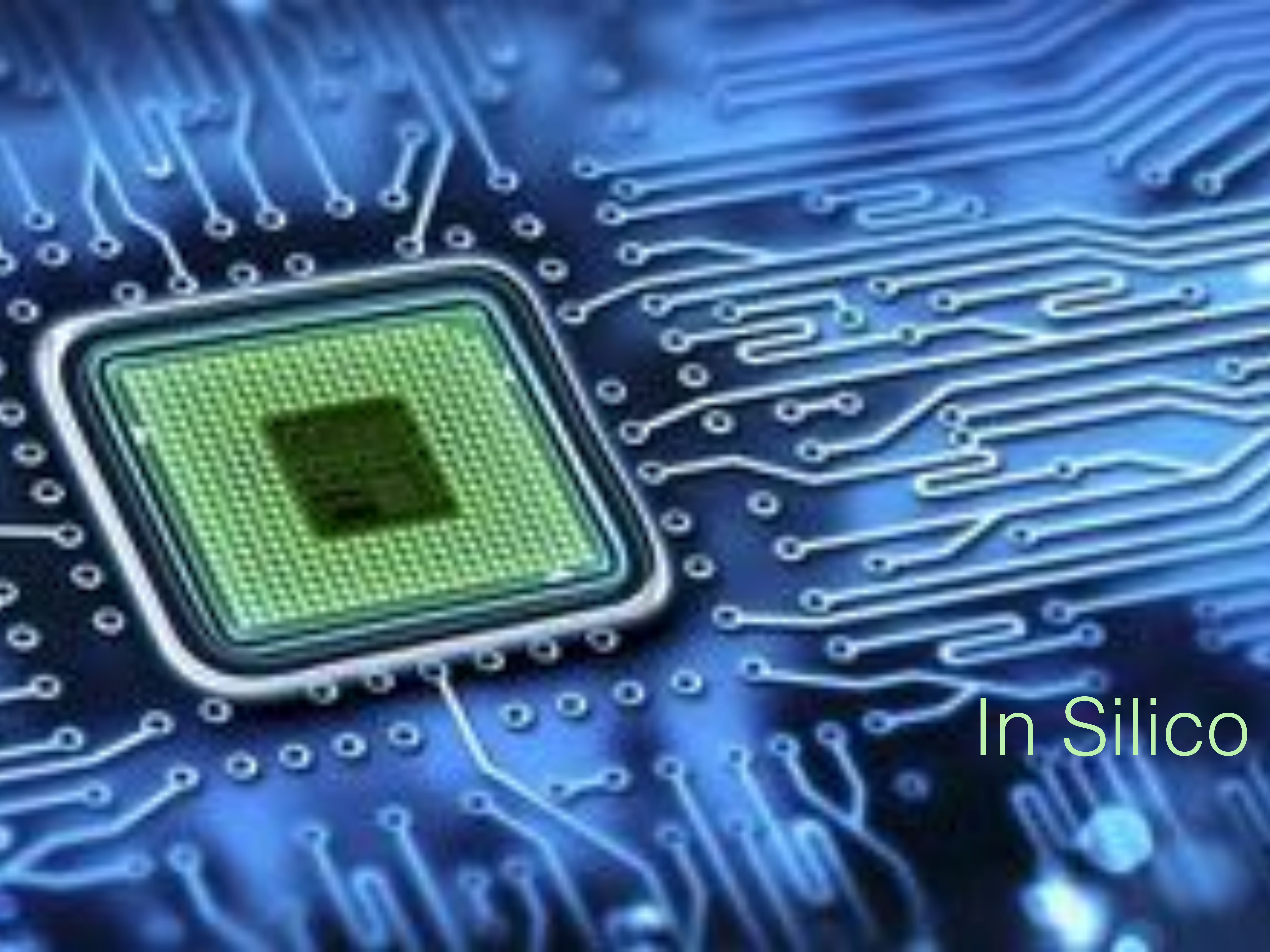
Effects

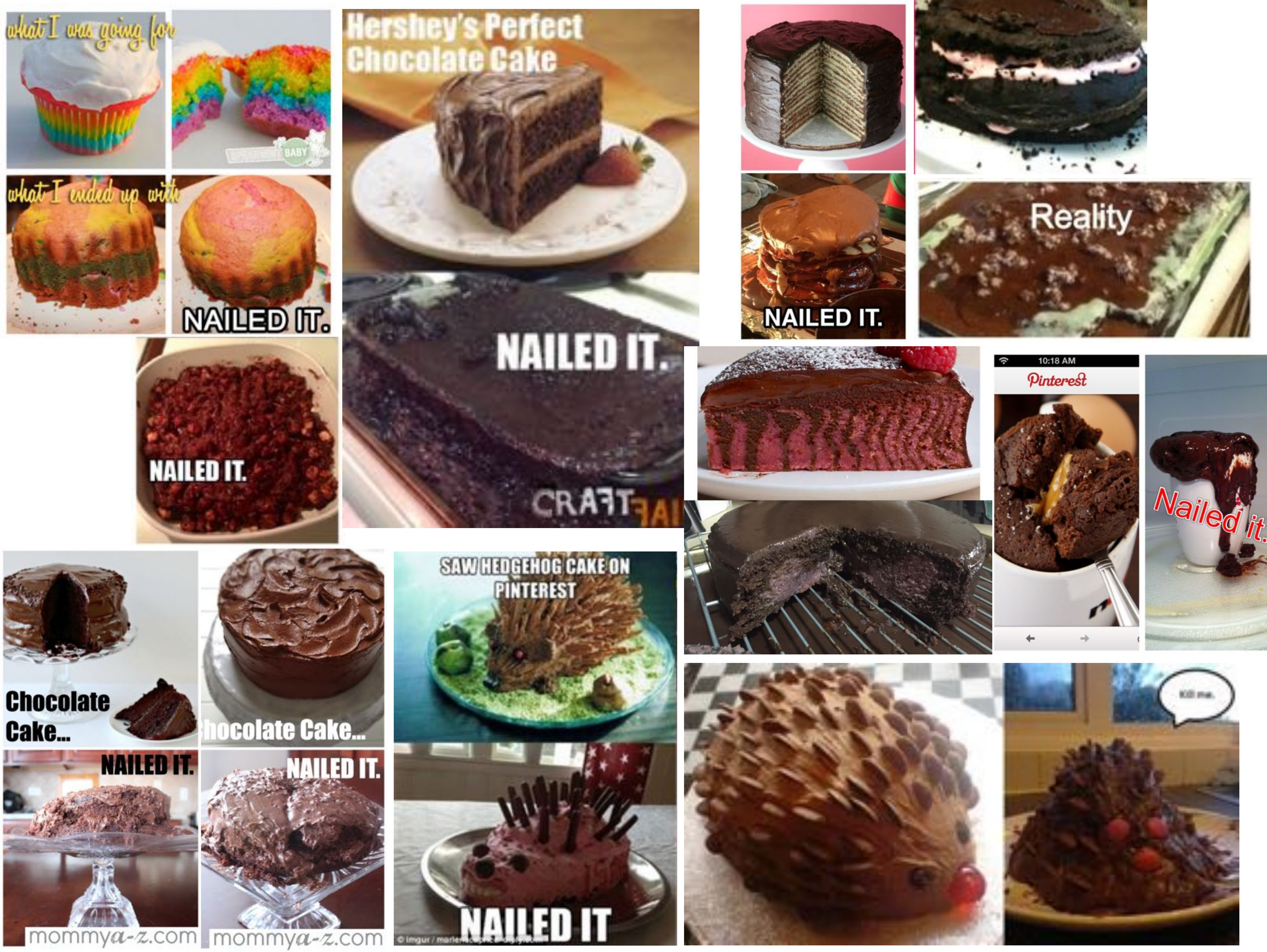
Model

Diagnostics



Diagnostics


In Silico

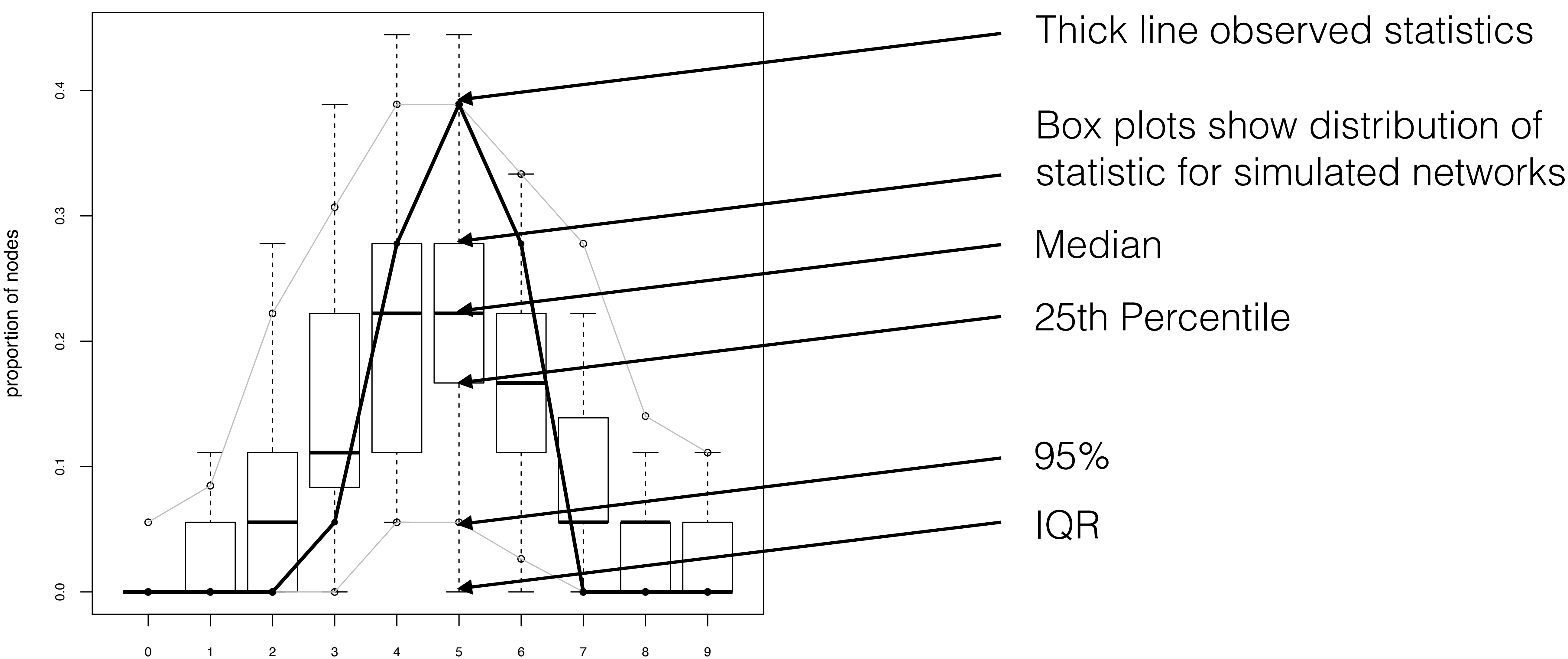


what I was going for
what I ended up with
NAILED IT.
Hershey's Perfect Chocolate Cake
NAILED IT.
CRAFT
NAILED IT.
NAILED IT.
Reality
10:18 AM
Pinterest
Nailed it.
Chocolate Cake...
NAILED IT.
NAILED IT.
mommya-z.com
mommya-z.com
SAW HEDGEHOG CAKE ON PINTEREST
NAILED IT

But is the converged model a *good* one?

- Goodness-of-fit (GOF) evaluates whether the simulated networks are similar to the observed one…
- In terms of statistics that are *not* explicitly modelled
  - degree distribution
  - triad census
  - geodesic distances
- Why does it have to be *other* statistics?
- GOFs can be considered equivalent to an R^2 statistic in regression models, though χ^2 and F tests are not available

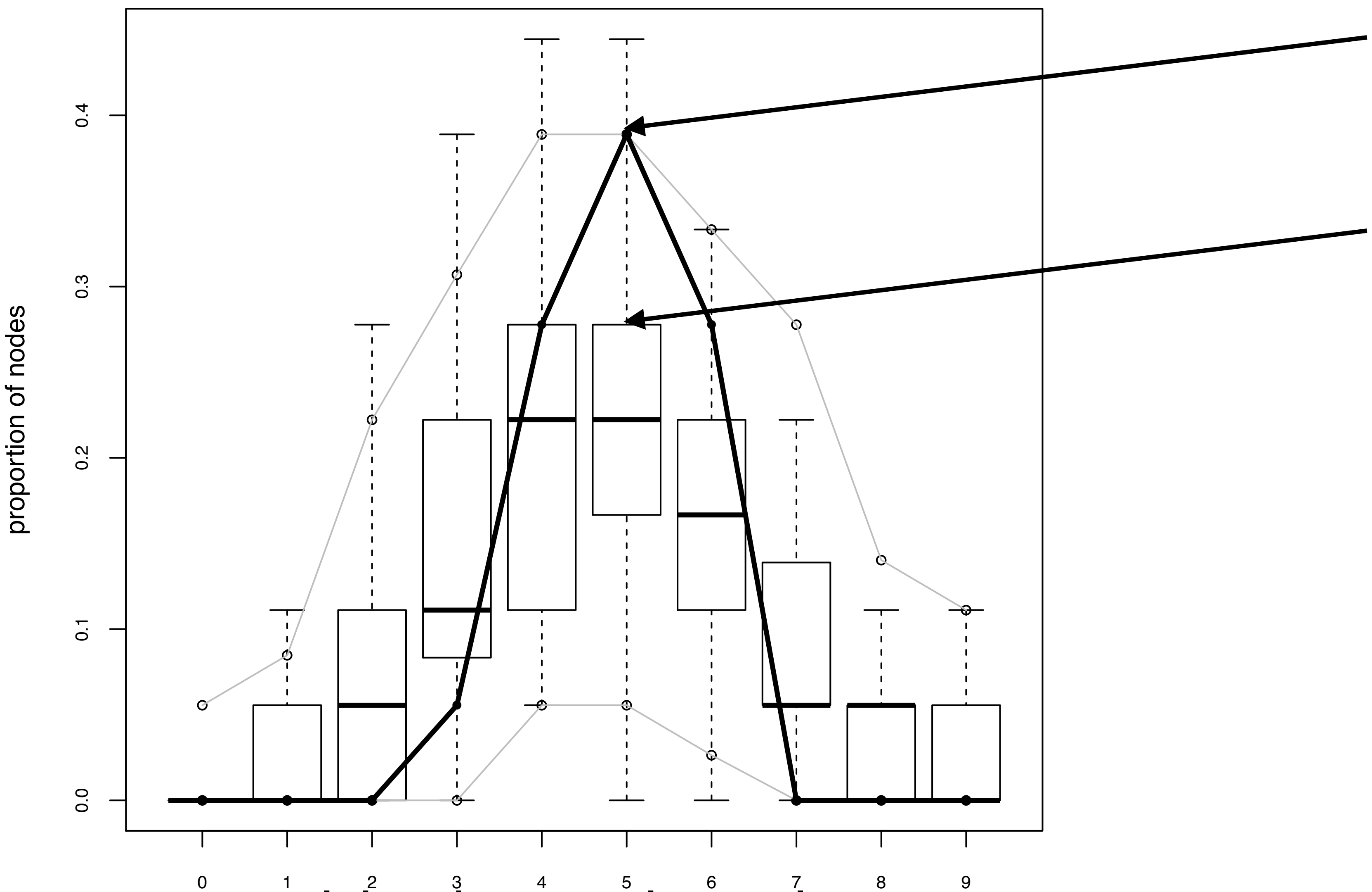
Goodness-of-fit diagnostics



Goodness-of-fit diagnostics

Overestimating low degrees

Overestimating high degrees



Underestimating medium degrees

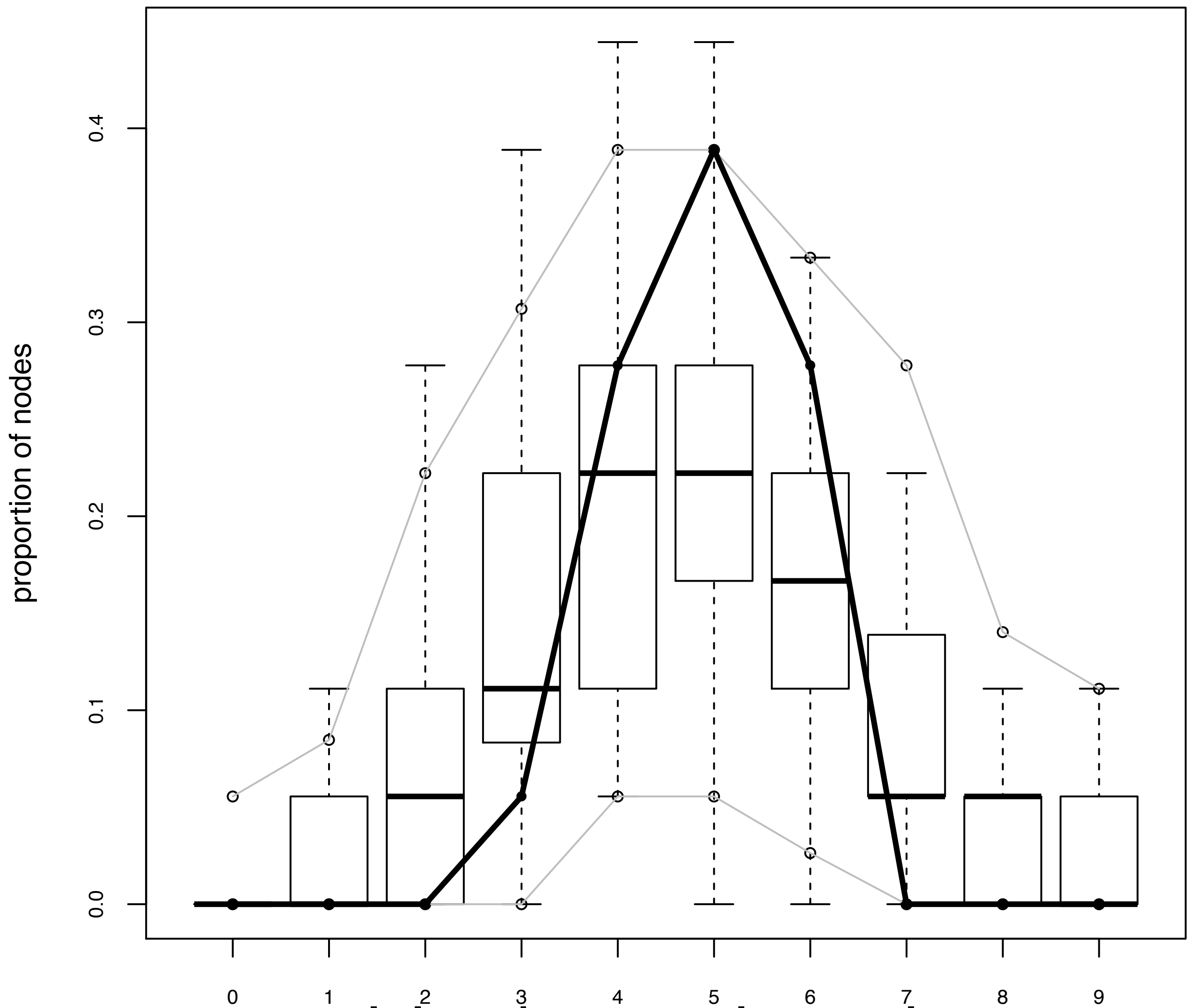
Thick line observed statistics

Box plots show distribution of statistic for simulated networks

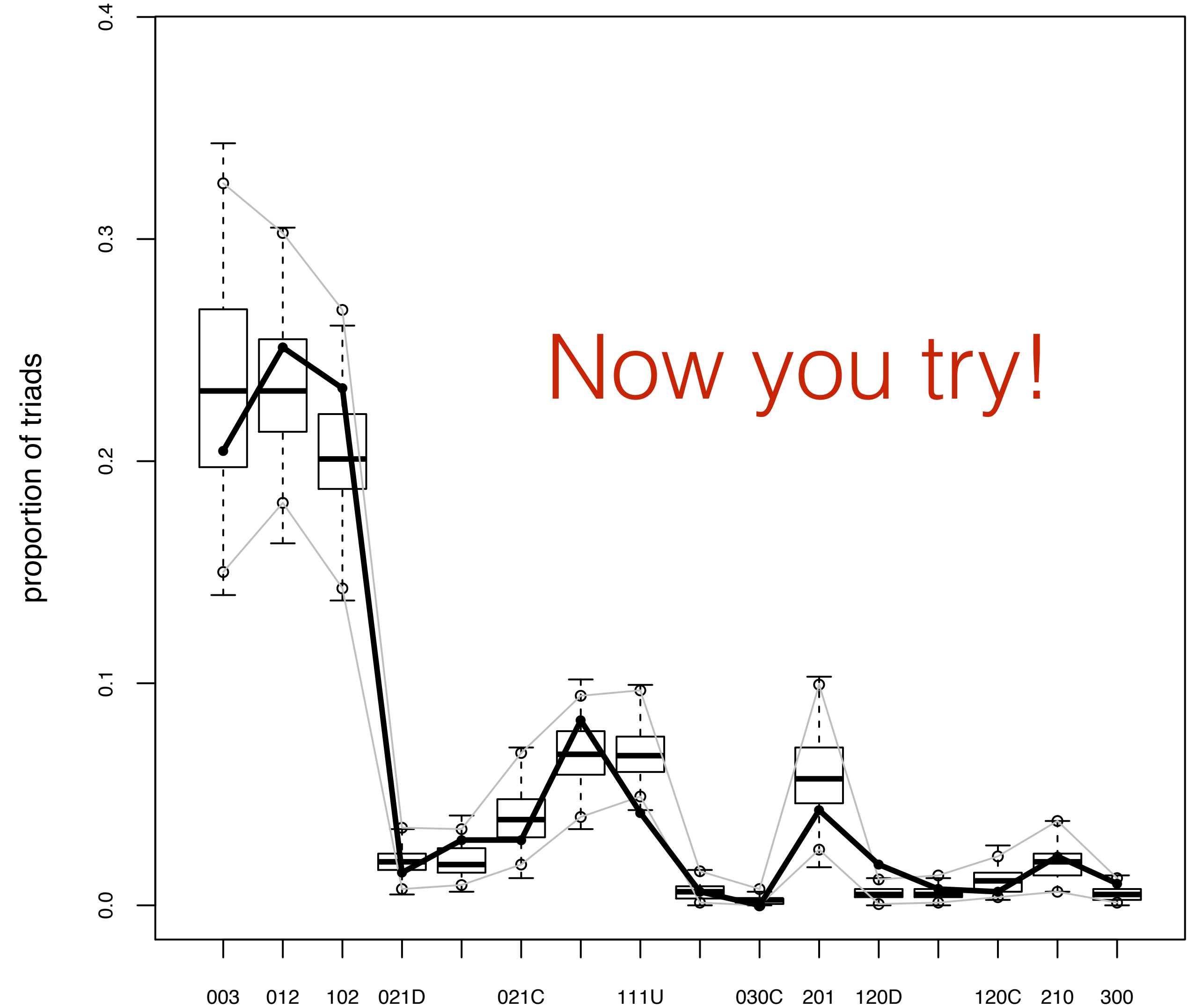
Goodness-of-fit diagnostics

Overestimating low degrees

Overestimating high degrees



Underestimating medium degrees



Summary: What ERGMs do

$$P(x;\theta) = \frac{\exp\left(\sum_k \theta_k z_k(x)\right)}{\kappa}$$

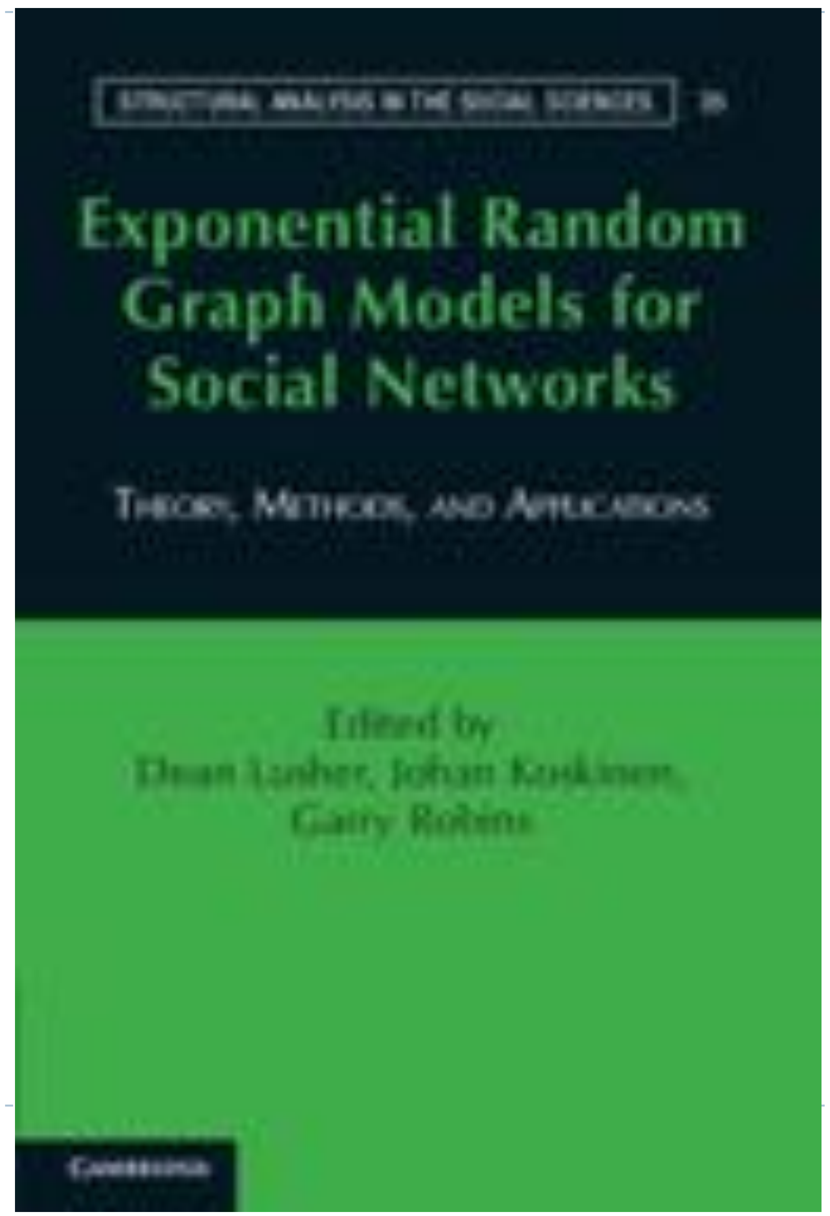
- Explains the probability of observing a specific graph/tie in a graph
- Dependence between observations is taken into account through statistic functions *z* that represent local patterns
  - Density
  - Reciprocity
  - Homophily
  - Transitivity
  - Similar institutional portfolios, …

Why ERGMs?

- ERGMs increasingly **understood** (sociology, political science, economics)
- ERGMs increasingly **used** (sociology, political science, economics)
- ERGMs increasingly **useful** (directed, bipartite, multilevel, valued, longitudinal, actor attributes, missing data, snowball designs)

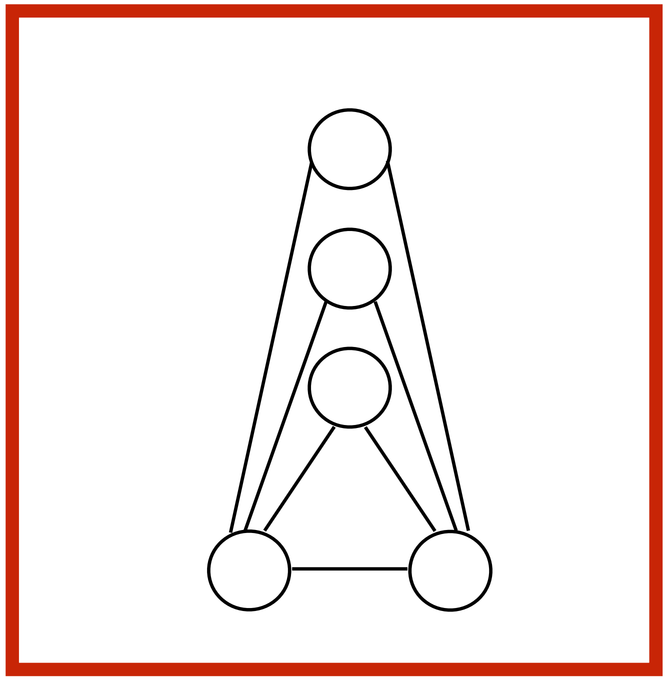

Fail
Cake

The real weapon…?



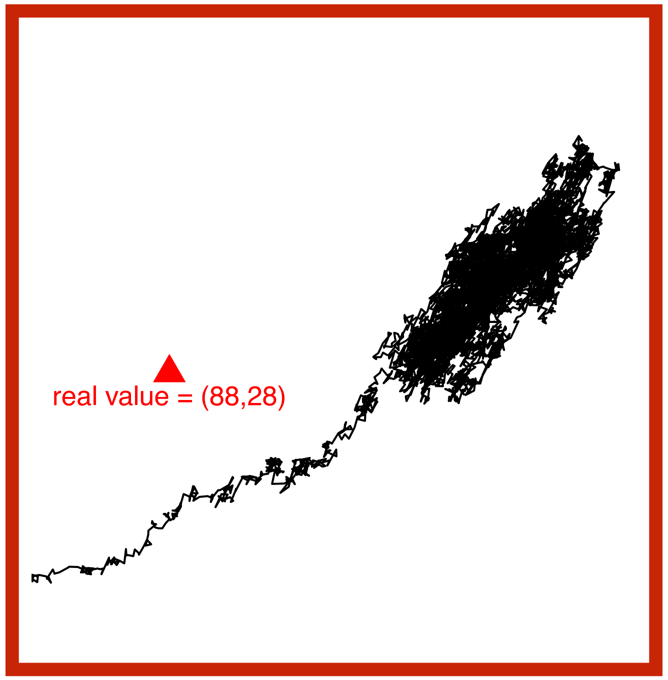
ERGM

Effects



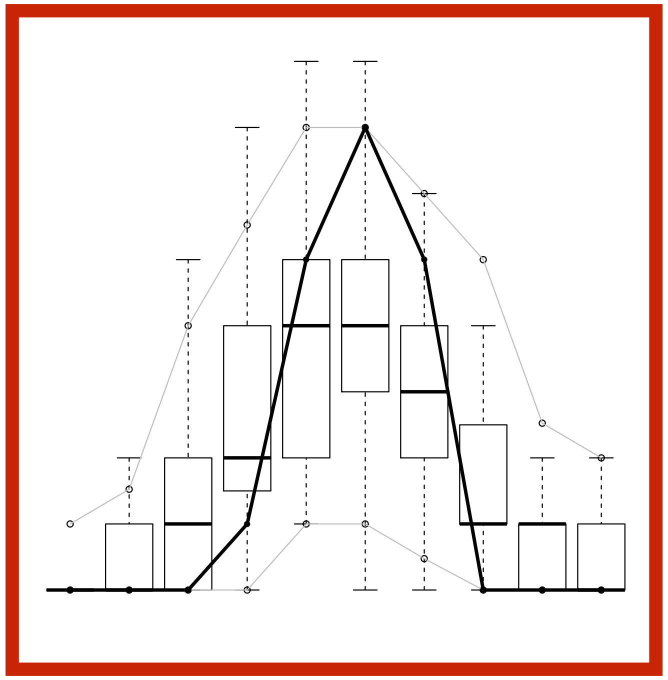
Explore various effects available

Model



Understand model intuition and estimation

Diagnostics



Recognise when a model converges and fits


Dig in!
Next week:
Tie-based -> Actor-oriented
Cross-sectional -> Longitudinal