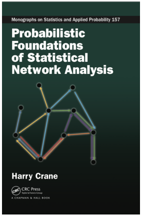
Probabilistic Foundations of Statistical Network Analysis
Chapter 5: Statistical modeling paradigm

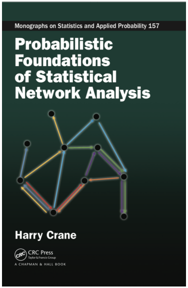
Harry Crane

Based on Chapter 5 of *Probabilistic Foundations of Statistical Network Analysis*



Book website: `http://www.harrycrane.com/networks.html`

Table of Contents



Chapters 3 and 4 highlight two primary contexts of network analysis:

- Chapter 3: modeling sampled network data.
- Chapter 4: modeling evolving networks.

Immediate observations:

- The concept of 'network' should not be conflated with the mathematical notion of 'graph' (Chapter 1).
- Sampling mechanism plays important role in model specification and statistical inference from sampled networks (Chapter 3).
- Statistical units are determined by the way in which the data is observed (Section 3.7).
- The explicit and implicit units should be aligned so that model-based inferences are compatible with their intended interpretation (Section 3.8).

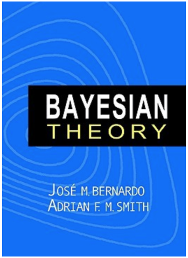
In this chapter, think of $\mathbf{Y}_N$ as generic 'network data' of 'size' N in space $\mathcal{N}_N$ of all such networks, where the interpretation of 'network' depends on context and 'size' is the number of units in that context.

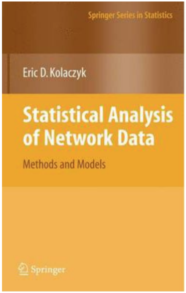
- In Section 2.4, $\mathcal{N}_N = \{0, 1\}^{N \times N}$ and the size is the number of vertices.
- In Section 3.6.1.1, $\mathcal{N}_N$ is the set of edge-labeled graphs with N edges and size is the number of edges.
- In Section 3.6.1.3, $\mathcal{N}_N$ is the set of path-labeled graphs with N paths and size is the number of paths.

What is a statistical model?

According to conventional wisdom in statistics literature:

A statistical model is a set of probability distributions on the sample space.



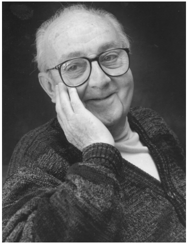


Questions:

- Just a set: $\{P_1, P_2, \ldots\}$?

All models are wrong ...

All models are wrong, but some are useful.



George Box (1919–2013)

A statistical model is a set of probability distributions on the sample space.

Questions:

- How can a *set* be 'wrong'?
- What determines whether this *set* is 'useful'?

Summary of Conclusions

(I) **What is a statistical model?**

$$\begin{array}{ccccc} \textit{Model} & = & \textit{Description} & + & \textit{Context} \\ & & \textit{`set'} & + & \textit{`inference rules'} \end{array}$$

(II) **All models are wrong, but some are useful.**

First step to being 'useful' is 'making sense'.

Coherence*: Model and inferences 'make sense' in a single context.*

(III) **Network Modeling**:

Sound theory for network analysis should be built on models that are

(i) *coherent and*

(ii) *account for realistic sampling schemes.*

Role of the model

All models are wrong, but some are useful.

A statistical model is a set of probability distributions on the sample space.

Role of the model in statistics:

- Sometimes exploratory data analysis (EDA)
- More often inference (out of sample) and prediction
- Asymptotic approximations

When is a model useful for these purposes?

Just one set?

Scenario:

$X_1, X_2, \ldots$ are i.i.d. $\mathcal{N}(\mu, 1)$.

Observe:

$X_1^*, \ldots, X_n^*$ for some finite $n \geq 1$.

Model:

Set of distributions $\{\mathcal{N}(\mu, 1) : -\infty < \mu < \infty\}$ on $\mathbb{R}$.

What can I do with this?

Estimate population parameter μ based on sample $X_1^*, \ldots, X_n^*$. (e.g., MLE, Bayesian posterior inference, ...)

What makes this possible?

Assumed: $X_1, X_2, \ldots$ i.i.d. $\mathcal{N}(\mu, 1)$ (population data).

Implicit: $X_1^*, \ldots, X_n^*$ i.i.d. $\mathcal{N}(\mu, 1)$ (sampled data).

Relationship between population and sample left implicit by convention.

Leaving relationship between inferential universe (population) and observed data (sample) ambiguous causes confusion in more complicated situations.

Modeling household sizes

Scenario:

$X_1, \ldots, X_N$ are sizes (i.e., # of residents) of N households in a population. Household sizes are i.i.d. from a '1-shifted Poisson':

$$Pr(X_i = k + 1; \lambda) = \lambda^k e^{-\lambda}/k!, \quad k = 0, 1, \ldots. \tag{1}$$

Observe:

$X_1^*, \ldots, X_n^*$ for some $n < N$.

Model: (Depends on context)

1. $X_1^*, \ldots, X_n^*$ obtained by sampling uniformly without replacement from $X_1, \ldots, X_N$. (Sampling households)

$$\Longrightarrow \qquad X_1^*, \ldots, X_n^* \quad \text{i.i.d. from (1).}$$

2. $X_1^*, \ldots, X_n^*$ obtained by sampling individuals in population and recording the size of their household. (Size-biased sampling)

$$\Pr(X_i^* = k + 1; \lambda) = \frac{(k+1)\lambda^k e^{-\lambda}}{(\lambda + 1)k!}, \quad k = 0, 1, \ldots.$$

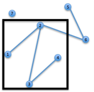
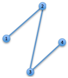
What is a statistical model?

A **statistical model** consists of

$\mathcal{M}$	*Description* of the observed data:	Set of candidate distributions
$\mathcal{C}$	*Context* under which data observed:	Relations among different sets

For each $n \geq 1$, the model $(\mathcal{M}, \mathcal{C})$ induces a set of candidate distributions $\mathcal{M}_n$ for sample of size n.

What makes a **model** $\mathcal{M}$ *"statistical" is that it can be used for statistical inference. Requires the* **context** $\mathcal{C}$ *under which the inference is performed.*

	Population	Observed network (sample)
	$\mathbf{Y}_N$	$\mathbf{Y}_n$
		
Model	$\mathcal{M}$	$\mathcal{M}_n$ (induced by context)

What is a statistical model?

A **statistical model** consists of

$\mathcal{M}$	*Description* of the observed data:	Set of candidate distributions
$\mathcal{C}$	*Context* under which data observed:	Relations among different sets

For each $n \geq 1$, the model $(\mathcal{M}, \mathcal{C})$ induces a set of candidate distributions $\mathcal{M}_n$ for sample of size n.

What makes a **model** $\mathcal{M}$ *"statistical" is that it can be used for statistical inference. Requires the* **context** $\mathcal{C}$ *under which the inference performed.*

Example (i.i.d. sequence):

$\mathcal{M} = \{\mathcal{N}(\mu, 1) : -\infty < \mu < \infty\}$

For $n \geq 1$, $(X_1^*, \ldots, X_n^*)$ modeled as $\mathcal{M}_n = \{\mathcal{N}^{\otimes n}(\mu, 1) : -\infty < \mu < \infty\}$

Example (household sizes):

$\mathcal{M} = \{\text{1-shifted Poisson}(\lambda) : \ \lambda > 0\}$

For $n \geq 1$, $(X_1^*, \ldots, X_n^*)$ modeled from size-biased distribution (assuming 2nd context of sampling individuals)

'Using' the model

Given: model $(\mathcal{M}, \mathcal{C})$ with induced sample models $\{\mathcal{M}_n\}_{n\geq 1}$.

1. Given data **D** of size $n \geq 1$.
2. Find optimal candidate distribution $\hat{P}_n$ in $\mathcal{M}_n$ based on **D** (according to some criteria).
3. Infer optimal distribution $\hat{P}_{\mathcal{M}}$ by interpreting $\hat{P}_n$ in context $\mathcal{C}$.

Example (i.i.d. sequence):

$\mathcal{M} = \{\mathcal{N}(\mu, 1) : -\infty < \mu < \infty\}$

For $n \geq 1$, $(X_1^*, \ldots, X_n^*)$ modeled as $\mathcal{M}_n = \{\mathcal{N}^{\otimes n}(\mu, 1) : -\infty < \mu < \infty\}$.

Given $\hat{P}_n = \mathcal{N}^{\otimes n}(\hat{\mu}, 1)$ infer $\hat{P}_{\mathcal{M}} = \mathcal{N}(\hat{\mu}, 1)$.

Example (household sizes):

$\mathcal{M} = \{\text{1-shifted Poisson}(\lambda) : \ \lambda > 0\}$

For $n \geq 1$, $(X_1^*, \ldots, X_n^*)$ modeled from size-biased distribution (assuming 2nd context of sampling individuals).

Given $\hat{P}_n$ from size-based with parameter $\hat{\lambda}_n$, infer population parameter through relationship $\hat{\lambda}_n \leftrightarrow \hat{\lambda}_n - 1$.

Sampling context (Example)

- For $m \leq n$ define **selection sampling**

$$\mathbf{S}_{m,n} : \mathbb{R}^n \to \mathbb{R}^m$$
$$(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_m)$$

- For a distribution F on $\mathbb{R}^n$, let $\mathbf{S}_{m,n} F$ denote distribution of $\mathbf{S}_{m,n} \mathbf{X}_n$ for $\mathbf{X}_n \sim F$. (Note: $\mathbf{S}_{m,n} F = F\mathbf{S}_{m,n}^{-1}$, usual induced distribution)
- Given set $\mathcal{M}_n$, we write set of all induced distributions as

$$\mathbf{S}_{m,n} \mathcal{M}_n = \{\mathbf{S}_{m,n} F : F \in \mathcal{M}_n\}.$$

	Population	Observed network (sample)
	$\mathbf{X}$	$\mathbf{X}_n$
	$(X_1, X_2, \ldots)$	$\mathbf{S}_{n,\mathbb{N}} \mathbf{X} = (X_1, \ldots, X_n)$
Model	$\mathcal{M} = \{\mathcal{N}^{\otimes\infty}(\mu, 1)\}$	$\mathbf{S}_{n,\mathbb{N}} \mathcal{M} = \mathcal{M}_n = \{\mathcal{N}^{\otimes n}(\mu, 1)\}$

- Sampling scheme $\mathbf{S}_{m,n}$ necessary to establish relationship between observation and population.

Sampling mechanism often (almost always) left out of model specification.

- For $m \leq n$ and injection $\psi : [m] \to [n]$, define ψ**-sampling** $\mathbf{S}^{\psi}_{m,n} : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\mathbf{S}^{\psi}_{m,n} : \mathbb{R}^n \to \mathbb{R}^m$$
$$(x_1, \ldots, x_n) \mapsto (x_{\psi(1)}, \ldots, x_{\psi(m)}).$$

- Let $\Sigma_{m,n}$ be random sampling map obtained by choosing $\psi : [m] \to [n]$ randomly and putting $\Sigma_{m,n} = \mathbf{S}^{\psi}_{m,n}$. (Distribution of ψ can depend on $\mathbf{X}_n$.)
- Write $\Sigma_{m,n}F$ to denote the distribution of $\mathbf{S}^{\psi}_{m,n}\mathbf{X}_n$ for this randomly chosen ψ and $\mathbf{X}_n \sim F$. Also write

$$\Sigma_{m,n}\mathcal{M}_n = \{\Sigma_{m,n}F : F \in \mathcal{M}_n\}.$$

Definition (Coherence)

A statistical model $(\{\mathcal{M}_n\}_{n\geq 1}, \{\Sigma_{m,n}\}_{n\geq m\geq 1})$ *is* ***coherent*** *if*

$$\begin{aligned} \Sigma_{m,n}\mathcal{M}_n &= \mathcal{M}_m \quad \textit{for all } n \geq m \geq 1 \\ \textit{induced} &= \textit{specified} \end{aligned}$$

Coherent $\Longrightarrow$ 'useful'

Definition (Coherence)

A statistical model $(\{\mathcal{M}_n\}_{n\geq 1}, \{\Sigma_{m,n}\}_{n\geq m\geq 1})$ *is* **coherent** *if*

$$\Sigma_{m,n}\mathcal{M}_n = \mathcal{M}_m \quad \textit{for all } n \geq m \geq 1.$$

Suppose $(\{\mathcal{M}_n\}_{n\geq 1}, \{\Sigma_{m,n}\}_{n\geq m\geq 1})$ is coherent.

- Given data **D** of size $m \geq 1$.
- Estimate $\hat{P}_m$ from $\mathcal{M}_m$ given **D**.
- For $n \geq m$, infer

$$\hat{P}_n = \{F \in \mathcal{M}_n : \Sigma_{m,n}F = \hat{P}_m\}.$$

  * This set is a singleton if model is identifiable.
- For smaller sample size ($\ell \leq m$) estimate

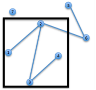
$$\hat{P}_\ell = \Sigma_{\ell,m}\hat{P}_m.$$

- Coherence needed to guarantee (i) $\hat{P}_n$ is non-empty and (ii) $\hat{P}_\ell \in \mathcal{M}_\ell$.

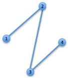
- These basic ideas are mostly ignored/invisible/unknown in the modern literature on network analysis.
- Frank and co-authors studied effects of sampling in social network analysis (1970s, 80s, 90s).
- Importance of sampling (and relevance of **context**) has not been emphasized in the modern statistics literature until very recently (Crane–Dempsey, 2015).
- Implications of exchangeability also seem to be poorly understood.

Assumed setting:

Population



Observed network (sample)



Guiding Question:

How to model network data in the presence of sampling?

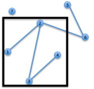
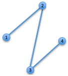
Given any sufficient statistics $(T_1, \ldots, T_k)$ and parameters $(\theta_1, \ldots, \theta_k)$, assign probability

$$\Pr(\mathbf{Y} = \mathbf{y}; \theta, T) \propto \exp\left\{\sum_{i=1}^{k} \theta_i T_i(\mathbf{y})\right\}, \quad \mathbf{y} = (y_{ij})_{1 \leq i,j \leq N} \in \{0,1\}^{N \times N}.$$

Holland and Leinhardt (1981), Frank and Strauss (1986), Wasserman and Pattison (1996), Wasserman and Faust (1994).

- **Typical approach**: Estimate θ by fitting ERGM (θ) to $\mathbf{Y}_n$, obtain $\hat{\theta}_n$ and use as estimate for θ in population.

→ Validity of this step depends on context (i.e., **coherence**).

	Population $\mathbf{Y}_N$	**Sample** $\mathbf{Y}_n$
		
Model	ERGM (θ)	???
Parameter	θ	θ

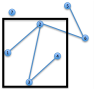
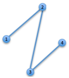
Coherence in ERGMs

Theorem (Shalizi–Rinaldo)

Model for $\mathbf{S}_{n,N}(\mathbf{Y}_n)$ *is ERGM* (θ) *if and only if sufficient statistics* T *have* separable increments.

$\Longrightarrow$ $(\{\mathcal{M}_n\}_{n\geq 1}, \{\mathbf{S}_{m,n}\}_{n\geq m\geq 1})$ *coherent if and only if* T *has "separable increments" (very strong condition).*

- In other words, given $\mathbf{Y}_n \sim \mathrm{ERGM}(\theta, T)$, the distribution of $\mathbf{S}_{m,n}\mathbf{Y}_n$ is also parameterized by 'θ', but distribution of $\mathbf{S}_{m,n}\mathbf{Y}_n$ is unknown (in general).
$\Longrightarrow$ Relationship between θ in two models unknown $\Longrightarrow$ Cannot do inference.

	Population $\mathbf{Y}_N$	**Sample** $\mathbf{Y}_n$
		
Model	ERGM (θ)	???
Parameter	θ	θ
Estimate	???	$\hat{\theta}_n$

Scenario 2: Vertex exchangeable models (graphons)

Let $\phi : [0,1] \times [0,1] \to [0,1]$ be a function (symmetric).

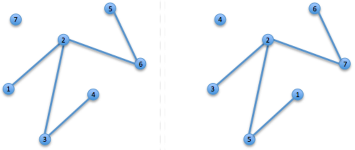
- Generate $U_1, U_2, \ldots$ i.i.d. Uniform[0, 1].
- Given $U_1, U_2, \ldots$, generate edges conditionally independently by

$$\Pr(Y_{ij} = 1 \mid U_1, U_2, \ldots) = \phi(U_i, U_j)$$
$$\Pr(Y_{ij} = 0 \mid U_1, U_2, \ldots) = 1 - \phi(U_i, U_j).$$

- Outcome $\mathbf{Y} = (Y_{ij})_{i,j \geq 1}$ satisfies

$$\Pr(\mathbf{Y}_n = (y_{ij})_{1 \leq i,j \leq n}) = \int_{[0,1]^n} \prod_{1 \leq i < j \leq n} \phi(u_i, u_j)^{y_{ij}} (1 - \phi(u_i, u_j))^{1 - y_{ij}} \, du_1 \cdots du_n.$$

- $\mathbf{Y}$ is **exchangeable**: $\mathbf{Y}^\sigma = (Y_{\sigma(i)\sigma(j)})_{i,j \geq 1} =_{\mathcal{D}} \mathbf{Y}$ for all permutations $\sigma : \mathbb{N} \to \mathbb{N}$.
  $\Rightarrow$ distribution of $\mathbf{Y}$ assigns equal probability to



Coherence of graphon models

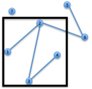
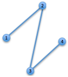
(Aldous–Hoover)

Let $\mathbf{Y} = (Y_{ij})_{i,j\geq 1}$ *be a vertex exchangeable random graph. Then* $\mathbf{Y}$ *is a mixture of graphon processes.*

(0) Sample $\phi \sim \varphi$ randomly from among functions $[0,1] \times [0,1] \to [0,1]$.

(1) Given ϕ, generate $\mathbf{Y}$ from the graphon model directed by ϕ.

$$\Pr(\mathbf{Y}_n = (y_{ij})_{1\leq i,j\leq n}) = \int_{[0,1]^2\to[0,1]} \int_{[0,1]^n} \phi(u_i,u_j)^{y_{ij}}(1-\phi(u_i,u_j))^{1-y_{ij}} du_1 \cdots du_n \varphi(d\phi).$$

	Population $\mathbf{Y}_N$	**Sample** $\mathbf{Y}_n$
		
Model	graphon (ϕ)	graphon (ϕ)
Parameter	ϕ	ϕ
Estimate	$\hat{\phi}_n$	$\hat{\phi}_n$

An impasse

Many real world networks exhibit:

(A) sparsity/power law

(B) exchangeability, consistency of finite sample distributions

Fact (Aldous (1981), Hoover (1979), Lovász–Szegedy (2006))

An infinite exchangeable random graph is dense or empty with probability 1.

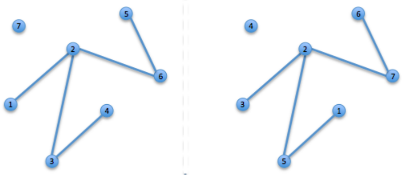
$\Longrightarrow$ Graphons cannot model (A) or (B).

- Often used to refute vertex exchangeability in networks applications, but empirical properties not even necessary to refute.
- The assumed context is off.

Implications of exchangeability assumption

Practical purpose of exchangeability assumption:

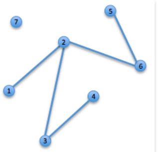
- Account for arbitrary labels assigned to sampled vertices by assigning equal probability to isomorphic graphs:



- Tractable class of models by incorporating symmetries.

Further implications of exchangeability:

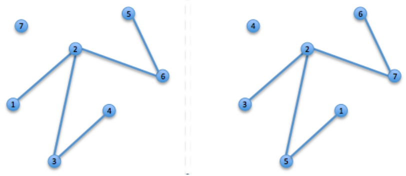
- Also implies sampled vertices interchangeable with unsampled vertices.



Implications of exchangeability assumption

Practical purpose of exchangeability assumption:

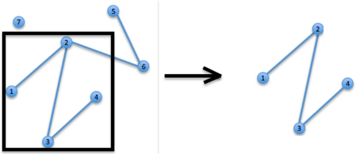
- Account for arbitrary labels assigned to sampled vertices by assigning equal probability to isomorphic graphs:



- Tractable class of models by incorporating symmetries.

Further implications of exchangeability:

- Also implies sampled vertices interchangeable with unsampled vertices.



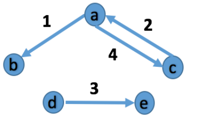
Scenario 3: Phone calls from a database

Entries are sampled uniformly at random from a large database of phone calls (or emails). Each observation (C_i, R_i) contains identity of the caller C_i and receiver R_i on the ith sampled call.

Interested in inferring the structure of connections among users in the database.

Caller	**Receiver**	**Time of Call**	...
555-7892 (a)	555-1243 (b)	15:34	...
550-9999 (c)	555-7892 (a)	15:38	...
555-1200 (d)	445-1234 (e)	16:01	...
555-7892 (c)	550-9999 (a)	15:38	...
⋮	⋮	⋮	⋱

Call sequence $X_1 = (a, b)$, $X_2 = (c, a)$, $X_3 = (d, e)$, $X_4 = (a, c)$ induces network:



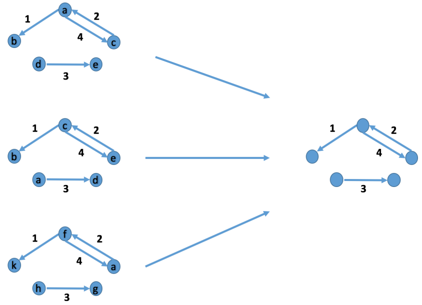
Interaction Networks

Dataset	vertices	**edges**
Actor collaborations	actors	**movies**
Enron email corpus	employees	**emails**
Karate club dataset	club members	**social interactions**
Wikipedia voting	Wikipedia admin.	**votes**
US Airport	airports	**flights**
Scientific collaborations	scientists	**articles**
UC Irvine online community	members	**online messages**
Political blogs	Websites	**hyperlinks**

- These datasets are driven by interactions
- Edges are the units — not represented as a (vertex-labeled) graph

Edge exchangeable models

- Vertices cannot be identified independently of their interactions with other vertices



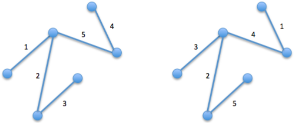
- Phone calls are sampled uniformly from the database $\Rightarrow$ exchangeable sequence of pairs $(C_1, R_1), (C_2, R_2), \ldots$.

- Edge-labeled graph contains ‘sufficient information’ about network structure.

Edge exchangeable models

- Phone calls are sampled uniformly from the database $\Rightarrow$ exchangeable sequence of pairs $(C_1, R_1), (C_2, R_2), \ldots$.

- Edge exchangeable model: Assign same probability to



Edge exchangeability $\Longleftrightarrow$ Size-biased vertex sampling

Other practical benefits (Hollywood model):

- Easy for estimation, prediction, and testing questions.
- Sparse with probability 1 for $1/2 < \alpha < 1$.
- Power law with exponent $\alpha + 1$ for $0 < \alpha < 1$.

H. Crane and W. Dempsey. (2016). Edge exchangeable models for interaction networks. Journal of the American Statistical Association, *in press.*

Sampling contexts for network models

- ERGM: none known
- Vertex exchangeable (graphons): representative sample of vertices
- Edge exchangeable: representative sample of edges (size-biased vertices)
- Relational exchangeability: representative sample of relations (Crane–Dempsey, 2016)
- Relative exchangeability: representative sample of vertices subject to heterogeneity in population (Crane–Towsner, 2015). Examples: stochastic blockmodel (Holland and Leinhardt)
- Completely random measures (graphex): representative sample edge patterns with respect to duration of time (Caron–Fox, 2017).

Summary of Conclusions

(I) **What is a statistical model?**

$$\begin{array}{rcccc} \textit{Model} & = & \textit{Description} & + & \textit{Context} \\ & & \textit{`set'} & + & \textit{`inference rules'} \end{array}$$

(II) **All models are wrong, but some are useful.**

First step to being ‘useful’ is ‘making sense’.

Coherence: *Model and inferences ‘make sense’ in a single context.*

(III) **Network Modeling**:

Sound theory for network analysis should be built on models that are

(i) *coherent and*

(ii) *account for realistic sampling schemes.*

Conclusions

What is a statistical model?

Model = Description + Context

A statistical model has two components:

- **Descriptive**: $\mathcal{M}_n$ – set of candidate distributions for each sample size $n \geq 1$.
- **Inferential**: $\mathcal{C}$ – context within which different sample sizes are related.

All models are wrong, but some are useful.

First step toward 'usefulness' is 'making sense' (coherence).

- Models aren't 'right' or 'wrong' but rather 'coherent' or 'incoherent'.
- **Coherence**: model $(\{\mathcal{M}_n\}_{n\geq 1}, \mathcal{C})$ 'makes sense' within a single context.
- Coherent models are 'useful' insofar as they 'make sense'.
- After coherence, other practical matters (e.g., computational tractability, accurate context) determined on a case-by-case basis.

Conclusions

Applications to Network Modeling:

Sound theory for network analysis should be build on models that are (i) coherent and (ii) account for realistic sampling schemes.

- Sampling mechanism should be accounted for in the context: edge sampling, hyperedge sampling, path sampling, snowball sampling, ...
- Current state of affairs: either no sampling context specified or vertex sampling taken as implicit (e.g., Shalizi–Rinaldo, 2013).
- Vertex sampling (selection, simple random sampling) usually not accurate reflection of context.

  $\Rightarrow$ Sound theory for network analysis should be built on models that are (i) coherent and (ii) account for realistic sampling schemes.
- Might this give clearer interpretation to asymptotics in network analysis?

- *H. Crane. (2018).* Foundations and Principles of Statistical Network Modeling. *Chapman–Hall.*
- *H. Crane and W. Dempsey. (2017). Edge exchangeable models for interaction networks.* Journal of the American Statistical Association.
- *H. Crane and W. Dempsey. (2015). A framework for statistical network modeling.*