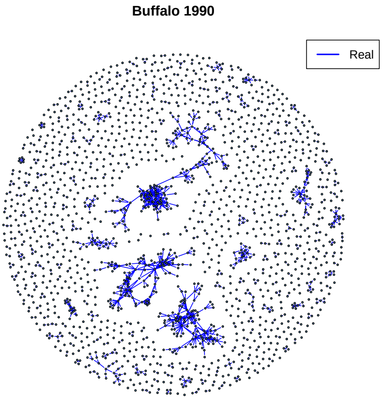
Sparsity Misspecification and Robust Covariate Effect Estimation for Sparse Social Networks

Alexander D'Amour
(Joint work with Edoardo Airoldi)

Harvard University
Department of Statistics

Joint Statistical Meetings 2014
August 3, 2014

Network Link Generation Problem



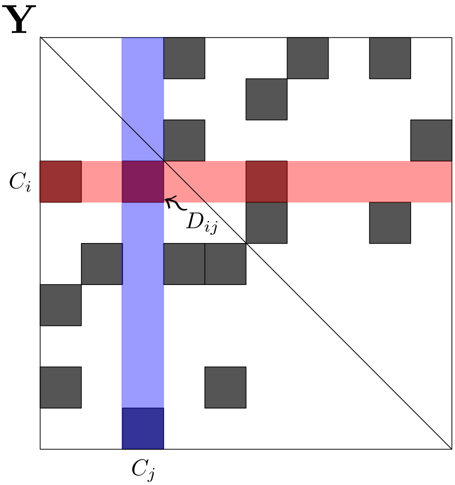
Sample of n actors.

Explain or predict pairwise outcomes Y using pairwise covariates X.

X may be observed or latent.

Network Sample Representation

Generalized Random Graph



Y entries in **arbitrary sample space** with some element $\mathbf{0}$.

Binarized graph A with

$$A_{ij} \equiv \mathbf{1}_{Y_{ij} \neq \mathbf{0}}.$$

Covariates X combine observed, latent attributes,

$$X_{ij} = f(C_i, C_j, D_{ij}).$$

Generative Network Models

Tradeoffs

Local structure: Homophily, Heterophily, Transitivity, etc.

Global structure: Sparsity, Percolation, etc.

Generative Network Models

Tradeoffs

Local structure: Homophily, Heterophily, Transitivity, etc.

Global structure: Sparsity, Percolation, etc.

Local structure dominates generative network modeling.

Generative Network Models

Common local approaches

Conditionally independent dyads (regression):

$$P(Y \mid X) = \prod_{i<j<n} P(Y_{ij} \mid X_{ij}).$$

Infinitely exchangeable dyads (Aldous-Hoover):

$$P(Y \mid X) = \int_{\mathcal{C}} \prod_{i<j<n} P(Y_{ij} \mid X_{ij}(C_i, C_j)) dF(C).$$

Generative Network Models

Common local approaches

Conditionally independent dyads (regression):

$$P(Y \mid X) = \prod_{i<j<n} P(Y_{ij} \mid X_{ij}).$$

Infinitely exchangeable dyads (Aldous-Hoover):

$$P(Y \mid X) = \int_{\mathcal{C}} \prod_{i<j<n} P(Y_{ij} \mid X_{ij}(C_i, C_j)) dF(C).$$

Do not capture global features, e.g., sparsity. Does this matter?

Inference Paradigms

Finite/Fixed Population

For a fixed set of actors

- Impute unmeasured links.
- Project forward in time.

Inference Paradigms

Finite/Fixed Population

For a fixed set of actors

- Impute unmeasured links.
- Project forward in time.

Superpopulation

For differing sets of actors

- Compare networks.
- Pool information or predict across networks.
- Scale local intuition to global network.

Inference Paradigms

Finite/Fixed Population

For a fixed set of actors

- Impute unmeasured links.
- Project forward in time.

Smoothing

Superpopulation

For differing sets of actors

- Compare networks.
- Pool information or predict across networks.
- Scale local intuition to global network.

Extrapolating

Network Superpopulation Inference

What is a network superpopulation?

Network Superpopulation Inference

What is a network superpopulation?

Intuiviely, require common generative process to “bridge” unlike samples.

Infinite network population defnied as stochastic process (Rinaldo and Shalizi 2013).

Observed samples are finite subgraphs of population graph.

Network Superpopulation Inference

What is a network superpopulation?

Intuiviely, require common generative process to "bridge" unlike samples.

Infinite network population defnied as stochastic process (Rinaldo and Shalizi 2013).

Observed samples are finite subgraphs of population graph.

Definition 1 (Generalized Random Graph Process).

A generalized random graph process $Y_{\mathbb{V}}$ is a stochastic process indexed by a countably infinite vertex set $\mathbb{V}$ whose finite-dimensional distribution for any finite subset $V \subset \mathbb{V}$ defines a generalized random graph Y_V with vertex set V.

Network Superpopulation Inference

Inferential procedure and assumptions

Likelihood inference procedure

1. Propose a model family $\mathcal{P}$ of models $P_{\beta,\gamma}$.
2. $\mathcal{P}$ implies a log-likelihood $L_{\beta,\gamma,n}$ on the sampled index set. Compute

$$\hat{\beta}_n, \hat{\gamma}_n = \arg\max_{B,\Gamma} L_{\beta,\gamma,n}(Y_n). \tag{1}$$

3. Interpret $\hat{\beta}_n$ as a population parameter estimate.

Likelihood inference procedure

1. Propose a model family $\mathcal{P}$ of models $P_{\beta,\gamma}$.
2. $\mathcal{P}$ implies a log-likelihood $L_{\beta,\gamma,n}$ on the sampled index set. Compute

$$\hat{\beta}_n, \hat{\gamma}_n = \arg\max_{B,\Gamma} L_{\beta,\gamma,n}(Y_n). \tag{1}$$

3. Interpret $\hat{\beta}_n$ as a population parameter estimate.

Step 3 requires **coherence** between inferences from different samples drawn from same population.

Network Superpopulation Inference

Assessing coherence

Network Superpopulation Inference

Assessing coherence

Intuitively, procedure is coherent if **object of estimation** is invariant to sampling.

Intuitively, procedure is coherent if **object of estimation** is invariant to sampling.

Effective estimand is the object of estimation for all n,

$$\bar{\beta}_n, \bar{\gamma}_n = \arg\max_{B,\Gamma} \mathbb{E}_0(L_{\beta,\gamma,n}(Y_n)). \tag{1}$$

where $\mathbb{E}_0$ is expectation with respect to the true process.

Intuitively, procedure is coherent if **object of estimation** is invariant to sampling.

Effective estimand is the object of estimation for all n,

$$\bar{\beta}_n, \bar{\gamma}_n = \arg\max_{B,\Gamma} \mathbb{E}_0(L_{\beta,\gamma,n}(Y_n)). \tag{1}$$

where $\mathbb{E}_0$ is expectation with respect to the true process.

For coherent procedures, effective estimand is invariant to sampling.

Intuitively, procedure is coherent if **object of estimation** is invariant to sampling.

Effective estimand is the object of estimation for all n,

$$\bar{\beta}_n, \bar{\gamma}_n = \arg\max_{B,\Gamma} \mathbb{E}_0(L_{\beta,\gamma,n}(Y_n)). \tag{1}$$

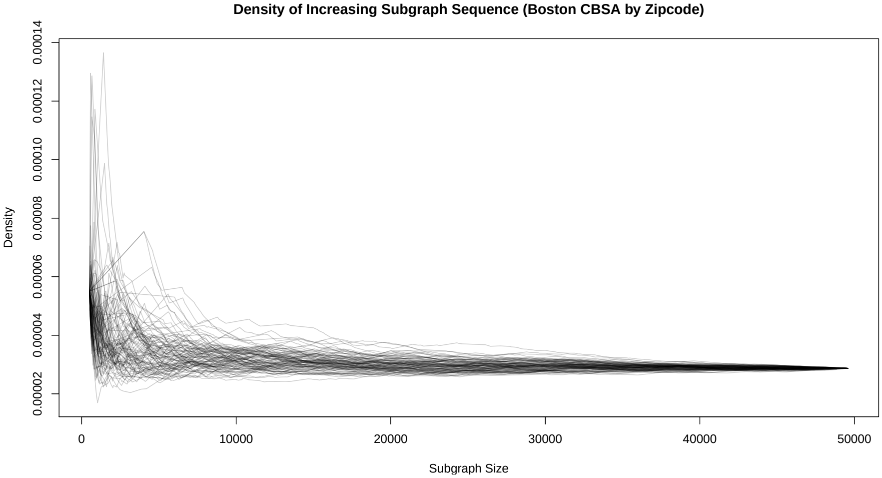
where $\mathbb{E}_0$ is expectation with respect to the true process.

For coherent procedures, effective estimand is invariant to sampling.

Under misspecified global structure?

Sparsity

Illustration



Sparsity

Formally

Sparsity

Formally

Define the **density operator**

$$D(Y_V) = \frac{\sum_{ij} A_{ij}}{\binom{|V|}{2}}.$$

Define the **density operator**

$$D(Y_V) = \frac{\sum_{ij} A_{ij}}{\binom{|V|}{2}}.$$

Definition 1 (Sparse Generalized Random Graph Process).

Let $Y_{\mathbb{V}}$ be a generalized random graph process on $\mathbb{V}$. $Y_{\mathbb{V}}$ is *sparse* if and only if for any $\epsilon > 0$, there exists an n such that for any subset of vertices $V \in \mathbb{V}$ with $|V| > n$ the corresponding finite dimensional generalized random graph Y_V has the property $\mathbb{E}(D(Y_V)) < \epsilon$.

Sparsity

Formally

Define the **density operator**

$$D(Y_V) = \frac{\sum_{ij} A_{ij}}{\binom{|V|}{2}}.$$

Definition 1 (Sparse Generalized Random Graph Process).

Let $Y_{\mathbb{V}}$ be a generalized random graph process on $\mathbb{V}$. $Y_{\mathbb{V}}$ is *sparse* if and only if for any $\epsilon > 0$, there exists an n such that for any subset of vertices $V \in \mathbb{V}$ with $|V| > n$ the corresponding finite dimensional generalized random graph Y_V has the property $\mathbb{E}(D(Y_V)) < \epsilon$.

Also, **sparsity rate** $\epsilon(n)$.

Sparsity Misspecification

Definition

A model family $\mathcal{P}_{\beta,\gamma}$ is **sparsity misspecfied** iff

$$\frac{\mathbb{E}_{\beta,\gamma}(D(Y_n))}{\mathbb{E}_0(D(Y_n))} \to 0 \text{ or } \infty.$$

Sparsity Misspecification

Definition

A model family $\mathcal{P}_{\beta,\gamma}$ is **sparsity misspecfied** iff

$$\frac{\mathbb{E}_{\beta,\gamma}(D(Y_n))}{\mathbb{E}_0(D(Y_n))} \to 0 \text{ or } \infty.$$

For example,

- For CID (under regularity) and exchangeable models, population extension is **dense** or **empty** (e.g., Orbanz and Roy, 2013).
- For process models, most lock in a given form for $\epsilon(n)$ (e.g., power law for preferential attachment).

Sparsity Misspecification

Consequences

Theorem 2 (Moving target theorem).

Suppose that the following hold:

1. *The inferential family $\mathcal{P}$ is sparsity misspecified for the true population process P_0.*
2. *The marginal distribution of the binarized data A identifies β in the presenece of nuisance parameters γ in $\mathcal{P}$.*
3. *The inferential model is* **responsive** *to the sample density $D(Y_n)$ under the true population process and*

$$|\mathbb{E}_{\bar{\beta}_n,\bar{\gamma}_n}(D(Y_n)) - \mathbb{E}_0(D(Y_n))| \in O(\epsilon_0(n)). \tag{2}$$

Then, for any n, there exists an $n' > n$ such that $\bar{\beta}_n \neq \bar{\beta}_{n'}$.

Example: Real Model Output

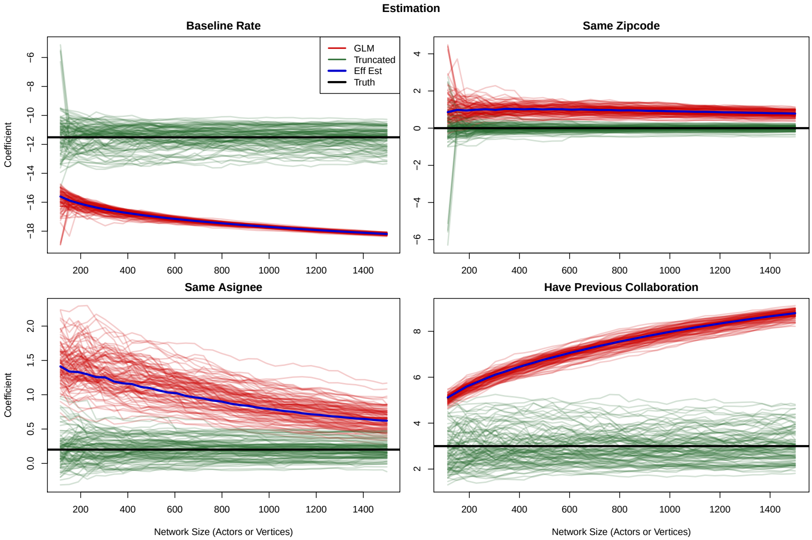
Cox PH regression. (Perry and Wolfe, 2013)
Inventor coauthorships in Michigan's motor industry 1982-1988.

Covariates (Coefs are log-ratios):

- `post85`: After 1985.
- `asgnum`: Work for same firm.
- `Ng0`: Have worked together before.

	lower	est	upper
post85	15.49	15.84	16.20
asgnum	4.65	4.83	5.02
Ng0	11.36	11.73	12.10
post85:asgnum	-4.77	-4.40	-4.03
post85:Ng0	-14.57	-14.00	-13.44
asgnum:Ng0	-5.56	-5.16	-4.76
post85:asgnum:Ng0	3.91	4.52	5.13

Example: Simulation



Partial Resolution

Salvaging conditional independence

Conditional independence assumptions are desirable.

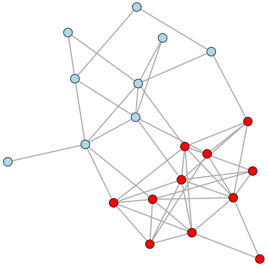
Can propose more complex latent structures R that induce conditional independence on conditional process.

Changes question:

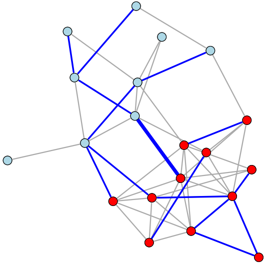
$$P_{\beta,\gamma}(Y \mid X) \to P_{\beta}(Y \mid R, X).$$

Partial Resolution

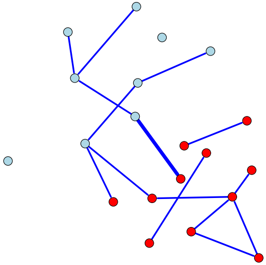
Conditionally Independent Relationship model



Unobservable Relationship/Risk



Both



Observable Interactions

Partial Resolution

Partial likelihood inference

Partial Resolution

Partial likelihood inference

Exploit conditional distribution $P_\beta(Y \mid A, X)$, or **zero-truncated likelihood**.

Partial Resolution

Partial likelihood inference

Exploit conditional distribution $P_\beta(Y \mid A, X)$, or **zero-truncated likelihood**.

Invariant to marginal distribution of R.

Partial Resolution

Partial likelihood inference

Exploit conditional distribution $P_\beta(Y \mid A, X)$, or **zero-truncated likelihood**.

Invariant to marginal distribution of R.

Under regularity conditions, recovers cohrent procedure for β.

Partial Resolution

Partial likelihood inference

Exploit conditional distribution $P_\beta(Y \mid A, X)$, or **zero-truncated likelihood**.

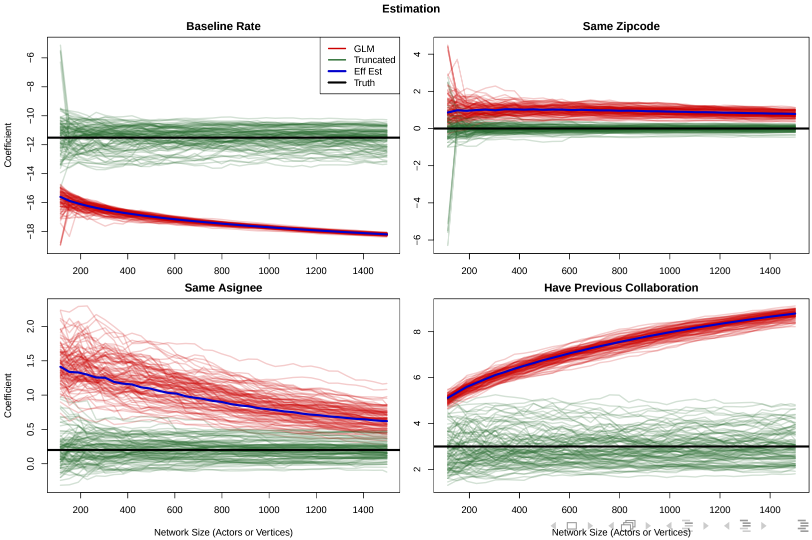
Invariant to marginal distribution of R.

Under regularity conditions, recovers cohrent procedure for β.

Bonus: Computation is $O(\sum_{ij} A)$.

Partial Resolution

Simulated success



Discussion

Network modeling is hard.

- Local intuition may contradict global structure.
- Misspecified global structure gives incoherent inference for interesting superpopulation questions.

Discussion

Network modeling is hard.

- Local intuition may contradict global structure.
- Misspecified global structure gives incoherent inference for interesting superpopulation questions.

Hope for network superpopulation inference?

- Invariance approaches for smaller questions.
- More flexible global models for general questions.

Discussion

Network modeling is hard.

- Local intuition may contradict global structure.
- Misspecified global structure gives incoherent inference for interesting superpopulation questions.

Hope for network superpopulation inference?

- Invariance approaches for smaller questions.
- More flexible global models for general questions.

Assessing coherence of sample-wise inferences is important.

- Box: "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."