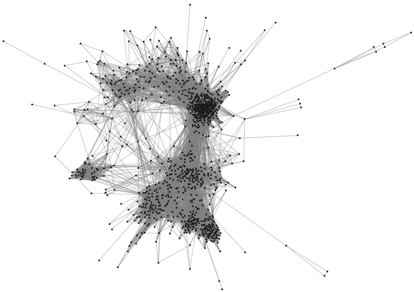
Statistical models of large graphs and networks

PETER ORBANZ

Columbia University

OBSERVED GRAPH



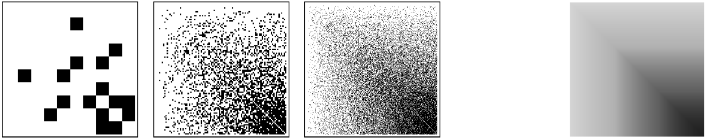
Problem statement

- There is a very large, unobserved graph g (the population).
- The observed graph G is sampled from g using some randomized algorithm.
- What can we learn about g from G?

How do we aggregate information over data that is clearly not i.i.d.?

EXCHANGEABLE GRAPHS

Graphon models [Borgs, Chayes, Lovasz, many others]



G_{10} $\quad$ G_{100} $\quad$ G_{300} $\quad \xrightarrow{\text{in distribution}} \quad$ G_w

Sampling from a graphon

For graphon $w : [0,1]^2 \to [0,1]$, draw

$$U_1, U_2, \ldots \sim_{\text{iid}} \text{Uniform}[0,1] \qquad X_{ij} | U_i, U_j \sim \text{Bernoulli}(w(U_i, U_j))$$

The graph $X = (X_{ij})_{i,j \in \mathbb{N}}$ is distributed as $X \stackrel{\text{d}}{=} G_w$

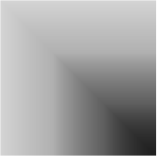
Exchangeable graphs

If $X = (X_{ij})_{i,j \in \mathbb{N}}$ is drawn from a graphon, then it is exchangeable, i.e.

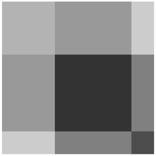
$$(X_{\pi(i)\pi(j)}) \stackrel{\text{d}}{=} (X_{ij}) \qquad \text{for every permutation } \pi \text{ of } \mathbb{N}$$

Machine Learning Examples

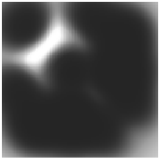
Undirected graphs



Toy example

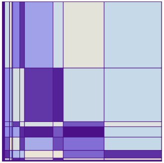


Stochastic block model

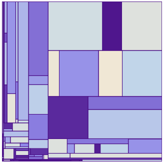


Gaussian process fit

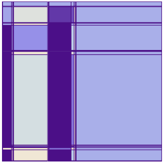
Directed graphs and matrices



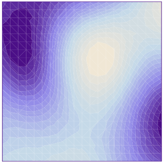
Infinite Relational Model
(Kemp et al, ’06)



Mondrian process
(Roy & Teh, ’08)



Latent feature relational model
(Miller et al, ’09)



Gaussian process

All of these models imply an exchangeability assumption.

STATISTICS IN THE I.I.D. WORLD

- $X = (X_1, X_2, \ldots)$ random sequence with i.i.d. entries.
- A sample of size n is summarized by the sample average

$$\mathbb{F}_n(f, X) = \frac{1}{n}\sum_{i\leq n} f(X_i) \qquad \text{for a function } f \, .$$

Convergence results

Law of large numbers ("consistency")

$$\mathbb{F}_n(f, X) - \mathbb{E}[f(X)] \qquad \xrightarrow{\text{p}} \qquad 0 \qquad \text{as} \quad n \to \infty$$

Central limit theorem ("asymptotic normality")

$$\sqrt{n}\,(\mathbb{F}_n(f, X) - \mathbb{E}[f(X)]) \qquad \xrightarrow{\text{d}} \qquad \eta Z \quad \text{for} \quad Z \sim N(0, 1)$$

η^2 is the *asymptotic variance*.

Berry-Esseen bound

$$d_{\text{W}}\Big(\frac{\sqrt{n}}{\eta}\,(\mathbb{F}_n(f, X) - \mathbb{E}[f(X)]),\, Z\Big) \quad < \quad \tau(n)$$

Concentration

$$P(\,|\mathbb{F}_n(f, X) - \mathbb{E}[f(X)]| > \varepsilon) \quad < \quad \rho(\varepsilon, n)$$

THE I.I.D. ASSUMPTION AND SAMPLING

Sampling from a finite population

- Fix an infinite sequence $x_1, x_2, \ldots$.
- Consider only the first n elements $x_1, \ldots, x_n$. Think of these as a "population".
- From these, select k elements uniformly (with replacement) and collect them in a random sequence $S_{n\to k} := (X_{n1}, \ldots, X_{nk})$.
- The variables X_{ni} are distributed i.i.d. according to some distribution P_n.

Sampling from an infinite population

- As $n \to \infty$, the distribution of $S_{n\to k}$ converges to that of a random sequence

$$S_k := (X_1, \ldots, X_n)$$

- The variables X_i are still i.i.d., according to some distribution P.

I.i.d. observations arise by independent sampling from a (possibly infinite) population.

GRAPH SUBSAMPLING

- Invent a randomized algorithm that can draw a random subgraph $G_{n \to k}$ of any size k from an input graph g_n of any size n
- Now fix an infinitely large graph g. Let g_n be its subgraph on the first n vertices.
- Consider the limit G_k in input size:

$$G_{n \to k} \xrightarrow{\text{d}} G_k \quad \text{as } n \to \infty$$

- Provided the limit exists, we can think of G_k as a sample of size k from the infinite population g

Uniform subsampling and graphons

i.) Sample k vertices w/o replacement from input graph g_n.
ii.) Extract the induced subgraph $G_{n \to k}$.

The limit G_k is distributed as a graph of size k drawn from some graphon w.

Graphons arise as limiting models from uniform subsampling.
All exchangeable graphs can be obtained by randomizing g.

A **symmetry property** is invariance under a group $\mathbb{G}$ of transformations,

$$\phi(X) \stackrel{\mathrm{d}}{=} X \qquad \text{for all } \phi \in \mathbb{G}$$

The most important example is exchangeability. For example:

- Exchangeable sequences (cf. de Finetti's theorem)
- Exchangeable graphs ($\rightarrow$ graphons)
- Exchangeable partitions (Chinese restaurant process etc.)

Symmetry from subsampling

sampling scheme	induced model class	symmetry
k independent vertices	graphon models	vertices exchangeable
select vertices by coin flips + delete isolated vertices	generalized Caron-Fox	underlying point process exchangeable
k independent edges (input multigraph)	"edge-exchangeable graphs"	edges exchangeable
neighborhood of random vertex	certain local weak limits	"involution invariance"

Beyond the i.i.d. case, sampling arguments yield symmetry properties.

Generalized sample averages

How do we average within a sample graph or structure?

- Suppose the symmetry group $\mathbb{G}$ is countably infinite.
- We approximate it by a suitable sequence of finite subsets $\mathbf{A}_1, \mathbf{A}_2, \ldots \subset \mathbb{G}$
- Define:

$$\mathbb{F}_n(f, X) := \frac{1}{|\mathbf{A}_n|} \sum_{\phi \in \mathbf{A}_n} f(\phi X)$$

Example: Exchangeable sequence $X = (X_1, X_2, \ldots)$

- $\mathbb{G} =$ finitely supported permutations of $\mathbb{N}$ and $\mathbf{A}_n$ = permutations of $\{1, \ldots, n\}$
- $f(X) = g(X_1)$ function of one variable

$$\mathbb{F}_n(f, X) \;=\; \frac{1}{|\mathbf{A}_n|} \sum_{\phi \in \mathbf{A}_n} g(X_{\phi(1)}) \;=\; \frac{(n-1)!}{n!} \sum_{i \leq n} g(X_i) \;=\; \frac{1}{n} \sum_{i \leq n} g(X_i)$$

Convergence in the symmetric world

with Morgane Austern

- X is an infintite random graph or structure with symmetry group $\mathbb{G}$.
- $\mathbb{F}_n(f,X)$ is the group average defined above. We want to estimate $\mathbb{E}[f(X)]$.
- We impose conditions on the sets $\mathbf{A}_n$ and on how f interacts with X.

Theorems [Austern & O. 2018]

Law of large numbers [E. Lindenstrauss, 2001]

$$\mathbb{F}_n(f,X) - \mathbb{E}[f(X)] \qquad \xrightarrow{\text{p}} \qquad 0 \qquad \text{as} \quad n \to \infty$$

Central limit theorem

$$\sqrt{n}\,(\mathbb{F}_n(f,X) - \mathbb{E}[f(X)]) \qquad \xrightarrow{\text{d}} \qquad \eta Z \quad \text{for} \quad Z \sim N(0,1)$$

η^2 is the *asymptotic variance*.

Berry-Esseen bound

$$d_{\text{W}}\Big(\tfrac{\sqrt{n}}{\eta}\,(\mathbb{F}_n(f,X) - \mathbb{E}[f(X)]),\, Z \Big) \quad < \quad \tau(n)$$

Concentration

$$P(\,|\mathbb{F}_n(f,X) - \mathbb{E}[f(X)]| > \varepsilon) \quad < \quad \rho(\varepsilon, n)$$

Example: Exchangeable graph

$X = (X_{ij})_{i,j\in\mathbb{N}}$ adjacency matrix, f real-valued function

$$\mathbb{F}_n(f,X) = \frac{1}{n!}\sum_{\phi\in\mathbf{A}_n} f\big((X_{\phi(i)\phi(j)})_{i,j\in\mathbb{N}}\big) \quad \text{averages over first } n \text{ vertices}$$

Theorem [Austern & O., 2018]

- X is an exchangeable random object.
- f is function with $\mathbb{E}[f(X)^2] < \infty$ and $\mathbb{E}[f(X)] = 0$.
- X and f satisfy $\sum_{i\in\mathbb{N}} \limsup_j \|f(X) - f(\tau_{ij}X)\|_{\mathbf{L}_2} < \infty$

Then:

$$\frac{\sqrt{n}}{n!}\sum_{\phi\in\mathbb{S}_n} f(\phi X) \xrightarrow{d} \eta Z \qquad \text{as } n\to\infty, \text{ for } Z\sim N(0,1) \text{ and } \eta^2 < \infty \text{ a.s.}$$

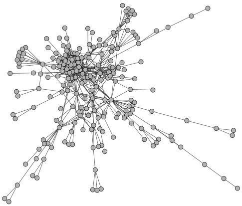
and

$$d_{\mathrm{W}}\Big(\frac{\sqrt{n}}{\eta n!}\sum_{\phi\in\mathbb{S}_n} f(\phi X)\ ,Z\Big) = O\Big(\min_{k\in\mathbb{N}}\big[\tfrac{k^2}{\sqrt{n}} + \sum_{i>k}\limsup_j \|f(X)-f(\tau_{ij}X)\|_2\big]\Big)$$

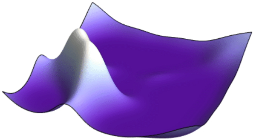
HICCUPS

- G_w (an infinite random graph generated by a graphon w) is dense almost surely.

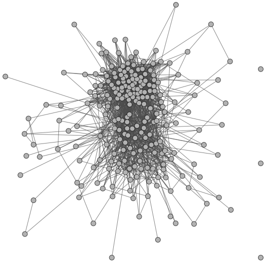
  Dense means: # edges $\approx$ (# vertices)^2
- The distance between any two vertices is 1, 2, or ∞
  (unless G_w is k-partite for $k \geq 2$)



Observed graph



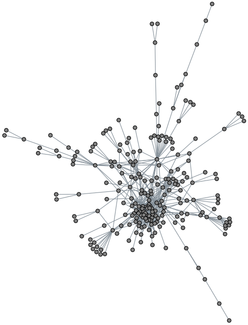
Estimated graphon



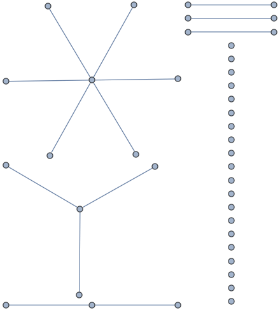
Reconstruction from estimate

EXPLANATION AS SELECTION BIAS

i.) Sample k vertices w/o replacement.
ii.) Extract the induced subgraph G_k.



Input graph g



G_{40}

APPLICATION: EMBEDDINGS

WITH VICTOR VEITCH, WENDA ZHOU, MORGANE AUSTERN, DAVID BLEI

A **graph embedding** attempts to assign to each vertex i a vector $w_i \in \mathbb{R}^d$ such that

$$\text{vertices } i \text{ and } j \text{ ``behave similarly''} \quad \Leftrightarrow \quad \langle w_i, w_j \rangle \text{ large}$$

Embeddings as empirical risk minimization

- Define a subsampling algorithm that generates a random subgraph G_k of size k
- Choose a loss function L and a predictor (say a classifier) f
- Define empirical risk $\mathbb{E}_{G_k}[L(f(\text{data}), \text{label})]$

Things we can show

- Various popular embeddings (such as node2vec) can be explained in this way.
- Stochastic gradients that use the same sampler for mini-batching are automatically unbiased.
- We have very fast implementations using automatic differentiation.
- Theory: If input graph is generated by graphon, embedding w_i asymptotically depends only on latent information at vertex i and graphon.

PROBLEMS, PROBLEMS, PROBLEMS

Random graphs + networks

- How do I have to sample from a large graph if I want to estimate ...?
- What model class is induced by, say, sampling “ego-networks”?
- Can one cross-validate in a network analysis problem?

Symmetry theorems: Other applications

- Network models (e.g. graphex convergence)
- Neural networks
- Cross validation