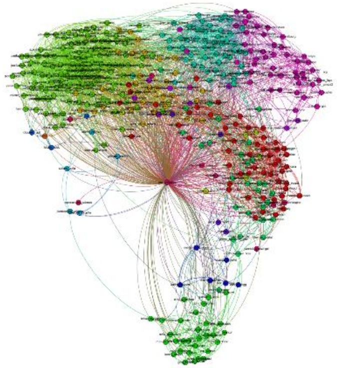
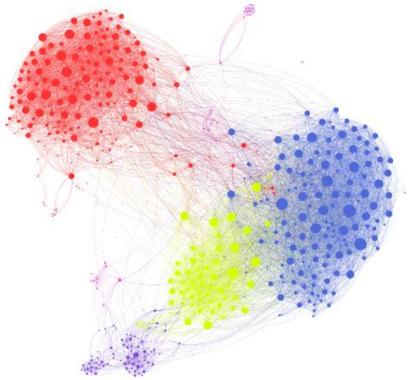
Graph Convergence and Estimation

Christian Borgs
Microsoft Research New England

Janke Festschrift
April 2016

Motivation: Questions

- When do we consider two large graphs similar?

$\Longrightarrow$ Graph convergence

- What is the limit?
- What is a good random model for large graphs?
- How do we estimate such a model from a single, observed graph?

Preview: Graphs and Graphons

Graphs

- Vertex set V
- Adjacency matrix $A: V \times V \to \{0,1\}$

↓ Graph Limits

↑ Random Graph Models

↓ Graphon Estimation

Graphons

- Probability space $(\Omega, \mathcal{F}, \mu)$.
- Symmetric, integrable function $W: \Omega \times \Omega \to \mathbb{R}_+$

Content

1. Notions of Convergence [BCLSV'06-'12, BCKL'13, BCCZ'14]
2. Limit objects [LS'06, BCCZ'14, BS'01, DJ'08]
3. Random Models and Estimation [BCCG'15]
4. Privacy Preserving Estimation [BCS'15]

B=Borgs, C=Chayes, C=Cohn, K=Kahn, L=Lovasz, S=Sos, S=Smith, V=Vestergombi, G=Gangly, S=Szegedy, Z=Zhao, B=Benjamini, D=Diaconis, J=Janson, S=Schramm

Notation

- Graph — $G = (V, E) = (V(G), E(G))$
- Degree of vertex x — $d_x = d_x(G) = \#\{y\colon xy \epsilon E(G)\}$
- Average degree — $\bar{d}(G)$
- Density of G — $\rho(G) = \frac{2|E(G)|}{|V(G)|^2} = \frac{\bar{d}(G)}{|V(G)|}$
- Sparse Graph — $\rho(G_n) \to 0$
- Dense Graph — $\liminf_{n\to\infty} \rho(G_n) > 0$
- Power Law Graph — $\Pr\{d_i \geq k\bar{d}\} \sim k^{-\sigma}$
- In this talk — $V(G_n) = [n] = \{1, \dots, n\}$
- Simplex in $\mathbb{R}_+^k$ — $\Delta_k = \{a \in \mathbb{R}_+^k \mid \sum_i a_i = 1\}$

1) Notions of Convergence

Examples we want to cover:

- Erdos-Renyi random graph $G_{n,p}$

  n vertices, each pair ij connected with probability p
- Stochastic block models:

  Given $B = (B_{\alpha\beta}) \epsilon [0,1]^{k\ x\ k}$ and $p \in \Delta_k$

  Choose colors $\alpha_1, \dots, \alpha_n$ i.i.d. in $[k]$ with $\Pr(\alpha_i = \alpha) = p_\alpha$

  Connect i and j with probability $B_{\alpha_i \alpha_j}$
- Fixed degree sequence:

  Given $d_1 \dots, d_n$, connect i and j with probability $d_i d_j / \sum d_i$

  Special case: Power-law graphs

$$d_i = n^\alpha \left(\frac{n}{i}\right)^\beta \rightarrow \Pr\{d_i \geq k\bar{d}\} \sim k^{-1/\beta}$$

1) Notions of Convergence

Task: compare two graphs on different vertex sets V, W

- Statistics involving subgraphs and colorings → Left and Right Convergence
- Decompose both V and W into q blocks and compare block averages → Convergence of quotients
- Blow up the points of both V and W, say into subintervals of $[0,1]$, and define a suitable metric on functions over $[0,1]^2$ → Metric convergence

1) Notions of Convergence
a) Left and Right Convergence

Def: A dense sequence of graphs G_n is left convergent if

$$\frac{1}{n^{|V(F)|}} \mathrm{N}(F, G_n)$$

converges for all connected graphs F, where $\mathrm{N}(F, G)$ is the # of subgraphs $F' \subset G$ that are isomorphic to F

Def: A sequence of graphs G_n with uniformly bounded degrees is left convergent if

$$\frac{1}{n} \mathrm{N}(F, G_n)$$

converges for all connected graphs F

Note: For sparse graphs with $\bar{d}(G_n) \to \infty$ left convergence does not give a nice notion of convergence, since we can, e.g., remove all triangles while removing only an $o(1)$ fraction of the edges if the graph is sufficiently sparse.

1) Notions of Convergence
a) Left and Right Convergence

- Given: a graph G on n vertices and a weighted graph H with edge weights β_{ij} and vertex weights α_i with $\sum_i \alpha_i = 1$
- Energy of a map $\phi: V(G) \to V(H)$ as

$$E_H(\phi) = \frac{1}{|E(G)|} \sum_{xy \in E(G)} \beta_{\phi(x)\phi(y)}$$

- Microcanonical Ensemble

$$\Omega_\alpha(G) = \left\{ \phi : \left| |\phi^{-1}(i)| - \alpha_i n \right| \leq 1 \right\}$$

- Micro canonical Gibbs distribution

$$\mu_{G,H}(\phi) = \frac{1}{Z_{MC}(G,H)} e^{-n\, E_H(\phi)}$$

where

$$Z_{MC}(G,H) = \sum_{\phi \in \Omega_\alpha(G)} e^{-n\, E_H(\phi)}$$

1) Notions of Convergence
a) Left and Right Convergence

- Micro canonical free energies

$$F_{MC}(G,H) = -\frac{1}{n}\log Z_{MC}(G_n,H)$$

Def: A sequence of graphs G_n is called right convergent if the micro canonical free energies $F_{MC}(G,H)$ converge for all H

Thm 1: For dense graph sequences, left and right convergence are equivalent.

Thm 2: For sequences with uniformly bounded degrees, right convergence is strictly stronger than left convergence

1) Notions of Convergence
b) Convergence of Quotients

- Given $\phi : V(G) \to [q]$, let V_i be the set of vertices of color i
- The quotient G/ϕ is a weighted graph on $[q]$ with vertex weights $\alpha_i = \frac{1}{n}|V_i|$ and edge weights

$$\beta_{ij} = \frac{1}{|E(G)|} \# \left\{ (u,v) \in V_i \times V_j,\ uv \in E(G) \right\}$$

- Set of q-quotients:

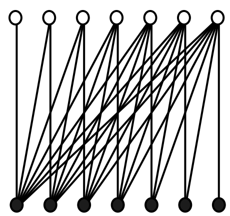
$$S_q(G) = \{ G/\phi \mid \phi : V(G) \to [q] \}$$

Rem: Given the quotients of G, we clearly can calculated ground state energies as well as microcanonical ground state energies

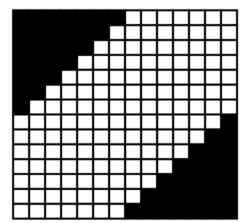
Def: G_n has convergent quotients if for all q, the sets $S_q(G_n)$ converge in the Hausdorff metric on subsets of $\mathbb{R}^{q+q^2}$

1) Notions of Convergence
c) Convergence in Metric

- Given G on $[n]$, let $I_1, \dots, , I_n$ be adjacent intervals of lengths $1/n$
- Define $W^G: [0,1]^2 \to [0,1]$ by setting W^G to $A_{ij}(G)$ on $I_i \times I_j$ where $A(G)$ is the adjacency matrix of G



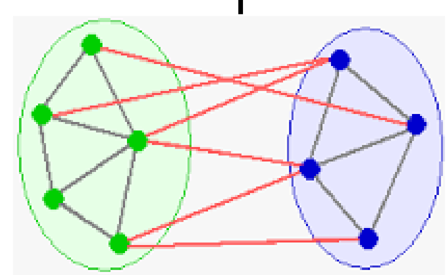
G



W^G

- Define the cut-norm

$$\|W\|_\square = \sup_{S,T \subset [0.1]} \left| \int_{S \times T} W(x,y)\,dx\,dy \right|$$



1) Notions of Convergence
c) Convergence in Metric

- Define the cut-norm

$$\|W\|_\square = \sup_{S,T\subset[0.1]} \left|\int_{S\times T} W(x,y)dxdy\right|$$

- Define cut distance

$$\delta_\square(W,W') = \inf_\phi \left\|W^\phi - W'\right\|_\square$$

where the infimum runs over bijections ϕ s.th. both ϕ and ϕ^{-1} are measure preserving

Def: G_n converges in metric if there exists a $W: [0,1]^2 \to \mathbb{R}_+$ s.t.

$$\delta_\square\left(\frac{1}{\rho(G_n)} W^{G_n}, W\right) \to 0$$

In this case, we say $G_n \to W$ in metric

2) Limit Object
a) Graphon

- A graphon over $[0,1]$ is an integrable function $W: [0,1]^2 \to \mathbb{R}_+$ such that $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$
- $F_{MC}(W,H)$
  $$= \min_{\rho} \sum \beta_{ij} \int W(x,y)\rho_i(x)\rho_j(y)dxdy + \sum \int \rho_i(x) \log \rho_i(x)dx$$
  where the min runs over all $\rho: [0,1] \to \Delta_q$ such that $\int \rho_i(x)dx = a_i$
- $S_q(W)$ is the set of all W/ρ, defined as the weighted graphs on $[q]$ with weights $\alpha_i = \int \rho_i(x)dx$ and $\beta_{ij} = \int W(x,y)\rho_i(x)\rho_j(y)dxdy$

Thm: Assume $\bar{d}(G_n) \to \infty$. Then $G_n \to W$ in metric $\Leftrightarrow F_{MC}(G_n, H) \to F_{MC}(W,H)$ for all $H \Leftrightarrow S_q(G_n) \to S_q(W)$ for all q

For dense graphs: Also equivalent to left convergence

2) Limit Object
b) Infinite Graphs

- [BS'01, BCKL'13] If a sequence G_n with uniformly bounded degrees is left convergent, then there exists an infinite, random rooted graph $(0, G_\infty)$ such that G_n equipped with a uniformly random root $x_n \in [n]$ converges to $(0, G_\infty)$ in distribution

3) Random Models and Estimation
a) Latent Position Graphs

Latent Position Graphs as "non-parametric graph models" [Hoff et al.'02, LS06, BR07]:

Let (Ω, F, π) be a probability space, and $W: \Omega \times \Omega \to \mathbb{R}_+$ be integrable, with $W(x,y) = W(y,x)$ for all $x, y \in \Omega$ and $\int W = 1$. Given a target density ρ_n

- Choose $x_1, \ldots, x_n \in \Omega$ independently according to π
- Connect i and j with probability $(\mathrm{P_n})_{\mathrm{ij}} = \min\{1, \rho_n W(x_i, x_j)\}$

Denote the resulting random graph by $G_n(\rho_n W)$

Task: Estimate W from a single sample of $G_n(\rho_n W)$

3) Random Models and Estimation
b) Identification Problem

Problem: If ϕ is a measure preserving map from (Ω', F', π') to (Ω, F, π) and $W' = W^{\phi}$, where $W^{\phi}(x', y') = W(\phi(x'), \phi(y'))$ then $G_n(\rho_n W)$ and $G_n(\rho_n W')$ are identically distributed $\Longrightarrow$ neither the underlying space, nor W can be identified from $G_n(\rho_n W)$

Thm [BCCG'15]: $G_n(\rho_n W)$ and $G_n(\rho_n W')$ are identically distributed iff there exists a graphon U over a third probability space such that $W = U^{\phi}$ and $W' = U^{\phi'}$ for two measure preserving maps ϕ and ϕ' (under these conditions, we call W and W' equivalent).

Thm [BCCG'15]: For each graphon, there exists an equivalent graphon over $[0,1]$ with the uniform distribution

3) Random Models and Estimation
b) Identification Problem (cont.)

- Rather than functions, graphons should be considered equivalence classes of functions, where w.l.o.g., we can choose a representative function $W: [0,1]^2 \to \mathbb{R}_+$

Goal: From a single observation of $G_n(\rho_n W)$, learn a good approximation of W in a distance between equivalence classes, say the δ_2-distance defined by

$$\delta_2(W, W') = \inf_{\phi} \left\| W^{\phi} - W' \right\|_2$$

where the infimum runs over bijections ϕ s.th. both ϕ and ϕ^{-1} are measure preserving

3) Random Models and Estimation
c) Previous Approaches

- Heuristic clustering approaches, e.g., Newman–Girvan modularity [NG '04]
- Strictly monotone degrees: assume there exists an equivalent graphon such that $\mathrm{D_W(y)} = \int W(x,y)dx$ is strictly monotone [assumed in most rigorous approaches]
- Maximum likelihood (ML): find a block model and a reordering that maximizes the probability of producing the observed graph [Wolfe, Olhede '13]. Gives consistency in δ_2 if W is bounded above, bounded away from zero, and Hoelder continuous
- All previous rigorous work assumes a bounded graphons

Goal: Replace these assumptions by integrability assumptions

3) Random Models and Estimation
d) Alignment and L_2-Estimation

For $B \in \mathbb{R}^{k\times k}$ and $A \in \mathbb{R}^{n\times n}$ define

$$\hat{\delta}_2(B, A) = \min_{\pi} \|B_\pi - A\|_2$$

where the min goes over all equipartitions, i.e., all $\pi: [n] \to [k]$ such that $\left| |\pi^{-1}(\{i\})| - \frac{n}{k} \right| < 1$ for all i and $(B_\pi)_{ij} = B_{\pi(i)\pi(j)}$.

Least Square (LS) Algorithm: Observing an instance G_n of a W-random graph, output the $k \times k$ block model

$$\hat{B}_n = \text{argmin}_B\, \hat{\delta}_2(B, G_n)$$

Main Thm A [BCCG '15]: Assume that $W \in L_2$, $\rho_n \to 0$ and $n\rho_n \to \infty$. If $k = k_n \to \infty$ sufficiently slowly with n, then

$$\delta_2\left(\frac{1}{\|W\|_1} W, \frac{1}{\rho(G_n)} W^{\hat{B}_n}\right) \to 0 \text{ a.s.}$$

4) Differentially Private Graphon Estimation

- Privacy Problem for Databases: Given a database with information about individuals, can we release statistical information which respects the privacy of individuals
- Relevant and difficult for graphs, since others (via their connection) reveal information about us.
- Examples of sensitive information contained in our social network graphs
  - sexual orientation [Jernigan, Mistree '09]
- Naïve anonymization can often be undone
- Need a more principled approach

4) Differentially Private Graphon Estimation, cont.

Def: A randomized algorithm A is ϵ-node-private if for all graphs G, G' that differ in at most one node, and all events E in the output space

$$e^{-\epsilon} \leq \frac{Pr(A(G') = E)}{Pr(A(G) = E)} \leq e^{\epsilon}$$

Previous work: Mostly individual graph statistics, and dense graphs

Idea: Use the exponential mechanism [McSherry-Talwar '07] to make the LS-algorithm private: given an input graph G_n, release an estimator $\hat{B}$ with

$$Pr\{\hat{B} = B\} = \frac{1}{Z} exp\left(-\frac{\epsilon}{2\Delta}\hat{\delta}_2(B, G_n)\right)$$

where Δ is chosen to guarantee ϵ-node-privacy

4) Differentially Private Graphon Estimation, cont.

Thm [BCS '15]: Let W be a bounded, graphon, let $\epsilon > 0$ and let $\rho_n = \omega(n^{-1} \log n)$ with $\rho_n \|W\|_\infty \leq 1$. Then the exponential algorithm can be modified in such a way that

a) it is ϵ-private on all input graphs G_n

b) If $G_n = G_n(\rho_n W)$, then

$$\delta_2\left(\frac{1}{\|W\|_1} W, \frac{1}{\|\hat{B}\|_1} W^{\hat{B}}\right) \to 0 \text{ in probability}$$

Thm [BCS '15]: If $G_n = G_n(\rho_n W)$, then the output of the modified exponential algorithm gives a consistent estimation of all multi-way cuts

Summary

- All notions of graph convergence for dense graphs except for left convergence generalize to sparse graphs with divergent average degrees, and are equivalent (for dense graphs, left convergence is equivalent as well)
- Inhomogeneous random graphs give non-parametric method to generate models of networks
- Methods from the theory of sparse graph limits allow us to give consistent estimators for any L_2-Graphon – for arbitrary slowly growing average degrees
- For bounded graphons and average degrees growing at least logarithmically, we solve the node-private graphon estimation problem without any additional assumptions