Phase transitions in inference: Tree reconstruction and community detection

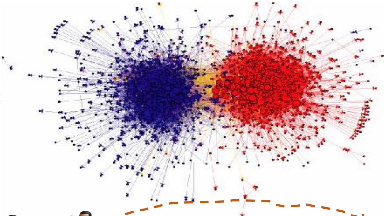
Laurent Massoulié

Inria, MSR-Inria Joint Centre

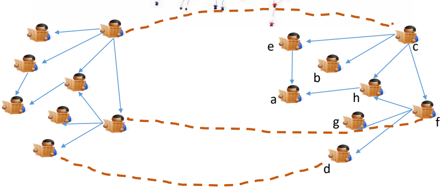
January 9, 2019

Context: inference of structure hidden in random noise

Community detection

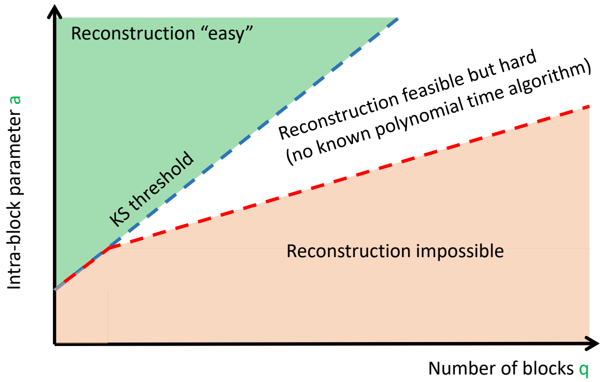


Graph alignment



Analysis of random instances
→ reveals phase transitions on **feasibility** and **hardness** of task
→ suggests new algorithms

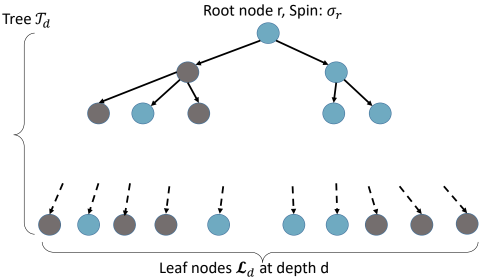
Phase diagram for community detection



Outline

- Tree reconstruction
    - Census reconstruction and Kesten-Stigum threshold
    - Belief Propagation and Information-Theoretic threshold
- Community detection
    - Relation to tree reconstruction
    - The hard phase
    - Above the Kesten-Stigum threshold
    - Links to random matrix theory

The tree reconstruction problem [Evans et al.'00]



Genealogy tree $\mathcal{T}$ with root r, "Spins" $\sigma_i \in [q]$ transmitted independently:
$\mathbb{P}(\sigma_i = t | \sigma_{p(i)} = s) = P_{st}$ for stochastic irreducible matrix P

Root spin $\sigma_r \sim \nu$ where ν: stationary distribution of P

$\rightarrow$ Can one infer σ_r non-trivially from $\mathcal{G}_d := (\mathcal{T}_d, \sigma_{\mathcal{L}_d})$ as $d \rightarrow \infty$?

Non-trivial reconstruction

Definition

Reconstructibility: $\lim_{d\to\infty} I(\sigma_r;\mathcal{G}_d) > 0$
where I: mutual information

Census at generation d:

$$X_d := \{X_{s,d}\}_{s\in[q]} \text{ where } X_{s,d} = \sum_{i\in\mathcal{L}_d} \mathbb{I}_{\sigma_i=s}$$

Definition

Census reconstructibility: $\lim_{d\to\infty} I(\sigma_r;X_d) > 0$

In the sequel: $\mathcal{T}$ a Galton-Watson branching tree with mean number of children α (e.g. $\mathcal{S}(i) \sim \mathrm{Poi}(\alpha)$)

Kesten-Stigum threshold for census reconstructibility [Mossel-Peres'03]

Spectrum of P: $\lambda_i(P)$, with $\lambda_1 = 1 \geq |\lambda_2| \cdots \geq |\lambda_q|$

Theorem

Census reconstructibility
holds if $\alpha|\lambda_2|^2 > 1$,
fails if $\alpha|\lambda_2|^2 < 1$.

[Proof elements for Poisson Galton-Watson tree]

Reconstruction and belief propagation [Pearl'82]

$\nu^i_{s,d} = \mathbb{P}(\sigma_i = s | \sigma_{\mathcal{L}_{i,d}}, \mathcal{T}_d)$ satisfy "belief propagation" recursion

$$\nu^i_{s,d} := \begin{cases} \frac{1}{Z^i}\nu_s \prod_{j\in\mathcal{S}(i)} \sum_{s_j\in[q]} \frac{\nu^j_{s_j}}{\nu_{s_j}} P_{ss_j}, & i \in V_d \setminus \mathcal{L}_d, \\ \mathbb{I}_{s=\sigma_i}, & i \in \mathcal{L}_d. \end{cases}$$

Both an algorithm and an analysis tool: let

$\pi_{t,d} = \text{ Law}\left((\nu^r_{s,d})_{s\in[q]} \middle| \sigma_r = t \right)$

Theorem

π_d *satisfies* **density evolution** *equation* $\pi_{d+1} = F(\pi_d)$*:*

$$\pi_{\tau,d+1} = Law\left(\left\{ U_s / (\textstyle\sum_{s'\in[q]} U_{s'}) \right\}_{s\in[q]} \right) \text{ where:}$$

$U_s = \nu_s \prod_{j=1}^D \sum_{s_j\in[q]} \frac{X^j_{s_j}}{\nu_{s_j}} P_{ss_j}$, $D \sim Poi(\alpha)$, $\sigma_1 \ldots, \sigma_D$ *i.i.d.* $\sim P_{\tau\cdot}$,
$X^1, \ldots, X^D$: *independent, and* $X^j \sim \pi_{\sigma_j,d}$, $j \in [D]$.

Reconstruction and belief propagation

Non-reconstructibility $\Leftrightarrow \forall t \in [q],\ \lim_{d\to\infty} \pi_{t,d} = \delta_\nu$

Hence: Uniqueness of fixed point of density evolution, $\pi = F(\pi)$ implies non-reconstructibility

Converse [Mézard-Montanari'06]: For ν uniform on $[q]$, non-uniqueness (for unconditional version of fixed point equation) implies reconstructibility

[Sly'09]: For symmetric Potts model with q, reconstruction differs from census reconstruction (3 distinct phases).

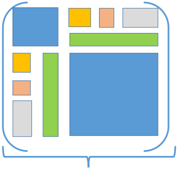
Stochastic block model and community detection

Matrix P assumed reversible ($\nu_s P_{st} = \nu_t P_{ts}$)

Definition

Stochastic block model $\mathcal{G} = \mathcal{G}(n, P, \alpha)$: random graph $\mathcal{G}$ on $[n]$,
s.t. $\sigma_{[n]}$ i.i.d. $\sim \nu$, and
Conditionally on $\sigma_{[n]}$, edges present independently,
$\mathbb{P}(i \sim j | \sigma_{[n]}) = \frac{R_{\sigma_i \sigma_j}}{n}$,
where $R_{st} = \alpha P_{st} / \nu_t$.

A=



+ W

Noise matrix

$\bar{A}$: block matrix (useful “signal”)

Block reconstruction

Definition

Block reconstruction feasible if can compute estimates $\hat{\sigma}_i, i \in [n]$ from observed graph $\mathcal{G}$ such that with high probability

$$\liminf_{n\to\infty} I_n(\sigma;\hat{\sigma}) > 0, \text{ where } I_n(\sigma;\hat{\sigma}) := \sum_{s,t\in[q]} p_n(s,t) \ln\left(\frac{p_n(s,t)}{q_n(s)\nu_t}\right).$$

Alternative notion:

Definition

$\text{overlap}(\hat{\sigma}_{[n]})$: max over permutations π of $[q]$ of $\frac{1}{n}\sum_{i\in[n]} \mathbb{I}_{\pi(\sigma_i)=\hat{\sigma}_i} - \sup_{s\in[q]} \nu_s$.
Strong reconstruction holds if with high probability for some $\hat{\sigma}_{[n]}$, $\liminf_{n\to\infty} \text{overlap}(\hat{\sigma}_{[n]}) > 0$

Strong reconstuction implies reconstruction; the two coincide for uniformly distributed ν.

Block reconstruction and tree reconstruction

Theorem (Mossel-Neeman-Sly'15)

Block reconstruction in $\mathcal{G}(n, P, \alpha)$ implies tree reconstruction in Poisson Galton-Watson tree with parameters P, α.

Proof ideas:

-local structure of $\mathcal{G}(n, P, \alpha)$

-approximate Markov random field property of $\sigma_{[n]}$ on $\mathcal{G}$

Existence of hard phase

Mean progeny matrix: $M = \alpha P$
Kesten-Stigum threshold: $|\lambda_2(M)|^2 > \lambda_1(M) = \alpha.$

Consider symmetric SBM:

$$R_{st} = \begin{cases} c_{in} & \text{if } s = t, \\ c_{out} & \text{if } s \neq t. \end{cases}, \ \alpha = \frac{c_{in}+(q-1)c_{out}}{q}.$$

Theorem (Banks et al.'16)

For the symmetric SBM with $q \geq 4$, strong reconstruction holds strictly below the Kesten-Stigum threshold for some $\hat{\sigma}_{[n]}$ that can be computed in exponential time.

Proof idea: **good partition** of $[n]$= balanced partition with inter- and intra-group edge counts close to target values;
show with first moment method sufficient condition for all good partitions to achieve positive overlap.

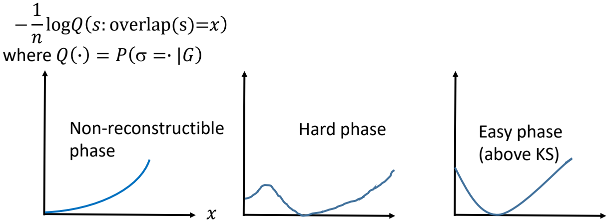
Above Kesten-Stigum threshold: belief propagation

Belief propagation iteration: $\psi_s^{i\to j} \propto \nu_s \prod_{k\sim i, k\neq j} \sum_{s_k\in[q]} \psi_{s_k}^{k\to i} R_{ss_k}$.

When converged, predicted distribution:
$\mathbb{P}(\sigma_i = s|\mathcal{G}) \approx \psi_s^i \propto \nu_s \prod_{j\sim i} \sum_{s_j\in[q]} \psi_{s_j}^{j\to i} R_{ss_j}$

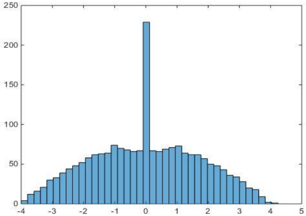
Conjecture (Decelle et al.'11)

Above Kesten-Stigum threshold, belief propagation on $\mathcal{G}(n,P,\alpha)$ *with random initialization converges, and fixed points* ψ^i *such that with high probability* $\liminf_{n\to\infty} I_n(\sigma;\psi) > 0$.



Failure of standard spectral methods

Standard spectral method: Associate $u \in [n]$ with $x_i(u), i \in [r]$ where eigenvector $x_i \leftrightarrow \lambda_i(A)$



$\rightarrow$ fails to correlate with σ_i:

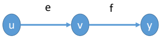
For $\alpha = O(1)$, largest eigenvalues $\lambda_i(A) \sim \sqrt{\frac{\log n}{\log\log n}}$ induced by highest degree nodes, $\sup_{j \in [n]} d_j \sim \frac{\log n}{\log\log n}$

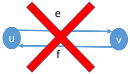
Corresponding eigenvector uncorrelated with $\sigma_{[n]}$

Spectral redemption [Krzakala et al'13]

Linearization of BP around trivial fixed point: $\psi_s^{i\to j} = \nu_s(1+\epsilon_s^{i\to j})$

$\to \epsilon = B \otimes P\epsilon$, where
B: non-backtracking matrix indexed by oriented edges $i \to j \in \vec{E}$ of $\mathcal{G}$,
$B_{u\to v, x\to y} = \mathbb{I}_{v=x}\mathbb{I}_{y\neq u}$





Asymmetric, such that B_{ef}^k= number of non-backtracking walks on $\mathcal{G}$ of $k+1$ edges starting with e and ending with f

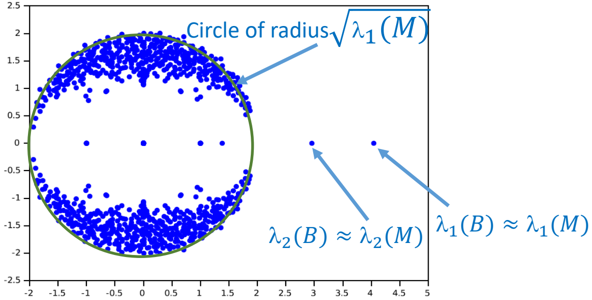
For $\lambda_i(M)$ and associated eigenvector x_i, let
$y_i \in \mathbb{R}^{\vec{E}}: \; y_i(u \to v) = x_i(\sigma_u)$
$z_i = B^\ell B^{\top \ell} y_i, \; \ell = \Theta(\log n)$
Let $r_0 = \sup\{i \in [q] : \lambda_i(M)^2 > \lambda_1(M)\}$.

Theorem (Bordenave, Lelarge, M.'15)

$i \in [r_0] \Rightarrow \lim_{n\to\infty} \lambda_i(B) = \lambda_i(M)$,
$i > r_0 \Rightarrow \limsup_{n\to\infty} |\lambda_i(B)| \le \sqrt{\lambda_1(M)}$
For $i \in [r_0]$ s.t. $\lambda_i(M)$ simple:
B admits eigenvector $\xi_i \leftrightarrow \lambda_i(B)$ s.t.
$\lim_{n\to\infty} \frac{\langle \xi_i, z_i \rangle}{\|\xi_i\| \cdot \|z_i\|} = 1$.
If moreover $i > 1$, then $\hat{\sigma}_u := \sqrt{n} \sum_{v \sim u} \xi_i(u \to v)$ s.t.
$\liminf_{n\to\infty} I_n(\sigma; \hat{\sigma}) > 0$.

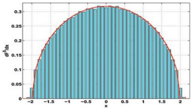
Non-backtracking spectrum of SBM $\mathcal{G}(n, P, \alpha)$

Illustration for symmetric SBM, $q = 2$, above Kesten-Stigum threshold:



Baik-Ben Arous-Péché phase transition

Low-rank deformation of random matrices:
$W_n \in \mathbb{R}^{n\times n}$ Wigner matrix, i.e. symmetric,
$\{W_{i,j}\}_{i<j}$ i.i.d. $\mathcal{N}(0, \sigma^2/n)$, $\{W_{ii}\}$ i.i.d. $\mathcal{N}(0, 2\sigma^2/n)$
Spectral measure $\to$ Wigner's semi-circle law
$\propto \sqrt{4\sigma^2 - x^2}\mathbb{I}_{|x|\le 2\sigma} dx$



Let $P_n \in \mathbb{R}^{n\times n}$ symmetric, fixed rank q and spectrum.
Let $r_0 = \{\sup\{i \in [q] : \lambda_i(P)^2 > \sigma^2\}$. Then:

Theorem (Benaych-Georges, Nadakuditi'11)

For $i \in [r_0]$, $\lim_{n\to\infty} \lambda_i(W_n + P_n) = \lambda_i(P_n) + \frac{\sigma^2}{\lambda_i(P_n)}$.
For $i > r_0$, $\limsup_{n\to\infty} |\lambda_i(W_n + P_n)| \le 2\sigma$

Parallel with SBM's non-backtracking spectrum

$$\sigma^2(BBP) \leftrightarrow \sum_{v\in[n]} \mathrm{var}(A_{uv}|\sigma_{[n]}) \to \alpha;$$

$$P_n \leftrightarrow \mathbb{E}(A|\sigma_{[n]}) \Rightarrow \mathrm{spectrum}(P_n) \to \mathrm{spectrum}(M)$$

Hence:

$$\text{KS condition } \lambda_i(M)^2 > \alpha \leftrightarrow \text{ BBP condition } \lambda_i(P_n)^2 > \sigma^2$$

Ihara-Bass formula

For any graph $\mathcal{G}$ with n nodes and m edges, any $z \in \mathbb{C}$,

$$(1-z^2)^{n-m}\det(I-zB) = \det(I - zA - z^2\mathrm{Diag}(\{d_i-1\}_{i\in[n]}))$$

Corollary

$\lambda \notin \{-1,0,1\}$ is an eigenvalue of B if and only if $det(\lambda^2 I - \lambda A - Diag(\{d_i - 1\}_{i\in[n]})) = 0$.

Corollary

If $\alpha \gg 1$ s.t. node degrees d_i concentrate, i.e. $\sup_{i\in[n]} |d_i - \alpha| = o(\alpha)$, then for $\lambda \in sp(B)$, $|\lambda| \ll \alpha$, one has $\lambda + \frac{\alpha}{\lambda} + o(1) \in sp(A)$.

Suggests that for non-sparse models $\alpha \gg 1$, spectral methods based on A succeed, and non-backtracking spectrum properties imply BBP transition

To conclude

- for denser models, “Approximate Message Passing” [Montanari] an approach of choice
- Finer phase transitions occur [Ricci-Tersenghi et al’18]
- Hard phase needs better understanding (basins of attraction, alternative dynamics)
- Statistical physics brings rich perspective on computational complexity