Partial Recovery

Consider the constant degree case

$$p = \frac{a}{n}, \quad q = \frac{b}{n}$$

and $a, b = O(1)$

Claim: Exact recovery is impossible

Proof: The degree of a node

$$\sim \text{Bin}(p, \tfrac{n}{2}) + \text{Bin}(q, \tfrac{n}{2})$$

$$\sim \text{Poi}\left(\frac{a+b}{2}\right)$$

where $X \sim \text{Poi}(\lambda)$ has $\mathbb{P}[X=k] = \dfrac{\lambda^k e^{-\lambda}}{k!}$

Thus $\Omega(1)$ fraction of nodes are isolated. $\blacksquare$

But can we still solve partial* recovery?

* i.e. our partition is $\frac{1}{2}+\varepsilon$ correlated with the true bisection

Conjecture [Decelle et al] If $(a-b)^2 > 2(a+b)$ (Kesten-Stigum Bound) then there is a polynomial time algorithm to solve partial recovery. Else if $(a-b)^2 < 2(a+b)$ it's information theoretically impossible

Where does this conjecture come from?

Belief Propagation

Let $\phi^i_{u \leftarrow v} \triangleq$ node u's belief that it is in community i, if v were not there

Assumption: neighbors of u's communities are independent conditioned on community of u

Hope: It's true on trees, and sparse random graphs are locally tree-like

We get the following update rules

$$\phi^i_{u \leftarrow v} \propto \prod_{\substack{w \neq v \\ \text{s.t. } (w,u) \in E}} \sum_{j=1}^{2} \phi^j_{w \leftarrow u} P_{i,j}$$

$P_{i,j}$: probability of edge btwn community i & j

Iterate until convergence, and compute marginals

$$\phi^i_u \propto \prod_{\substack{w \text{ s.t.} \\ (w,u) \in E}} \sum_{j=1}^{2} \phi^j_{w \leftarrow u} P_{i,j}$$

Some effect of missing edges — global interaction

But there is a trivial fixed point

$$\phi^i_{u\leftarrow v} = \frac{1}{2} \quad \forall i, u, v$$

i.e. no one knows anything

Decelle et al: The trivial fixed point is unstable iff $(a-b)^2 > 2(a+b)$

① If BP doesn't get stuck here, maybe it solves partial recovery?

② If BP does get stuck, maybe the problem is impossible?

Thm [Mossel, Neeman, Sly] [Massoulie]
Both parts of the conjecture are true

However spectral partitioning does not work — the maximum degree is $\Theta\left(\frac{\log n}{\log\log n}\right)$ and top eigenvectors are localized

Non Backtracking walks

Simpler approach following [Hopkins, Steurer]

Let $d = \frac{a+b}{2}$ = avg. degree

Then $p = \frac{(1+\varepsilon)d}{n}$, $q = \frac{(1-\varepsilon)d}{n}$

Now the Kesten-Stigum Bound becomes $\varepsilon^2 d > 1$

Goal: Find a polynomial in A_{ij}'s that can be used to estimate $x_i x_j$

$\uparrow \ \uparrow$

± 1 community membership

Consider $P_{ij}(A) = \frac{n}{\varepsilon d}\left(A_{ij} - \frac{d}{n}\right)$

We have $\mathbb{E}[P_{ij}(A)] = x_i x_j$ but its variance

$$\sim \mathbb{E}[P_{ij}(A)^2] \approx \left(\frac{n}{\varepsilon d}\right)^2 \left(\frac{d}{n}\right) = \frac{n}{\varepsilon^2 d}$$

is too large

Main Idea: Average over many walks

$$P_\alpha(A) = \prod_{(a,b)\in\alpha} P_{ab}(A)$$

It is still an unbiased estimator since

$$\mathbb{E}[P_\alpha(A)] = \prod_{(a,b)\in\alpha} \mathbb{E}[P_{ab}(A)]$$

$$= \prod_{(a,b)\in\alpha} x_a x_b = x_i x_j$$

It's variance is even larger

$$\mathbb{E}[P_\alpha(A)^2] \approx \left(\frac{n}{\varepsilon^2 d}\right)^\ell$$

where α is a length ℓ path

But if they were pairwise independent we could take

$$\frac{1}{|P_{ij}^{\ell}|} \sum_{\alpha \in P_{ij}^{\ell}} P_\alpha(A)$$

paths of length ℓ from i to j

which would have variance

$$\sim \left(\frac{1}{n^{\ell-1}}\right)\left(\frac{n}{\varepsilon^2 d}\right)^{\ell}$$

which is $o(1)$ when $\varepsilon^2 d > 1$

<u>Comment</u>: This does <u>not</u> work because up to scaling, we are computing the (i,j) entry of

$$\left(A - \frac{d}{n}\vec{1}\vec{1}^T\right)^{\ell}$$

and we know spectral methods fail b/c of high degree nodes!

But if you use selfavoiding walks, they're close enough to pairwise independent

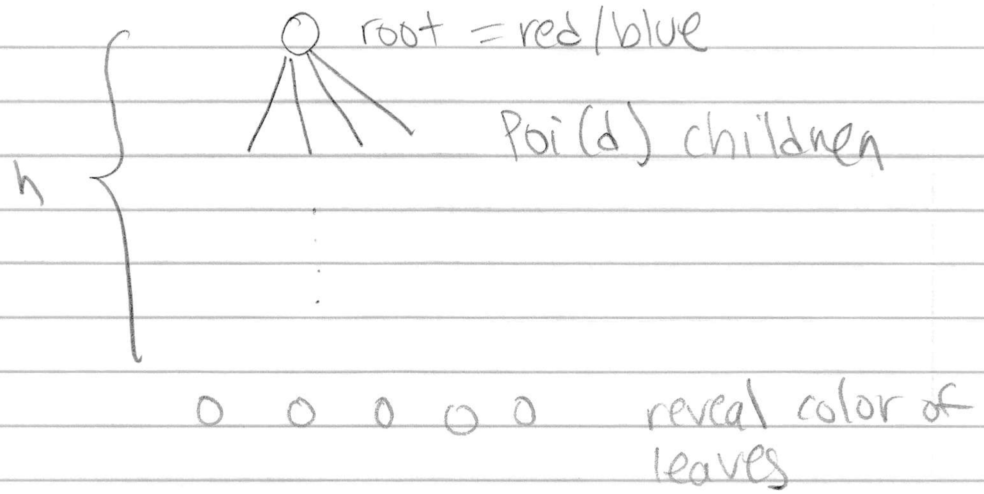
Essentially, want to bound

$$\sum_{\alpha,\beta \in P_{ij}^{\ell}} \mathbb{E}[P_\alpha(A) P_\beta(A)]$$

and for SAW, if α and β share r edges they must share at least r vertices, which reduces the number of possibilities by n^r

Broadcast Tree

Another view



Each child has probability $\frac{a}{a+b}$ of being same color, o.w different

Main Question: when can you guess the label of the root $\geq \frac{1}{2}+\varepsilon$ probability independent of h?

Thm [Kesten-Stigum] Can solve partial recovery if $(a-b)^2 > 2(a+b)$

Thm [Evans et al]: If $(a-b)^2 \leq 2(a+b)$ it's impossible

For upperbound, can take majority vote

[Kesten-Stigum] proved CLT for multi-type branching processes

Aside: Majority vote does not achieve optimal accuracy but BP/dynamic programming does

For lower bound, intuition is

$$\underbrace{I}_{\text{mutual information}}\big(\sigma(\rho);\sigma(x)\big) = \left(\frac{a-b}{a+b}\right)^{2h}$$

color of root/leaf

There are d^h leaves, so we think

$$I\big(\sigma(x);\underbrace{\sigma(x_1),\ldots\sigma(x_{d^h})}_{\text{all leaves}}\big) \approx d^h\left(\frac{a-b}{a+b}\right)^{2h}$$

$$= \left(\frac{a+b}{2}\right)^h\left(\frac{a-b}{a+b}\right)^{2h}$$

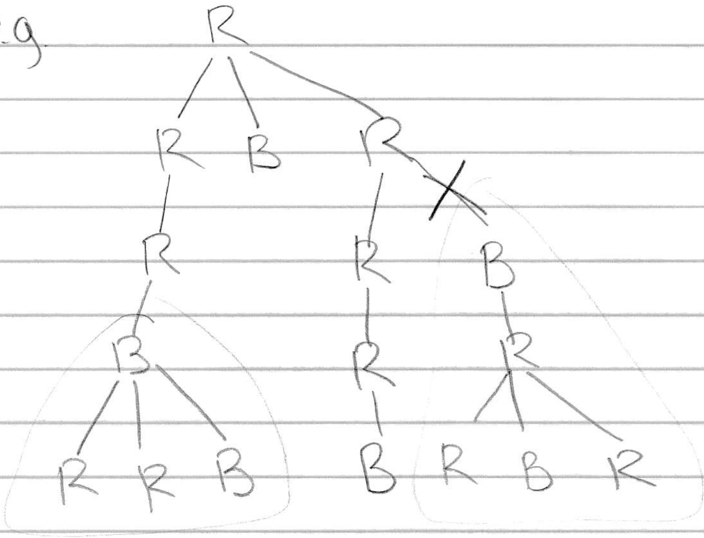
$$= \left[\frac{(a-b)^2}{2(a+b)}\right]^h$$

Mutual information doesn't satisfy subadditivity, but [Evans et al] give a coupling argument

How robust is the Kesten-Stigum bound?

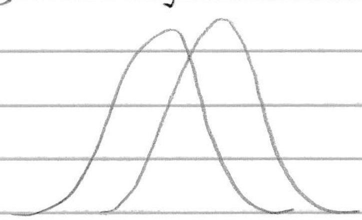
def: A monotone adversary can cut edges btwn nodes of opposite colors, remove subtree

e.g.

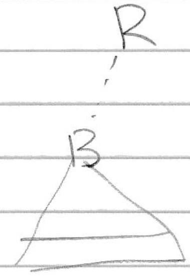


Claim: The adversary can whp flip the majority vote near the Kesten-Stigum bound

Proof: [sketch] # red - blue distributed as

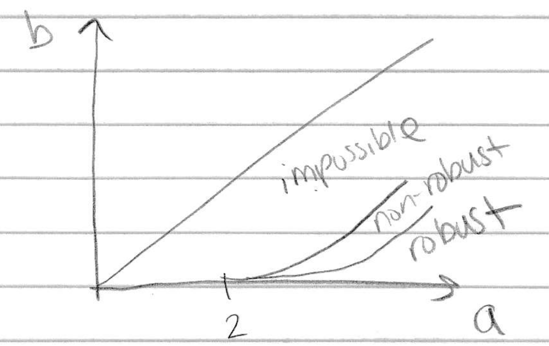


Likely to have subtrees where



more red than blue $\square$

Theorem [Moitra, Perry, Wein] Reconstruction in semirandom broadcast tree model is impossible for $(a-b)^2 \leq C_{a,b}(a+b)$ for some $C_{a,b} > 2$



These are the first random vs. semirandom separations

Main Complication: Need carefully designed adversary that maps into nice distribution

Theorem [Moitra, Perry, Wein] Recursive majority succeeds in semirandom broadcast tree model if $(a-b)^2 > (2+o(1))(a+b)\log\frac{a+b}{2}$

And also random vs. semirandom separations for community detection

Aside: Above Average-Case Analysis

[Spielman, Teng '01]:

"Explain why algorithms work well in practice, despite bad worst-case behavior"

usually called Beyond Worst-Case Analysis

Another example where semirandom models are informative

The Netflix Problem

$$M = \begin{array}{c} \text{users} \\ \text{500k} \end{array} \underbrace{\begin{bmatrix} \quad & \quad & \quad \\ & & \\ & & \end{bmatrix}}_{\substack{\text{movies} \\ \text{20k}}}$$

rating from 1-5 stars

<u>Assume</u>: M is low rank, e.g.

$$M = \underbrace{\begin{bmatrix} \; \\ \; \\ \; \end{bmatrix}\begin{bmatrix} \quad & \quad & \quad \end{bmatrix}}_{\text{drama}} + \underbrace{\begin{bmatrix} \; \\ \; \\ \; \end{bmatrix}\begin{bmatrix} \quad & \quad & \quad \end{bmatrix}}_{\text{Comedy}} + \cdots$$

and <u>incoherent</u>, ie its singular vectors far from standard basis vectors

<u>Theorem</u>: [Candes, Tao]. Given $m \gtrsim nr\log^2 n$ u.a.r observations from M, there is a polynomial time algorithm that whp recovers M <u>exactly</u>

Their approach was based on convex programming

Let Ω = observed entries

$$\min \|X\|_*$$

$$\text{s.t. } X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega$$

where $\|X\|_* = \sum_{i=1}^{n} \sigma_i(X)$ is the <u>nuclear norm</u>

<u>Remark</u>: This is a non-commutative generalization of ℓ_1-norm minimization in compressed sensing

sparsity : ℓ_1 :: rank : $\| \ \|_*$

Another powerful approach is alternating minimization

Repeat

$$U \leftarrow \underset{U}{\operatorname{argmin}} \sum_{(i,j)\in\Omega} \left| (UV^T)_{ij} - M_{ij} \right|^2$$

$$V \leftarrow \underset{V}{\operatorname{argmin}} \sum_{(i,j)\in\Omega} \left| (UV^T)_{ij} - M_{ij} \right|^2$$

<u>Theorem</u> [Keshavan et al] [Jain et al] [Hardt]
Alternating minimization with proper initialization succeeds whp given

$$Cnr^2 \frac{\|M\|_F^2}{\sigma_r^2} \text{ observations}$$

What if a monotone adversary reveals more of M?

Claim: Nuclear norm minimization still succeeds

Proof: It's just more constraints. $\boxtimes$

Observation: Alternating minimization fails in the semirandom model

[Cheng, Ge] give a nearly linear time preprocessing step to fix the nonconvex approach

---

Another application: GMMs

def: A semirandom GMM proceeds as follows

(1) samples $x_i \sim F$

(2) adversary can move point in the direction of the center

$$y_i = (1-\lambda)x_i + \lambda \mu_j$$

Can we still find an accurate clustering?

[Awasthi, Vijayaraghavan] Yes if the separation satisfies

$$\|\mu_i - \mu_j\| \geq \Delta \sigma$$

where $\|\Sigma_i\| \leq \sigma$ and

$$\Delta \geq C\sqrt{\min(k,d)\log n}$$

moreover this is information-theoretically tight

Are there other stochastic models where semirandomness offers new insights?