Recovering a Hidden Community in Networks

Jiaming Xu

Department of Statistics, The Wharton School
University of Pennsylvania

jiamingx@wharton.upenn.edu

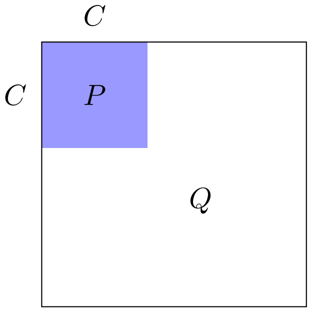
Joint work with Bruce Hajek and Yihong Wu

KAIST, October 16, 2015

A hidden community model [Deshpande-Montanari '13]

Given two distributions P, Q

- A community $C \subset [n]$ of size K is chosen uniformly

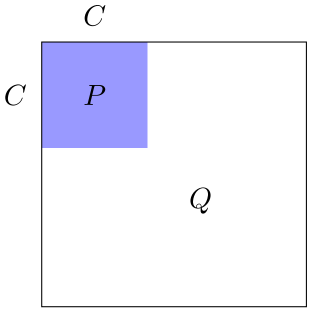


A hidden community model [Deshpande-Montanari '13]

Given two distributions P, Q

- A community $C \subset [n]$ of size K is chosen uniformly
- $A_{ij} = A_{ji}$,

$$A_{ij} \sim \begin{cases} P & \text{if } i, j \in C \\ Q & \text{otherwise} \end{cases}$$

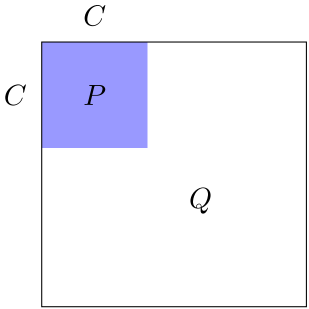


A hidden community model [Deshpande-Montanari '13]

Given two distributions P, Q

- A community $C \subset [n]$ of size K is chosen uniformly
- $A_{ij} = A_{ji}$,

$$A_{ij} \sim \begin{cases} P & \text{if } i,j \in C \\ Q & \text{otherwise} \end{cases}$$



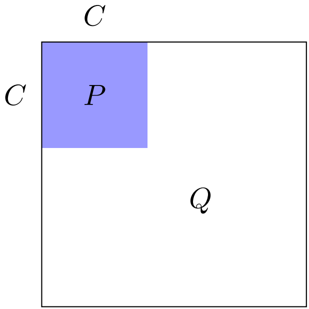
Goal: recover C given A

A hidden community model [Deshpande-Montanari '13]

Given two distributions P, Q

- A community $C \subset [n]$ of size K is chosen uniformly
- $A_{ij} = A_{ji}$,

$$A_{ij} \sim \begin{cases} P & \text{if } i, j \in C \\ Q & \text{otherwise} \end{cases}$$



Goal: recover C given A

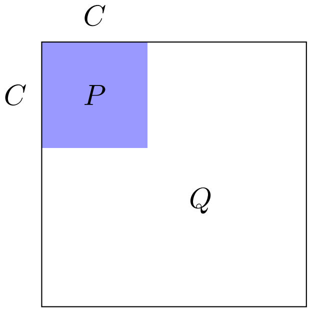
- $P = \mathrm{Bern}(p), Q = \mathrm{Bern}(q) \to$ Bruce's plenary talk at ITW 2015

A hidden community model [Deshpande-Montanari '13]

Given two distributions P, Q

- A community $C \subset [n]$ of size K is chosen uniformly
- $A_{ij} = A_{ji}$,

$$A_{ij} \sim \begin{cases} P & \text{if } i,j \in C \\ Q & \text{otherwise} \end{cases}$$



Goal: recover C given A

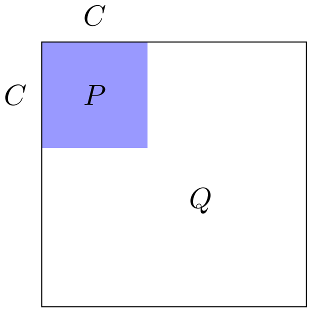
- $P = \mathrm{Bern}(p), Q = \mathrm{Bern}(q) \to$ Bruce's plenary talk at ITW 2015
- $P = \mathcal{N}(\mu, 1), Q = \mathcal{N}(0,1)$ with $\mu > 0 \to$ This talk

A hidden community model [Deshpande-Montanari '13]

Given two distributions P, Q

- A community $C \subset [n]$ of size K is chosen uniformly
- $A_{ij} = A_{ji}$,

$$A_{ij} \sim \begin{cases} P & \text{if } i, j \in C \\ Q & \text{otherwise} \end{cases}$$



Goal: recover C given A

- $P = \mathrm{Bern}(p), Q = \mathrm{Bern}(q) \to$ Bruce's plenary talk at ITW 2015
- $P = \mathcal{N}(\mu, 1), Q = \mathcal{N}(0, 1)$ with $\mu > 0 \to$ This talk
- $A = \mu \mathbf{1}_C \mathbf{1}_C^\top + Z$, where Z is Wigner (submatrix localization)

Exact versus weak recovery

Model parameters (K, μ) as functions of n

$$C \quad \longrightarrow \quad A \quad \longrightarrow \quad \widehat{C}$$

- Exact recovery:

$$\mathbb{P}\{\widehat{C} = C\} \xrightarrow{n \to \infty} 1$$

- Weak recovery:

$$\mathbb{E}\left[|\widehat{C} \Delta C|\right] = o(K)$$

Exact versus weak recovery

Model parameters (K, μ) as functions of n

$$C \quad \longrightarrow \quad A \quad \longrightarrow \quad \widehat{C}$$

- Exact recovery:

$$\mathbb{P}\{\widehat{C} = C\} \xrightarrow{n\to\infty} 1$$

- Weak recovery:

$$\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$$

Two fundamental Limits

- What are the information-theoretic limits of recovery?
- What are the computational limits of recovery in polynomial, or linear time?

Computational gap when $\mu = \Theta(1)$

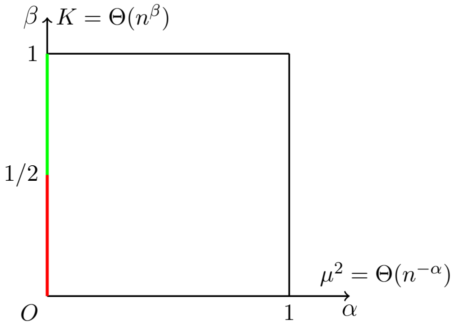
- $K = \Omega(\log n)$: exact recovery is possible via ML estimator
- $K \geq \sqrt{n/\mathrm{e}}$: message passing succeeds in nearly linear-time [Deshpande-Montanari '13]
- $K = o(\sqrt{n})$: exact recovery is believed to be hard [Deshpande-Montanari '15]

Computational gap when $\mu = \Theta(1)$

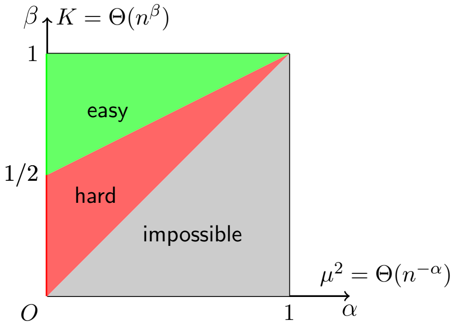
- $K = \Omega(\log n)$: exact recovery is possible via ML estimator
- $K \geq \sqrt{n/\mathrm{e}}$: message passing succeeds in nearly linear-time [Deshpande-Montanari '13]
- $K = o(\sqrt{n})$: exact recovery is believed to be hard [Deshpande-Montanari '15]

There is a computational gap: $K \in [\Omega(\log n), o(\sqrt{n})]$!

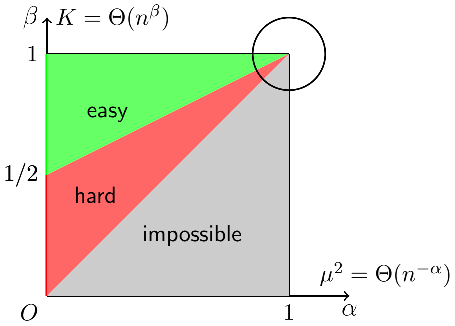
Computational gap in general [Chen-X. '14]



Computational gap in general [Chen-X. '14]

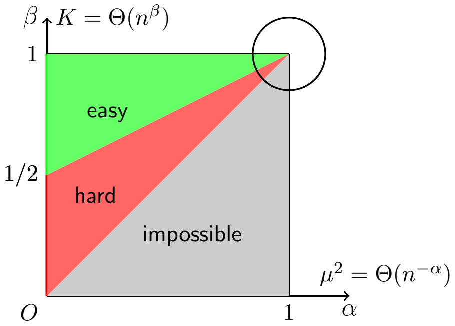


- exact recovery in the red regime is at least as hard as recovering a clique of size $K = o(\sqrt{n})$ [Cai-Liang-Rahklin '15]



- exact recovery in the red regime is at least as hard as recovering a clique of size $K = o(\sqrt{n})$ [Cai-Liang-Rahklin '15]

Computational gap in general [Chen-X. '14]

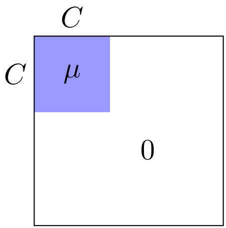


- exact recovery in the red regime is at least as hard as recovering a clique of size $K = o(\sqrt{n})$ [Cai-Liang-Rahklin '15]
- Question 1: Can we obtain the information limits with sharp constants when $K = n^{1+o(1)}$?
- Question 2: How about weak recovery?

Outline of the remainder of the talk

1. Performance of algorithms for weak recovery (i.e., $\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$):
   - degree thresholding
   - optimized message passing
   - spectral limit
2. Algorithms for exact recovery (i.e. $\mathbb{P}\{\widehat{C} = C\} \xrightarrow{n\to\infty} 1$)
   - weak recovery plus voting

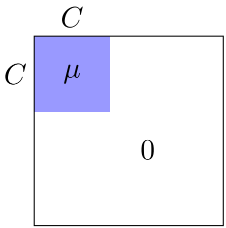
A naïve row-sum-thresholding algorithm



$$r_i = \sum_j A_{ij} \sim \begin{cases} \mathcal{N}(K\mu, n) & i \in C \\ \mathcal{N}(0, n) & i \notin C \end{cases}$$

$$\lambda \triangleq \frac{K^2\mu^2}{n}$$ Signal-to-noise ratio

A naïve row-sum-thresholding algorithm

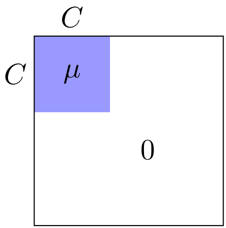


$$r_i = \sum_j A_{ij} \sim \begin{cases} \mathcal{N}(K\mu, n) & i \in C \\ \mathcal{N}(0, n) & i \notin C \end{cases}$$

$$\lambda \triangleq \frac{K^2\mu^2}{n} \qquad \text{Signal-to-noise ratio}$$

- If $K \asymp n$, needs $\lambda \to \infty$ for weak recovery (coincides with information limit) [Hajek-Wu-X. 'Info15]

A naïve row-sum-thresholding algorithm



$$r_i = \sum_j A_{ij} \sim \begin{cases} \mathcal{N}(K\mu, n) & i \in C \\ \mathcal{N}(0, n) & i \notin C \end{cases}$$

$$\lambda \triangleq \frac{K^2\mu^2}{n}$$ Signal-to-noise ratio

- If $K \asymp n$, needs $\lambda \to \infty$ for weak recovery (coincides with information limit) [Hajek-Wu-X. 'Info15]
- If $K = o(n)$, can we do better than naïve thresholding?

Message passing framework [Deshpande-Montanari '13]

- $\theta_{i \to j}^t$: message from i to j at iteration t
- θ_i^t: i's belief at iteration t

Message passing framework [Deshpande-Montanari '13]

- $\theta^t_{i\to j}$: message from i to j at iteration t
- θ^t_i: i's belief at iteration t

Let $W = A/\sqrt{n}$

$$\theta^{t+1}_{i\to j} = \sum_{\ell\in[n]\setminus\ \{i,j\}} W_{\ell i} f(\theta^t_{\ell\to i}, t), \quad \forall j \neq i \in [n]$$

$$\theta^{t+1}_{i} = \sum_{\ell\in[n]\setminus\{i\}} W_{\ell i} f(\theta^t_{\ell\to i}, t)$$

$$\theta^{0}_{i\to j} \equiv 0$$

Message passing framework [Deshpande-Montanari '13]

- $\theta^t_{i\to j}$: message from i to j at iteration t
- θ^t_i: i's belief at iteration t

Let $W = A/\sqrt{n}$

$$\theta^{t+1}_{i\to j} = \sum_{\ell\in[n]\setminus \{i,j\}} W_{\ell i} f(\theta^t_{\ell\to i}, t), \quad \forall j \neq i \in [n]$$

$$\theta^{t+1}_{i} = \sum_{\ell\in[n]\setminus\{i\}} W_{\ell i} f(\theta^t_{\ell\to i}, t)$$

$$\theta^0_{i\to j} \equiv 0$$

- Non-backtracking property: $\theta^{t+1}_{i\to j}$ does not depend on $\theta^t_{j\to i}$
- At iteration 1

$$\theta^1_{i\to j} = \sum_{\ell\in[n]\setminus \{i,j\}} W_{\ell i} f(0,0) \to \text{row-sum-thresholding}$$

- Q: how to optimize over $f(\cdot, t)$?

Optimal choice of f

Suppose

$$\theta_{i\to j}^t \sim \begin{cases} \mathcal{N}(\mu_t, \tau_t^2) & \text{if } i \in C, \\ \mathcal{N}(0, \tau_t^2) & \text{if } i \notin C. \end{cases}$$

Optimal choice of f

Suppose

$$\theta_{i\to j}^t \sim \begin{cases} \mathcal{N}(\mu_t, \tau_t^2) & \text{if } i \in C, \\ \mathcal{N}(0, \tau_t^2) & \text{if } i \notin C. \end{cases}$$

State evolution equations: $Z \sim \mathcal{N}(0,1)$, $\lambda = \frac{K^2\mu^2}{n}$,

$$\mu_{t+1} = \sqrt{\lambda}\mathbb{E}\left[f(\mu_t + \sqrt{\tau_t}Z, t)\right],$$
$$\tau_{t+1} = \mathbb{E}\left[f(\sqrt{\tau_t}Z, t)^2\right],$$

Optimal choice of f

Suppose

$$\theta^t_{i\to j} \sim \begin{cases} \mathcal{N}(\mu_t, \tau_t^2) & \text{if } i \in C, \\ \mathcal{N}(0, \tau_t^2) & \text{if } i \notin C. \end{cases}$$

State evolution equations: $Z \sim \mathcal{N}(0,1)$, $\lambda = \frac{K^2\mu^2}{n}$,

$$\mu_{t+1} = \sqrt{\lambda}\mathbb{E}\left[f(\mu_t + \sqrt{\tau_t}Z, t)\right],$$
$$\tau_{t+1} = \mathbb{E}\left[f(\sqrt{\tau_t}Z, t)^2\right],$$

Maximize signal-to-noise ratio: $\max_f \frac{\mu_{t+1}}{\tau_{t+1}}$

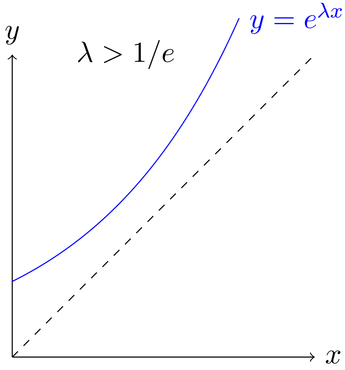
$$\Rightarrow f(x,t) = \exp(\mu_t(x-\mu_t))$$
$$\tau_{t+1} = 1$$
$$\mu_{t+1}^2 = \lambda e^{\mu_t^2}$$

Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda \mathrm{e}^{\mu_t^2}$$

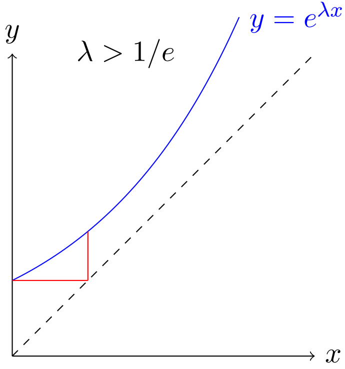
Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda e^{\mu_t^2}$$



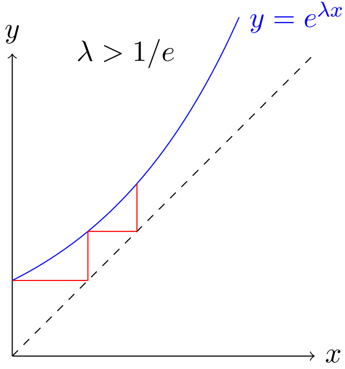
Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda e^{\mu_t^2}$$



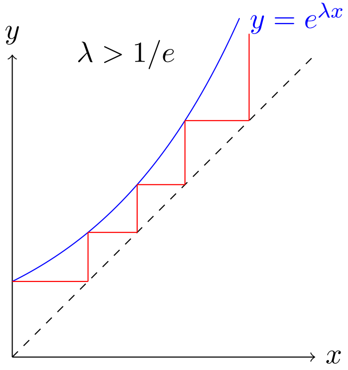
Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda e^{\mu_t^2}$$



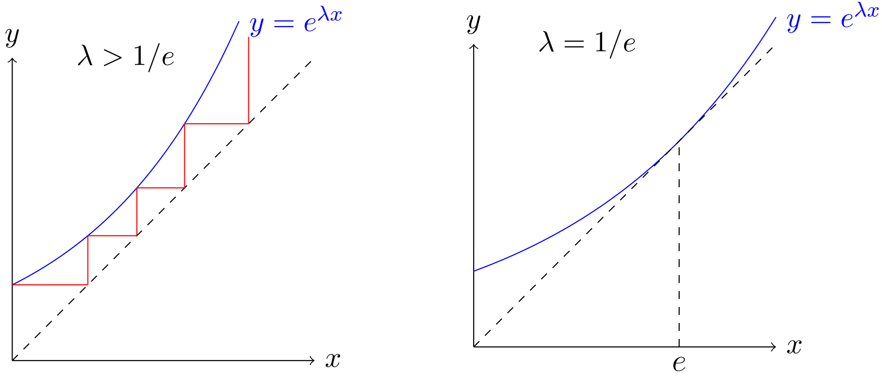
Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda \mathrm{e}^{\mu_t^2}$$



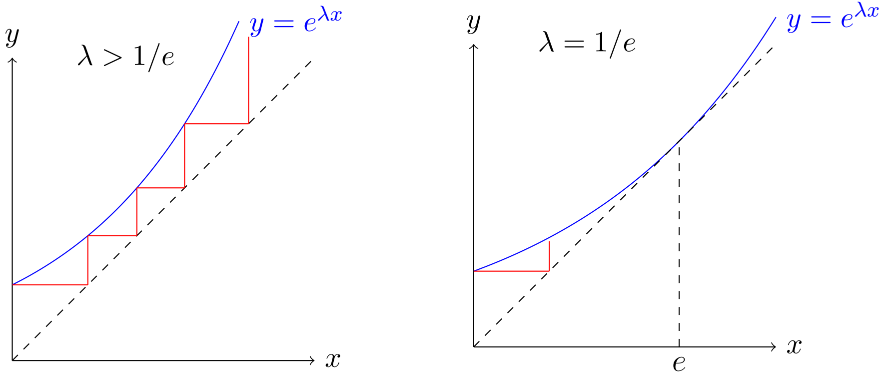
Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda \mathrm{e}^{\mu_t^2}$$



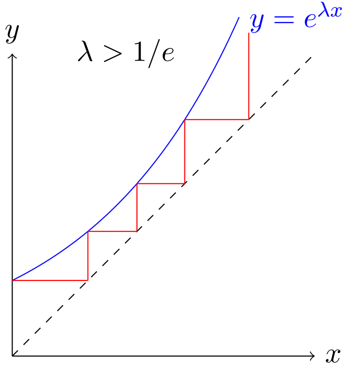
Critical threshold for optimized message passing

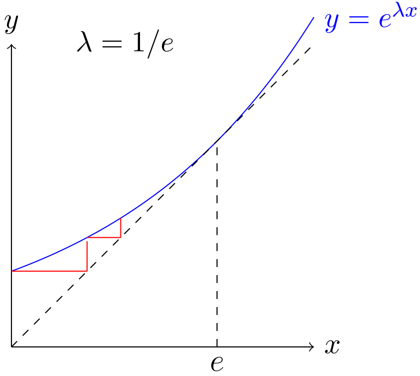
$$\mu_{t+1}^2 = \lambda e^{\mu_t^2}$$



Critical threshold for optimized message passing

$$\mu_{t+1}^2 = \lambda e^{\mu_t^2}$$





Weak recovery by optimized message passing

Theorem (Hajek-Wu-X. 'submatrix15)

Assume $\Omega(\sqrt{n}) \le K \le o(n)$. *Fix*

$$\lambda \triangleq \frac{K^2\mu^2}{n} > 1/\mathrm{e}.$$

Then the optimized message passing (+ spectral clean-up) achieves weak recovery.

Weak recovery by optimized message passing

Theorem (Hajek-Wu-X. 'submatrix15)

Assume $\Omega(\sqrt{n}) \le K \le o(n)$. *Fix*

$$\lambda \triangleq \frac{K^2\mu^2}{n} > 1/\mathrm{e}.$$

Then the optimized message passing (+ spectral clean-up) achieves weak recovery.

Remarks

- Use method of moments; choose f to be poly. to approx. $\exp(x)$
- The case of $K = \Theta(\sqrt{n})$ is proved in [Deshpande-Montanari '13]
- Main challenge: a larger $K \to$ a larger portion of messages is sent between vertices in community

Weak recovery by linear message passing (spectral method)

$$\theta_{i \to j}^{t+1} = \sum_{\ell \in [n] \setminus \{i,j\}} W_{\ell i} f(\theta_{\ell \to i}^{t}, t), \quad \forall j \neq i \in [n]$$

Choose $f(x,t) \equiv x \to$ Power method for computing leading eigenvector

Weak recovery by linear message passing (spectral method)

$$\theta_{i\to j}^{t+1} = \sum_{\ell \in [n] \backslash \{i,j\}} W_{\ell i} f(\theta_{\ell \to i}^{t}, t), \quad \forall j \neq i \in [n]$$

Choose $f(x,t) \equiv x \to$ Power method for computing leading eigenvector

Corollary (Hajek-Wu-X. 'submatrix15)

Assume $\Omega(\sqrt{n}) \leq K \leq o(n)$. *Fix*

$$\lambda \triangleq \frac{K^2 \mu^2}{n} > 1.$$

Then the linear message passing achieves weak recovery.

Belief: spectral limit is $\lambda \geq 1$ [Deshpande-Montanari '13]

Computational gap for weak recovery

Theorem (Hajek-Wu-X. 'info15)

Assume that $K = o(n)$ *and* $K \to \infty$. *If*

$$\liminf_{n\to\infty} \frac{K\mu^2}{4\log\frac{n}{K}} > 1,$$

then weak recovery is possible. If weak recovery is possible, then

$$\liminf_{n\to\infty} \frac{K\mu^2}{4\log\frac{n}{K}} \geq 1.$$

Computational gap for weak recovery

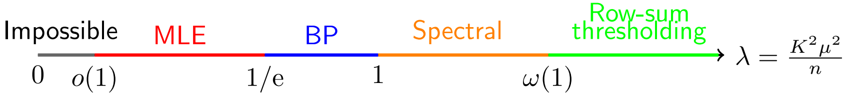
Theorem (Hajek-Wu-X. 'info15)

Assume that $K = o(n)$ and $K \to \infty$. If

$$\liminf_{n \to \infty} \frac{K\mu^2}{4 \log \frac{n}{K}} > 1,$$

then weak recovery is possible. If weak recovery is possible, then

$$\liminf_{n \to \infty} \frac{K\mu^2}{4 \log \frac{n}{K}} \geq 1.$$



There is a computational gap for weak recovery: $\lambda \in [o(1), 1/\mathrm{e}]$!

Recall the outline of the talk

1. Performance of algorithms for weak recovery (i.e., $\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$):
   - degree thresholding
   - optimized message passing
   - spectral limit
2. Algorithms for exact recovery (i.e. $\mathbb{P}\{\widehat{C} = C\} \xrightarrow{n\to\infty} 1$)
   - weak recovery plus voting

Voting for exact recovery given weak recovery

Given $\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$, let "votes" received by i: $v_i = \sum_{j \in \widehat{C}} A_{ij}$

Voting for exact recovery given weak recovery

Given $\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$, let "votes" received by i: $v_i = \sum_{j \in \widehat{C}} A_{ij}$

Hypothesis testing for a single vertex

	H_0	$vs.$	H_1
priors:	$\pi_0 = 1 - \frac{K}{n}$		$\pi_1 = \frac{K}{n}$
distributions:	$\mathcal{N}(K\mu, K)$		$\mathcal{N}(0, K)$

Voting for exact recovery given weak recovery

Given $\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$, let "votes" received by i: $v_i = \sum_{j\in\widehat{C}} A_{ij}$

Hypothesis testing for a single vertex

$$\begin{array}{lccc} & H_0 & vs. & H_1 \\ \text{priors:} & \pi_0 = 1 - \frac{K}{n} & & \pi_1 = \frac{K}{n} \\ \text{distributions:} & \mathcal{N}(K\mu, K) & & \mathcal{N}(0, K) \end{array}$$

- Exact recovery is guaranteed if $p_e = o(1/n)$, i.e.,

$$\liminf_{n\to\infty} \frac{\sqrt{K}\mu}{\sqrt{2\log K} + \sqrt{2\log n}} > 1 \text{ (voting condition)}$$

Voting for exact recovery given weak recovery

Given $\mathbb{E}\left[|\widehat{C}\Delta C|\right] = o(K)$, let "votes" received by i: $v_i = \sum_{j \in \widehat{C}} A_{ij}$

Hypothesis testing for a single vertex

	H_0	$vs.$	H_1
priors:	$\pi_0 = 1 - \frac{K}{n}$		$\pi_1 = \frac{K}{n}$
distributions:	$\mathcal{N}(K\mu, K)$		$\mathcal{N}(0, K)$

- Exact recovery is guaranteed if $p_e = o(1/n)$, i.e.,

$$\liminf_{n\to\infty} \frac{\sqrt{K}\mu}{\sqrt{2\log K} + \sqrt{2\log n}} > 1 \text{ (voting condition)}$$

- Linear additional complexity (beyond weak recovery)
- Analysis of weak recovery is key; exact recovery an afterthought

Voting for exact recovery given weak recovery

Given $\mathbb{E}\left[|\widehat{C} \Delta C|\right] = o(K)$, let "votes" received by i: $v_i = \sum_{j \in \widehat{C}} A_{ij}$

Hypothesis testing for a single vertex

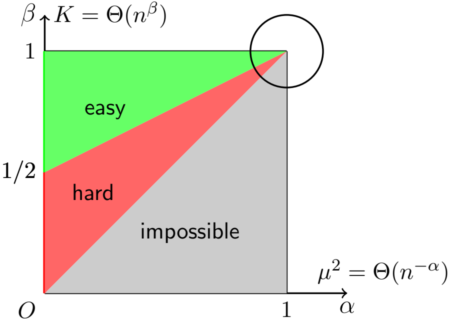
	H_0	$vs.$	H_1
priors:	$\pi_0 = 1 - \frac{K}{n}$		$\pi_1 = \frac{K}{n}$
distributions:	$\mathcal{N}(K\mu, K)$		$\mathcal{N}(0, K)$

- Exact recovery is guaranteed if $p_e = o(1/n)$, i.e.,

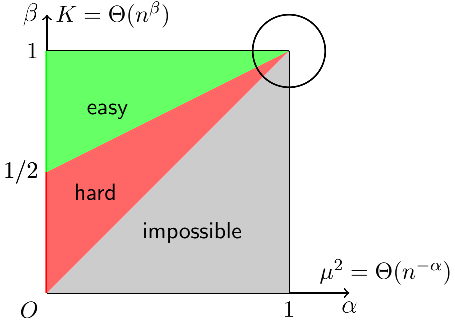
$$\liminf_{n \to \infty} \frac{\sqrt{K}\mu}{\sqrt{2 \log K} + \sqrt{2 \log n}} > 1 \text{ (voting condition)}$$

- Linear additional complexity (beyond weak recovery)
- Analysis of weak recovery is key; exact recovery an afterthought
- Gain independence by method of successive withholding

Computational gap for exact recovery

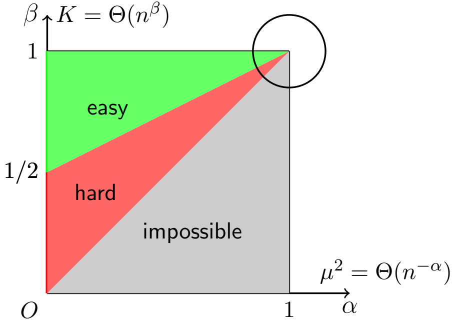


Computational gap for exact recovery



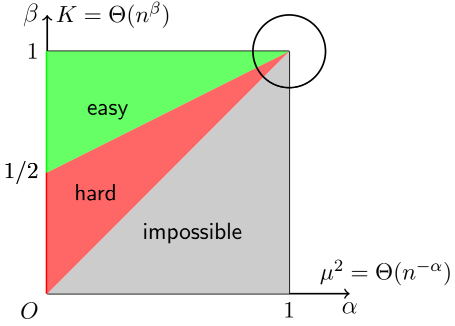
- If $K = \omega(n/\log n)$, $\lambda > 1/\mathrm{e}$ is implied by the voting condition and hence no computational gap for exact recovery!

Computational gap for exact recovery



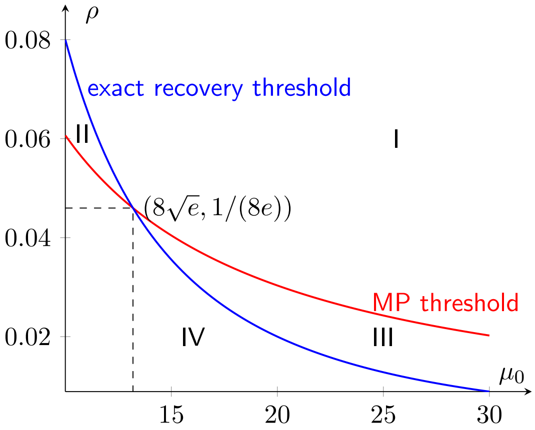
- If $K = \omega(n/\log n)$, $\lambda > 1/\mathrm{e}$ is implied by the voting condition and hence no computational gap for exact recovery!
- Computational gap for exact recovery emerges at $K = O(n/\log n)$

Computational gap for exact recovery



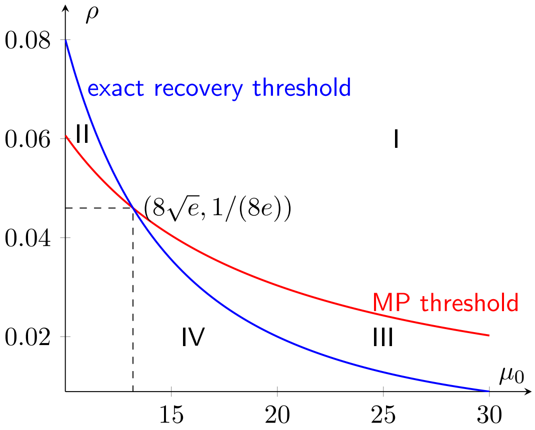
- If $K = \omega(n/\log n)$, $\lambda > 1/\mathrm{e}$ is implied by the voting condition and hence no computational gap for exact recovery!
- Computational gap for exact recovery emerges at $K = O(n/\log n)$
- Semidefinite programming relaxations provably cannot close the gap!

$$K = \rho n / \log n, \quad \mu^2 = \mu_0^2 \log^2 n / n$$



weak recovery by MLE is possible throughout open positive quadrant

$$K = \rho n / \log n, \quad \mu^2 = \mu_0^2 \log^2 n / n$$



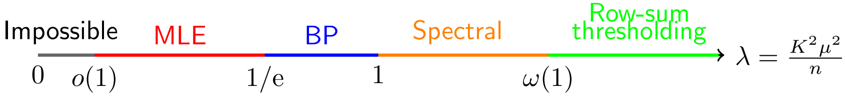
weak recovery by MLE is possible throughout open positive quadrant

Regions III $\cup$ IV $=$ conjectured computational gap for weak recovery

Region III $=$ conjectured computational gap for exact recovery

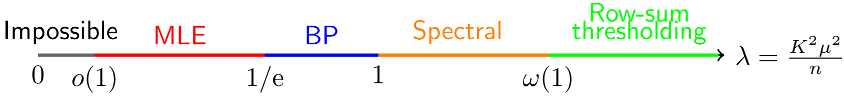
Conclusions

- Computational gap for weak recovery emerges at $K = o(n)$



Conclusions

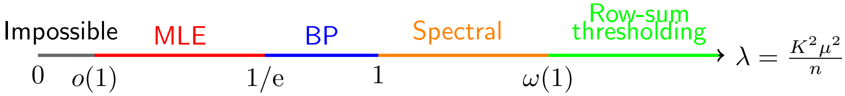
- Computational gap for weak recovery emerges at $K = o(n)$



- Computational gap for exact recovery emerges at $K = \frac{n}{8e \log n}$

Conclusions

- Computational gap for weak recovery emerges at $K = o(n)$



- Computational gap for exact recovery emerges at $K = \frac{n}{8\mathrm{e}\log n}$
- Can be generalized to the bicluster case

Conclusions

- Computational gap for weak recovery emerges at $K = o(n)$

- Computational gap for exact recovery emerges at $K = \frac{n}{8e \log n}$
- Can be generalized to the bicluster case

Reference:

1. B. Hajek, Y. Wu, and J. Xu, "Submatrix localization via message passing." arXiv, October 2015.
2. B. Hajek, Y. Wu, and J. Xu, "Information limits for recovering a hidden community." arXiv 1509.07859, September 2015.