Sharp Phase Transitions in Estimation with Low-Degree Polynomials

Alex Wein

University of California, Davis

Based on 2 papers…

- “Sharp Phase Transitions in Estimation with Low-Degree Polynomials” (with Youngtak Sohn)
- “Stochastic Block Models with Many Communities and the Kesten-Stigum Bound” (with Byron Chin, Elchanan Mossel, Youngtak Sohn)

Stochastic Block Model (SBM)

Stochastic Block Model (SBM)

- n vertices partitioned into q hidden “communities” of equal size

Stochastic Block Model (SBM)

- n vertices partitioned into q hidden “communities” of equal size
- Edge (i, j) occurs with probability…
  - a/n if i, j same community
  - b/n if i, j different community

Stochastic Block Model (SBM)

- n vertices partitioned into q hidden "communities" of equal size
- Edge (i, j) occurs with probability…
  - a/n if i, j same community
  - b/n if i, j different community
- Goal: given the graph, estimate the hidden partition

Stochastic Block Model (SBM)

- n vertices partitioned into q hidden "communities" of equal size
- Edge (i, j) occurs with probability...
  - a/n if i, j same community
  - b/n if i, j different community
- Goal: given the graph, estimate the hidden partition

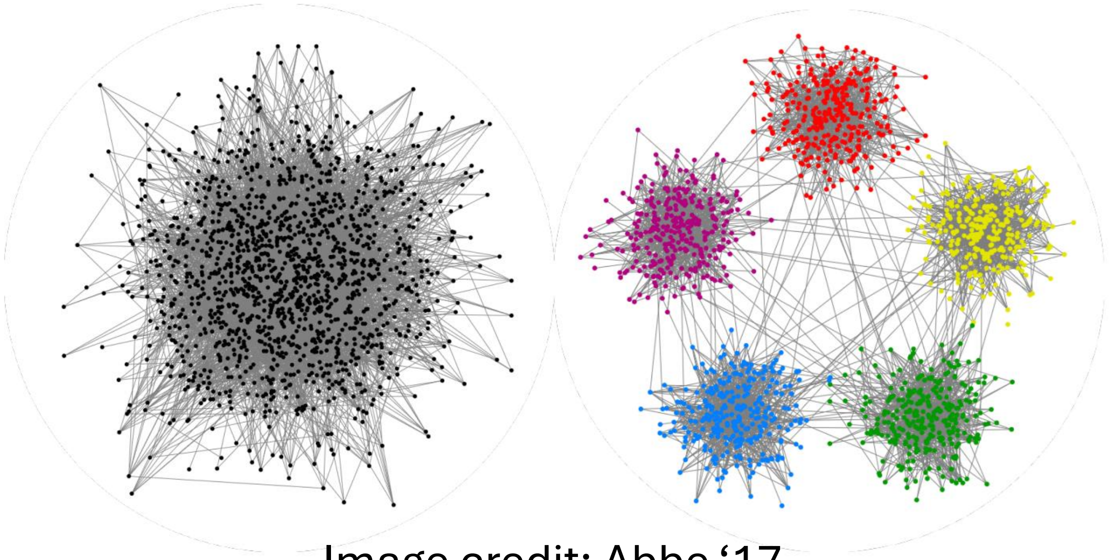

Image credit: Abbe '17

Stochastic Block Model (SBM)

- n vertices partitioned into q hidden "communities" of equal size
- Edge (i,j) occurs with probability...
  - a/n if i,j same community
  - b/n if i,j different community
- Goal: given the graph, estimate the hidden partition
- Alternative parametrization:
  - $d = [a + (q-1)b]/q$ (average degree)
  - $\lambda = (a-b)/[a+(1-q)b]$ (SNR)

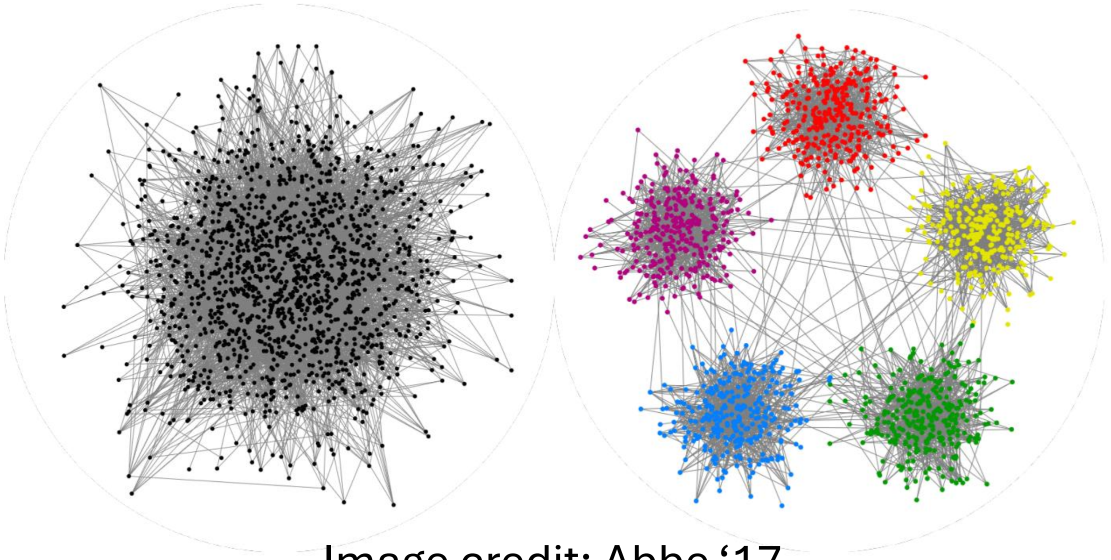

Image credit: Abbe '17

Kesten-Stigum (KS) Threshold

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$
- Landmark prediction: computational threshold is $d\lambda^2 = 1$ (KS bound) [Decelle, Krzakala, Moore, Zdeborová '11]

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$
- Landmark prediction: computational threshold is $d\lambda^2 = 1$ (KS bound) [Decelle, Krzakala, Moore, Zdeborová '11]
  - If $d\lambda^2 > 1$, there is a poly-time algorithm (belief propagation)
  - If $d\lambda^2 < 1$, no poly-time algorithm

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$
- Landmark prediction: computational threshold is $d\lambda^2 = 1$ (KS bound) [Decelle, Krzakala, Moore, Zdeborová '11]
  - If $d\lambda^2 > 1$, there is a poly-time algorithm (belief propagation)
  - If $d\lambda^2 < 1$, no poly-time algorithm
  - Sharp threshold

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$
- Landmark prediction: computational threshold is $d\lambda^2 = 1$ (KS bound)
  [Decelle, Krzakala, Moore, Zdeborová ‘11]
  - If $d\lambda^2 > 1$, there is a poly-time algorithm (belief propagation)
  - If $d\lambda^2 < 1$, no poly-time algorithm
  - Sharp threshold
- Poly-time algorithms rigorously reach KS bound
  [Massoulié ‘13; Mossel, Neeman, Sly ‘13; Abbe, Sandon ‘15]

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$
- Landmark prediction: computational threshold is $d\lambda^2 = 1$ (KS bound) [Decelle, Krzakala, Moore, Zdeborová '11]
  - If $d\lambda^2 > 1$, there is a poly-time algorithm (belief propagation)
  - If $d\lambda^2 < 1$, no poly-time algorithm
  - Sharp threshold
- Poly-time algorithms rigorously reach KS bound [Massoulié '13; Mossel, Neeman, Sly '13; Abbe, Sandon '15]
- For large enough q, exponential-time algorithms work below KS [Abbe, Sandon '15; Banks, Moore, Neeman, Netrapalli '16]

Kesten-Stigum (KS) Threshold

- For now: parameters q, d, λ are fixed as $n \to \infty$
- Landmark prediction: computational threshold is $d\lambda^2 = 1$ (KS bound)
  [Decelle, Krzakala, Moore, Zdeborová ‘11]
  - If $d\lambda^2 > 1$, there is a poly-time algorithm (belief propagation)
  - If $d\lambda^2 < 1$, no poly-time algorithm
  - Sharp threshold
- Poly-time algorithms rigorously reach KS bound
  [Massoulié ‘13; Mossel, Neeman, Sly ‘13; Abbe, Sandon ‘15]
- For large enough q, exponential-time algorithms work below KS
  [Abbe, Sandon ‘15; Banks, Moore, Neeman, Netrapalli ‘16]
- Statistical-computation gap

Tasks: Detection vs Recovery

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:
- **(Strong) detection**: distinguish with prob $1 - o(1)$ between...
  - Null hypothesis: Erdős–Rényi graph $G(n, d/n)$
  - Alternative hypothesis: SBM with parameters q, d, λ

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:
- **(Strong) detection**: distinguish with prob $1 - o(1)$ between...
  - Null hypothesis: Erdős–Rényi graph $G(n, d/n)$
  - Alternative hypothesis: SBM with parameters q, d, λ
- **(Weak) recovery**: recover a partition with $1/q + \epsilon$ fraction of vertices correct (beat random guessing) for a constant $\epsilon > 0$

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:
- **(Strong) detection**: distinguish with prob $1 - o(1)$ between...
  - Null hypothesis: Erdős–Rényi graph $G(n, d/n)$
  - Alternative hypothesis: SBM with parameters q, d, λ
- **(Weak) recovery**: recover a partition with $1/q + \epsilon$ fraction of vertices correct (beat random guessing) for a constant $\epsilon > 0$
  - Can't hope for strong recovery in sparse graphs

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:
- **(Strong) detection**: distinguish with prob $1 - o(1)$ between…
  - Null hypothesis: Erdős–Rényi graph $G(n, d/n)$
  - Alternative hypothesis: SBM with parameters q, d, λ
- **(Weak) recovery**: recover a partition with $1/q + \epsilon$ fraction of vertices correct (beat random guessing) for a constant $\epsilon > 0$
  - Can't hope for strong recovery in sparse graphs
- How are the two tasks related?

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:
- **(Strong) detection**: distinguish with prob $1 - o(1)$ between…
  - Null hypothesis: Erdős–Rényi graph $G(n, d/n)$
  - Alternative hypothesis: SBM with parameters q, d, λ
- **(Weak) recovery**: recover a partition with $1/q + \epsilon$ fraction of vertices correct (beat random guessing) for a constant $\epsilon > 0$
  - Can't hope for strong recovery in sparse graphs
- How are the two tasks related?
  - Detection "feels" easier…

Tasks: Detection vs Recovery

- KS is the (conjectured) threshold for 2 different tasks:
- **(Strong) detection**: distinguish with prob $1 - o(1)$ between…
  - Null hypothesis: Erdős–Rényi graph $G(n, d/n)$
  - Alternative hypothesis: SBM with parameters q, d, λ
- **(Weak) recovery**: recover a partition with $1/q + \epsilon$ fraction of vertices correct (beat random guessing) for a constant $\epsilon > 0$
  - Can't hope for strong recovery in sparse graphs
- How are the two tasks related?
  - Detection "feels" easier…
  - Made rigorous (in some sense) by [Ding, Hua, Slot, Steurer '25]

Average-Case Hardness?

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, ...]

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, …]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, ...]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)
  - Edge count: $\sum_{i<j} Y_{ij}$ (degree-1 polynomial)

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, ...]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)
  - Edge count: $\sum_{i<j} Y_{ij}$ (degree-1 polynomial)
  - Triangle count: $\sum_{i<j<k} Y_{ij} Y_{ik} Y_{jk}$ (degree-3 polynomial)

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, …]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)
  - Edge count: $\sum_{i<j} Y_{ij}$ (degree-1 polynomial)
  - Triangle count: $\sum_{i<j<k} Y_{ij} Y_{ik} Y_{jk}$ (degree-3 polynomial)
  - Surprisingly good track record at predicting computational thresholds

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, …]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)
  - Edge count: $\sum_{i<j} Y_{ij}$ (degree-1 polynomial)
  - Triangle count: $\sum_{i<j<k} Y_{ij} Y_{ik} Y_{jk}$ (degree-3 polynomial)
  - Surprisingly good track record at predicting computational thresholds
  - "Conjecture": if degree-$\omega(\log n)$ polynomials fail $\Rightarrow$ no poly-time alg

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, ...]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)
  - Edge count: $\sum_{i<j} Y_{ij}$ (degree-1 polynomial)
  - Triangle count: $\sum_{i<j<k} Y_{ij} Y_{ik} Y_{jk}$ (degree-3 polynomial)
  - Surprisingly good track record at predicting computational thresholds
  - "Conjecture": if degree-$\omega(\log n)$ polynomials fail ⇒ no poly-time alg
- Question (HS17): Can low-degree polynomials recover?

Average-Case Hardness?

- How do we know there's no poly-time alg for $d\lambda^2 < 1$?
- Belief propagation fails [Decelle, Krzakala, Moore, Zdeborová '11]
- Low-degree tests fail [Hopkins, Steurer '17, ...]
  - Input variable $Y_{ij} \in \{0,1\}$ for each pair $i < j$ (edge indicator)
  - Edge count: $\sum_{i<j} Y_{ij}$ (degree-1 polynomial)
  - Triangle count: $\sum_{i<j<k} Y_{ij} Y_{ik} Y_{jk}$ (degree-3 polynomial)
  - Surprisingly good track record at predicting computational thresholds
  - "Conjecture": if degree-$\omega(\log n)$ polynomials fail ⇒ no poly-time alg
- Question (HS17): Can low-degree polynomials recover?
- Other frameworks for average-case hardness (less applicable here)
  - Reductions, statistical query model, sum-of-squares, overlap gap, ...

Growing Number of Communities

Growing Number of Communities

- SBM is well-understood for fixed q (number of communities)

Growing Number of Communities

- SBM is well-understood for fixed q (number of communities)
- Now take $q \to \infty$ (where d, λ are still fixed)
  - Similar to [Chen, Xu '14]

Growing Number of Communities

- SBM is well-understood for fixed q (number of communities)
- Now take $q \to \infty$ (where d, λ are still fixed)
  - Similar to [Chen, Xu '14]
- For now, say $1 \ll q \ll \sqrt{n}$

Growing Number of Communities

- SBM is well-understood for fixed q (number of communities)
- Now take $q \to \infty$ (where d, λ are still fixed)
  - Similar to [Chen, Xu '14]
- For now, say $1 \ll q \ll \sqrt{n}$
- We show: There is still a poly-time alg for weak recovery above KS bound, $d\lambda^2 > 1$ [Chin, Mossel, Sohn, W '25]

Growing Number of Communities

- SBM is well-understood for fixed q (number of communities)
- Now take $q \to \infty$ (where d, λ are still fixed)
  - Similar to [Chen, Xu '14]
- For now, say $1 \ll q \ll \sqrt{n}$
- We show: There is still a poly-time alg for weak recovery above KS bound, $d\lambda^2 > 1$ [Chin, Mossel, Sohn, W '25]
  - Decide whether two vertices are in the same community by counting (weighted) non-backtracking walks of length $\sim \log n$ between them

Growing Number of Communities

- SBM is well-understood for fixed q (number of communities)
- Now take $q \to \infty$ (where d, λ are still fixed)
  - Similar to [Chen, Xu '14]
- For now, say $1 \ll q \ll \sqrt{n}$
- We show: There is still a poly-time alg for weak recovery above KS bound, $d\lambda^2 > 1$ [Chin, Mossel, Sohn, W '25]
  - Decide whether two vertices are in the same community by counting (weighted) non-backtracking walks of length $\sim \log n$ between them
- Detection becomes easy, even below KS!
  - Triangle count works for any fixed $d, \lambda \neq 0$, as long as $q \to \infty$

Can We Do Better (For Recovery)?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no… So how do we show recovery is hard?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no... So how do we show recovery is hard?
- What framework to use?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no... So how do we show recovery is hard?
- What framework to use?
- Statistical physics / belief propagation?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no… So how do we show recovery is hard?
- What framework to use?
- Statistical physics / belief propagation?
  - Unclear if applicable for this scaling regime?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no… So how do we show recovery is hard?
- What framework to use?
- Statistical physics / belief propagation?
  - Unclear if applicable for this scaling regime?
- Low-degree polynomials?

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no… So how do we show recovery is hard?
- What framework to use?
- Statistical physics / belief propagation?
  - Unclear if applicable for this scaling regime?
- Low-degree polynomials?
  - Seems more reliable for different scaling regimes (e.g. tensor PCA)

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no… So how do we show recovery is hard?
- What framework to use?
- Statistical physics / belief propagation?
    - Unclear if applicable for this scaling regime?
- Low-degree polynomials?
    - Seems more reliable for different scaling regimes (e.g. tensor PCA)
    - But we can’t leverage hardness of testing (“detection-recovery gap”)

Can We Do Better (For Recovery)?

- $1 \ll q \ll \sqrt{n}$
- Can do detection below KS, so can we do the same for recovery?
- Seems like no... So how do we show recovery is hard?
- What framework to use?
- Statistical physics / belief propagation?
  - Unclear if applicable for this scaling regime?
- Low-degree polynomials?
  - Seems more reliable for different scaling regimes (e.g. tensor PCA)
  - But we can't leverage hardness of testing ("detection-recovery gap")
  - Need to address recovery directly...

Low-Degree Recovery Framework

Low-Degree Recovery Framework

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar

$$x \coloneqq \mathbf{1}\{\text{vertices 1,2 same community}\}$$

Low-Degree Recovery Framework

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar
$$x \coloneqq \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\}$$
- Estimator $f: \mathbb{R}^{\binom{n}{2}} \to \mathbb{R}$, polynomial of degree $\leq D$

Low-Degree Recovery Framework

- Goal: Given $Y \sim SBM$, estimate scalar

$$x := \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\}$$

- Estimator $f: \mathbb{R}^{\binom{n}{2}} \to \mathbb{R}$, polynomial of degree $\leq D$
  - $\mathrm{MMSE}_{\leq D} := \inf_{\deg(f) \leq D} \mathrm{E}[(f(Y) - x)^2]$

Low-Degree Recovery Framework

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar
$$x := \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\}$$
- Estimator $f : \mathbb{R}^{\binom{n}{2}} \to \mathbb{R}$, polynomial of degree $\leq D$
  - $\mathrm{MMSE}_{\leq D} := \inf_{\deg(f) \leq D} \mathrm{E}[(f(Y) - x)^2]$
- [Schramm, W '20] "Coarse" bounds on $\mathrm{MMSE}_{\leq D}$, no sharp thresholds

Low-Degree Recovery Framework

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar
$$x \coloneqq \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\}$$
- Estimator $f : \mathbb{R}^{\binom{n}{2}} \to \mathbb{R}$, polynomial of degree $\leq D$
  - $\mathrm{MMSE}_{\leq D} \coloneqq \inf_{\deg(f) \leq D} \mathrm{E}[(f(Y) - x)^2]$
- [Schramm, W '20] "Coarse" bounds on $\mathrm{MMSE}_{\leq D}$, no sharp thresholds
- [Sohn, W '25] Sharp thresholds for $\mathrm{MMSE}_{\leq D}$: SBM with fixed q, planted submatrix, ...

Low-Degree Recovery Framework

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar
$$x := \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\}$$
- Estimator $f : \mathbb{R}^{\binom{n}{2}} \to \mathbb{R}$, polynomial of degree $\leq D$
  - $\mathrm{MMSE}_{\leq D} := \inf_{\deg(f) \leq D} \mathrm{E}[(f(Y) - x)^2]$
- [Schramm, W '20] "Coarse" bounds on $\mathrm{MMSE}_{\leq D}$, no sharp thresholds
- [Sohn, W '25] Sharp thresholds for $\mathrm{MMSE}_{\leq D}$: SBM with fixed q, planted submatrix, …
- [Chin, Mossel, Sohn, W '25] SBM with $q \ll \sqrt{n}$, $\mathrm{MMSE}_{\leq D}$ trivial below KS

Low-Degree Recovery Framework

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar
$$x \coloneqq \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\}$$
- Estimator $f: \mathbb{R}^{\binom{n}{2}} \to \mathbb{R}$, polynomial of degree $\leq D$
  - $\mathrm{MMSE}_{\leq D} \coloneqq \inf_{\deg(f) \leq D} \mathrm{E}[(f(Y) - x)^2]$
- [Schramm, W '20] "Coarse" bounds on $\mathrm{MMSE}_{\leq D}$, no sharp thresholds
- [Sohn, W '25] Sharp thresholds for $\mathrm{MMSE}_{\leq D}$: SBM with fixed q, planted submatrix, ...
- [Chin, Mossel, Sohn, W '25] SBM with $q \ll \sqrt{n}$, $\mathrm{MMSE}_{\leq D}$ trivial below KS
- Conclusion: KS bound remains the computational threshold for weak recovery when q grows... as long as $q \ll \sqrt{n}$

Low-Degree Phase Transition

Low-Degree Phase Transition

- Goal: Given $Y \sim \mathrm{SBM}$, estimate scalar
$$x := \mathbf{1}\{\text{vertices } 1,2 \text{ same community}\} - 1/q$$

Low-Degree Phase Transition

- Goal: Given $Y \sim SBM$, estimate scalar
$$x := \mathbf{1}\{\text{vertices } 1{,}2 \text{ same community}\} - 1/q$$
- $\mathrm{MMSE}_{\leq D}$ is directly related to "correlation":
$$\mathrm{Corr}_{\leq D} := \frac{\mathrm{E}[f(Y) \cdot x]}{\sqrt{\mathrm{E}[f(Y)^2] \cdot \mathrm{E}[x^2]}} \in [0,1]$$

Low-Degree Phase Transition

- Goal: Given $Y \sim SBM$, estimate scalar
$$x := \mathbf{1}\{\text{vertices } 1,2 \text{ same community}\} - 1/q$$
- $\text{MMSE}_{\leq D}$ is directly related to "correlation":
$$\text{Corr}_{\leq D} := \frac{\mathrm{E}[f(Y) \cdot x]}{\sqrt{\mathrm{E}[f(Y)^2] \cdot \mathrm{E}[x^2]}} \in [0,1]$$
- **Theorem** [Chin, Mossel, Sohn, W '25] Consider SBM with $q \ll \sqrt{n}$.
    - If $d\lambda^2 > 1$ then $\text{Corr}_{\leq C \log n} = \Omega(1)$ for some constant $C > 0$
      (Above KS bound, non-trivial recovery)
    - If $d\lambda^2 < 1$ then $\text{Corr}_{\leq n^\delta} = o(1)$ for some constant $\delta > 0$
      (Below KS bound, trivial recovery)

Beating KS when $q \gg \sqrt{n}$

Beating KS when $q \gg \sqrt{n}$

- $q = n^{\chi}$ for a constant $\chi \in (1/2, 1)$

Beating KS when $q \gg \sqrt{n}$

- $q = n^{\chi}$ for a constant $\chi \in (1/2, 1)$
- We show: There is now a poly-time alg for weak recovery when $d\lambda^2 > 1$ or $d\lambda^{1/\chi} > C(\log d)^2$ [Chin, Mossel, Sohn, W '25]

Beating KS when $q \gg \sqrt{n}$

- $q = n^\chi$ for a constant $\chi \in (1/2, 1)$
- We show: There is now a poly-time alg for weak recovery when $d\lambda^2 > 1$ or $d\lambda^{1/\chi} > C(\log d)^2$ [Chin, Mossel, Sohn, W '25]
  - Can beat KS: $d\lambda^2 > 1$

Beating KS when $q \gg \sqrt{n}$

- $q = n^\chi$ for a constant $\chi \in (1/2,1)$
- We show: There is now a poly-time alg for weak recovery when $d\lambda^2 > 1$ or $d\lambda^{1/\chi} > C(\log d)^2$ [Chin, Mossel, Sohn, W '25]
  - Can beat KS: $d\lambda^2 > 1$
  - Algorithm is again based on non-backtracking walks

Beating KS when $q \gg \sqrt{n}$

- $q = n^{\chi}$ for a constant $\chi \in (1/2,1)$
- We show: There is now a poly-time alg for weak recovery when $d\lambda^2 > 1$ or $d\lambda^{1/\chi} > C(\log d)^2$ [Chin, Mossel, Sohn, W '25]
  - Can beat KS: $d\lambda^2 > 1$
  - Algorithm is again based on non-backtracking walks
- Open: Is this the right threshold? Lower bounds?

Beating KS when $q \gg \sqrt{n}$

- $q = n^\chi$ for a constant $\chi \in (1/2, 1)$
- We show: There is now a poly-time alg for weak recovery when $d\lambda^2 > 1$ or $d\lambda^{1/\chi} > C(\log d)^2$ [Chin, Mossel, Sohn, W '25]
  - Can beat KS: $d\lambda^2 > 1$
  - Algorithm is again based on non-backtracking walks
- Open: Is this the right threshold? Lower bounds?
- Similar phenomenon in Gaussian mixture models: behavior changes when number of clusters passes $\sqrt{n}$ [Even, Giraud, Verzelen '24]