How Robust are Thresholds for Community Detection?

Ankur Moitra (MIT)

Robust Statistics Summer School
Let me tell you a story about the success of belief propagation and statistical physics...
THE STOCHASTIC BLOCK MODEL

Introduced by Holland, Laskey and Leinhardt (1983):

- k communities
- connection probabilities

\[ Q = \begin{bmatrix}
 Q_{11} & Q_{12} & Q_{13} \\
 Q_{12} & Q_{22} & Q_{32} \\
 Q_{13} & Q_{32} & Q_{33}
\end{bmatrix} \]

- edges independent
THE STOCHASTIC BLOCK MODEL

Introduced by Holland, Laskey and Leinhardt (1983):

- $k$ communities
- Connection probabilities $Q_{ij}$
- Edges independent

Ubiquitous model studied in **statistics**, **computer science**, **information theory**, and **statistical physics**
Testbed for diverse range of algorithms

(1) Combinatorial Methods
   e.g. degree counting [Bui, Chaudhuri, Leighton, Sipser ‘87]
Testbed for diverse range of algorithms

(1) Combinatorial Methods
e.g. degree counting [Bui, Chaudhuri, Leighton, Sipser ‘87]

(2) Spectral Methods e.g. [McSherry ‘01]
Testbed for diverse range of algorithms

(1) **Combinatorial Methods**
   e.g. degree counting [Bui, Chaudhuri, Leighton, Sipser ‘87]

(2) **Spectral Methods** e.g. [McSherry ‘01]

(3) **Markov chain Monte Carlo (MCMC)** e.g. [Jerrum, Sorkin ‘98]
Testbed for diverse range of algorithms

(1) **Combinatorial Methods**
   e.g. degree counting [Bui, Chaudhuri, Leighton, Sipser ‘87]

(2) **Spectral Methods** e.g. [McSherry ‘01]

(3) **Markov chain Monte Carlo (MCMC)** e.g. [Jerrum, Sorkin ‘98]

(4) **Semidefinite Programs** e.g. [Boppana ‘87]
Testbed for diverse range of algorithms

(1) Combinatorial Methods
e.g. degree counting [Bui, Chaudhuri, Leighton, Sipser ‘87]

(2) Spectral Methods e.g. [McSherry ‘01]

(3) Markov chain Monte Carlo (MCMC) e.g. [Jerrum, Sorkin ‘98]

(4) Semidefinite Programs e.g. [Boppana ‘87]

These algorithms succeed in some ranges of parameters
Testbed for diverse range of algorithms

(1) **Combinatorial Methods**
   e.g. degree counting [Bui, Chaudhuri, Leighton, Sipser ‘87]

(2) **Spectral Methods** e.g. [McSherry ‘01]

(3) **Markov chain Monte Carlo (MCMC)** e.g. [Jerrum, Sorkin ‘98]

(4) **Semidefinite Programs** e.g. [Boppana ‘87]

These algorithms succeed in some ranges of parameters

Can we reach the fundamental limits of the SBM?
Following Decelle, Krzakala, Moore and Zdeborová (2011), let’s study the **sparse** regime:

\[
\frac{a}{n} \quad \frac{b}{n} \quad \frac{a}{n}
\]

where \( a, b = O(1) \) so that there are \( O(n) \) edges
Following Decelle, Krzakala, Moore and Zdeborová (2011), let’s study the **sparse** regime:

where $a, b = O(1)$ so that there are $O(n)$ edges

**Remark:** The degree of each node is $\text{Poi}(a/2+b/2)$ hence there are many isolated nodes whose community we cannot find
Following Decelle, Krzakala, Moore and Zdeborová (2011), let’s study the *sparse* regime:

\[
\frac{a}{n}, \quad \frac{b}{n}, \quad \frac{a}{n}
\]

where \(a, b = O(1)\) so that there are \(O(n)\) edges

**Remark:** The degree of each node is \(\text{Poi}(a/2+b/2)\) hence there are many isolated nodes whose community we cannot find

**Goal (Partial Recovery):** Find a partition that has agreement better than \(\frac{1}{2}\) with true community structure
Following Decelle, Krzakala, Moore and Zdeborová (2011), let’s study the **sparse** regime:

\[
\frac{a}{n}, \quad \frac{b}{n}, \quad \frac{a}{n}
\]

where \( a, b = O(1) \) so that there are \( O(n) \) edges

**Conjecture:** Partial recovery is possible iff \((a-b)^2 > 2(a+b)\)
Following Decelle, Krzakala, Moore and Zdeborová (2011), let’s study the \textbf{sparse} regime:

\[
\frac{a}{n}, \frac{b}{n}, \frac{a}{n}
\]

where \(a, b = O(1)\) so that there are \(O(n)\) edges

**Conjecture:** Partial recovery is possible iff \((a-b)^2 > 2(a+b)\)

Conjecture is based on fixed points of \textbf{belief propagation}...
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction
  • The Stochastic Block Model
  • Belief Propagation and its Predictions
  • Semi-Random Models
  • Our Results

Part II: Broadcast Tree Model
  • The Kesten-Stigum Bound
  • A First Semi-Random vs. Random Separation
  • Our Results, continued

Part III: Above Average-Case?
BELIEF PROPAGATION

Introduced by Judea Pearl (1982):

“For fundamental contributions ... to probabilistic and causal reasoning”
Adapted to community detection:

Message $v \rightarrow u$

- Probability I think I am community #1, community #2, ...

Do same for all nodes
Adapted to community detection:

Message $v \rightarrow u$

Probability I think I am community #1, community #2, ...

Do same for all nodes
Adapted to community detection:

Message $v \rightarrow u$
- Probability I think I am community #1, community #2, ...

Message $u \rightarrow v$
- New probability I think I am community #1, community #2, ...

Do same for all nodes

Update beliefs
THE TRIVIAL FIXED POINT

Belief propagation has a trivial fixed point where it gets stuck.
Belief propagation has a trivial fixed point where it gets stuck

\[
\begin{align*}
\Pr[\text{red}] &= \frac{1}{2} \\
\Pr[\text{blue}] &= \frac{1}{2}
\end{align*}
\]
THE TRIVIAL FIXED POINT

Belief propagation has a trivial fixed point where it gets stuck.

**Claim:** No one knows anything, so you never have to update your beliefs.
THE TRIVIAL FIXED POINT

Belief propagation has a trivial fixed point where it gets stuck.

Fact: If \((a-b)^2 > 2(a+b)\) then the trivial fixed point is unstable.
THE TRIVIAL FIXED POINT

Belief propagation has a trivial fixed point where it gets stuck

**Fact:** If \((a-b)^2 > 2(a+b)\) then the trivial fixed point is unstable

**Hope:** Whatever it finds, solves partial recovery
Belief propagation has a trivial fixed point where it gets stuck

Fact: If \((a-b)^2 > 2(a+b)\) then the trivial fixed point is unstable

Hope: Whatever it finds, solves partial recovery

Evidence based on simulations
THE TRIVIAL FIXED POINT

Belief propagation has a trivial fixed point where it gets stuck

Fact: If \((a-b)^2 > 2(a+b)\) then the trivial fixed point is unstable

Hope: Whatever it finds, solves partial recovery

Evidence based on simulations

And if \((a-b)^2 \leq 2(a+b)\) and it does get stuck, then maybe partial recovery is information theoretically impossible?
CONJECTURE IS PROVED!

Mossel, Neeman and Sly (2013) and Massoulie (2013):

**Theorem:** It is possible to find a partition that is correlated with true communities iff \((a-b)^2 > 2(a+b)\)
CONJECTURE IS PROVED!

Mossel, Neeman and Sly (2013) and Massoulie (2013):

**Theorem:** It is possible to find a partition that is correlated with true communities iff \((a-b)^2 > 2(a+b)\)

Later attempts based on SDPs only get to

\[(a-b)^2 > C(a+b), \text{ for some } C > 2\]
CONJECTURE IS PROVED!

Mossel, Neeman and Sly (2013) and Massoulie (2013):

**Theorem:** It is possible to find a partition that is correlated with true communities iff \((a-b)^2 > 2(a+b)\)

Later attempts based on SDPs only get to

\[(a-b)^2 > C(a+b), \text{ for some } C > 2\]

Are nonconvex methods **better** than convex programs?
CONJECTURE IS PROVED!

Mossel, Neeman and Sly (2013) and Massoulie (2013):

**Theorem:** It is possible to find a partition that is correlated with true communities iff \((a-b)^2 > 2(a+b)\)

Later attempts based on SDPs only get to

\[(a-b)^2 > C(a+b), \text{ for some } C > 2\]

Are nonconvex methods **better** than convex programs?

How do predictions of statistical physics and SDPs compare?
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
SEMI-RANDOM MODELS

SEMI-RANDOM MODELS


(1) Sample graph from SBM
SEMI-RANDOM MODELS


(1) Sample graph from SBM

(2) Adversary can add edges within community and delete edges crossing
SEMI-RANDOM MODELS


(1) Sample graph from SBM

(2) Adversary can add edges within community and delete edges crossing
SEMI-RANDOM MODELS


(1) Sample graph from SBM

(2) Adversary can add edges within community and delete edges crossing
SEMI-RANDON MODELS


1. Sample graph from SBM

2. Adversary can add edges within community and delete edges crossing

Algorithms can no longer over tune to distribution
A NON-ROBUST ALGORITHM

Consider the following SBM:
A NON-ROBUST ALGORITHM

Consider the following SBM:

Nodes from same community: \( \left( \frac{1}{2} \right)^2 \frac{n}{2} + \left( \frac{1}{4} \right)^2 \frac{n}{2} \)

Number of common neighbors
A NON-ROBUST ALGORITHM

Consider the following SBM:

Number of common neighbors

Nodes from same community: \( \left( \frac{1}{2} \right)^2 \frac{n}{2} + \left( \frac{1}{4} \right)^2 \frac{n}{2} \)

Nodes from diff. community: \( \left( \frac{1}{2} \right) \left( \frac{1}{4} \right)^n \)
A NON-ROBUST ALGORITHM

Consider the following SBM:

Nodes from same community:

Nodes from diff. community:

Number of common neighbors

Nodes from same community: \( \left( \frac{1}{2} \right)^2 \frac{n}{2} + \left( \frac{1}{4} \right)^2 \frac{n}{2} \)

Nodes from diff. community: \( \left( \frac{1}{2} \right) \left( \frac{1}{4} \right)^n \)
A NON-ROBUST ALGORITHM

Semi-random adversary: Add clique to red community
A NON-ROBUST ALGORITHM

Semi-random adversary: Add clique to red community

Number of common neighbors

Nodes from blue community: \( \left( \frac{1}{2} \right)^2 \frac{n}{2} + \left( \frac{1}{4} \right)^2 \frac{n}{2} \)
A NON-ROBUST ALGORITHM

Semi-random adversary: Add clique to red community

Number of common neighbors

Nodes from blue community: \[ \left( \frac{1}{2} \right)^2 \frac{n}{2} + \left( \frac{1}{4} \right)^2 \frac{n}{2} \]

Nodes from diff. community: \[ \left( \frac{1}{2} \right) \left( \frac{1}{4} \right) \frac{n}{2} + \left( \frac{1}{4} \right) \frac{n}{2} \]
A NON-ROBUST ALGORITHM

Semi-random adversary: Add clique to **red** community

Number of common neighbors

Nodes from **blue** community:
\[
\left( \frac{1}{2} \right)^2 \frac{n}{2} + \left( \frac{1}{4} \right)^2 \frac{n}{2}
\]

Nodes from diff. community:
\[
\left( \frac{1}{2} \right) \left( \frac{1}{4} \right) \frac{n}{2} + \left( \frac{1}{4} \right) \frac{n}{2}
\]
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
OUR RESULTS

“Helpful” changes can hurt:

**Theorem:** Community detection in semirandom model is impossible for 
\[(a-b)^2 \leq C_{a,b}(a+b)\] for some \(C_{a,b} > 2\)
OUR RESULTS

“Helpful” changes can hurt:

**Theorem:** Community detection in semirandom model is impossible for \((a-b)^2 \leq C_{a,b}(a+b)\) for some \(C_{a,b} > 2\)

But SDPs continue to work in semirandom model
OUR RESULTS

“Helpful” changes can hurt:

**Theorem:** Community detection in semirandom model is impossible for \((a-b)^2 \leq C_{a,b}(a+b)\) for some \(C_{a,b} > 2\)

But SDPs continue to work in semirandom model

Follows same blueprint as [Guedon, Vershynin]
OUR RESULTS

“Helpful” changes can hurt:

**Theorem:** Community detection in semirandom model is impossible for $(a-b)^2 \leq C_{a,b} (a+b)$ for some $C_{a,b} > 2$

But SDPs continue to work in semirandom model

Follows same blueprint as [Guedon, Vershynin]

See [Makarychev, Makarychev, Vijayaraghavan] for SDP-based robustness guarantees for $k > 2$ communities
OUR RESULTS

“Helpful” changes can hurt:

**Theorem:** Community detection in semirandom model is impossible for \((a-b)^2 \leq C_{a,b}(a+b)\) for some \(C_{a,b} > 2\)

But SDPs continue to work in semirandom model

Reaching the information theoretic threshold requires exploiting the **structure of the noise**
OUR RESULTS

“Helpful” changes can hurt:

**Theorem:** Community detection in semirandom model is impossible for \((a-b)^2 \leq C_{a,b}(a+b)\) for some \(C_{a,b} > 2\)

But SDPs continue to work in semirandom model

Reaching the information theoretic threshold requires exploiting the **structure of the noise**

This is first **separation** between what is possible in random vs. semirandom models
OUTLINE

Part I: Introduction
- The Stochastic Block Model
- Belief Propagation and its Predictions
- Semi-Random Models
- Our Results

Part II: Broadcast Tree Model
- The Kesten-Stigum Bound
- A First Semi-Random vs. Random Separation
- Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
Let’s start with a simpler model originating from genetics...
(1) Root is either red/blue
(1) Root is either red/blue

(2) Each node gives birth to $\text{Poi}(a/2)$ nodes of same color and $\text{Poi}(b/2)$ nodes of opposite color
BROADCAST TREE MODEL

(1) Root is either red/blue

(2) Each node gives birth to $\text{Poi}(a/2)$ nodes of same color and $\text{Poi}(b/2)$ nodes of opposite color
BROADCAST TREE MODEL

(1) Root is either red/blue

(2) Each node gives birth to $\text{Poi}(a/2)$ nodes of same color and $\text{Poi}(b/2)$ nodes of opposite color
BROADCAST TREE MODEL

(1) Root is either red/blue

(2) Each node gives birth to $\text{Poi}(a/2)$ nodes of same color and $\text{Poi}(b/2)$ nodes of opposite color
BROADCAST TREE MODEL

(1) Root is either red/blue

(2) Each node gives birth to \( \text{Poi}(a/2) \) nodes of same color and \( \text{Poi}(b/2) \) nodes of opposite color
BROADCAST TREE MODEL

(1) Root is either red/blue

(2) Each node gives birth to Poi(a/2) nodes of same color and Poi(b/2) nodes of opposite color

(3) Goal: From leaves and unlabeled tree, guess color of root with > ½ prob. indep. of n (# of levels)
(1) Root is either red/blue

(2) Each node gives birth to $\text{Poi}(a/2)$ nodes of same color and $\text{Poi}(b/2)$ nodes of opposite color

(3) **Goal:** From leaves and unlabeled tree, guess color of root with $>\frac{1}{2}$ prob. indep. of $n$ (# of levels)

This is the natural analogue for partial recovery
BROADCAST TREE MODEL

(1) Root is either red/blue

(2) Each node gives birth to $\text{Poi}(a/2)$ nodes of same color and $\text{Poi}(b/2)$ nodes of opposite color

(3) **Goal:** From leaves and unlabeled tree, guess color of root with $> \frac{1}{2}$ prob. indep. of $n$ (# of levels)

For what values of $a$ and $b$ can we guess the root?
THE KESTEN STIGUM BOUND

“Best way to reconstruct root from leaves is majority vote”
THE KESTEN STIGUM BOUND

“Best way to reconstruct root from leaves is majority vote”

Theorem [Kesten, Stigum, ‘66]: Majority vote of the leaves succeeds with probability >½ iff \((a-b)^2 > 2(a+b)\)
THE KESTEN STIGUM BOUND

“Best way to reconstruct root from leaves is majority vote”

**Theorem [Kesten, Stigum, ‘66]:** Majority vote of the leaves succeeds with probability $> \frac{1}{2}$ iff $(a-b)^2 > 2(a+b)$

More generally, gave a limit theorem for multi-type branching processes
THE KESTEN STIGUM BOUND

“Best way to reconstruct root from leaves is majority vote”

**Theorem [Kesten, Stigum, ‘66]:** Majority vote of the leaves succeeds with probability $> \frac{1}{2}$ iff $(a-b)^2 > 2(a+b)$

More generally, gave a limit theorem for multi-type branching processes

**Theorem [Evans et al., ‘00]:** Reconstruction is information theoretically impossible if $(a-b)^2 \leq 2(a+b)$
“Best way to reconstruct root from leaves is majority vote”

**Theorem [Kesten, Stigum, ‘66]:** Majority vote of the leaves succeeds with probability > ½ iff \((a-b)^2 > 2(a+b)\)

More generally, gave a limit theorem for multi-type branching processes

**Theorem [Evans et al., ‘00]:** Reconstruction is information theoretically impossible if \((a-b)^2 \leq 2(a+b)\)

---

Local view in SBM = Broadcast Tree
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
SEMIRANDOM BROADCAST TREE MODEL

**Definition:** A semirandom adversary can cut edges between nodes of opposite colors and remove entire subtree.
SEMIRANDOM BROADCAST TREE MODEL

**Definition:** A semirandom adversary can cut edges between nodes of opposite colors and remove entire subtree
Definition: A semirandom adversary can cut edges between nodes of opposite colors and remove entire subtree

e.g.
SEMIRANDOM BROADCAST TREE MODEL

**Definition:** A semirandom adversary can cut edges between nodes of opposite colors and remove entire subtree.

Analogous to cutting edges between communities, and changing the local neighborhood in the SBM.
SEMIRANDOM BROADCAST TREE MODEL

**Definition:** A semirandom adversary can cut edges between nodes of opposite colors and remove entire subtree.

Analogous to cutting edges between communities, and changing the local neighborhood in the SBM.

Can the adversary usually flip the majority vote?
Key Observation: Some node’s descendants vote opposite way
Key Observation: Some node’s descendants vote opposite way
Key Observation: Some node’s descendants vote opposite way

Near the Kesten-Stigum bound, this happens everywhere
Key Observation: Some node’s descendants vote **opposite** way

By cutting these edges, adversary can usually flip majority vote
This breaks majority vote, but how do we move the information theoretic threshold?
This breaks majority vote, but how do we move the information theoretic threshold?

Need carefully chosen adversary where we can prove things about the distribution we get after he’s done
This breaks majority vote, but how do we move the **information theoretic threshold**?

Need carefully chosen adversary where we can prove things about the distribution we get after he’s done.

E.g. If we cut every subtree where this happens, would mess up independence properties.

More likely to have red children, given his parent is red and he was not cut.
This breaks majority vote, but how do we move the information theoretic threshold?

Need carefully chosen adversary where we can prove things about the distribution we get after he’s done

Need to design adversary that puts us back into nice model

e.g. a model on a tree where a sharp threshold is known
This breaks majority vote, but how do we move the information theoretic threshold?

Need carefully chosen adversary where we can prove things about the distribution we get after he’s done.

Need to design adversary that puts us back into nice model
e.g. a model on a tree where a sharp threshold is known

Following [Mossel, Neeman, Sly] we can embed the lower bound for semi-random BTM in semi-random SBM
This breaks majority vote, but how do we move the information theoretic threshold?

Need carefully chosen adversary where we can prove things about the distribution we get after he’s done.

Need to design adversary that puts us back into nice model.

- e.g. a model on a tree where a sharp threshold is known.

Following [Mossel, Neeman, Sly] we can embed the lower bound for semi-random BTM in semi-random SBM.

- e.g. Usual complication: once I reveal colors at boundary of neighborhood, need to show there’s little information you can get from rest of graph.
OUTLINE

Part I: Introduction
- The Stochastic Block Model
- Belief Propagation and its Predictions
- Semi-Random Models
- Our Results

Part II: Broadcast Tree Model
- The Kesten-Stigum Bound
- A First Semi-Random vs. Random Separation
- Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction

- The Stochastic Block Model
- Belief Propagation and its Predictions
- Semi-Random Models
- Our Results

Part II: Broadcast Tree Model

- The Kesten-Stigum Bound
- A First Semi-Random vs. Random Separation
- Our Results, continued

Part III: Above Average-Case?
“Helpful” changes can hurt:

**Theorem:** Reconstruction in semi-random broadcast tree model is impossible for \((a-b)^2 \leq C_{a,b}(a+b)\) for some \(C_{a,b} > 2\)
“Helpful” changes can hurt:

**Theorem:** Reconstruction in semi-random broadcast tree model is impossible for 
\[(a-b)^2 \leq C_{a,b}(a+b)\]  
for some \(C_{a,b} > 2\)

Is there any algorithm that succeeds in semirandom BTM?
SEMIRANDOM BROADCAST TREE MODEL

“Helpful” changes can hurt:

**Theorem:** Reconstruction in semi-random broadcast tree model is impossible for \((a-b)^2 \leq C_{a,b}(a+b)\) for some \(C_{a,b} > 2\)

Is there any algorithm that succeeds in semirandom BTM?

**Theorem:** *Recursive majority* succeeds in semi-random broadcast tree model if
\[
(a-b)^2 > (2 + o(1))(a+b) \log \frac{a+b}{2}
\]
OUTLINE

Part I: Introduction

• The Stochastic Block Model
• Belief Propagation and its Predictions
• Semi-Random Models
• Our Results

Part II: Broadcast Tree Model

• The Kesten-Stigum Bound
• A First Semi-Random vs. Random Separation
• Our Results, continued

Part III: Above Average-Case?
OUTLINE

Part I: Introduction
- The Stochastic Block Model
- Belief Propagation and its Predictions
- Semi-Random Models
- Our Results

Part II: Broadcast Tree Model
- The Kesten-Stigum Bound
- A First Semi-Random vs. Random Separation
- Our Results, continued

Part III: Above Average-Case?
Recursive majority is used in practice, despite the fact that it is known not to achieve the KS bound, why?
Recursive majority is used in practice, despite the fact that it is known not to achieve the KS bound, **why?**

Models are a measuring stick to compare algorithms, but are we studying the right ones?
Recursive majority is used in practice, despite the fact that it is known not to achieve the KS bound, why?

Models are a measuring stick to compare algorithms, but are we studying the right ones?

**Average-case models:** When we have many algorithms, can we find the best one?
Recursive majority is used in practice, despite the fact that it is known not to achieve the KS bound, **why?**

Models are a measuring stick to compare algorithms, but are we studying the right ones?

**Average-case models:** When we have many algorithms, can we find the *best* one?

**Semi-random models:** When recursive majority works, it’s not exploiting the structure of the noise.
Recursive majority is used in practice, despite the fact that it is known not to achieve the KS bound, **why?**

Models are a measuring stick to compare algorithms, but are we studying the right ones?

**Average-case models:** When we have many algorithms, can we find the *best* one?

**Semi-random models:** When recursive majority works, it’s not exploiting the structure of the noise

This is an axis on which recursive majority is superior
BETWEEN WORST-CASE AND AVERAGE-CASE

Spielman and Teng (2001):

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

Usually called *Beyond Worst-Case Analysis*
BETWEEN WORST-CASE AND AVERAGE-CASE

Spielman and Teng (2001):

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

Usually called *Beyond Worst-Case Analysis*

Semirandom models as *Above Average-Case Analysis*?
BETWEEN WORST-CASE AND AVERAGE-CASE

Spielman and Teng (2001):

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

Usually called *Beyond Worst-Case Analysis*

Semirandom models as *Above Average-Case Analysis*?

What else are we missing, if we only study problems in the average-case?
Let $M$ be an unknown, low-rank matrix

\[ M \approx \text{drama} + \text{comedy} + \ldots + \text{sports} \]
THE NETFLIX PROBLEM

Let $M$ be an unknown, low-rank matrix

\[ M \approx \text{drama} + \text{comedy} + \ldots + \text{sports} \]

**Model:** We are given random observations $M_{i,j}$ for all $i,j \in \Omega$
THE NETFLIX PROBLEM

Let \( M \) be an unknown, low-rank matrix

\[ \begin{align*}
M & \approx \begin{pmatrix}
\text{drama} \\
\text{comedy} \\
\text{sports}
\end{pmatrix} + \ldots + \begin{pmatrix}
\text{drama} \\
\text{comedy} \\
\text{sports}
\end{pmatrix}
\end{align*} \]

**Model:** We are given random observations \( M_{i,j} \) for all \( i,j \in \Omega \)

Is there an efficient algorithm to recover \( M \)?
CONVEX PROGRAMMING APPROACH

\[
\min \|X\|_* \quad \text{s.t.} \quad \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}| \leq \eta \quad (P)
\]

Here \( \|X\|_* \) is the **nuclear norm**, i.e. sum of the singular values of \( X \)

[Fazel], [Srebro, Shraibman], [Recht, Fazel, Parrilo], [Candes, Recht], [Candes, Tao], [Candes, Plan], [Recht],
CONVEX PROGRAMMING APPROACH

$$\min \|X\|_* \text{ s.t. } \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}| \leq \eta \quad (P)$$

Here $\|X\|_*$ is the nuclear norm, i.e. sum of the singular values of $X$

[Fazel], [Srebro, Shraibman], [Recht, Fazel, Parrilo], [Candes, Recht],
[Candes, Tao], [Candes, Plan], [Recht],

**Theorem:** If $M$ is $n \times n$ and has rank $r$, and is $C$-incoherent then $(P)$
recovers $M$ exactly from $C^6nr\log^2 n$ observations
ALTERNATING MINIMIZATION

Repeat:

\[ U \leftarrow \arg\min_U \sum_{(i,j) \in \Omega} |(UV^T)_{i,j} - M_{i,j}|^2 \]

\[ V \leftarrow \arg\min_V \sum_{(i,j) \in \Omega} |(UV^T)_{i,j} - M_{i,j}|^2 \]

[Keshavan, Montanari, Oh], [Jain, Netrapalli, Sanghavi], [Hardt]
ALTERNATING MINIMIZATION

Repeat:

\[
U \leftarrow \arg\min_U \sum_{(i,j) \in \Omega} |(UV^T)_{i,j} - M_{i,j}|^2
\]

\[
V \leftarrow \arg\min_V \sum_{(i,j) \in \Omega} |(UV^T)_{i,j} - M_{i,j}|^2
\]

[Keshavan, Montanari, Oh], [Jain, Netrapalli, Sanghavi], [Hardt]

**Theorem:** If M is n x n and has rank r, and is C-incoherent then alternating minimization approximately recovers M from

\[
Cn r^2 \frac{\|M\|^2_F}{\sigma_r^2} \text{ observations}
\]
ALTERNATING MINIMIZATION

Repeat:

\[
U \leftarrow \arg \min_U \sum_{(i,j) \in \Omega} |(UV^T)_{i,j} - M_{i,j}|^2
\]

\[
V \leftarrow \arg \min_V \sum_{(i,j) \in \Omega} |(UV^T)_{i,j} - M_{i,j}|^2
\]

[Keshavan, Montanari, Oh], [Jain, Netrapalli, Sanghavi], [Hardt]

**Theorem:** If \( M \) is \( n \times n \) and has rank \( r \), and is \( C \)-incoherent then alternating minimization approximately recovers \( M \) from

\[
Cnr^2 \frac{\|M\|_F^2}{\sigma_r^2}
\]

observations

Running time and space complexity are better
What if an adversary reveals more entries of M?
What if an adversary reveals more entries of $M$?

Convex program:

\[ \min \| X \|_* \quad \text{s.t.} \quad \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}| \leq \eta \quad (P) \]

still works, it’s just more constraints
What if an adversary reveals more entries of M?

Convex program:

\[
\min \|X\|_* \quad \text{s.t.} \quad \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}| \leq \eta \quad (P)
\]

still works, it’s just more constraints

Alternating minimization:

Analysis completely breaks down

observed matrix is no longer good spectral approx. to M
What if an adversary reveals more entries of $M$?

**Convex program:**

$$
\min \|X\|_* \quad \text{s.t.} \quad \sum_{(i,j) \in \Omega} |X_{i,j} - M_{i,j}| \leq \eta \quad (P)
$$

still works, it’s just more constraints

---

**Alternating minimization:**

Are there variants that work in semi-random models?
Summary:

• “Helpful” adversaries can make the problem harder
• Gave first random vs. semi-random separations
• Can we go above average-case analysis?
Summary:

• “Helpful” adversaries can make the problem harder
• Gave first random vs. semi-random separations
• Can we go above average-case analysis?

Thanks! Any Questions?