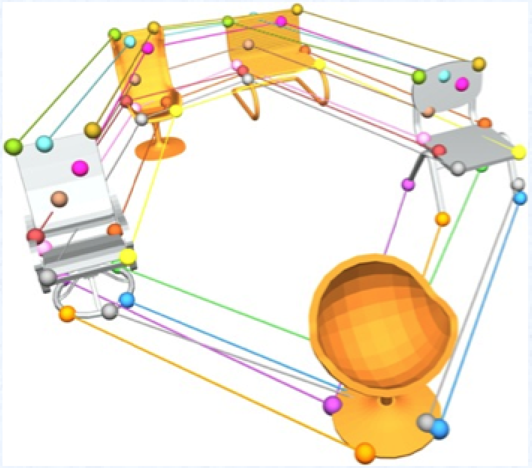
Information Recovery from Pairwise Measurements

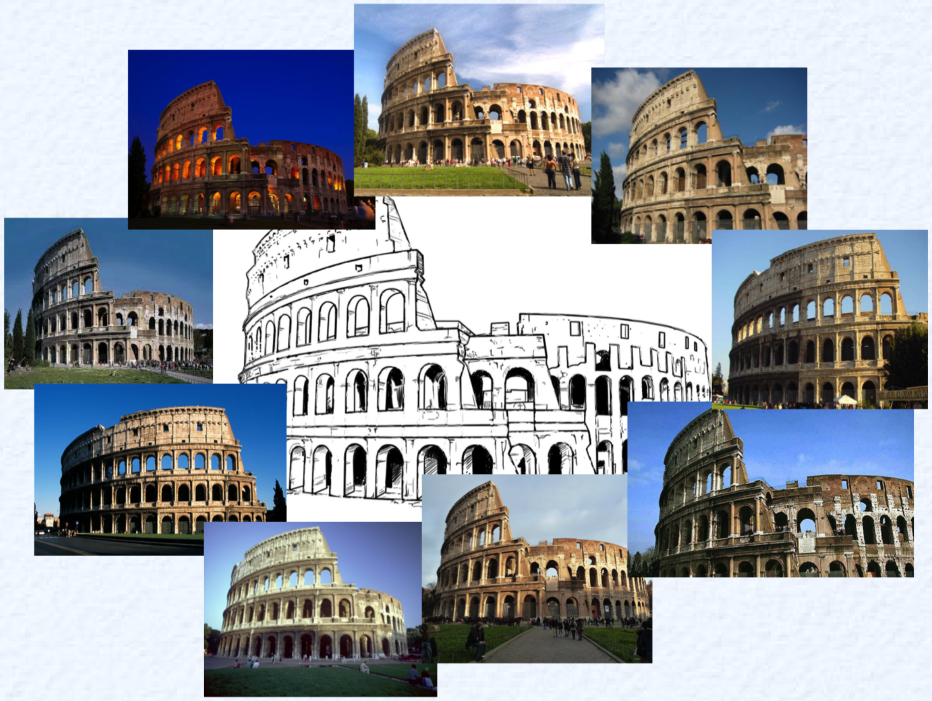
Yuxin Chen, Andrea Goldsmith

Electrical Engineering, Stanford University

Joint Data Analysis

- **Data-Intensive Applications**



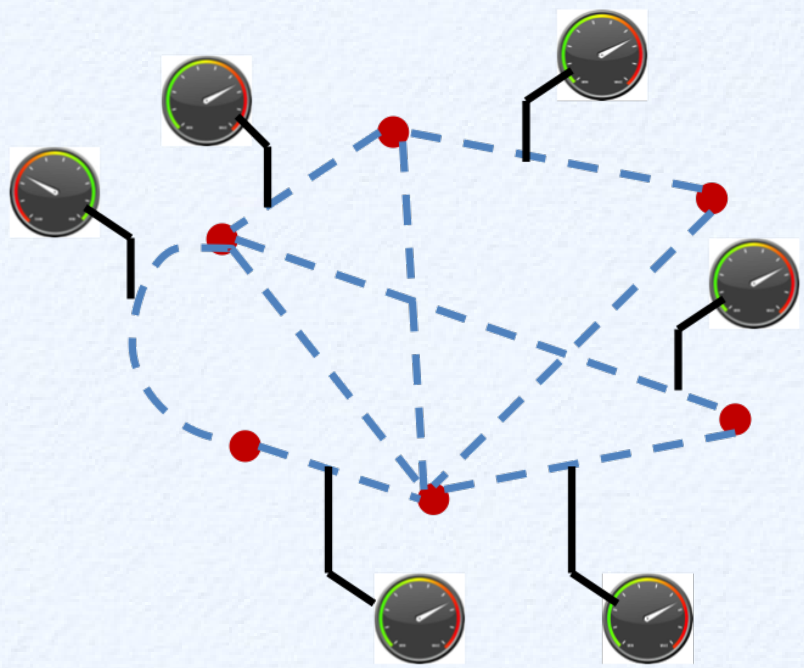


  - many **relevant** but **heterogeneous** data
  - exploit **data correlation** to improve individual task

This Talk: Info Recovery from Graph-Based Measurements

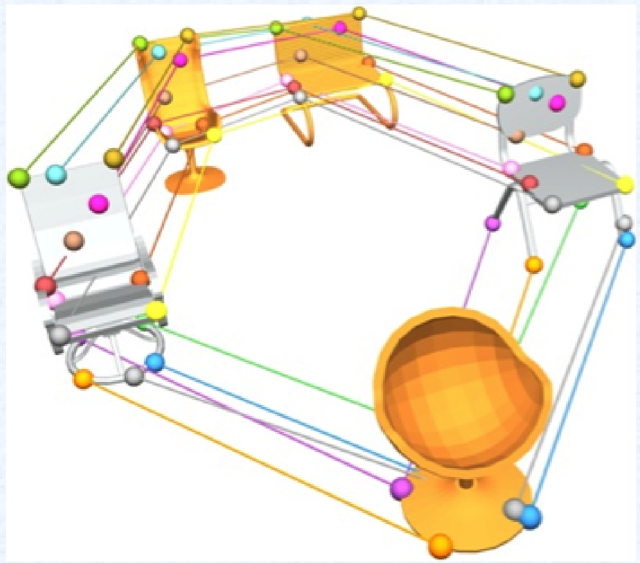
- **Information Network**
  - n vertices

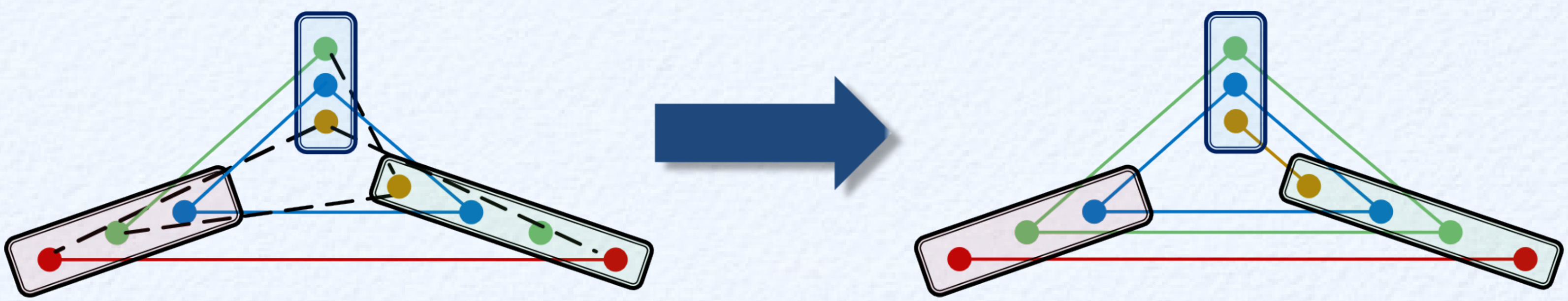
- **Measurements**
  - *taken on a few edges*



Motivating Application: Joint Graph Matching

- **Joint Graph Matching**: identify feature correspondences across instances

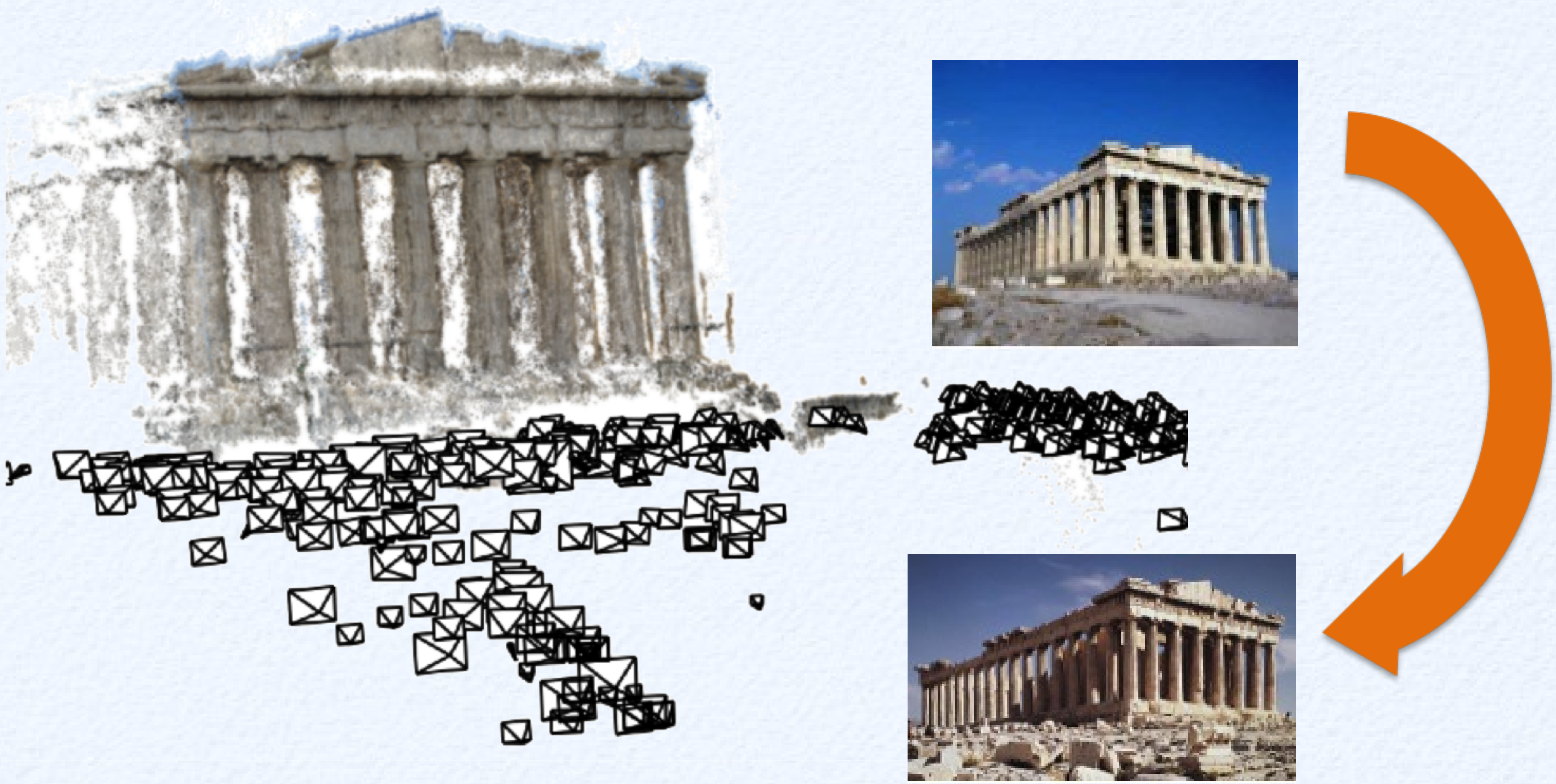




Input Output

Motivating Application: Multi-Image Registration

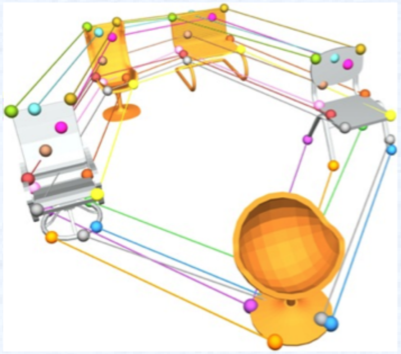
- **Structure from Motion**: estimate 3D structures from 2D image sequences



  - jointly align all images given relative camera poses between each pair

Lack of Benchmark

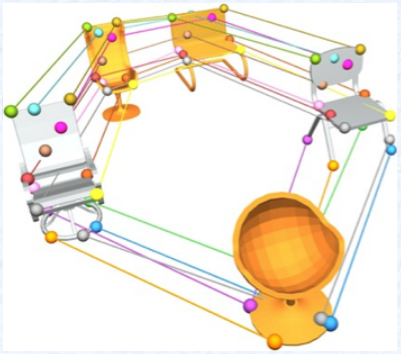
- **A Flurry of Activity in Algorithm Design**
  - Joint Matching (HuangGuibas'2013)
    - works if input error rate $\leq 50\%$



  - Rotation Registration (WangSinger'2012)
    - works if input error rate $\leq 50\%$

Lack of Benchmark

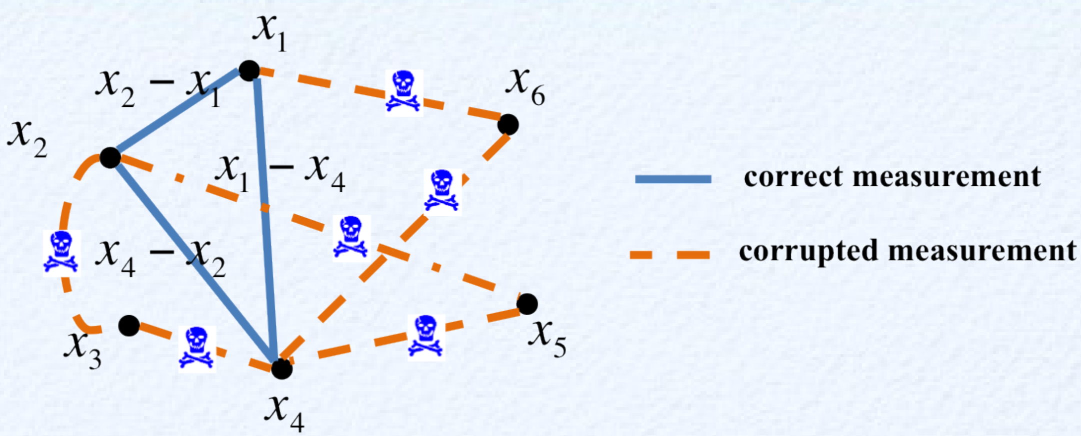
- **A Flurry of Activity in Algorithm Design**
  - Joint Matching (HuangGuibas'2013)
    - works if input error rate $\leq 50\%$



  - Rotation Registration (WangSinger'2012)
    - works if input error rate $\leq 50\%$

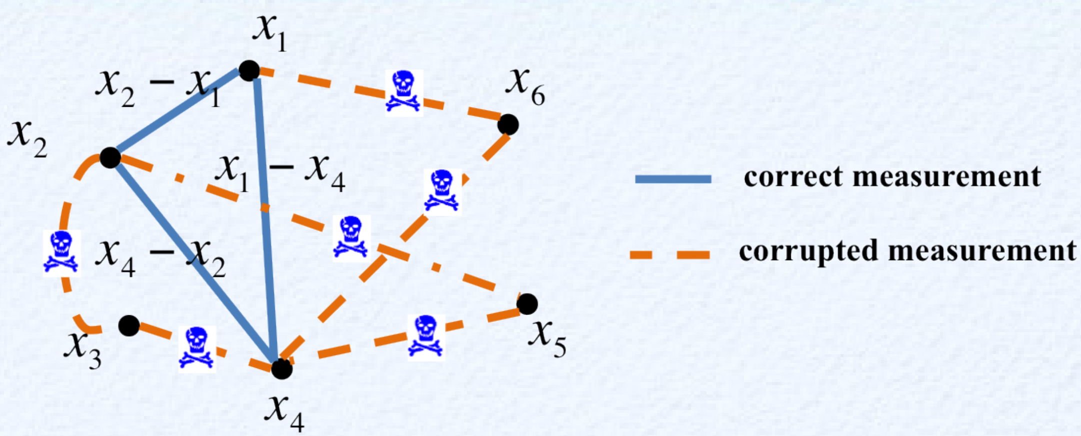
- **Lack of Benchmark:**
  - What are the fundamental performance limits?

Problem Setup: A Unified Formulation



- **Information Network**
  - n vertices
  - vertex i takes value $x_i \in \{1, \cdots, M\}$

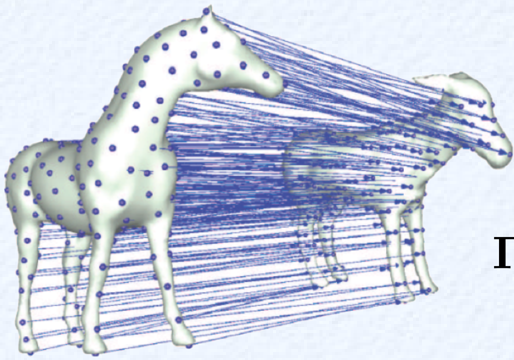
Problem Setup: A Unified Formulation



- **Information Network**
  - n vertices
  - vertex i takes value $x_i \in \{1, \cdots, M\}$
- **This Talk: Pairwise Measurements**
  - *measure certain pairwise relation function* $f(x_i, x_j)$

Example: Pairwise Difference

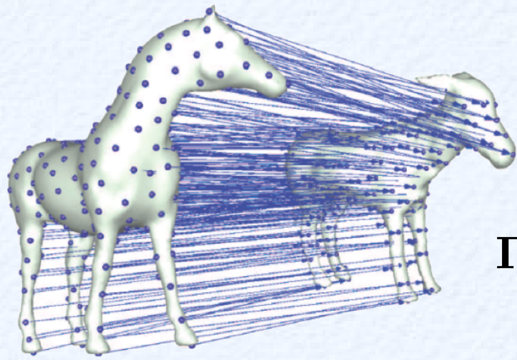
- **Pairwise Difference** $x_i - x_j$
  - pairwise matching: $\mathbf{\Pi}_j^{-1}\mathbf{\Pi}_i$



$$\mathbf{\Pi}_i = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

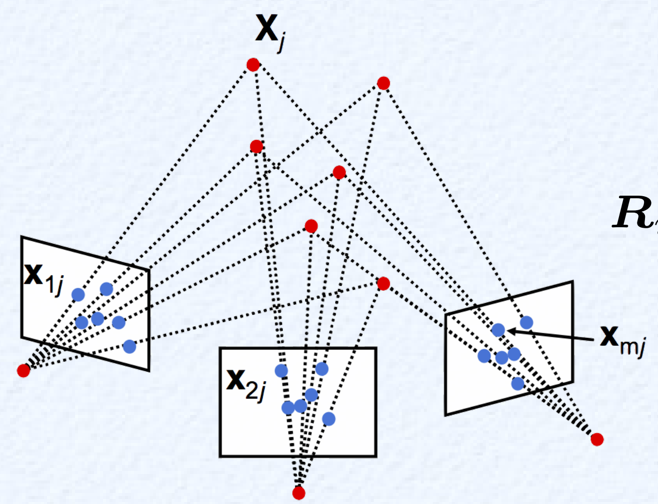
Example: Pairwise Difference

- **Pairwise Difference** $x_i - x_j$
  - pairwise matching: $\mathbf{\Pi}_j^{-1}\mathbf{\Pi}_i$



$$\mathbf{\Pi}_i = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

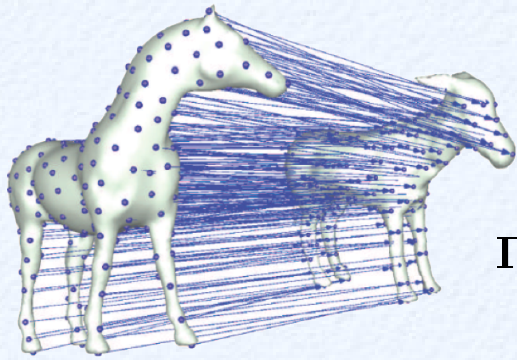
  - relative rotation: $\boldsymbol{R}_j^{-1}\boldsymbol{R}_i$



$$\boldsymbol{R}_i = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

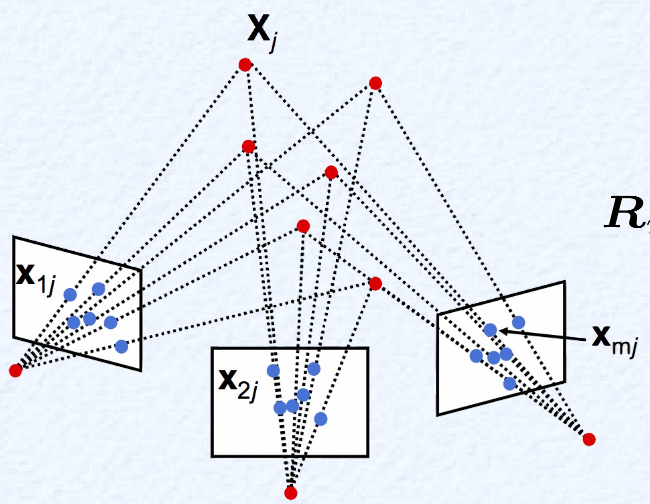
Example: Pairwise Difference

- **Pairwise Difference** $x_i - x_j$
  - pairwise matching: $\mathbf{\Pi}_j^{-1}\mathbf{\Pi}_i$



$$\mathbf{\Pi}_i = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

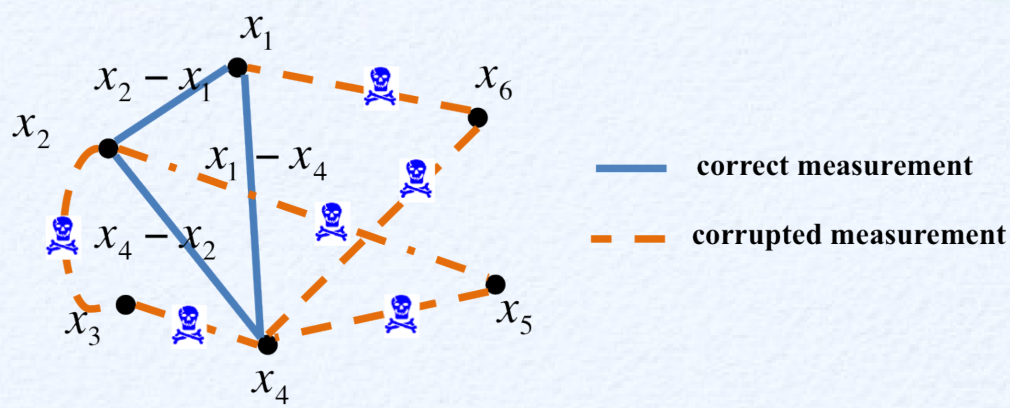
  - relative rotation: $\boldsymbol{R}_j^{-1}\boldsymbol{R}_i$



$$\boldsymbol{R}_i = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

  - angular difference ...

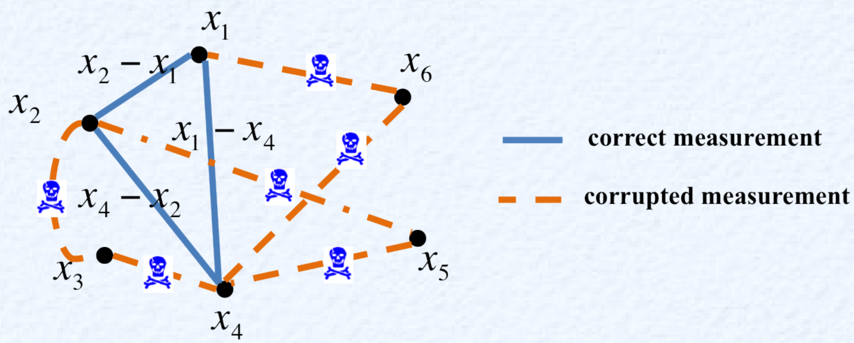
Graph-Based Measurement Model



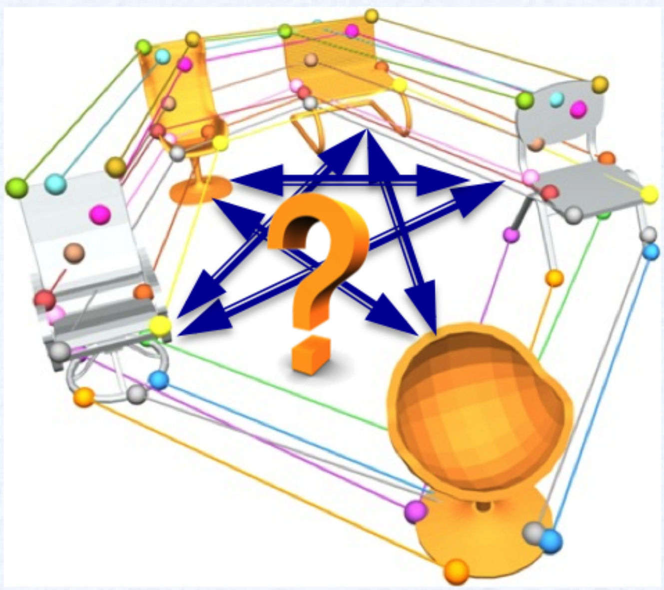
- **Measurement Graph**
  - $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - measurements taken over all edges $(i, j) \in \mathcal{E}$

- *Graphical channels* [AbbeMontanari'2013]
  - *community detection, graph-based coding, constrained satisfaction ...*

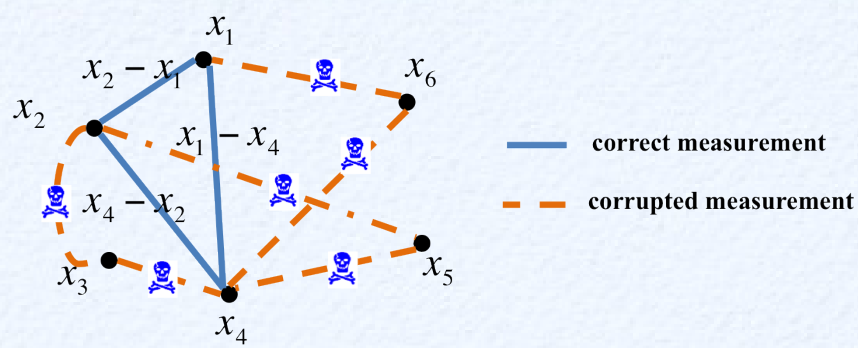
Challenge 1: Missing Measurements



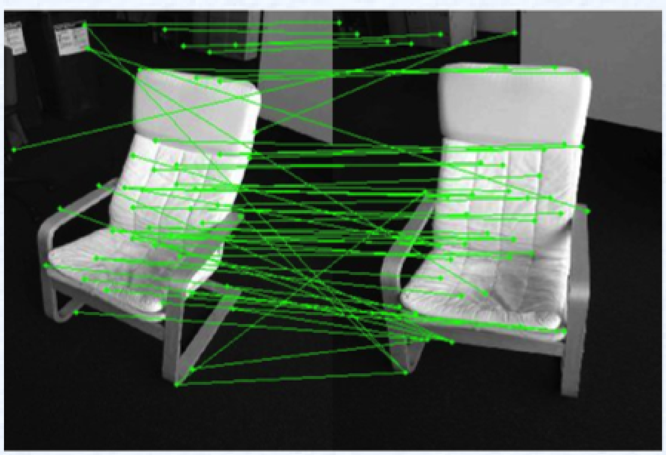
- **Incompleteness**
  - measurement graph $\mathcal{G}$ is sparse



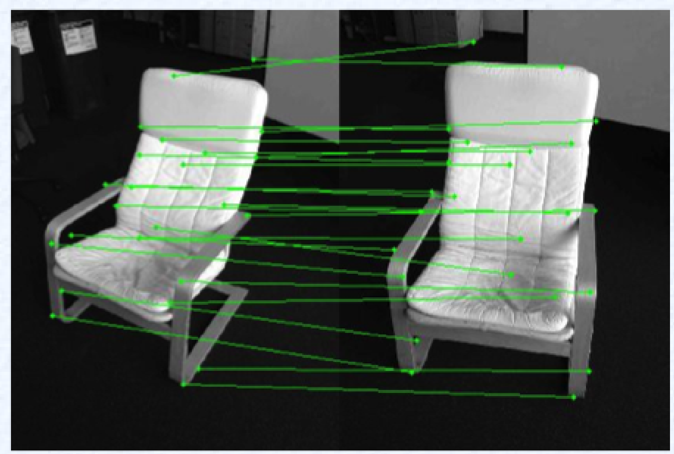
Challenge 2: Corrupted Measurements



- **Dense Measurement Errors**
  - A majority of measurements can be significantly corrupted



Measurements



Ground Truth

Fundamental Limit for Information Recovery

- **Random Corruption Model (for Pairwise Difference)**

$$\forall (i,j) \in \mathcal{G} \quad \text{measurement}(i,j) = \begin{cases} x_i - x_j, & \text{w.p. } \boldsymbol{p}_{\text{true}} \qquad \text{accurate} \\ \text{uniform}\,(M)\,, & \text{else} \qquad \text{corrupted} \end{cases}$$

  - $\mathcal{G}$: measurement graph
  - M: # possible values
  - $\boldsymbol{p}_{\text{true}}$: non-corruption rate

Fundamental Limit for Information Recovery

- **Random Corruption Model (for Pairwise Difference)**

$$\forall (i,j) \in \mathcal{G} \quad \text{measurement}(i,j) = \begin{cases} x_i - x_j, & \text{w.p. } \boldsymbol{p}_{\text{true}} \qquad \text{accurate} \\ \text{uniform}\,(M)\,, & \text{else} \qquad \text{corrupted} \end{cases}$$

  - $\mathcal{G}$: measurement graph
  - M: $\#$ possible values
  - $\boldsymbol{p}_{\text{true}}$: non-corruption rate

- **Goal:** identify the threshold on p_{true} below which recovery is impossible

Fundamental Limit for Information Recovery

- **Random Corruption Model (for Pairwise Difference)**

$$\forall(i,j) \in \mathcal{G} \quad \text{measurement}(i,j) = \begin{cases} x_i - x_j, & \text{w.p. } \boldsymbol{p}_{\text{true}} \qquad \text{accurate} \\ \text{uniform}\,(M)\,, & \text{else} \qquad\qquad \text{corrupted} \end{cases}$$

  - $\mathcal{G}$: measurement graph
  - M: # possible values
  - $\boldsymbol{p}_{\text{true}}$: non-corruption rate

Theorem. For various homogeneous graphs, exact info recovery is possible iff

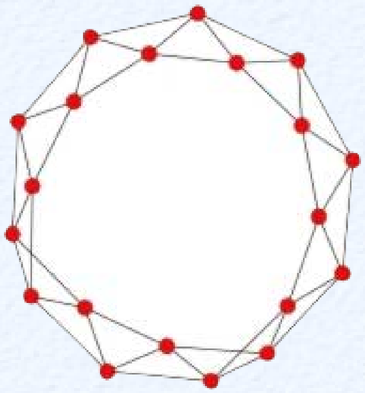
$$p_{\text{true}} \gtrsim \begin{cases} \frac{1}{\sqrt{M \cdot \text{avg-degree}}}, & \text{if } M \leq \text{avg-degree} \\ \frac{1}{\text{avg-degree}}, & \text{else} \end{cases}$$

$\gtrsim$: order-wise bound up to a polylog factor

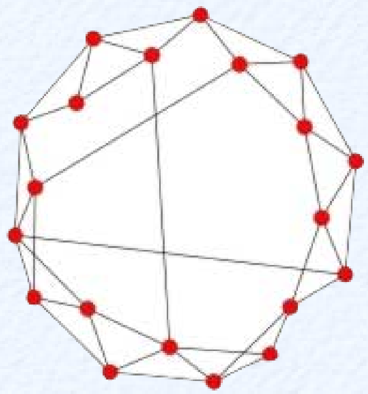
Key Metrics: Edge Sparsity

Theorem. For various homogeneous graphs, exact info recovery is possible iff

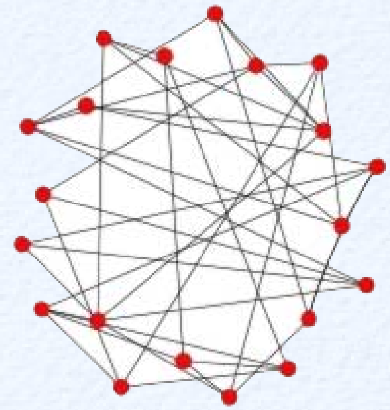
$$p_{\text{true}} \gtrsim \begin{cases} \frac{1}{\sqrt{M \cdot \text{avg-degree}}}, & \text{if } M \le \text{avg-degree} \\ \frac{1}{\text{avg-degree}}, & \text{else} \end{cases}$$



geometric graphs



small world graphs

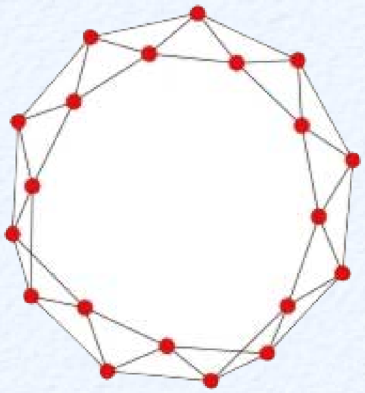


Erdos Renyi random graphs

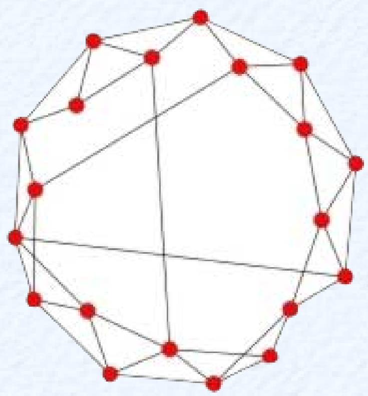
Key Metrics: Edge Sparsity

Theorem. For various homogeneous graphs, exact info recovery is possible iff

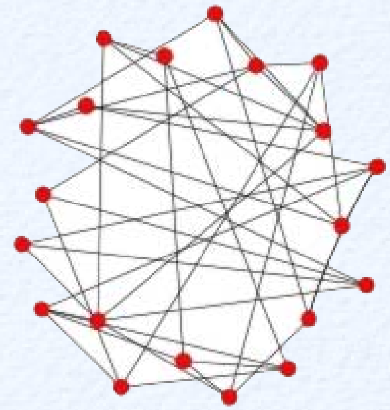
$$p_{\text{true}} \gtrsim \begin{cases} \frac{1}{\sqrt{M \cdot \text{avg-degree}}}, & \text{if } M \leq \text{avg-degree} \\ \frac{1}{\text{avg-degree}}, & \text{else} \end{cases}$$



geometric graphs



small world graphs



Erdos Renyi random graphs

- **Depend almost only on edge sparsity**
  - *irrespective of other graphical metrics (e.g. spectral gaps)*

More Precisely: Which homogeneous graphs?

- **Key Metric**

$$\mathcal{N}_k \triangleq \left\{ \text{Vertex Set } S : \underbrace{|\partial S|}_{\#\text{ edges crossing from } S \text{ to } S^c} \leq k \right\} \tag{1}$$

- Our results work for all graphs such that

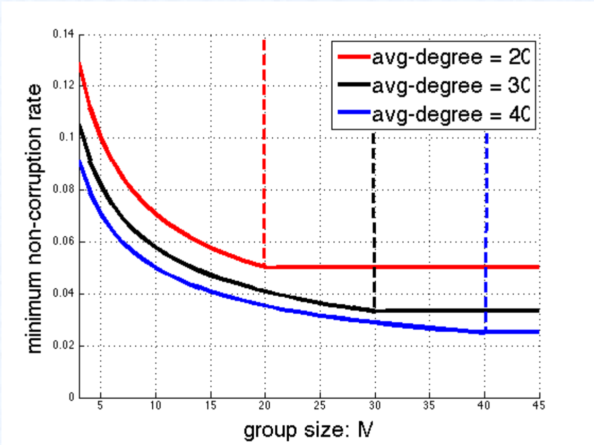
$$\forall k : \quad \frac{\log |\mathcal{N}_k|}{k} = O\left(\text{poly}\log(n)\right)$$

  - Erdos-Renyi Graphs
  - Random Geometric Graphs
  - Small-World Graphs
  - Other Expander Graphs ...

Key Metrics: Universe Size

Theorem. For various homogeneous graphs, exact info recovery is possible iff

$$p_{\text{true}} \gtrsim \begin{cases} \frac{1}{\sqrt{M} \cdot \text{avg-degree}}, & \text{if } M \le \text{avg-degree} \\ \frac{1}{\text{avg-degree}}, & \text{else} \end{cases}$$

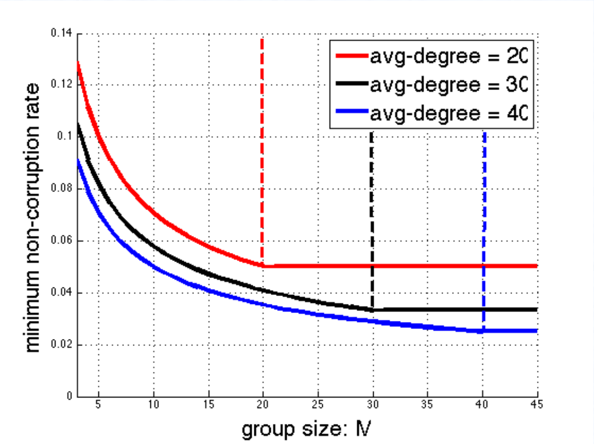


p_{true} v.s. M

Key Metrics: Universe Size

Theorem. For various homogeneous graphs, exact info recovery is possible iff

$$p_{\text{true}} \gtrsim \begin{cases} \frac{1}{\sqrt{M}\cdot\text{avg-degree}}, & \text{if } M \leq \text{avg-degree} \\ \frac{1}{\text{avg-degree}}, & \text{else} \end{cases}$$



p_{true} v.s. M

- **Improve with the universe size M**
  - *before entering a connectivity-limited regime*

Generalization

- **Same order-wise limits hold for general pairwise measurements!**
  - Any relation $f(x,y)$ such that

$$\forall y, \quad f(\cdot, y) \text{ is bijective}$$
$$\forall x, \quad f(x, \cdot) \text{ is bijective}$$

— *Pairwise Difference*

— *Pairwise Sum*

— *Many Other Linear Relations ...*

- Abbe et al (ISIT2014) — exact limit for Erdos-Renyi graphs when $M = 2$

Computational Feasibility?

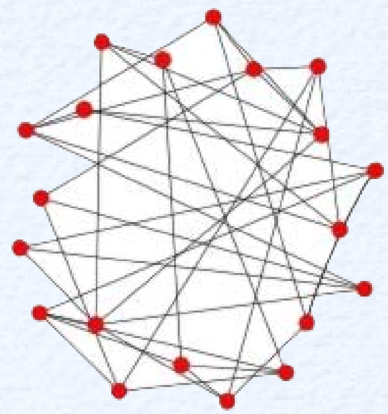
- Is the fundamental limit achievable by tractable algorithms?

Computational Feasibility?

- Is the fundamental limit achievable by tractable algorithms?

- **Consider Convex Relaxation**

- **Applied to the following example:**
  - pairwise difference: $x_i - x_j \bmod M$
  - Erdos-Renyi random graphs $\mathcal{G}(n, p_{\text{obs}})$



Erdos-Renyi random graphs

Matrix Representation

- **Object Representation**
  - cyclic shift of $\boldsymbol{I}$

$$x_i = k \quad \text{represented by} \quad \boldsymbol{X}_i = \begin{array}{c} \begin{array}{ccccc} \multicolumn{3}{c}{\overbrace{\hphantom{0\quad 0\quad 0}}^{k\ \text{columns}}} & & \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$x_i - x_j \bmod M \quad \text{represented by} \quad \boldsymbol{X}_i \boldsymbol{X}_j^\top$$

Matrix Representation

- **Object Representation**
  - cyclic shift of $\boldsymbol{I}$

$$x_i = k \quad \text{represented by} \quad \boldsymbol{X}_i = \begin{bmatrix} \overbrace{0 \quad 0 \quad 0}^{k \text{ columns}} & 1 & 0 \\ 0 \quad 0 \quad 0 & 0 & 1 \\ 1 \quad 0 \quad 0 & 0 & 0 \\ 0 \quad 1 \quad 0 & 0 & 0 \\ 0 \quad 0 \quad 1 & 0 & 0 \end{bmatrix}$$

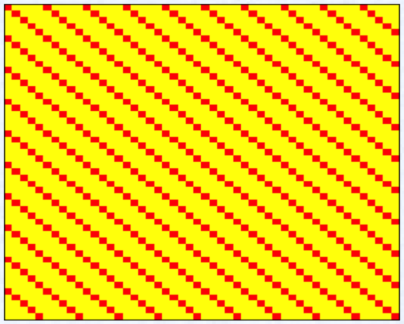
$$x_i - x_j \text{ mod } M \quad \text{represented by} \quad \boldsymbol{X}_i \boldsymbol{X}_j^\top$$

- **Low-Rank P.S.D. Matrix**

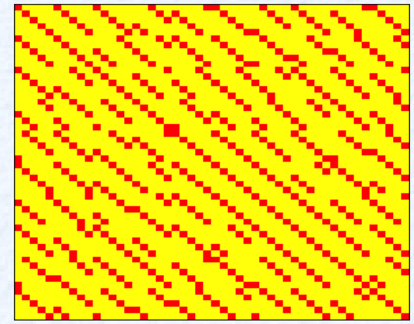
$$\boldsymbol{X} := \begin{bmatrix} \boldsymbol{I} & \boldsymbol{X}_{12} & \cdots & \boldsymbol{X}_{1n} \\ \boldsymbol{X}_{21} & \boldsymbol{I} & \cdots & \boldsymbol{X}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_{n1} & \boldsymbol{X}_{n2} & \cdots & \boldsymbol{I} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix}}_{M \text{ columns}} \begin{bmatrix} \boldsymbol{X}_1^\top & \boldsymbol{X}_2^\top & \cdots & \boldsymbol{X}_n^\top \end{bmatrix} \succeq 0$$

Exact Recovery via Convex Relaxation

Erdos-Renyi graph: $\mathcal{G}(n, p_{\text{obs}})$



$=$



ground truth

$\boldsymbol{X}^{\text{input}}$

$$
\begin{array}{ll}
\text{minimize}_{\boldsymbol{X}} & -\left\langle \boldsymbol{X}, \boldsymbol{X}^{\text{input}} \right\rangle + \lambda \left\langle \boldsymbol{X}, \boldsymbol{1}\boldsymbol{1}^{\top} \right\rangle \\
\text{subject to} & \boldsymbol{X} \geq \boldsymbol{0}, \\
& \begin{bmatrix} M & \boldsymbol{1}^{\top} \\ \boldsymbol{1} & \boldsymbol{X} \end{bmatrix} \succeq \boldsymbol{0}, \\
& \boldsymbol{X}_{ii} = \boldsymbol{I}.
\end{array}
$$

Theorem. Convex program with $\lambda = \sqrt{p_{\text{obs}}}$ is exact with high probability if

$$p_{\text{true}} \gtrsim \frac{\log^2(Mn)}{\sqrt{p_{\text{obs}} n}}$$

Optimality of Convex Relaxation?

Convex Relaxation v.s. Fundamental Limits.

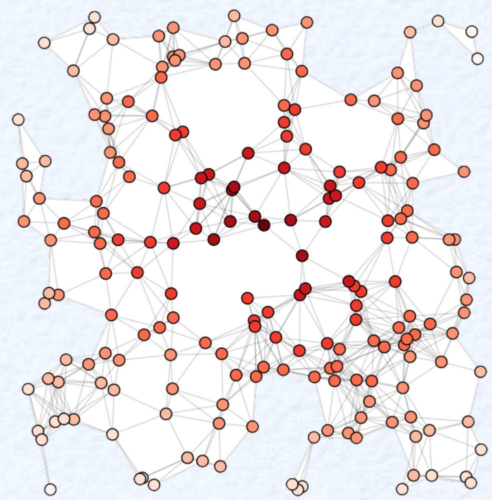
$$p_{\mathsf{true}} \gtrsim \frac{\log^2(Mn)}{\sqrt{p_{\mathsf{obs}} n}} \qquad \text{v.s.} \qquad p_{\mathsf{true}} \gtrsim \begin{cases} \frac{\log n}{\sqrt{M p_{\mathsf{obs}} n}}, & \text{if } M \leq p_{\mathsf{obs}} n \\ \frac{\log n}{p_{\mathsf{obs}} n}, & \text{else} \end{cases}$$

- **Convex relaxation is near-optimal**
  - when $M = \Theta(1)$
  - Erdos-Renyi random graph

Optimality of Convex Relaxation?

Convex Relaxation v.s. Fundamental Limits.

$$p_{\text{true}} \gtrsim \frac{\log^2(Mn)}{\sqrt{p_{\text{obs}} n}} \qquad \text{v.s.} \qquad p_{\text{true}} \gtrsim \begin{cases} \frac{\log n}{\sqrt{M p_{\text{obs}} n}}, & \text{if } M \leq p_{\text{obs}} n \\ \frac{\log n}{p_{\text{obs}} n}, & \text{else} \end{cases}$$



- **Convex relaxation is near-optimal**
  - when $M = \Theta(1)$
  - Erdos-Renyi random graph
- **Convex relaxation might be sub-optimal**
  - if M scales with n
  - general measurement graphs (e.g. geometric graphs)

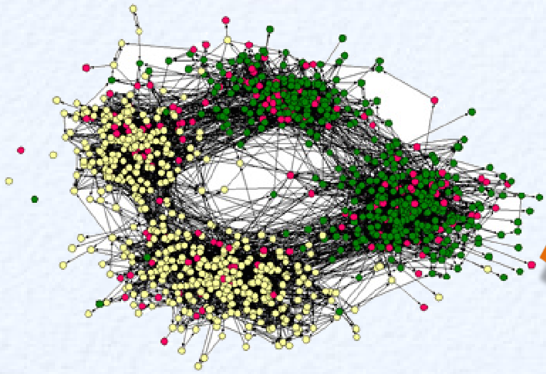
Summary: Graph-Based Joint Information Recovery

- **Fundamental Minimax Limits**
  - Depend almost only on edge sparsity
  - 2 operating regimes

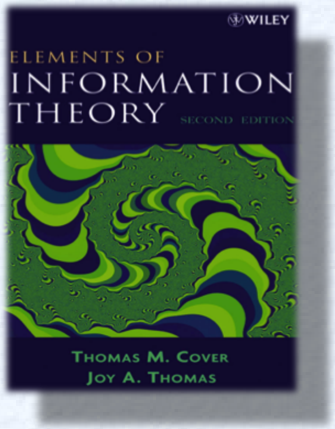
- **Computational Feasibility?**
  - SDP is near-optimal in some regime
  - A large regime remains unknown (computational barrier?)

- *Thanks to A. Montanari, E. Abbe, A. Bandeira, and A. Singer*

Concluding Remarks: Learn from Data Correlation

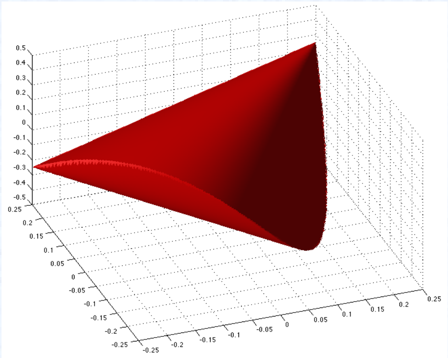


random graphs



information theory

matrix completion



convex optimization