Information Recovery from Pairwise Measurements

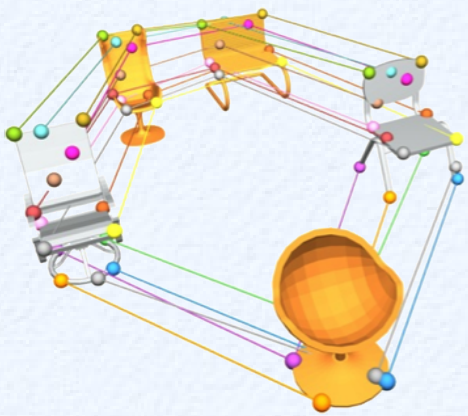
A Shannon-Theoretic Approach

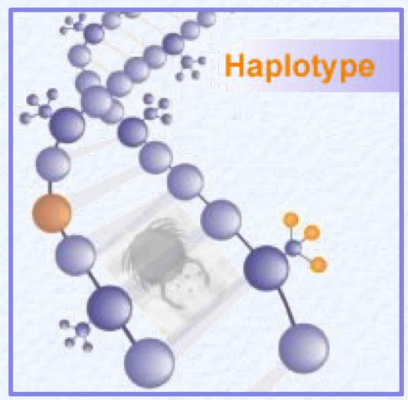
Yuxin Chen[†], **Changho Suh**[*], **Andrea Goldsmith**[†]

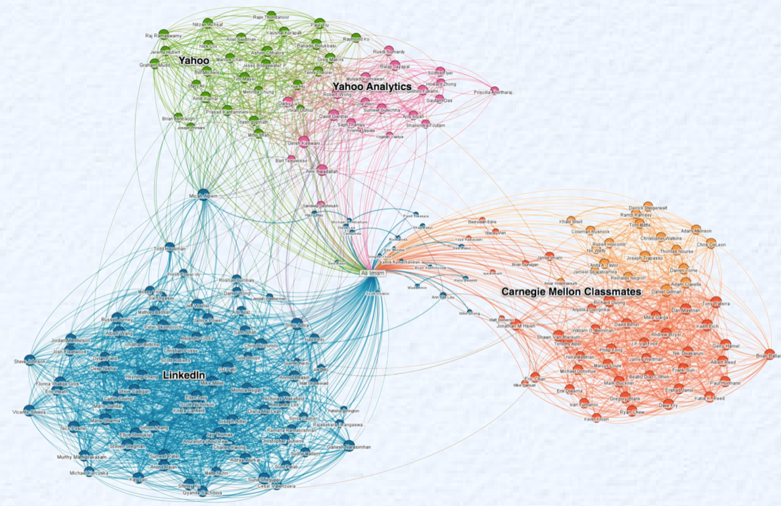
Stanford University[†] *KAIST*[*]

Recovering data from correlation measurements

- **A large collection of data instances**







- In many applications, it is
  - difficult/infeasible to measure each variable directly
  - feasible to measure pairwise correlation

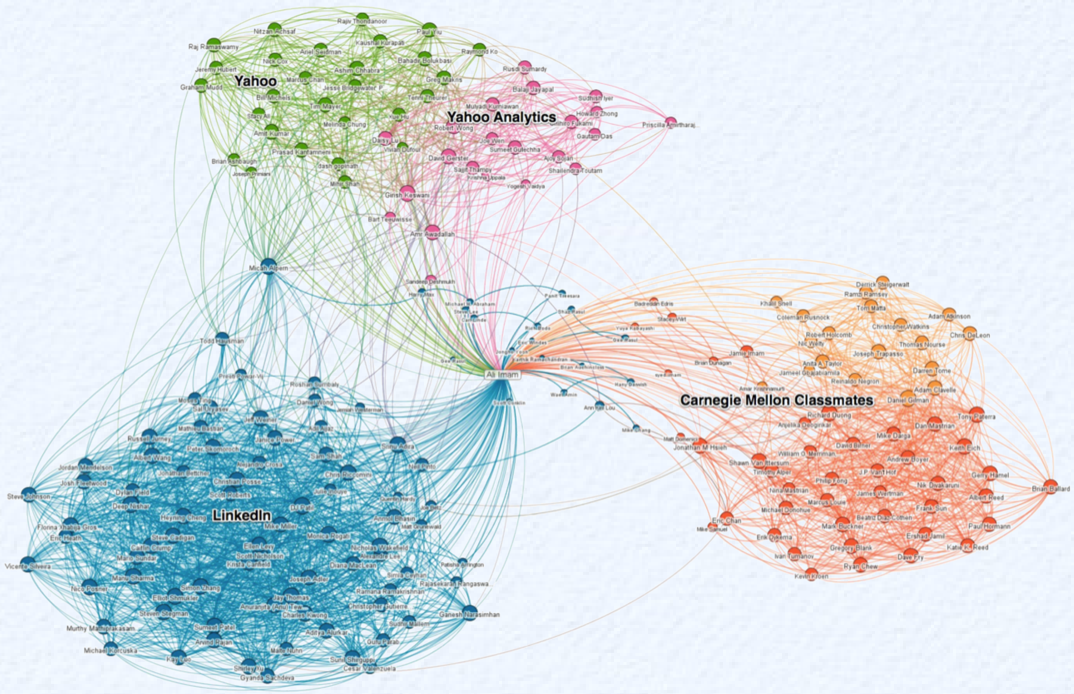
Motivating application: multi-image alignment

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

  - **Key step: joint alignment**
    - input: (noisy) estimates of relative camera poses
    - goal: jointly recover all camera poses

Motivating application: graph clustering

- **Real-world networks exhibit community structures**



  - input: pairwise similarities between members
  - goal: uncover hidden clusters

This talk: recovery from pairwise difference measurements

- Goal: recover a collection of variables $\{x_i\}$
- Can only measure several **pairwise difference** $x_i - x_j$ (broadly defined)

This talk: recovery from pairwise difference measurements

- Goal: recover a collection of variables $\{x_i\}$
- Can only measure several **pairwise difference** $x_i - x_j$ (broadly defined)
  - Examples:

— **joint alignment**

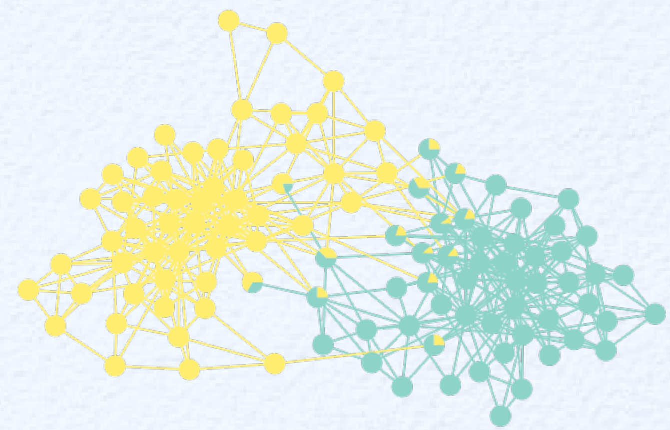
- x_i: (angle $\boldsymbol{\theta}_i$, position $\boldsymbol{z}_i$)
- relative rotation/translation $(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{z}_i - \boldsymbol{z}_j)$

This talk: recovery from pairwise difference measurements

- Goal: recover a collection of variables $\{x_i\}$
- Can only measure several **pairwise difference** $x_i - x_j$ (broadly defined)
  - Examples:

— **joint alignment**

- x_i: (angle $\boldsymbol{\theta}_i$, position $\boldsymbol{z}_i$)
- relative rotation/translation $(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{z}_i - \boldsymbol{z}_j)$

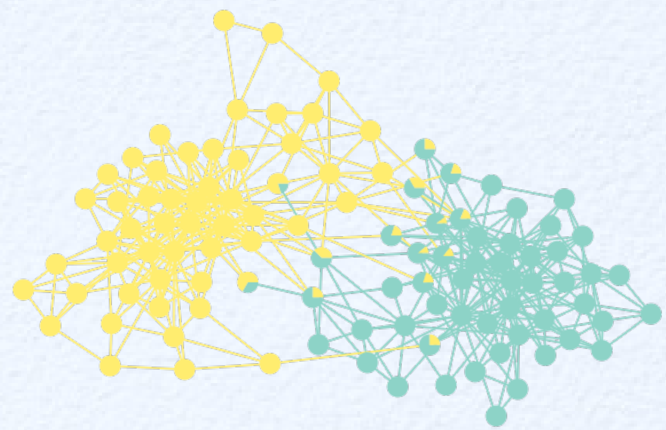


— **graph partition**

- x_i: membership (which partition it belongs to)
- cluster agreement: $x_i - x_j = \begin{cases} 1, & \text{if } i, j \in \text{ same partition} \\ 0, & \text{else.} \end{cases}$

This talk: recovery from pairwise difference measurements

- Goal: recover a collection of variables $\{x_i\}$
- Can only measure several **pairwise difference** $x_i - x_j$ (broadly defined)
  - Examples:



— **joint alignment**

- x_i: (angle $\boldsymbol{\theta}_i$, position $\boldsymbol{z}_i$)
- relative rotation/translation $(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j, \boldsymbol{z}_i - \boldsymbol{z}_j)$

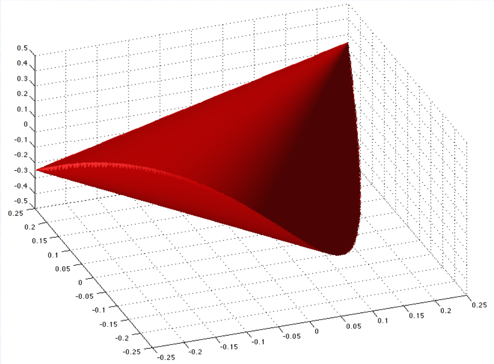
— **graph partition**

- x_i: membership (which partition it belongs to)
- cluster agreement: $x_i - x_j = \begin{cases} 1, & \text{if } i, j \in \text{ same partition} \\ 0, & \text{else.} \end{cases}$

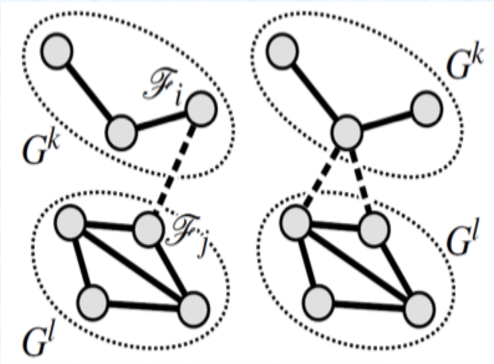
— **pairwise maps, parity reads, ...**

A fundamental-limit perspective?

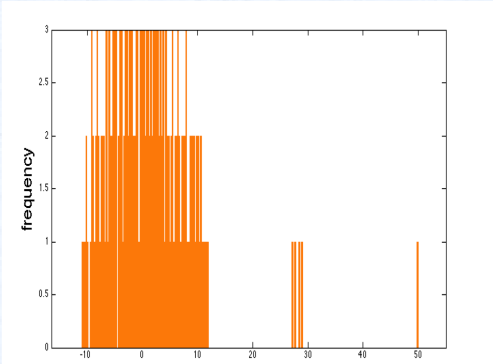
- A flurry of activity in recovery algorithm design



convex program



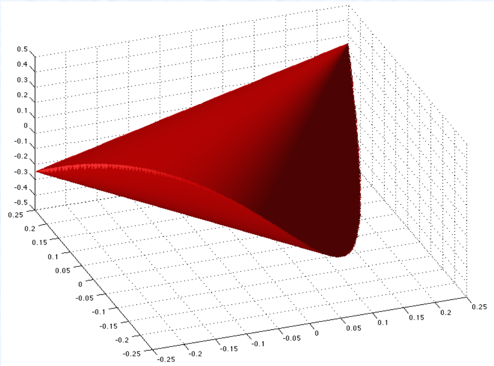
combinatorial



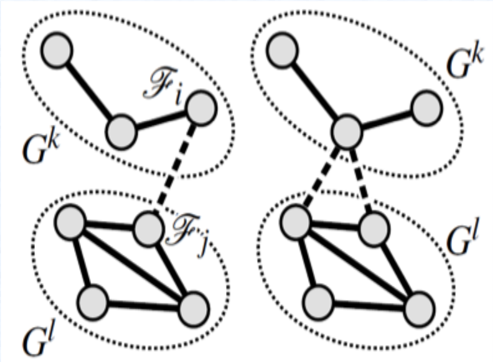
spectral method

A fundamental-limit perspective?

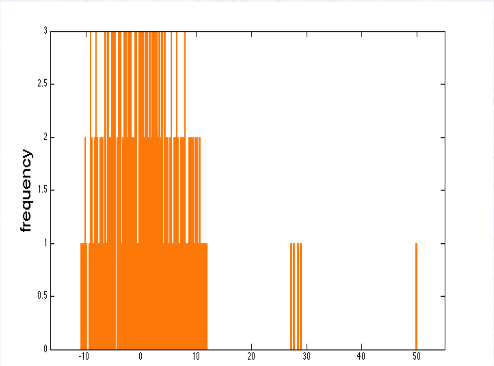
- A flurry of activity in recovery algorithm design



convex program



combinatorial



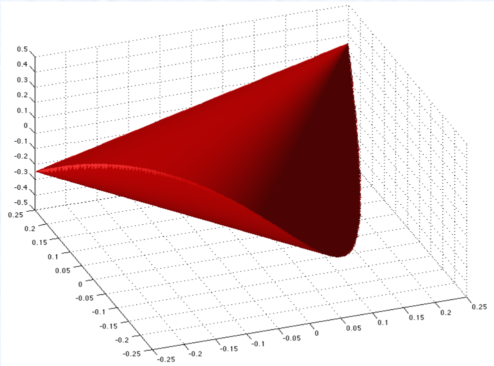
spectral method

  ◦ **What are the fundamental recovery limits?**

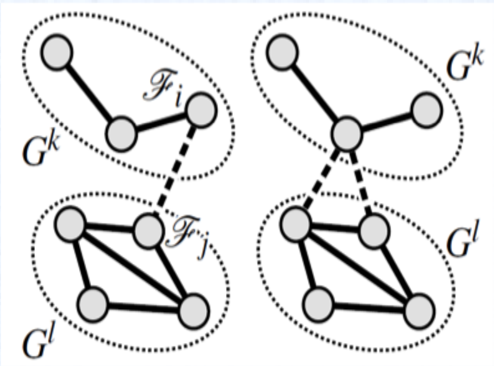
— min. sample complexity? how noisy the measurements can be?

A fundamental-limit perspective?

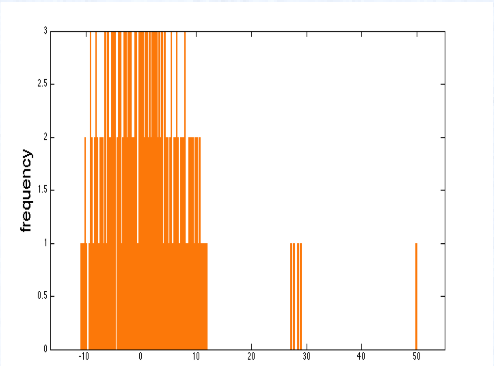
- A flurry of activity in recovery algorithm design



convex program



combinatorial



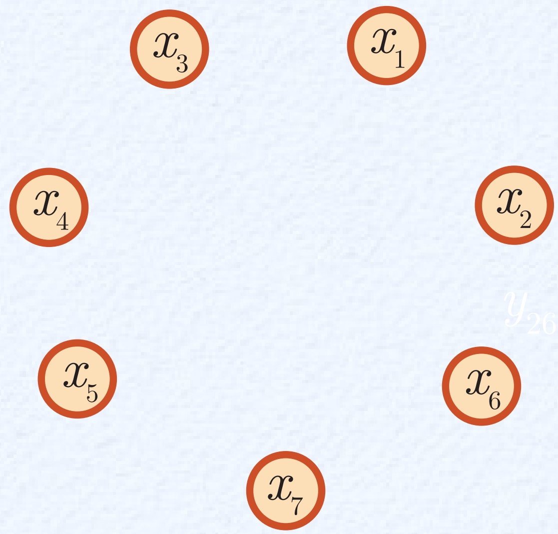
spectral method

  - **What are the fundamental recovery limits?**

  — min. sample complexity? how noisy the measurements can be?

- So far mostly studied in a model-specific manner
  - **Seek a more unified framework**

Problem setup: a Shannon-theoretic framework



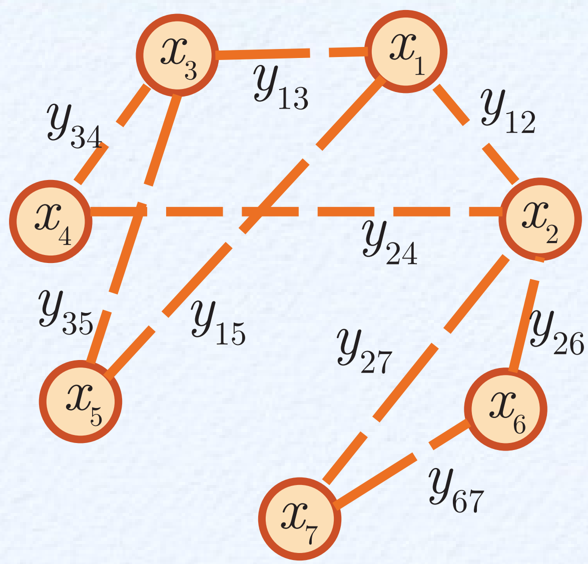
$$x_i \in \{0, \cdots, M-1\}$$

- **Information network**
  - n vertices
  - discrete inputs w/ **alphabet size:** M

  — could scale with n

Problem setup: a Shannon-theoretic framework

measurement graph $\mathcal{G}$

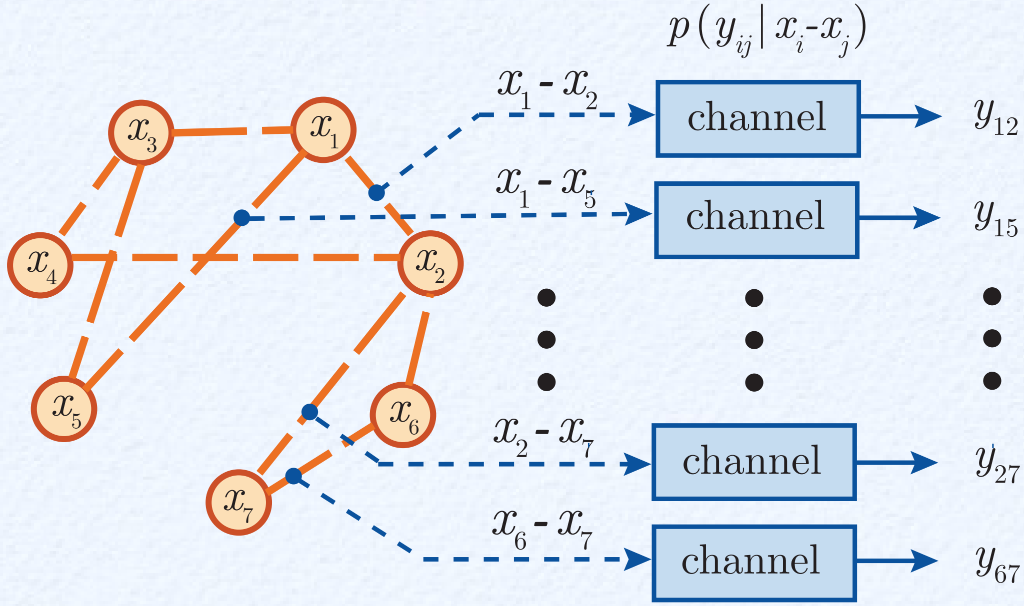


measurements of
$x_1 - x_2,\ x_1 - x_3,\ x_1 - x_5,\ \cdots$

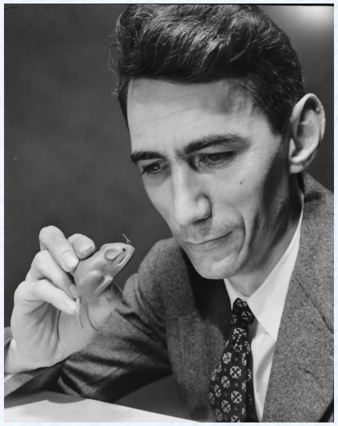
- **Pairwise difference measurements**
  - truth: $x_i - x_j$
  - measurements: y_{ij} (arbitrary alphabet)
    - $*$ can be corrupted by noise, distortion, ...

- **Graphical representation**
  - observe $y_{ij} \quad \Longleftrightarrow \quad (i,j) \in \mathcal{G}$

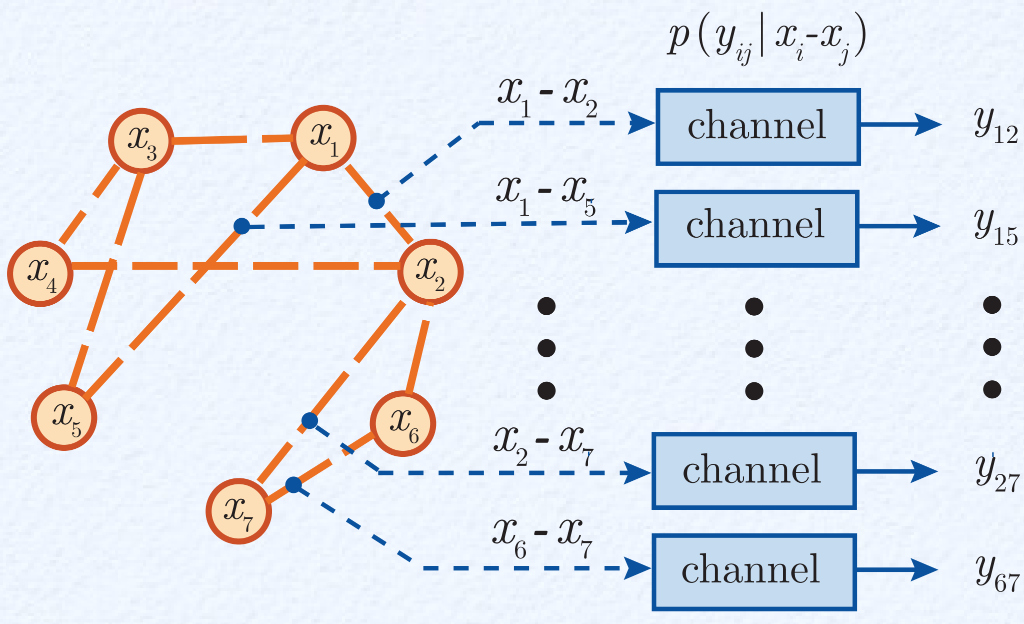
Problem setup: a Shannon-theoretic framework

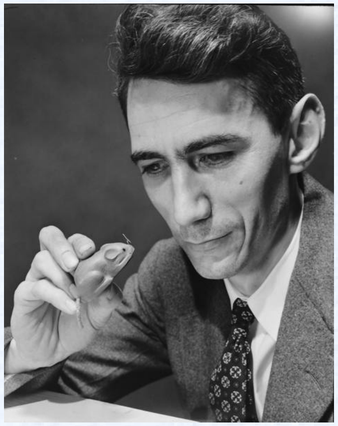


- **Channel-decoding perspective**
  - each measurement is modeled by **an i.i.d. channel**
  - transition prob. $P(y_{ij} \mid x_i - x_j)$



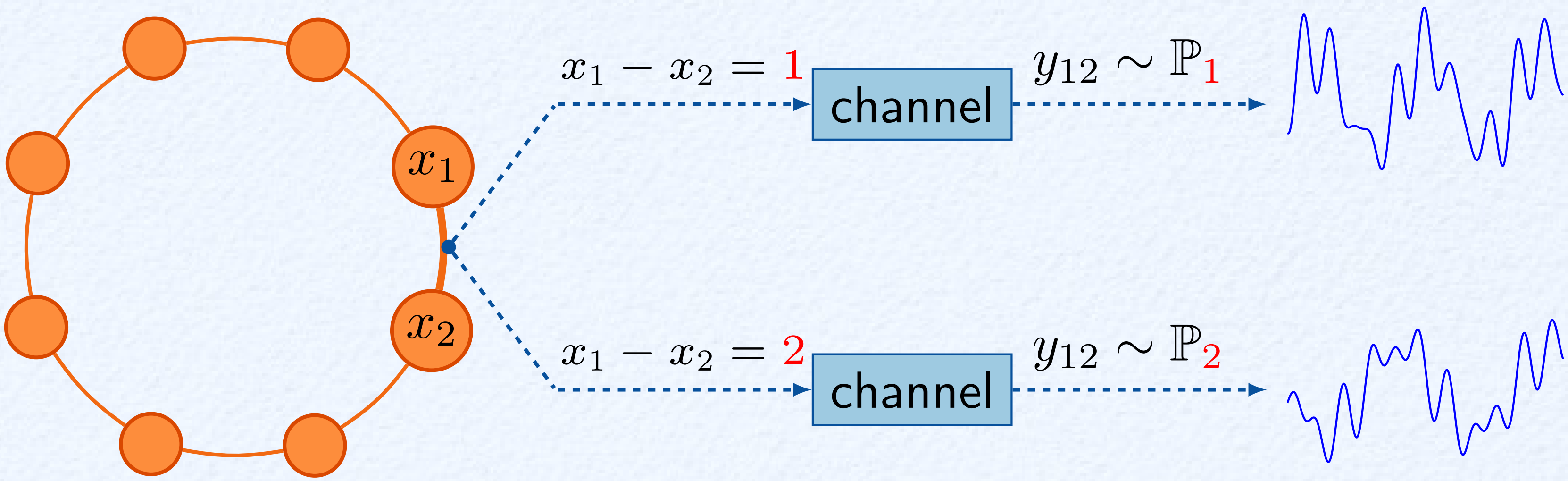
Problem setup: a Shannon-theoretic framework





- **Goal: recover $\{x_i\}$** exactly (up to global offset)
- **Unified framework for decoding model**
  - *capture similarities among various applications*

What factors dictate hardness of recovery?

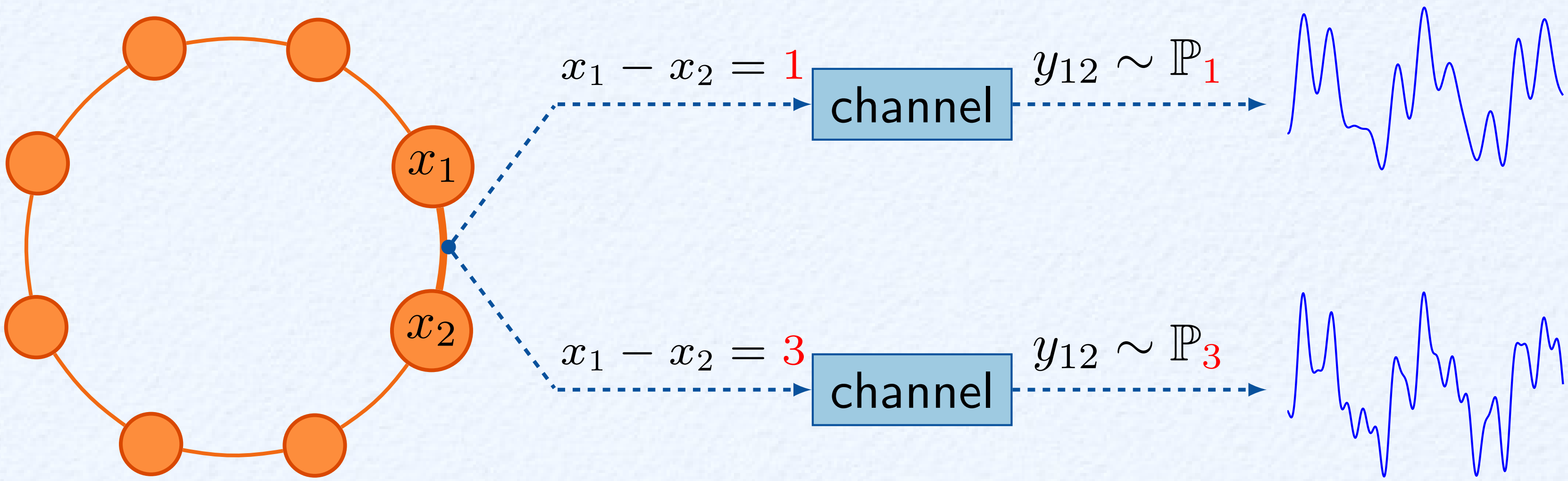


$$\mathbb{P}_l := \mathbb{P}(\, y_{ij} \mid x_i - x_j = l)$$

- **Channel distance/resolution**
  - Captured by

$$\mathrm{KL}(\, \mathbb{P}_l \,\|\, \mathbb{P}_k) \quad \text{or} \quad \text{Hellinger}(\, \mathbb{P}_l \,\|\, \mathbb{P}_k) \quad \text{or} \quad \ldots$$

What factors dictate hardness of recovery?



$$\mathbb{P}_l := \mathbb{P}(\, y_{ij} \mid x_i - x_j = l)$$

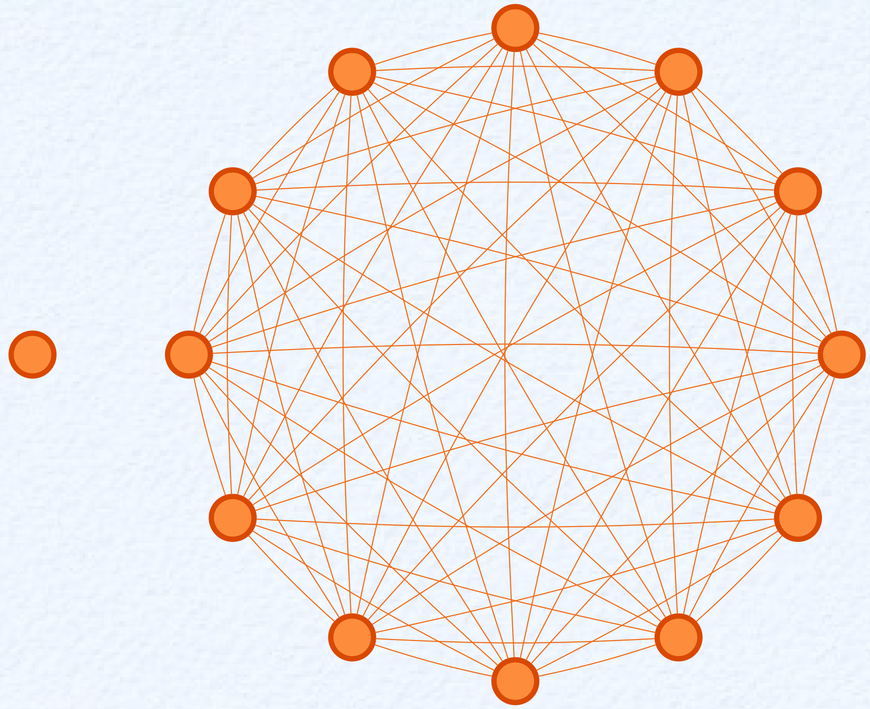
- ***Minimum* channel distance/resolution**

$$\min_{l \neq k} \mathsf{KL}(\, \mathbb{P}_l \,\|\, \mathbb{P}_k) := \mathsf{KL}^{\min} \quad \text{or}$$

$$\min_{l \neq k} \mathsf{Hellinger}(\, \mathbb{P}_l \,\|\, \mathbb{P}_k) := \mathsf{Hellinger}^{\min} \quad \text{or} \quad \ldots$$

  - Uncoded input

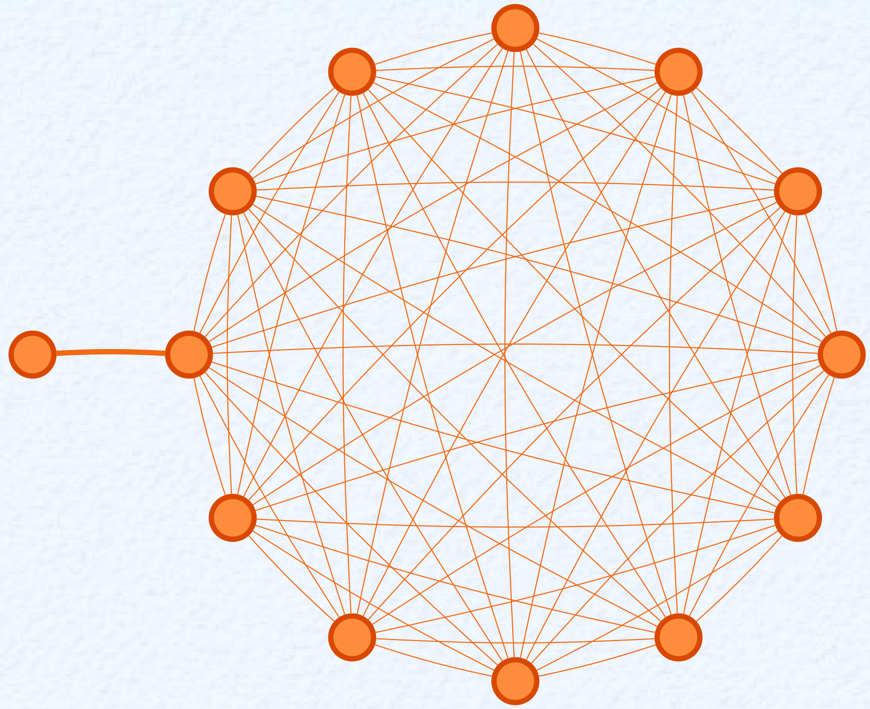
What factors dictate hardness of recovery?



measurement graph $\mathcal{G}$

- **Graph connectivity**
  - Impossible to recover isolated vertices

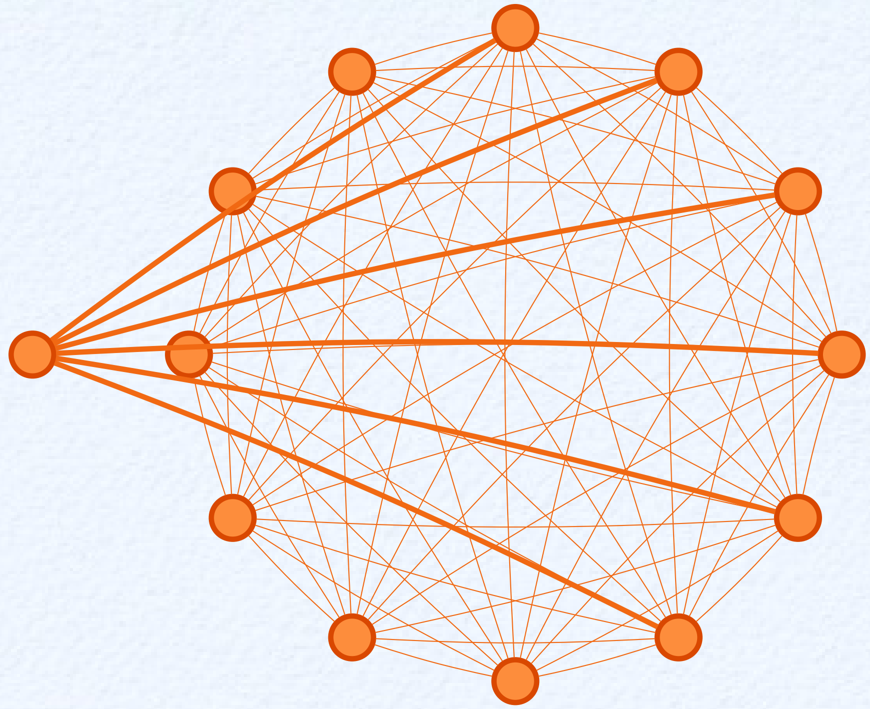
What factors dictate hardness of recovery?



measurement graph $\mathcal{G}$

- **Graph connectivity**
  - Over-sparse connectivity is fragile

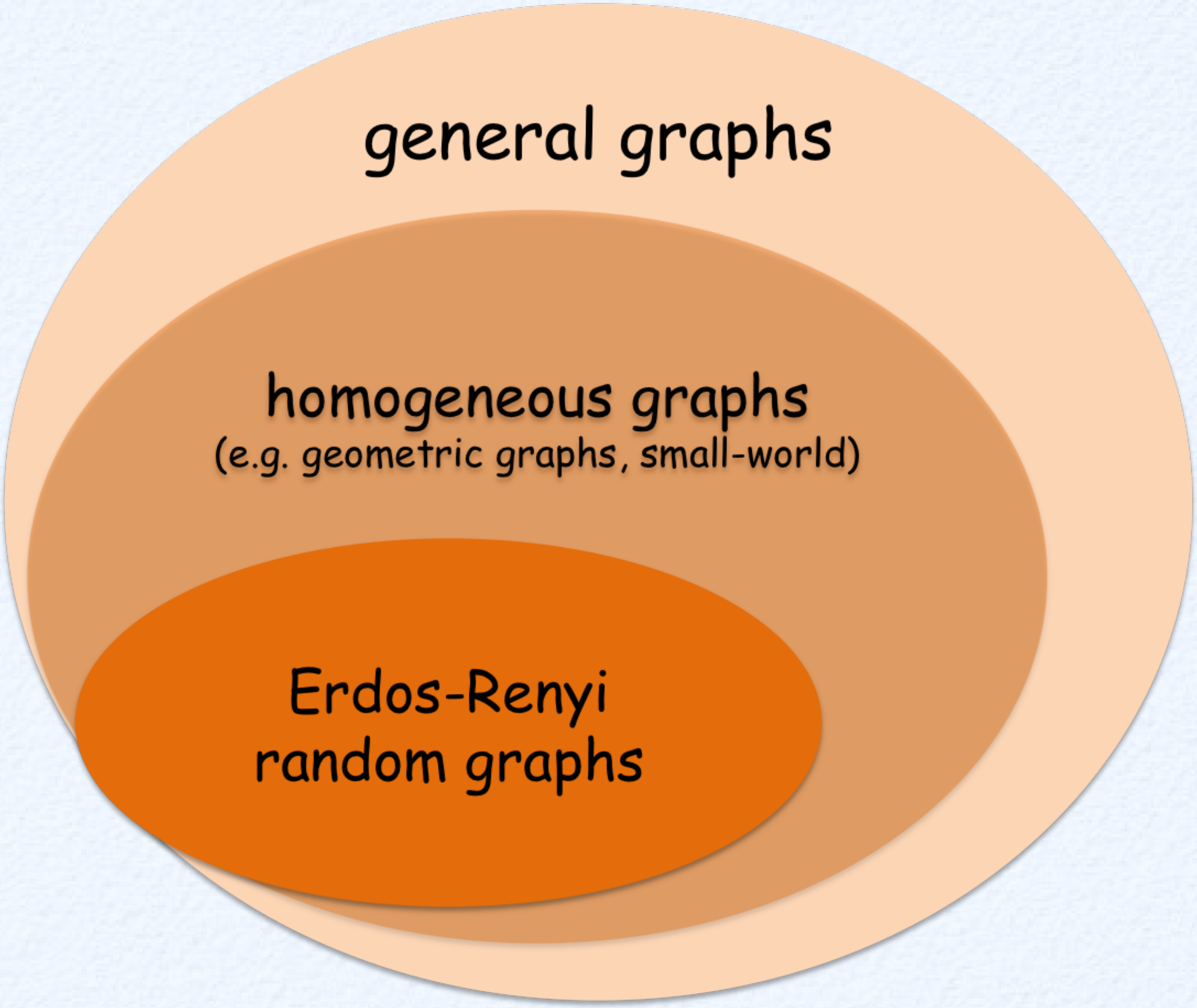
What factors dictate hardness of recovery?



measurement graph $\mathcal{G}$

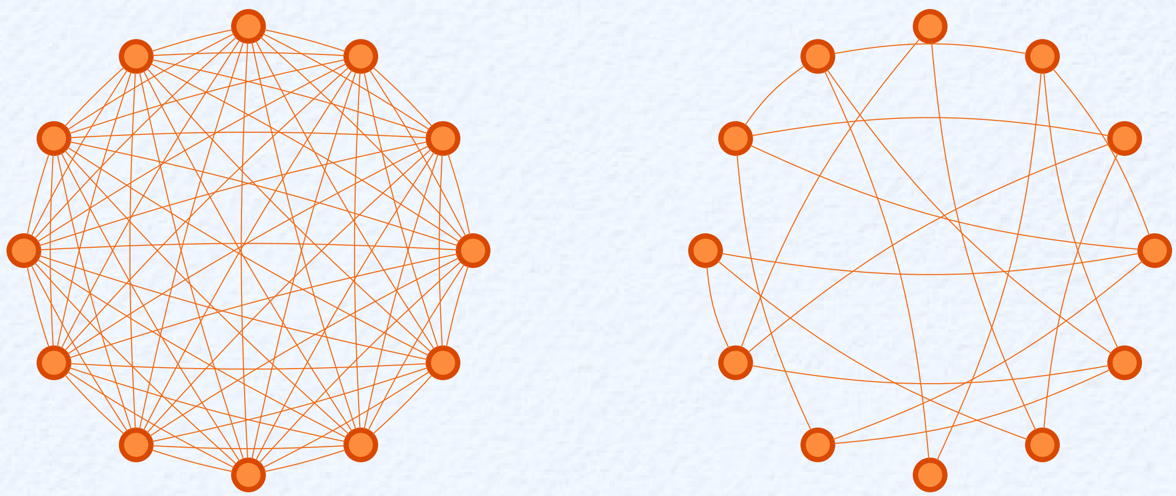
- **Graph connectivity**
  - Sufficient connectivity removes fragility!

Agenda


general graphs
homogeneous graphs
(e.g. geometric graphs, small-world)
Erdos-Renyi
random graphs

Main result: Erdos-Renyi random graph

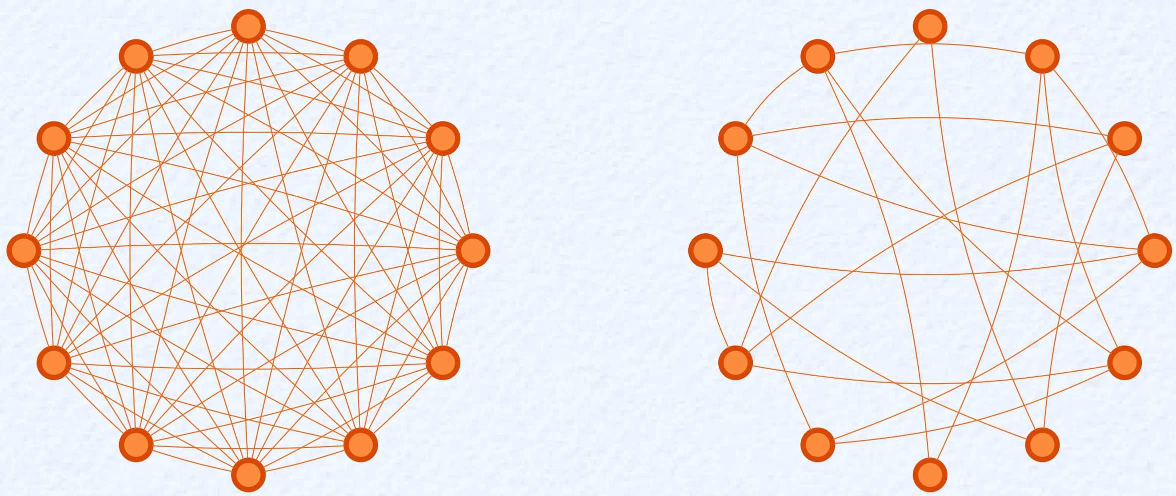
Erdos-Renyi graph $\mathcal{G}(n, p_{\text{obs}})$. Each edge (i, j) is present independently w.p. p_{obs}



$(p_{\text{obs}} = 1)$ $(p_{\text{obs}} = 0.3)$

Main result: Erdos-Renyi random graph

Erdos-Renyi graph $\mathcal{G}(n, p_{\mathsf{obs}})$. Each edge (i, j) is present independently w.p. p_{obs}



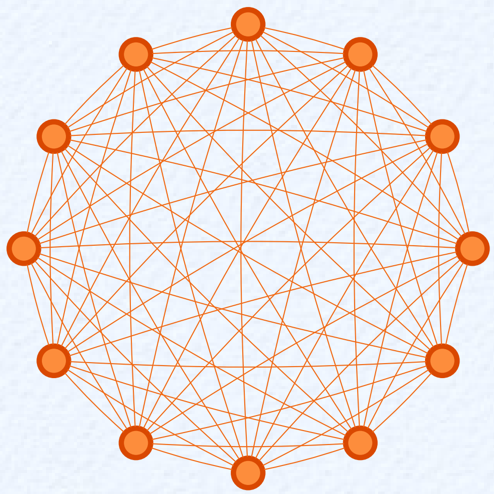
$(p_{\mathsf{obs}} = 1)$ $(p_{\mathsf{obs}} = 0.3)$

- **ML decoding works if**

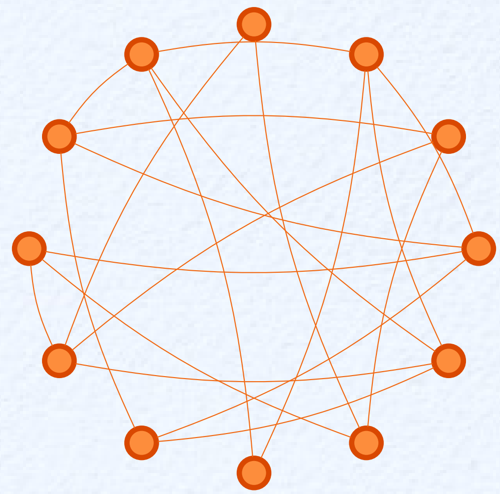
$$\text{Hellinger}^{\min} > \frac{2 \log n + 4 \log M}{p_{\text{obs}} n}$$

Main result: Erdos-Renyi random graph

Erdos-Renyi graph $\mathcal{G}(n, p_{\text{obs}})$. Each edge (i,j) is present independently w.p. p_{obs}



$(p_{\text{obs}} = 1)$



$(p_{\text{obs}} = 0.3)$

- **ML decoding works if**

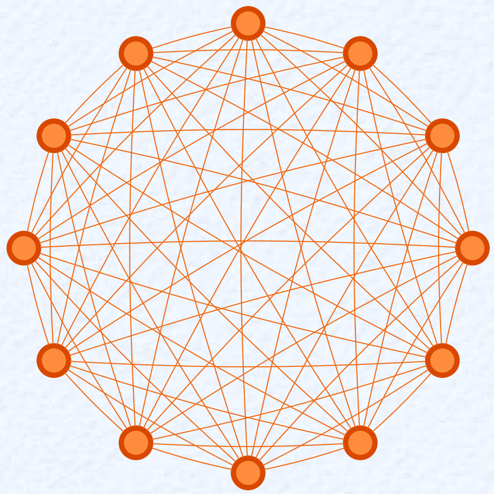
$$\text{Hellinger}^{\min} \quad > \quad \frac{2\log n + 4\log M}{p_{\text{obs}} n}$$

- **Converse: no method works if**

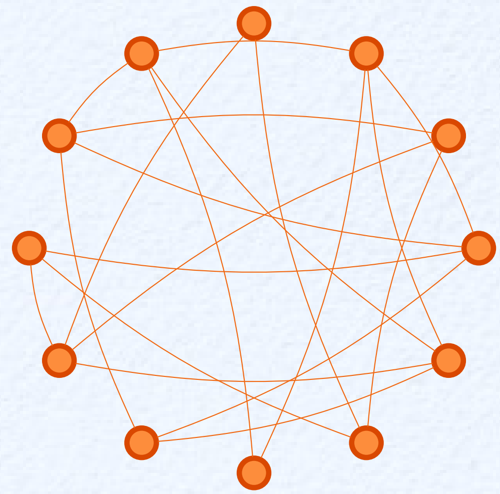
$$\text{KL}^{\min} \quad < \quad \frac{\log n}{p_{\text{obs}} n}$$

Main result: Erdos-Renyi random graph

Erdos-Renyi graph $\mathcal{G}(n, p_{\text{obs}})$. Each edge (i, j) is present independently w.p. p_{obs}



$(p_{\text{obs}} = 1)$



$(p_{\text{obs}} = 0.3)$

- **ML decoding works if**

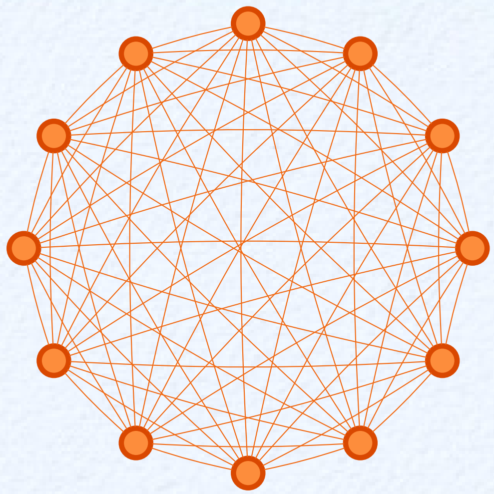
$$\text{Hellinger}^{\min} > \frac{2\log n + 4\log M}{p_{\text{obs}} n}$$

- **Converse: no method works if**

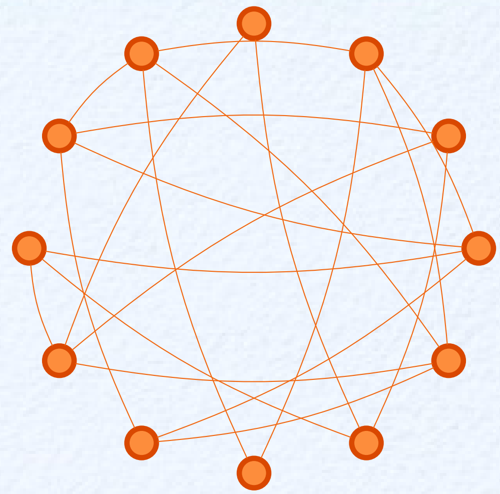
$$\text{KL}^{\min} < \frac{\log n}{p_{\text{obs}} n}$$

non-asymptotic!

Main result: Erdos-Renyi random graph

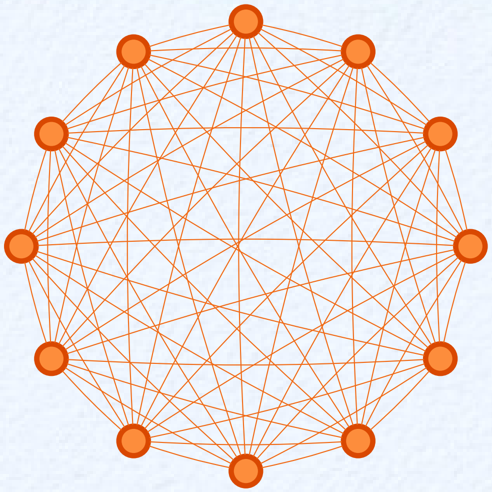


$(p_{\mathsf{obs}} = 1)$

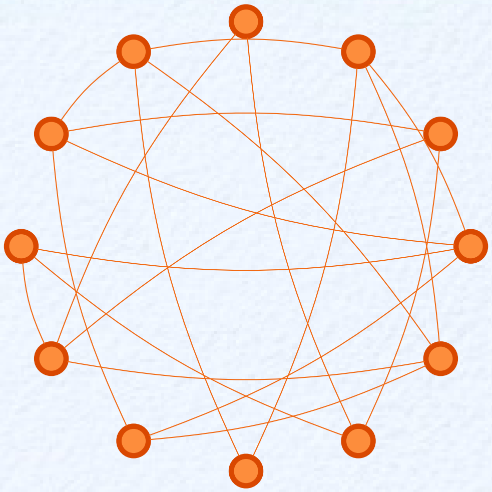


$(p_{\mathsf{obs}} = 0.3)$

Main result: Erdos-Renyi random graph



$(p_{\mathsf{obs}} = 1)$

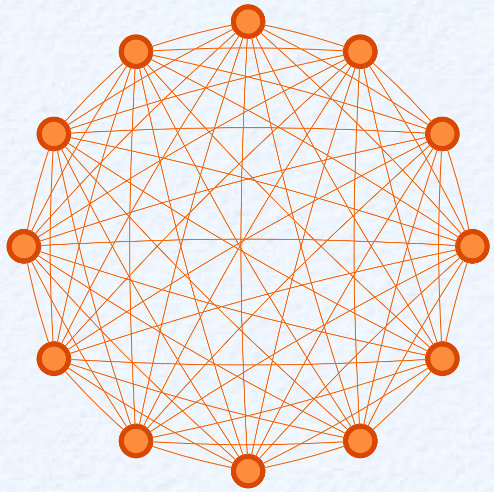


$(p_{\mathsf{obs}} = 0.3)$

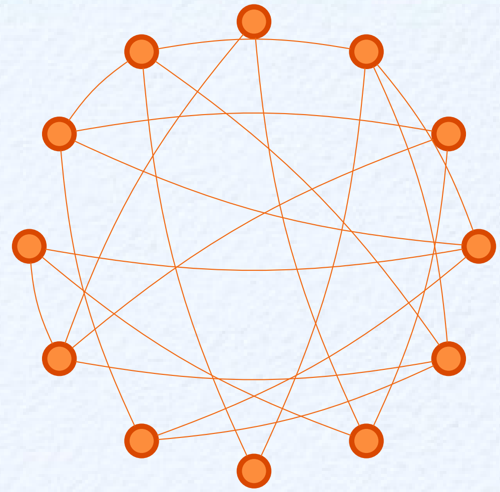
- In the **hard regime** where $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_k} \approx 1$:

$$\mathsf{KL}^{\min} \;\approx\; 2 \cdot \mathsf{Hellinger}^{\min}$$

Main result: Erdos-Renyi random graph



$(p_{\text{obs}} = 1)$



$(p_{\text{obs}} = 0.3)$

- In the **hard regime** where $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_k} \approx 1$:

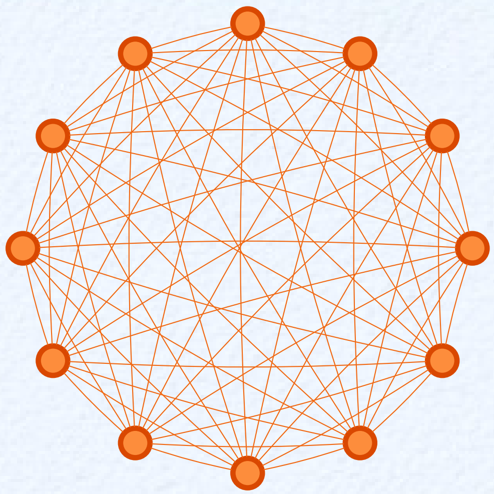
$$\mathsf{KL}^{\min} \;\approx\; 2 \cdot \mathsf{Hellinger}^{\min}$$

- **Recovery conditions**

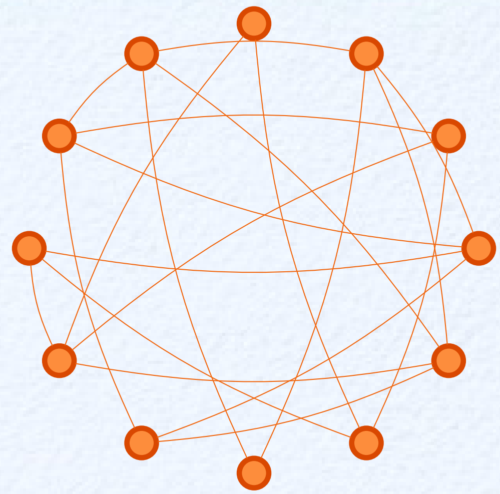
$$\text{ML works if} \qquad \mathsf{Hellinger}^{\min} > \frac{2\log n + 4\log M}{p_{\text{obs}} n}$$

$$\text{Impossible if} \qquad \mathsf{Hellinger}^{\min} < \frac{\log n}{2 p_{\text{obs}} n}$$

Main result: Erdos-Renyi random graph



$(p_{\mathsf{obs}} = 1)$



$(p_{\mathsf{obs}} = 0.3)$

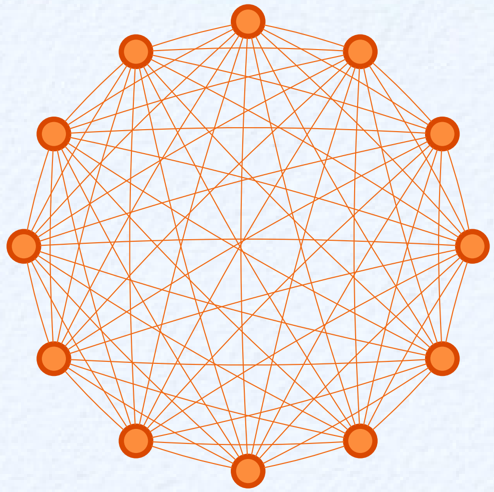
- In the **hard regime** where $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_k} \approx 1$:

$$\mathsf{KL}^{\min} \;\approx\; 2 \cdot \mathsf{Hellinger}^{\min}$$

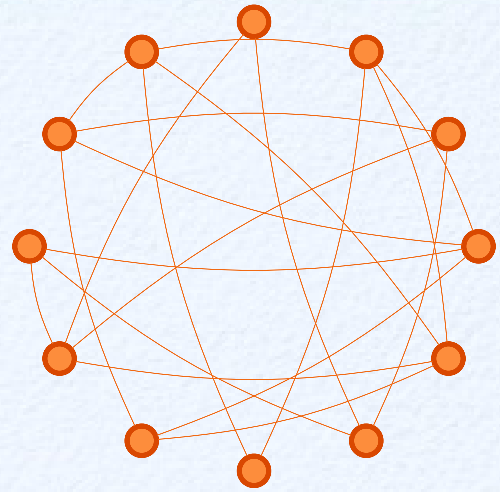
- **Fundamental recovery condition (assuming $M \lesssim \mathrm{poly}(n)$)**

$$\mathsf{Hellinger}^{\min} \;\gtrsim\; \frac{\log n}{p_{\mathrm{obs}} n}$$

Main result: Erdos-Renyi random graph



$(p_{\mathsf{obs}} = 1)$



$(p_{\mathsf{obs}} = 0.3)$

- In the **hard regime** where $\frac{\mathrm{d}\mathbb{P}_l}{\mathrm{d}\mathbb{P}_k} \approx 1$:

$$\mathsf{KL}^{\min} \;\approx\; 2 \cdot \mathsf{Hellinger}^{\min}$$

- **Fundamental recovery condition (assuming $M \lesssim \mathrm{poly}(n)$)**

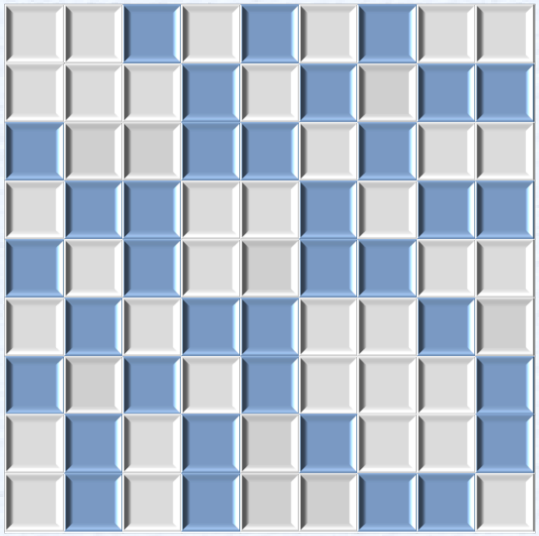
$$\mathsf{Hellinger}^{\min} \;\gtrsim\; \frac{\log n}{p_{\mathrm{obs}} n} \qquad \Longleftrightarrow \qquad \text{avg-degree} \times \mathsf{Hellinger}^{\min} \;\gtrsim\; \log n$$

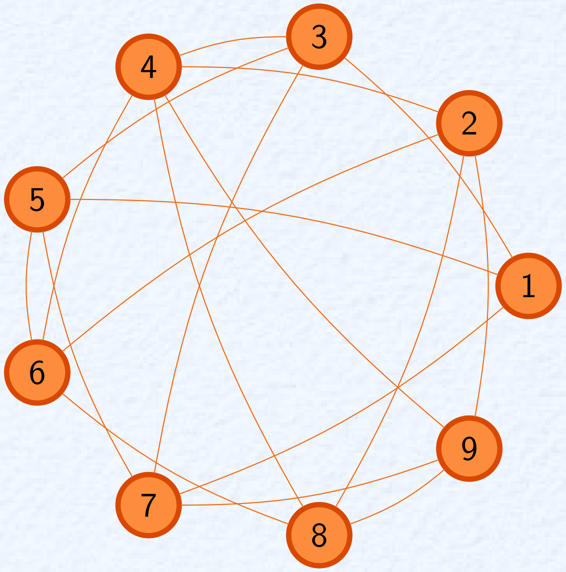
Intuition

Fundamental recovery condition (Erdos-Renyi graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

$[x_i - x_j]_{1 \le i,j \le n}$



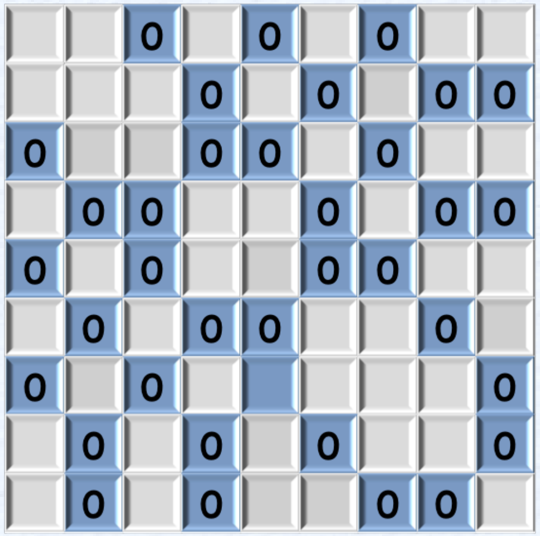


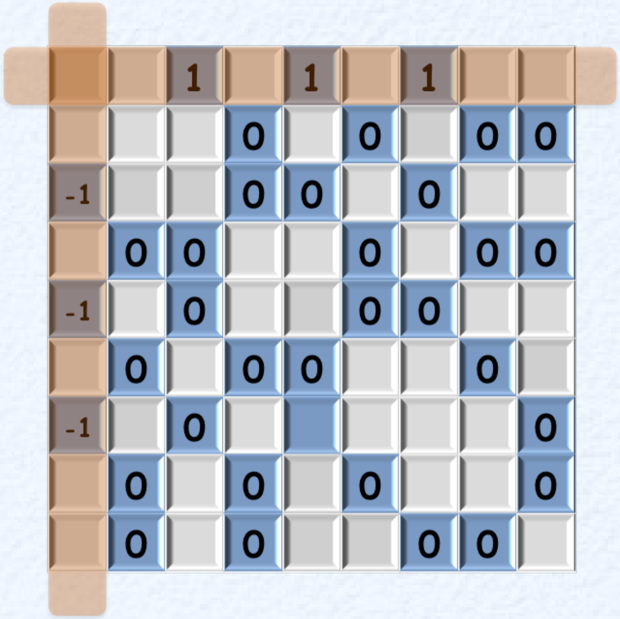
Intuition

Fundamental recovery condition (Erdos-Renyi graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

$[x_i - x_j]_{1 \le i,j \le n}$





hypotheses: H_0: $\boldsymbol{x} = [0, 0, \cdots, 0]$ H_1: $\boldsymbol{x} = [1, 0, \cdots, 0]$

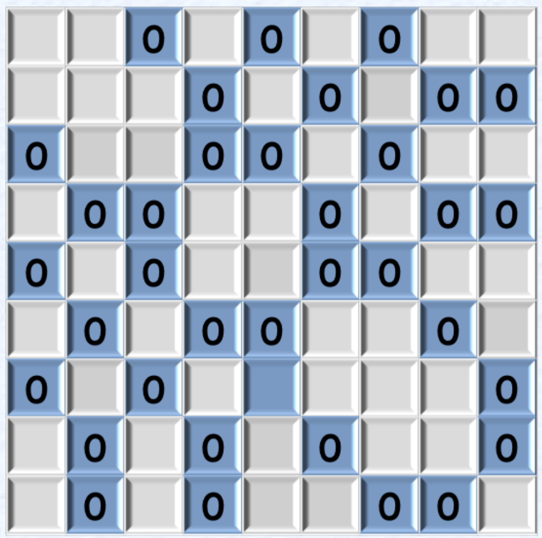
- H_0 and H_1 differ only at the highlighted region ($\approx$ avg-degree pieces of info)

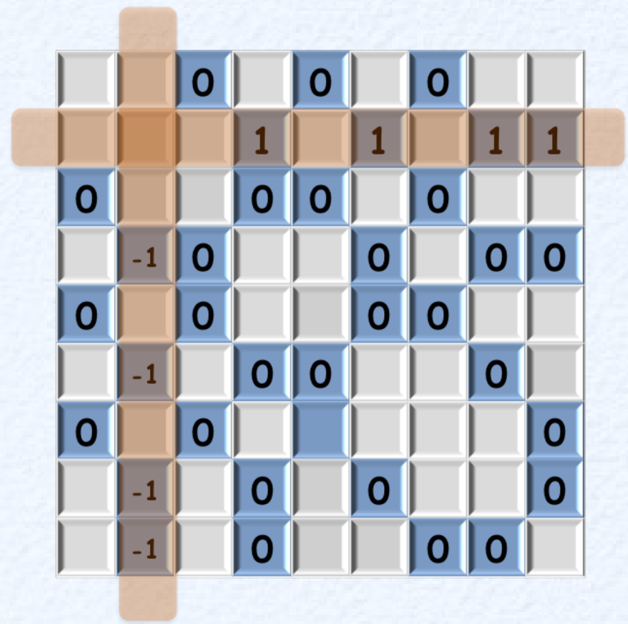
Intuition

Fundamental recovery condition (Erdos-Renyi graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

$[x_i - x_j]_{1 \le i,j \le n}$





hypotheses: H_0: $\boldsymbol{x} = [0, 0, \cdots, 0]$ H_2: $\boldsymbol{x} = [0, 1, \cdots, 0]$

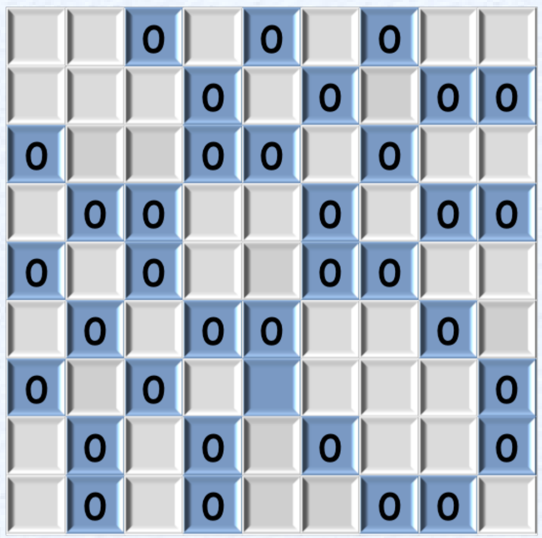
- H_0 and H_2 differ only at the highlighted region ($\approx$ avg-degree pieces of info)

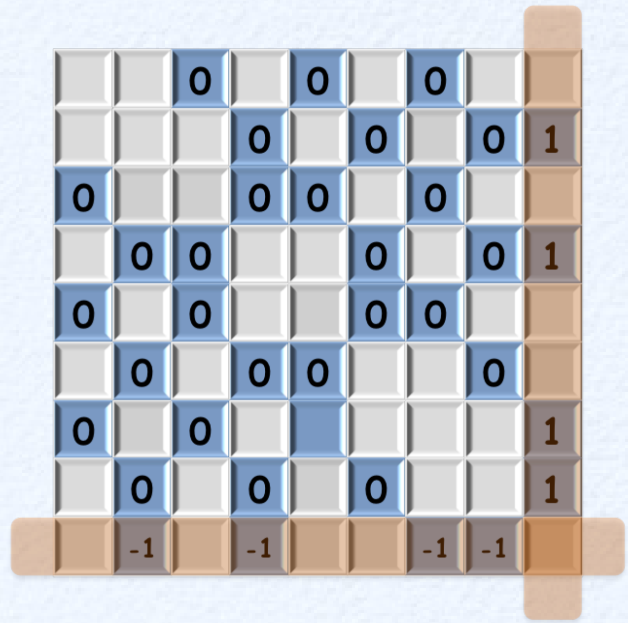
Intuition

Fundamental recovery condition (Erdos-Renyi graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n \qquad (1)$$

$[x_i - x_j]_{1 \le i,j \le n}$





hypotheses: H_0: $\boldsymbol{x} = [0, 0, \cdots, 0]$ H_n: $\boldsymbol{x} = [0, 0, \cdots, 1]$

- n minimally-separated hypotheses $\Rightarrow$ needs at least $\log n$ bits
  - the consequence of **uncoded inputs**

Minimal sample complexity

Fundamental recovery condition (Erdos-Renyi graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

- Sample complexity: $n \cdot$ avg-degree

Minimal sample complexity

Fundamental recovery condition (Erdos-Renyi graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \;\gtrsim\; \log n$$

- Sample complexity: $\quad n \cdot \text{avg-degree}$

$$\text{Min sample complexity} \;\asymp\; \frac{n \log n}{\text{Hellinger}^{\min}}$$

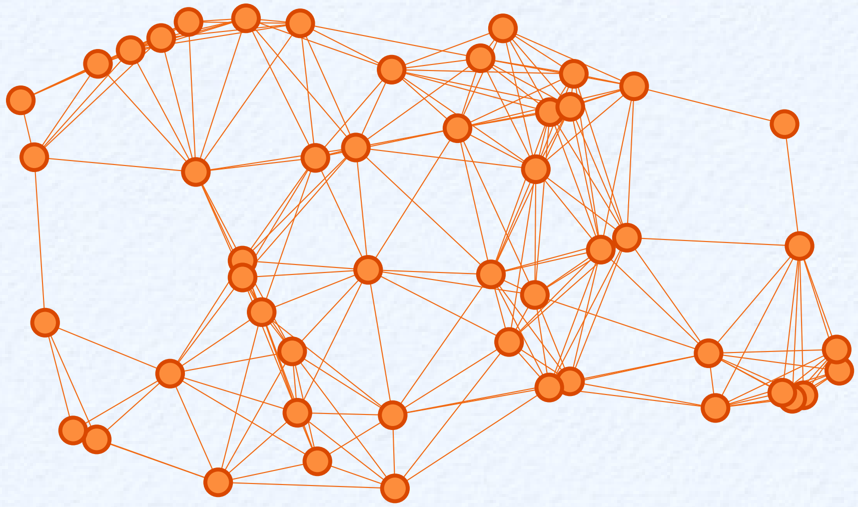
How general this limit is?

Fundamental recovery condition (Erdos-Renyi graphs).

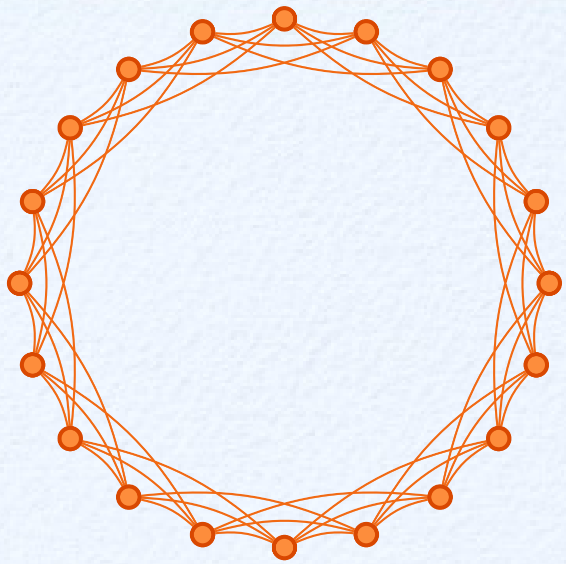
$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

- Can we go beyond Erdos-Renyi graphs?

Main results: homogeneous graphs

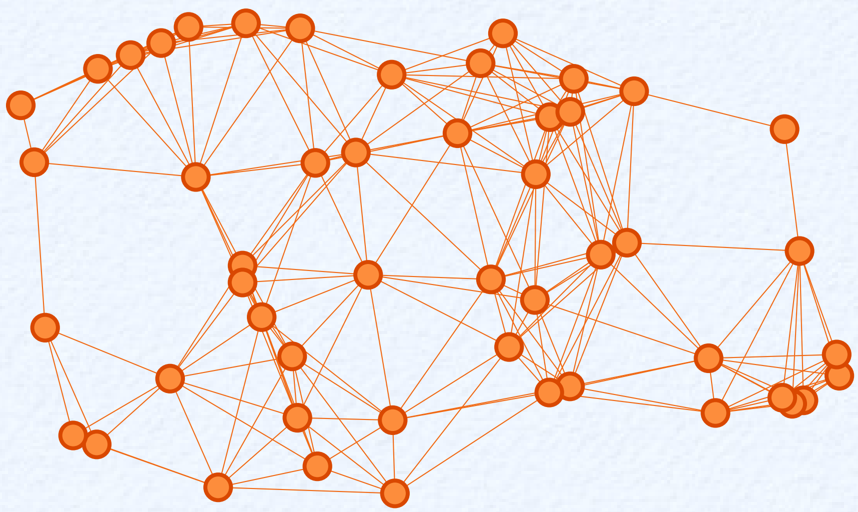


random geometric graph

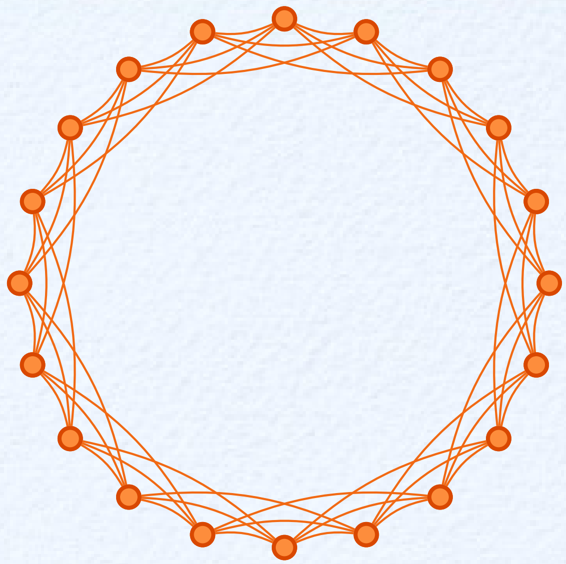


(generalized) ring

Main results: homogeneous graphs



random geometric graph



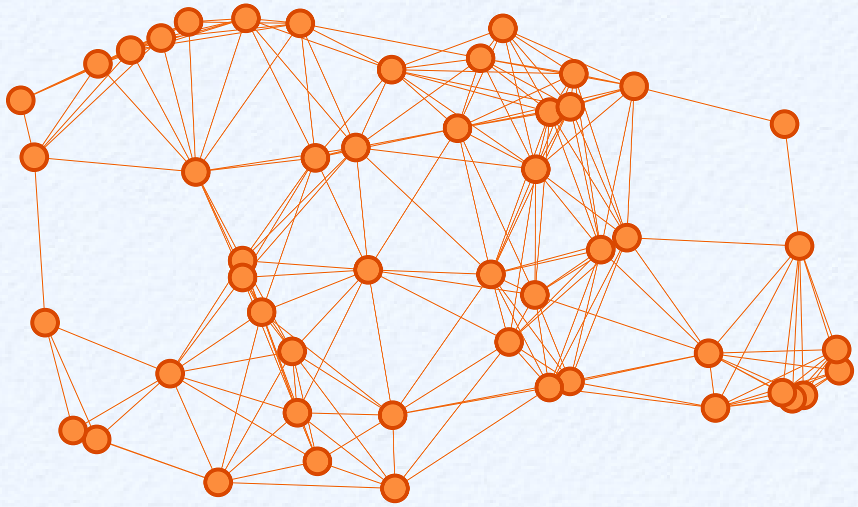
(generalized) ring

Fundamental recovery condition (various homogeneous graphs).

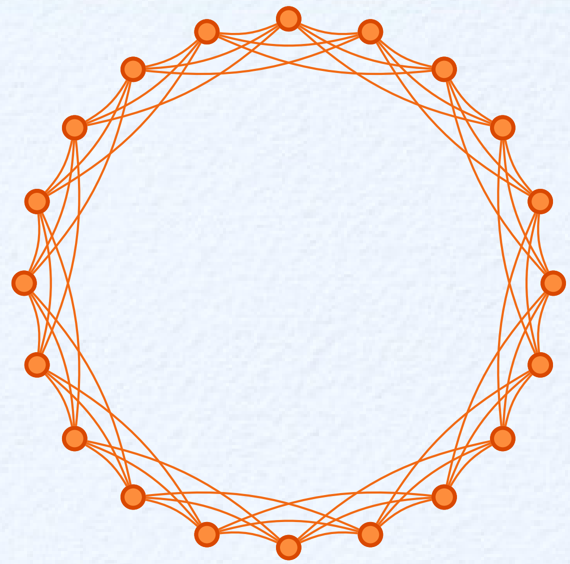
$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

- Homogeneous graphs:
  - min-degree $\asymp$ max-degree $\asymp$ mincut
  - *balanced cut-set distributions*

Main results: homogeneous graphs



random geometric graph



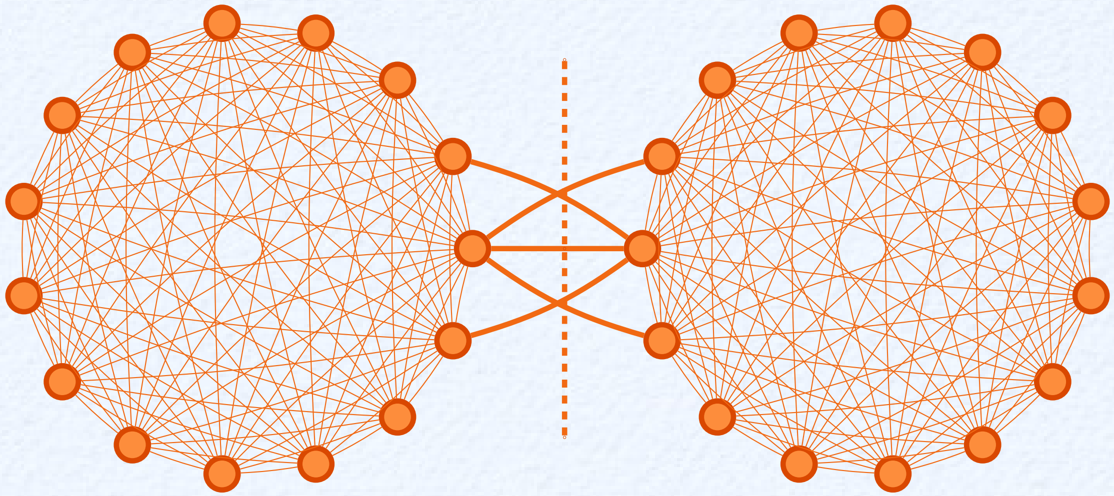
(generalized) ring

Fundamental recovery condition (various homogeneous graphs).

$$\text{avg-degree} \times \text{Hellinger}^{\min} \gtrsim \log n$$

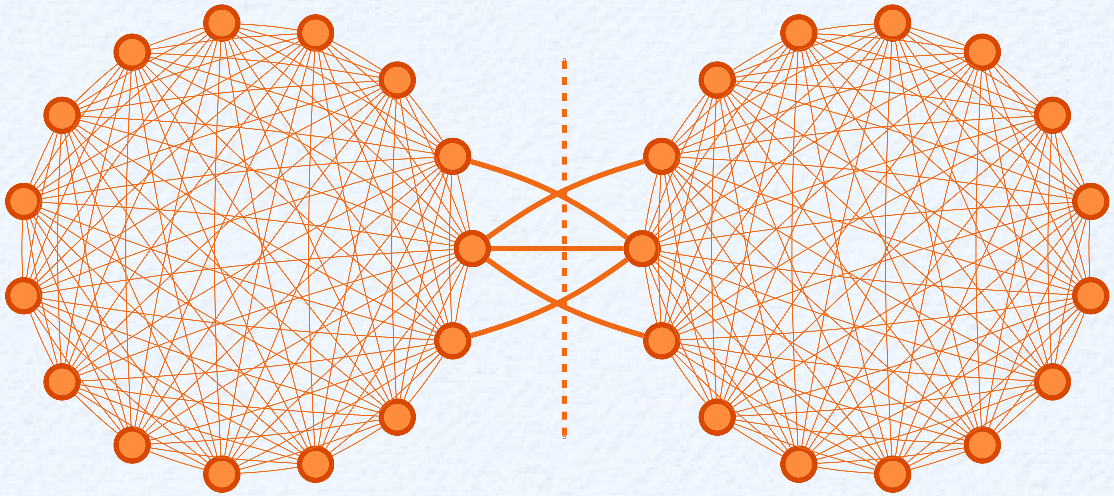
- Depend almost only on graph sparsity

Main results: general graphs



mincut $= 5$

Main results: general graphs

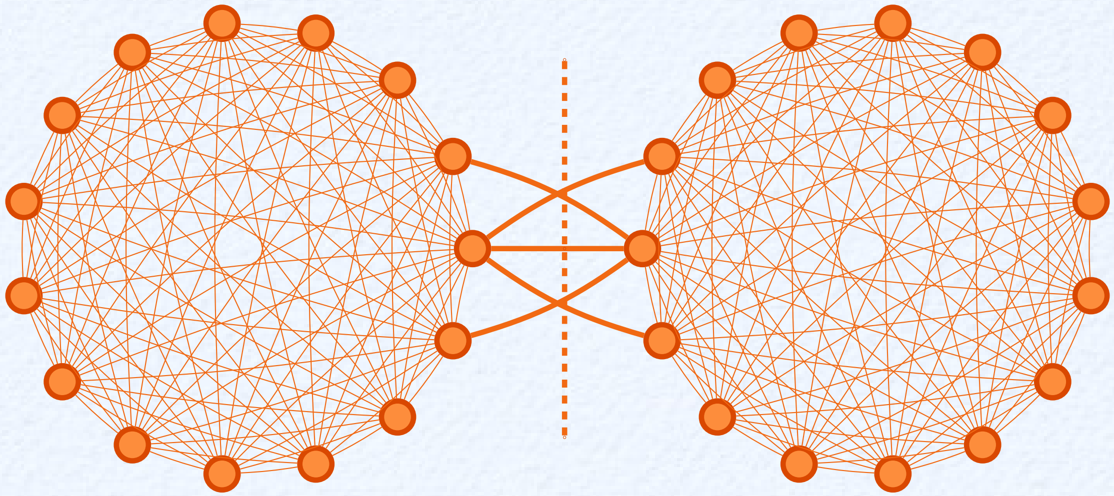


mincut $= 5$

- Information across the minimum cut set:

$$\text{mincut} \cdot \text{Hellinger}^{\min}$$

Main results: general graphs



mincut $= 5$

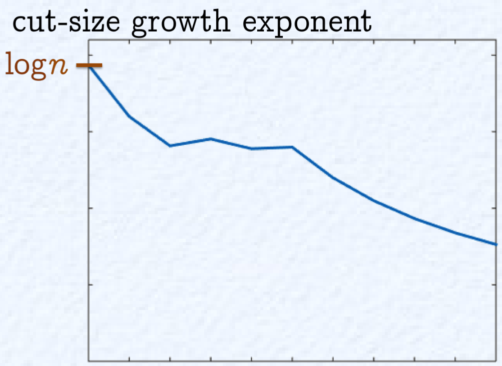
- **Recovery conditions**

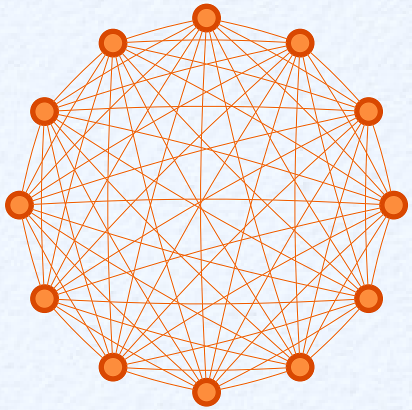
$$\text{ML works if} \qquad \text{mincut} \cdot \text{Hellinger}^{\min} \gtrsim \tau^{\text{cut}} + \log n + \log M$$

$$\text{Impossible if} \qquad \text{mincut} \cdot \text{Hellinger}^{\min} \lesssim \tau^{\text{cut}} + \frac{\text{mincut}}{\text{max-degree}} \log n$$

Cut-homogeneity exponent

- τ^{cut} captures[1]
  - growth rate of the cut-set distribution
  - the ratio $\frac{\text{mincut}}{\text{avg-degree}}$

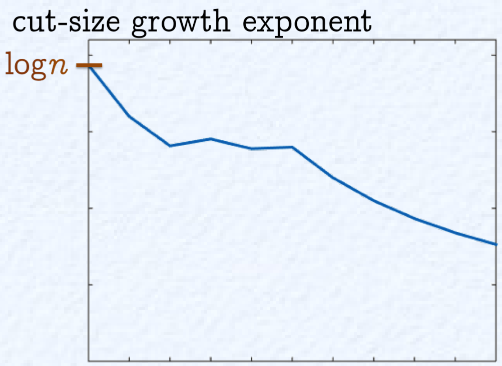


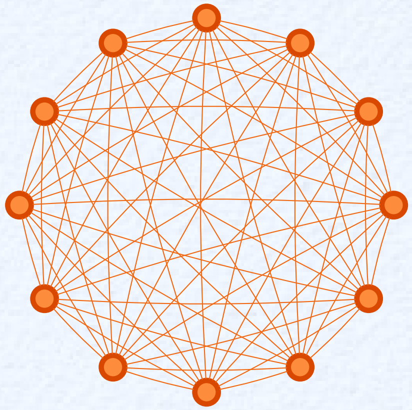


[1] $\tau^{\text{cut}} := \max_k \frac{1}{k} |\mathcal{N}(k \cdot \text{mincut})|$, where $\mathcal{N}(K) := |\{\text{cut} : \text{cut-size} \leq K\}|$

Cut-homogeneity exponent

- τ^{cut} captures[1]
  - growth rate of the cut-set distribution
  - the ratio $\frac{\text{mincut}}{\text{avg-degree}}$

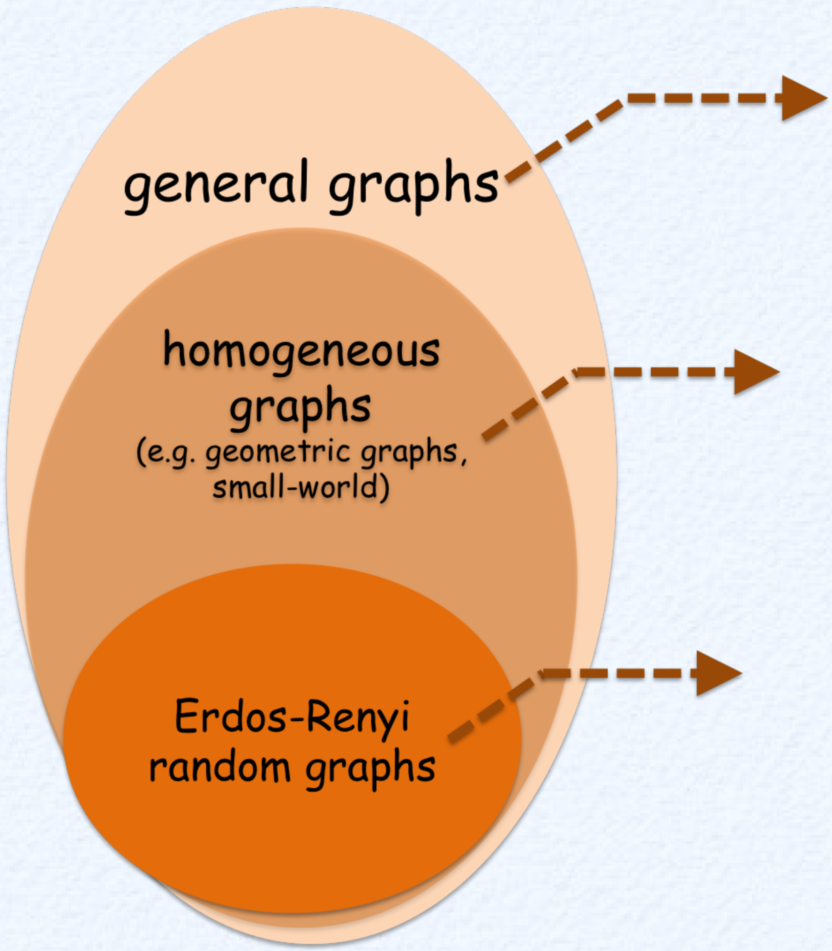




- In general: $\tau^{\text{cut}} \gtrsim 1$
- For homogeneous graphs: $\tau^{\text{cut}} \lesssim \log n$

[1] $\tau^{\text{cut}} := \max_k \frac{1}{k} |\mathcal{N}(k \cdot \text{mincut})|$, where $\mathcal{N}(K) := |\{\text{cut} : \text{cut-size} \leq K\}|$

Summary of main results



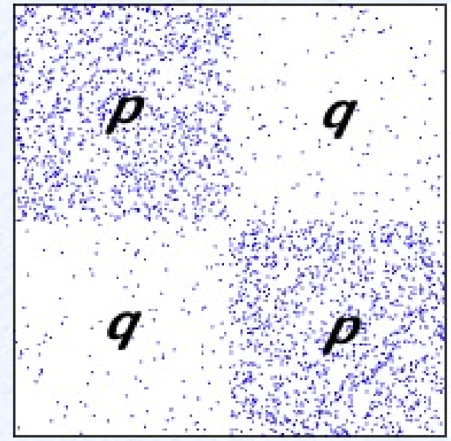
$$\text{mincut} \times \text{Hellinger}^{\min} \gtrsim (1 \sim \log n)$$
$$\text{gap} \lesssim \log n$$

$$\text{avg-deg} \times \text{Hellinger}^{\min} \gtrsim \log n$$
$$\text{gap} \asymp 1$$

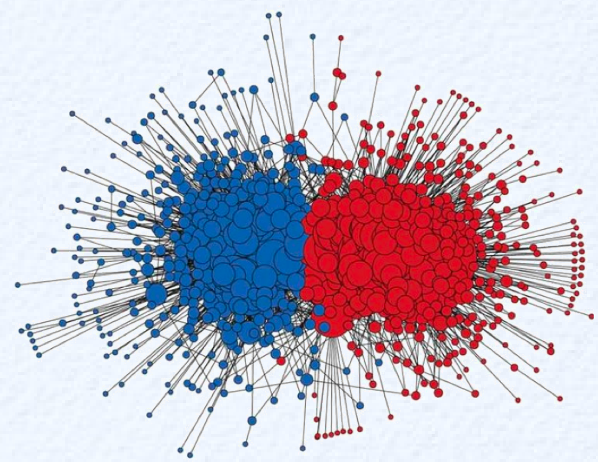
$$\text{avg-deg} \times \text{Hellinger}^{\min} \gtrsim \log n$$
$$\text{gap} \leq 4\left(1 + \frac{2 \log M}{\log n}\right)$$

Concrete application: stochastic block model

- **Stochastic block model**:
  - 2 clusters
  - edge densities:
    — within-cluster: $p = \frac{\alpha \log n}{n}$
    — across-cluster: $q = \frac{\beta \log n}{n}$ $(q < p)$

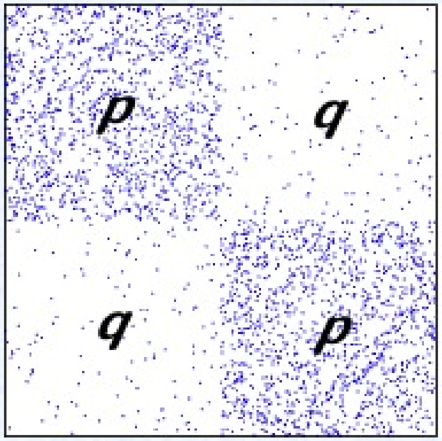


adjacency matrix

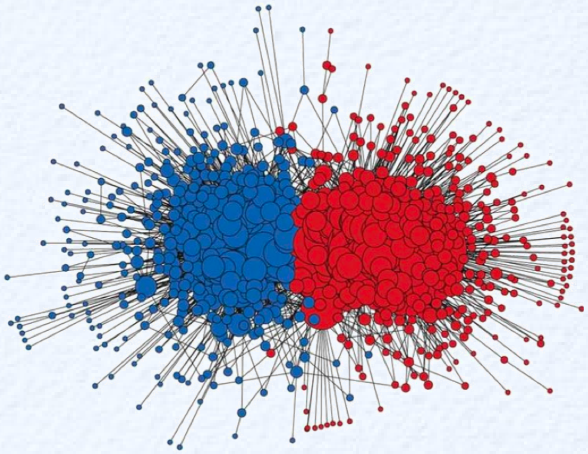


Concrete application: stochastic block model

- **Stochastic block model**:
  - 2 clusters
  - edge densities:
  - — within-cluster: $p = \frac{\alpha \log n}{n}$
  - — across-cluster: $q = \frac{\beta \log n}{n}$ $(q < p)$



adjacency matrix



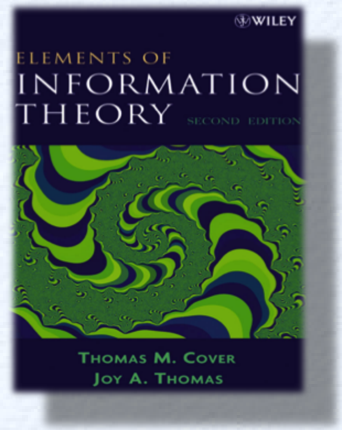
- Our theory:

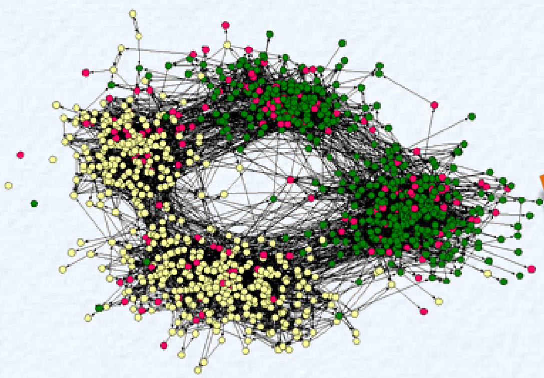
$$\text{feasible if} \quad \sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$$
$$\text{impossible if} \quad \sqrt{\alpha} - \sqrt{\beta} < 1/2$$

- Fundamental limit (Abbe et al. and Mossel et al.): $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$

Concluding remarks

- A unified framework to determine recovery limits
- Interplay between IT and graph theory
- Tighten the pre-constants?





Arxiv: http://arxiv.org/abs/1504.01369