Complexity Theoretic Lower Bounds for Sparse Principal Component Detection

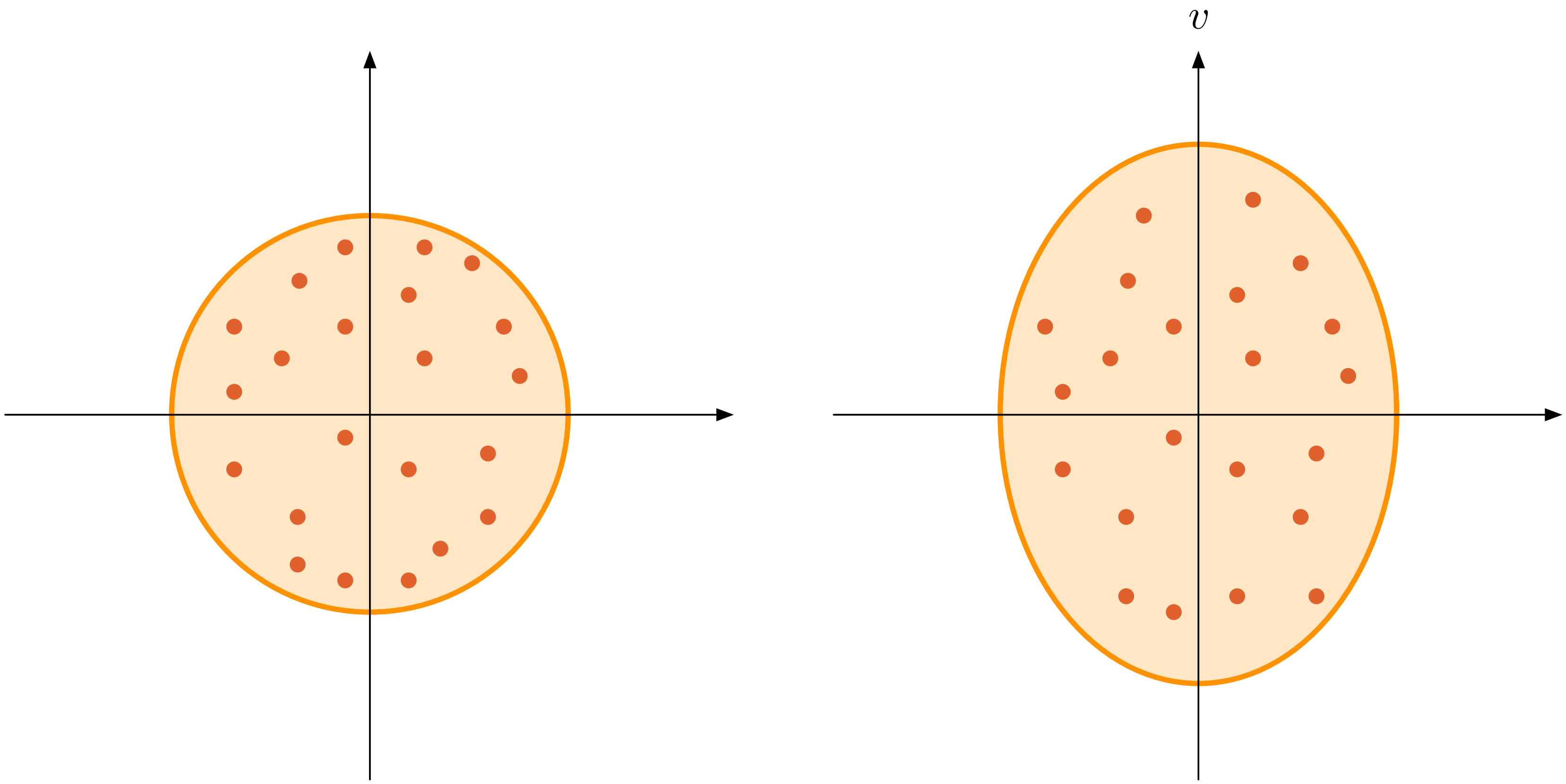
Conference On Learning Theory - 2013, Princeton

Quentin Berthet, Philippe Rigollet

Princeton University

Sparse principal component detection

$X_1, \ldots, X_n \in \mathbf{R}^d$ independent, centered Gaussian with unknown covariance.



Isotropy: $\mathcal{N}(0, I_d)$

Sparse PC: $\mathcal{N}(0, I_d + \theta v v^\top)$

Active topic: **Amini and Wainwright (09)**, **Vu and Lei (12)**, **Cai and Ma (13)**

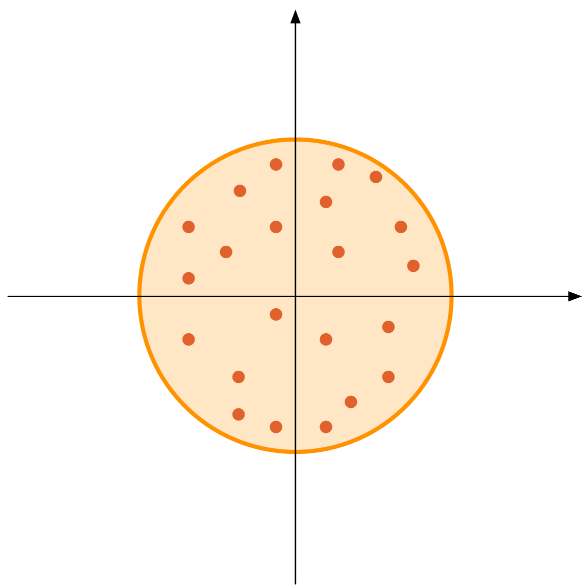
Detection problem: Is there a sparse principal component?

Sparse principal component detection

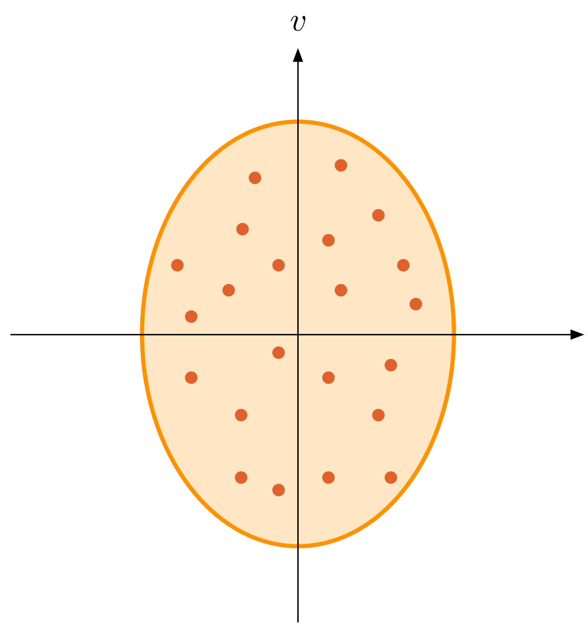
Testing problem between two hypotheses.

$$\begin{cases} H_0 : X \sim \mathcal{N}(0, I_d) \\ H_1 : X \sim \mathcal{N}(0, I_d + \theta vv^\top), \quad v \in \mathcal{B}_0(k) \end{cases}$$

v is a k-sparse unit vector. $\mathcal{B}_0(k) = \{v \in \mathbf{R}^p \, : \, |v|_2 = 1\, , \, |v|_0 \leq k\}$.



Isotropy: $\mathcal{N}(0, I_d)$



Sparse PC: $\mathcal{N}(0, I_d + \theta\, vv^\top)$

Optimal testing - Gaussian setting

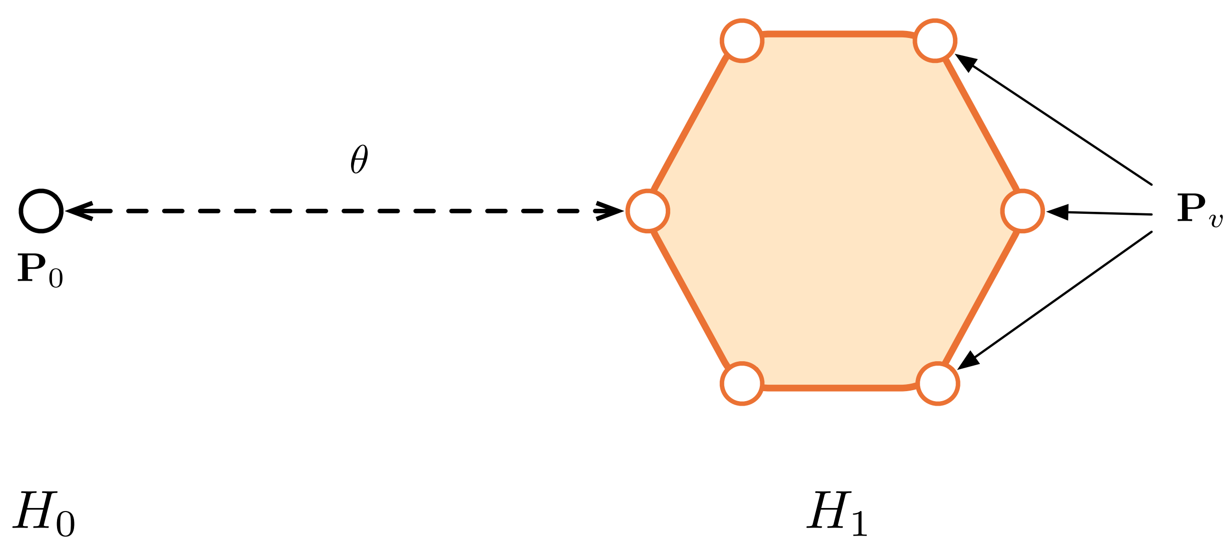
Minimax setting: θ^* is the **optimal rate of detection** when

- For $\theta > \bar{c}\,\theta^*$, testing is possible, with test ψ

$$\mathbf{P}_0^{\otimes n}(\psi = 1) \vee \max_{v \in \mathcal{B}_0(k)} \mathbf{P}_v^{\otimes n}(\psi = 0) \leq \delta\,.$$

- For $\theta < \underline{c}\,\theta^*$, testing is impossible, for all tests ϕ

$$\mathbf{P}_0^{\otimes n}(\phi = 1) \vee \max_{v \in \mathcal{B}_0(k)} \mathbf{P}_v^{\otimes n}(\phi = 0) \geq \delta\,.$$



Optimal testing - Robust setting

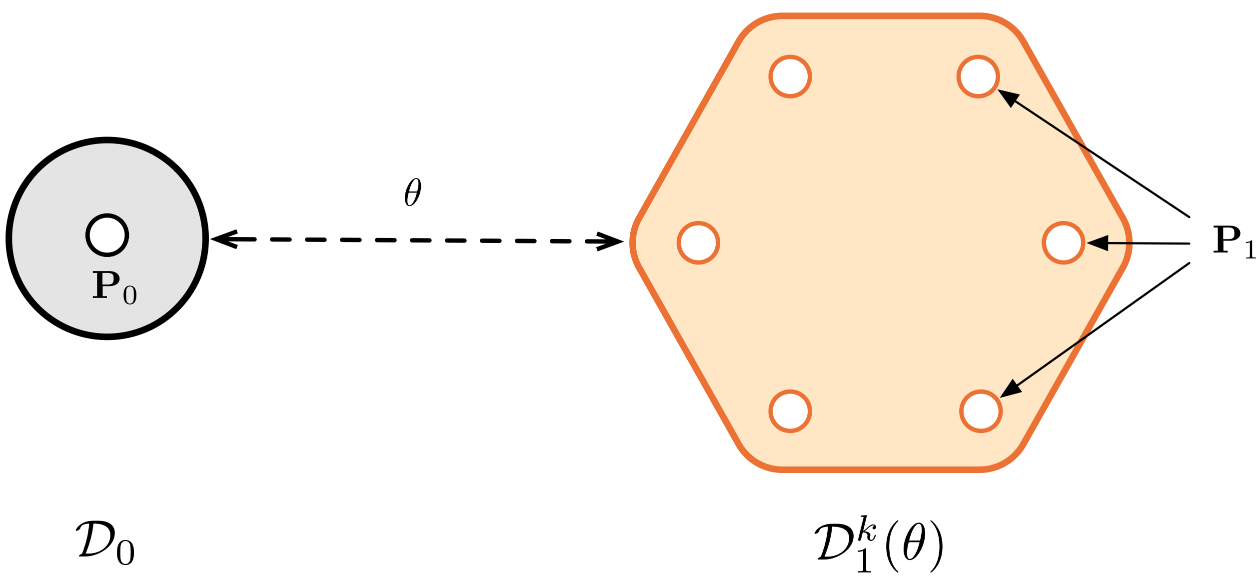
Minimax setting: θ^* is the **optimal rate of detection** when

- For $\theta > \bar{c}\,\theta^*$, testing is possible, with test ψ

$$\sup_{\mathbf{P}_0,\mathbf{P}_1} \left\{\mathbf{P}_0^{\otimes n}(\psi = 1) \vee \mathbf{P}_1^{\otimes n}(\psi = 0)\right\} \leq \delta\,.$$

- For $\theta < \underline{c}\,\theta^*$, testing is impossible, for **all** tests ϕ

$$\sup_{\mathbf{P}_0,\mathbf{P}_1} \left\{\mathbf{P}_0^{\otimes n}(\phi = 1) \vee \mathbf{P}_1^{\otimes n}(\phi = 0)\right\} \geq \delta\,.$$



Optimal testing - Robust setting

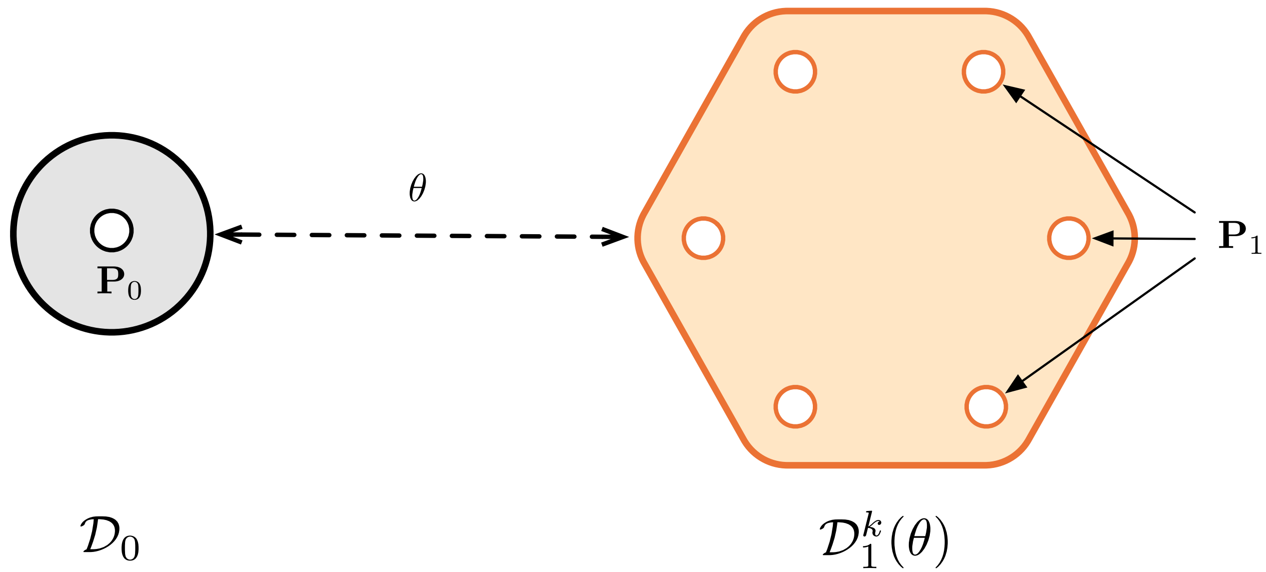
Minimax setting: θ^* is the **optimal rate of detection** over the class $\mathcal{T}$ when

- For $\theta > \bar{c}\,\theta^*$, testing is possible, with test $\psi \in \mathcal{T}$

$$\sup_{\mathbf{P}_0, \mathbf{P}_1} \left\{ \mathbf{P}_0^{\otimes n}(\psi = 1) \vee \mathbf{P}_1^{\otimes n}(\psi = 0) \right\} \leq \delta \,.$$

- For $\theta < \underline{c}_\phi\,\theta^*$, testing is impossible, for **all** tests $\phi \in \mathcal{T}$

$$\sup_{\mathbf{P}_0, \mathbf{P}_1} \left\{ \mathbf{P}_0^{\otimes n}(\phi = 1) \vee \mathbf{P}_1^{\otimes n}(\phi = 0) \right\} \geq \delta \,.$$



Minimax testing

Classical theory, no restriction on testing function.

$$\theta^* = \sqrt{\frac{k \log(d)}{n}}$$

- For $\theta > \bar{c}\,\theta^*$, for an explicit test ψ

$$\sup_{\mathbf{P}_0,\mathbf{P}_1} \left\{ \mathbf{P}_0^{\otimes n}(\psi = 1) \vee \mathbf{P}_1^{\otimes n}(\psi = 0) \right\} \le \delta\,.$$

- For $\theta < \underline{c}\,\theta^*$, testing is impossible, for **all** tests ϕ

$$\sup_{\mathbf{P}_0,\mathbf{P}_1} \left\{ \mathbf{P}_0^{\otimes n}(\phi = 1) \vee \mathbf{P}_1^{\otimes n}(\phi = 0) \right\} \ge \delta\,.$$

Test ψ based on a sparse eigenvalue statistic $\lambda^k_{\max}(\hat{\Sigma})$, $\mathtt{NP}$-hard to compute.

Polynomial-time testing

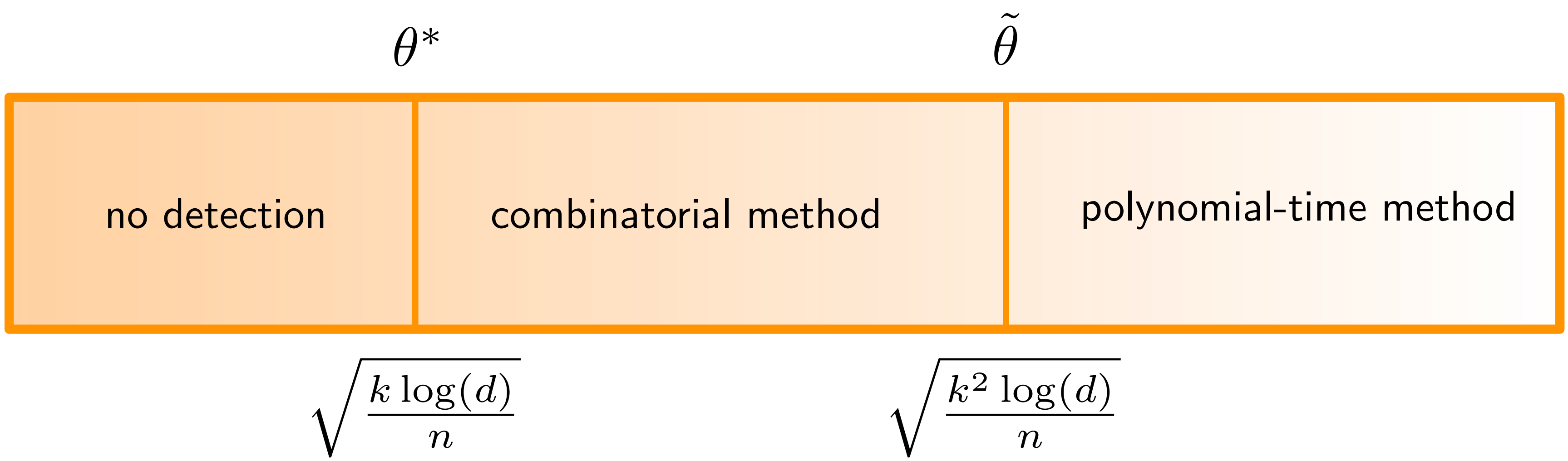
We only consider tests in $\mathcal{T} = \mathrm{P}$, running in polynomial time.

$$\tilde{\theta} = \sqrt{\frac{k^2 \log(d)}{n}}$$

- For $\theta > \bar{c}\,\tilde{\theta}$, for several explicit tests ψ

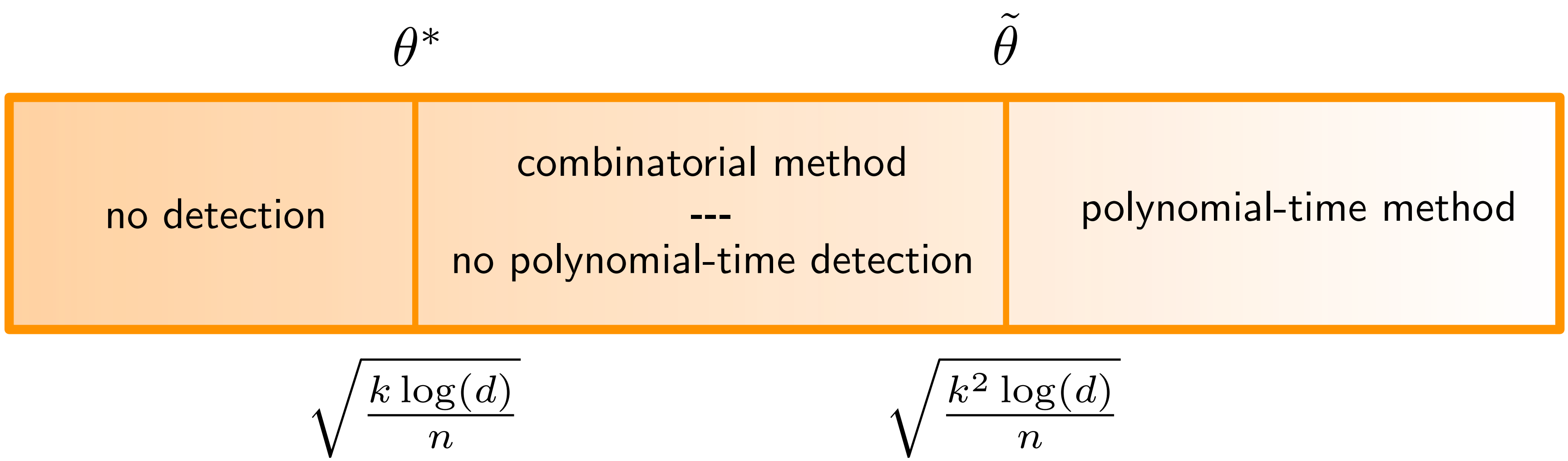
$$\sup_{\mathbf{P}_0, \mathbf{P}_1} \left\{ \mathbf{P}_0^{\otimes n}(\psi = 1) \vee \mathbf{P}_1^{\otimes n}(\psi = 0) \right\} \leq \delta \,.$$

Tests: Diagonal method - **Johnstone (01)**, SDP - **d'Aspremont et al. (07)**, MDP - **B. and R. (12)**, other heuristics.



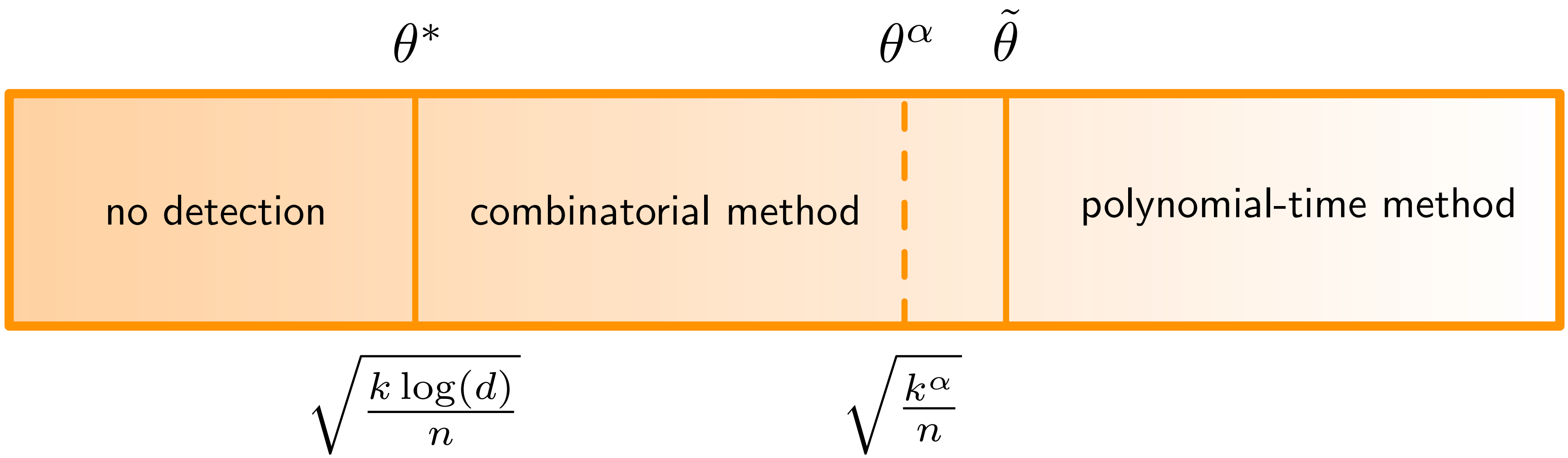
Detection levels

Situation suggested by those results



Detection rates

So far, only upper bounds, suggestions



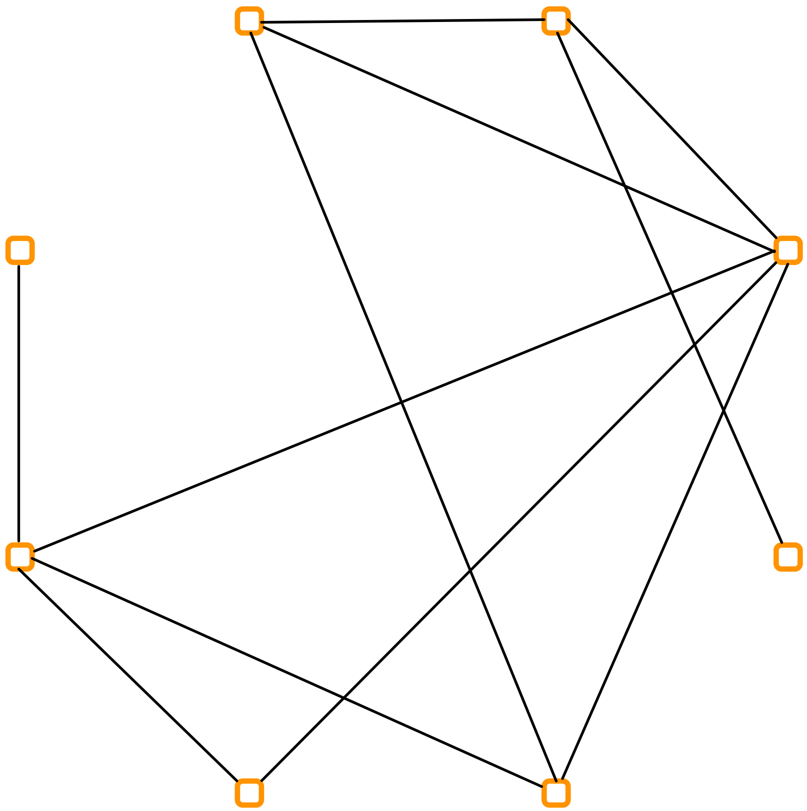
Situation could be very different.

Need for **Complexity Theoretic Lower Bounds**

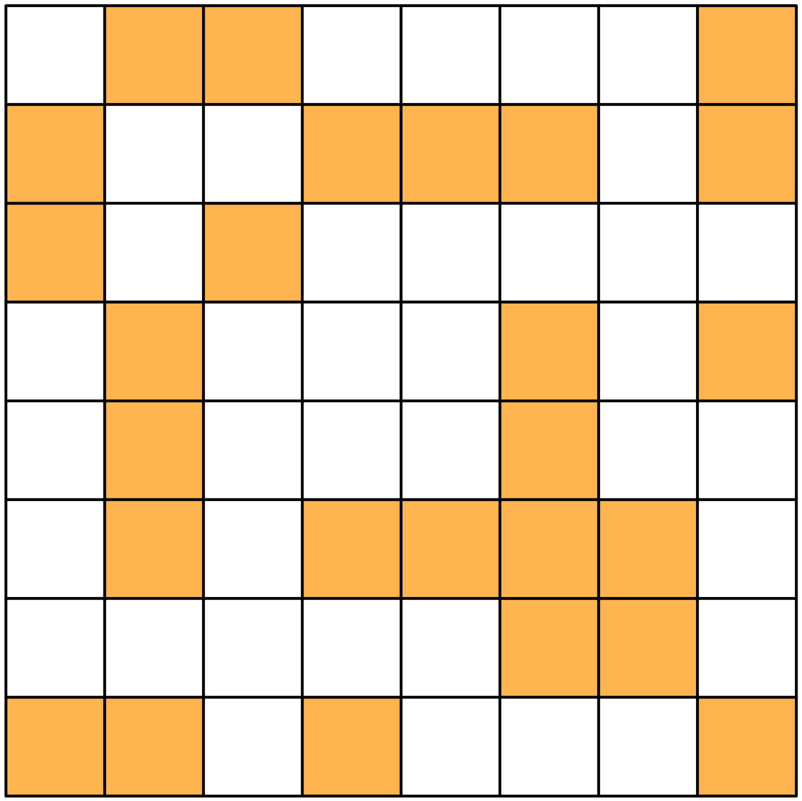
Planted clique problem

Erdős-Rényi graphs

$\mathcal{G}(m, 1/2)$: Each edge is randomly connected, with probability $1/2$, independently.



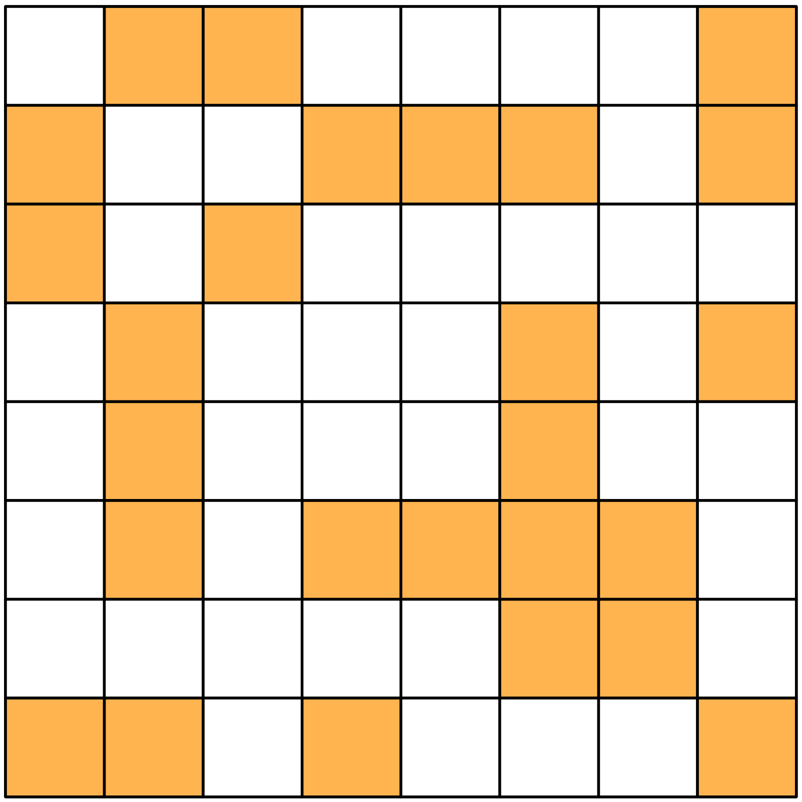
Graph



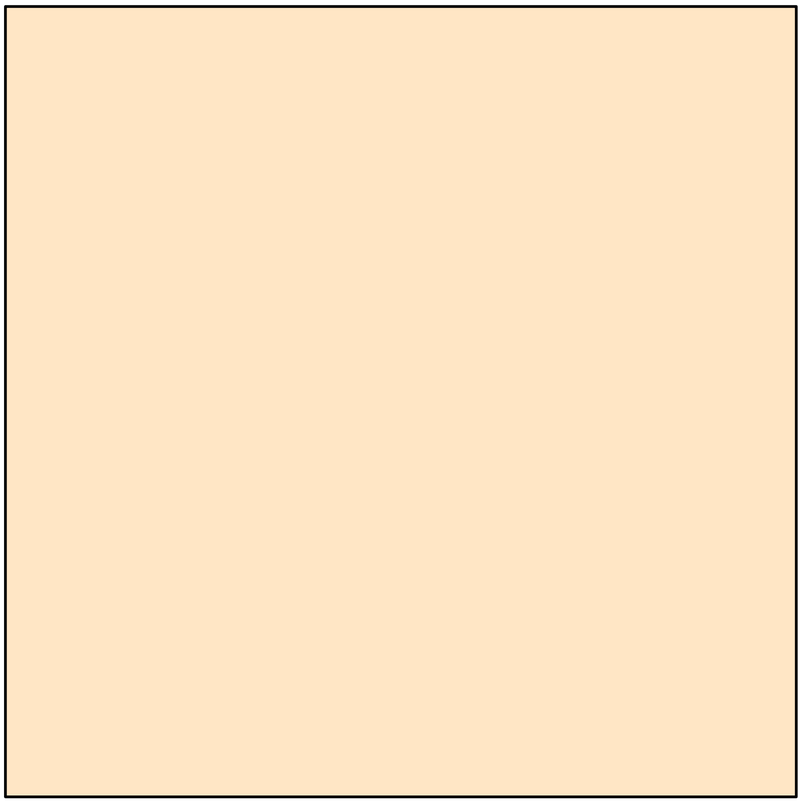
Adjacency matrix

Erdős-Rényi graphs

The expectation of the adjacency matrix is constant: pure noise setting.



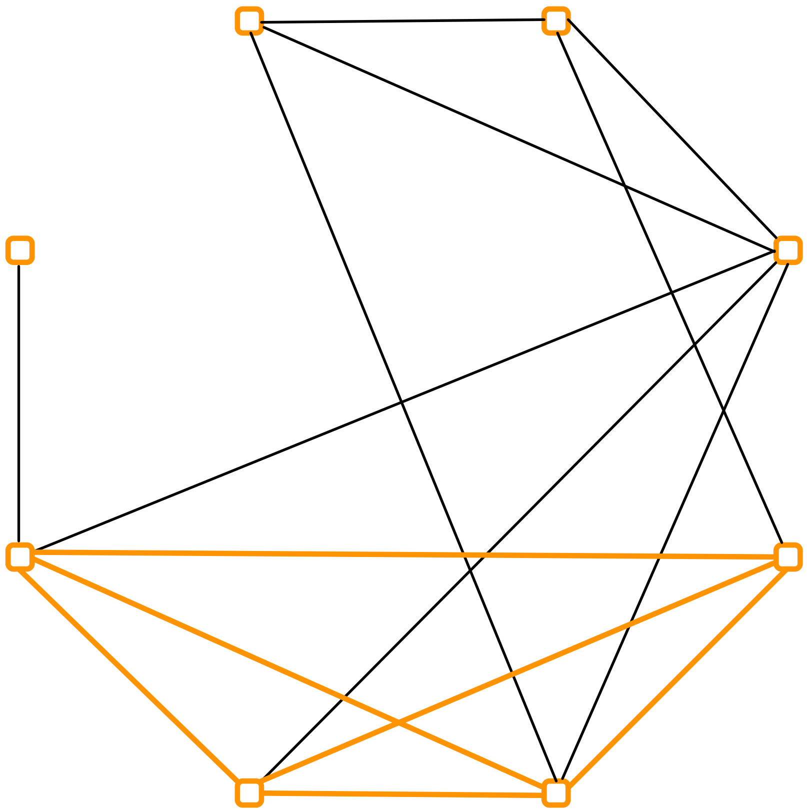
Random instance



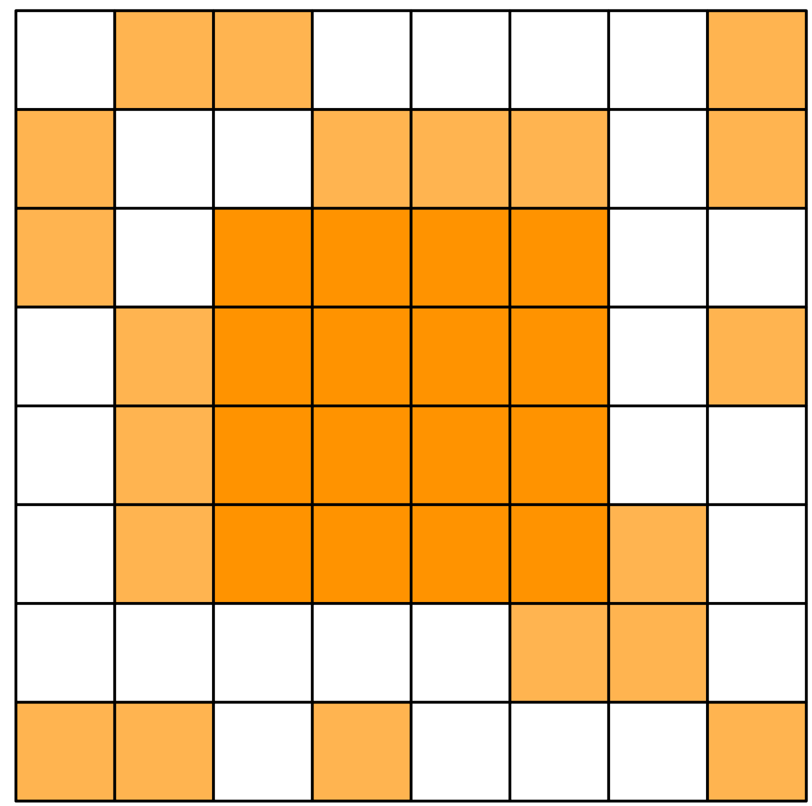
Expectation

Erdős-Rényi graphs

$\mathcal{G}(m, 1/2, \kappa)$: A clique of size κ is planted in a graph from $\mathcal{G}(m, 1/2)$.



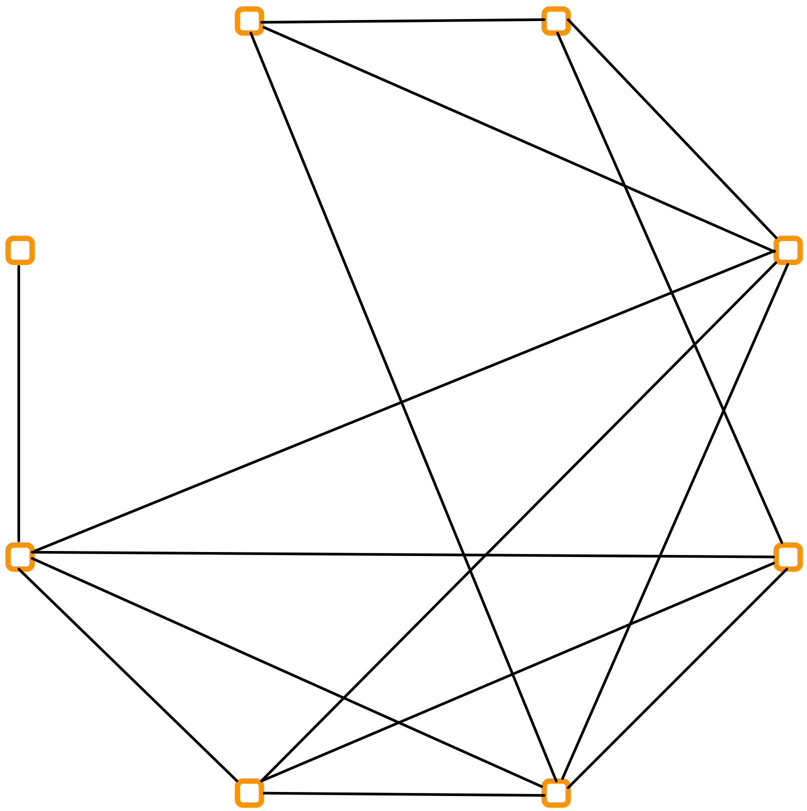
Graph



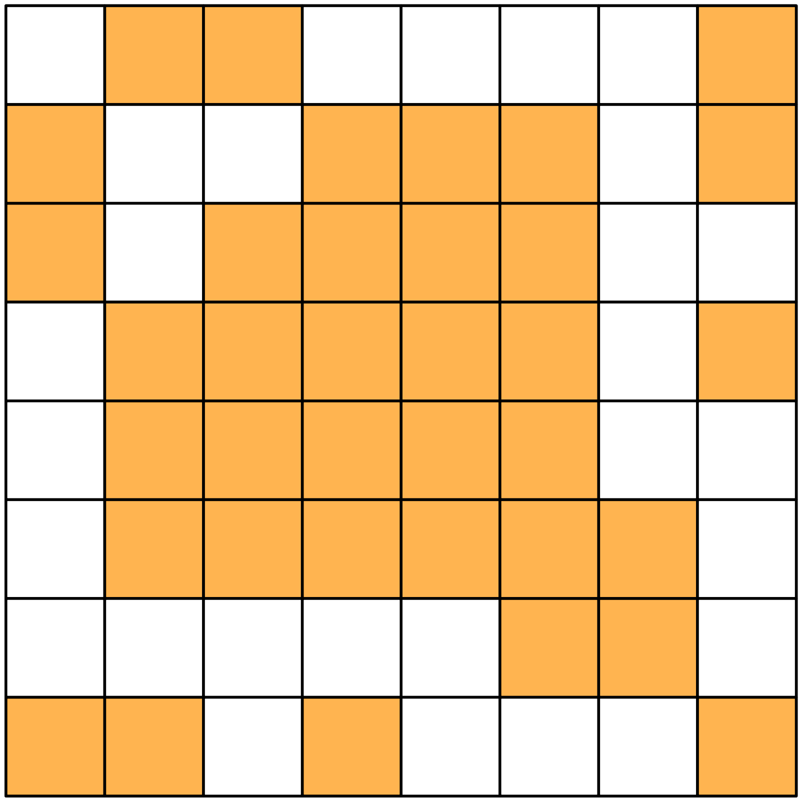
Adjacency matrix

Erdős-Rényi graphs

$\mathcal{G}(m, 1/2, \kappa)$: A clique of size κ is planted in a graph from $\mathcal{G}(m, 1/2)$.



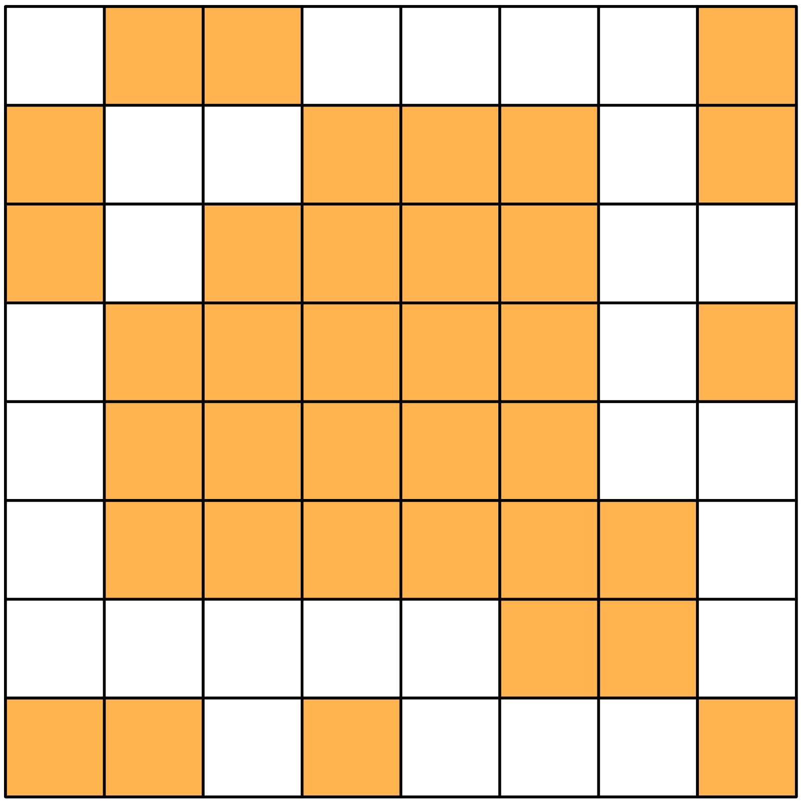
Graph



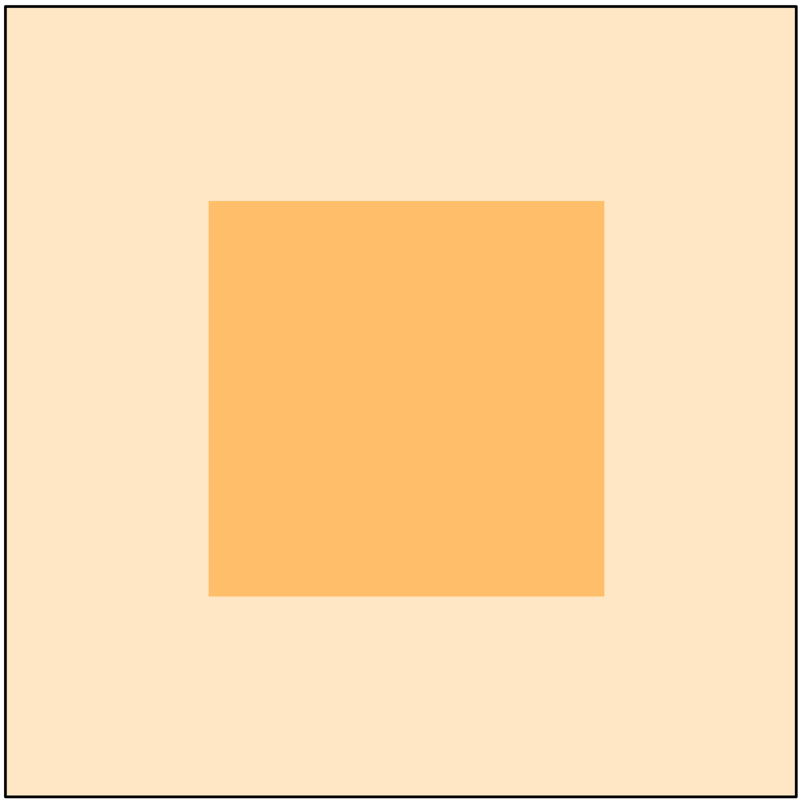
Adjacency matrix

Erdős-Rényi graphs

The expectation of the adjacency matrix has a sparse signal structure



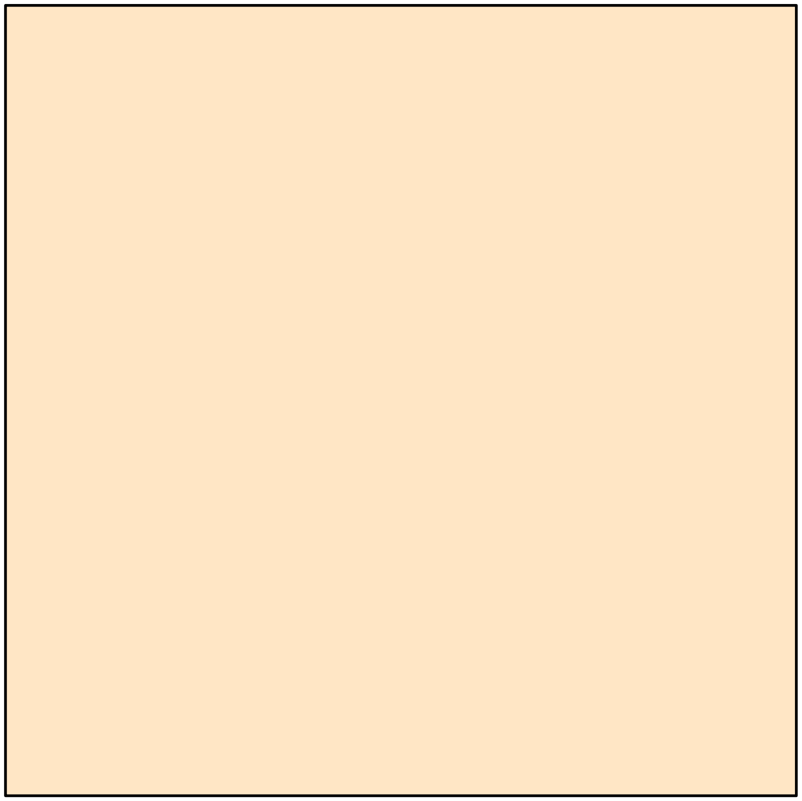
Random instance



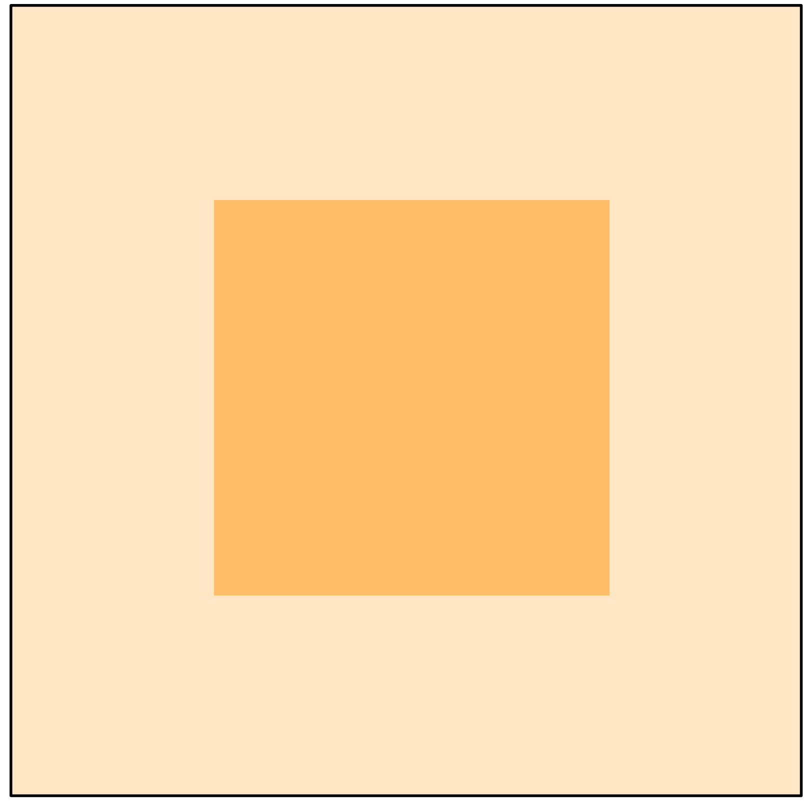
Expectation

Planted clique problem

Detection of a structured signal of sparsity κ in a random graph of size m.



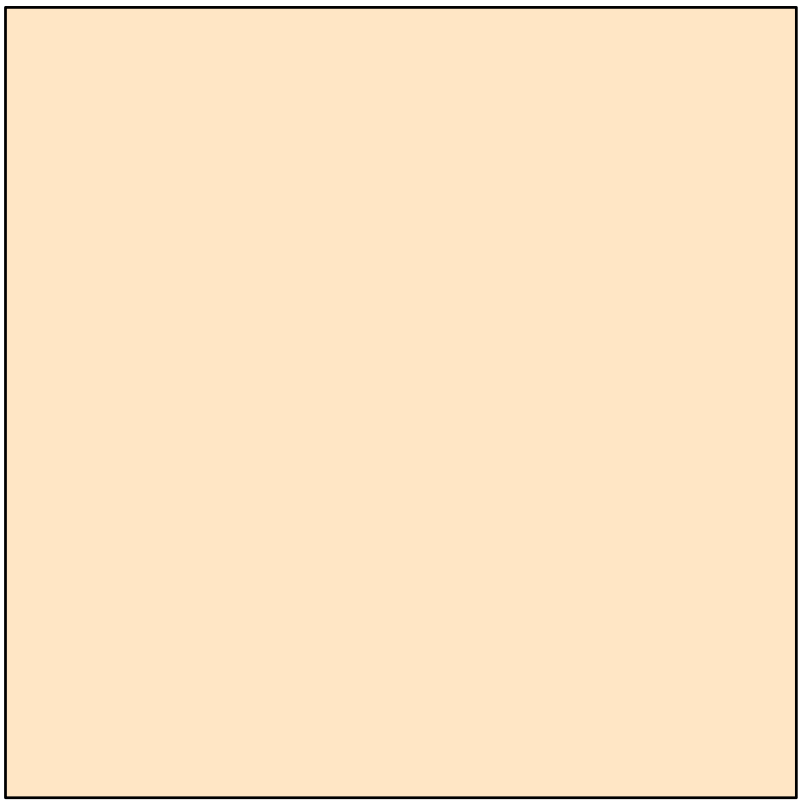
$$\mathcal{G}(m, 1/2) = \mathbf{P}_0^{(G)}$$



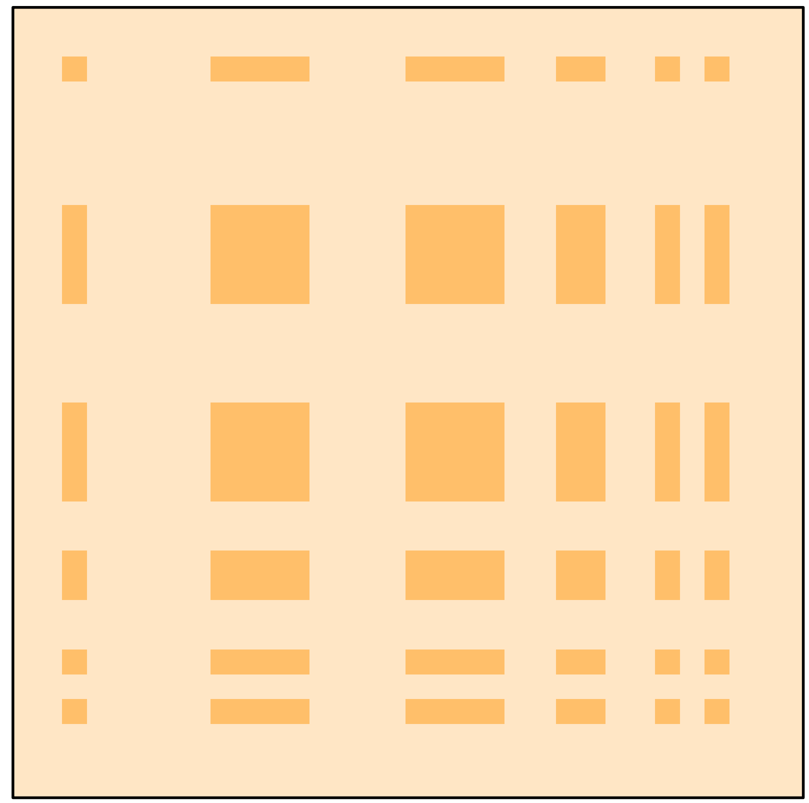
$$\mathcal{G}(m, 1/2, \kappa) = \mathbf{P}_1^{(G)}$$

Planted clique problem

Detection of a structured signal of sparsity κ in a random graph of size m.



$$\mathcal{G}(m, 1/2) = \mathbf{P}_0^{(G)}$$



$$\mathcal{G}(m, 1/2, \kappa) = \mathbf{P}_1^{(G)}$$

Planted clique problem

Distinguishing those two distributions is called the Planted Clique problem

$$\begin{cases} H_0^{\mathsf{PC}} : G \sim \mathcal{G}(m, 1/2) = \mathbf{P}_0^{(G)} \\ H_1^{\mathsf{PC}} : G \sim \mathcal{G}(m, 1/2, \kappa) = \mathbf{P}_1^{(G)} . \end{cases}$$

- Detection for $\kappa > 2\log_2(m)$, $\mathbb{NP}$-hard method (Max-clique) **Pencer (94)**.
- Polynomial time detection for $\kappa = O(\sqrt{m})$ **Alon et al. (98)**.
- Strong reasons to believe impossible detection in polynomial time for $\kappa = O(m^c)$, $c < 1/2$.

  Ames and Vavasis (11), Dekel et al. (10); Feige and Krauthgamer (00); Feige and Ron (10)

  Jerrum (92), Feige and Krauthgamer (92), Rossman (10), Feldman et al. (12)

  Juels and Peinado (00), Alon et al. (07), Hazan and Krauthgamer (11), Alon et al. (2011)

Hypothesis A_{PC}

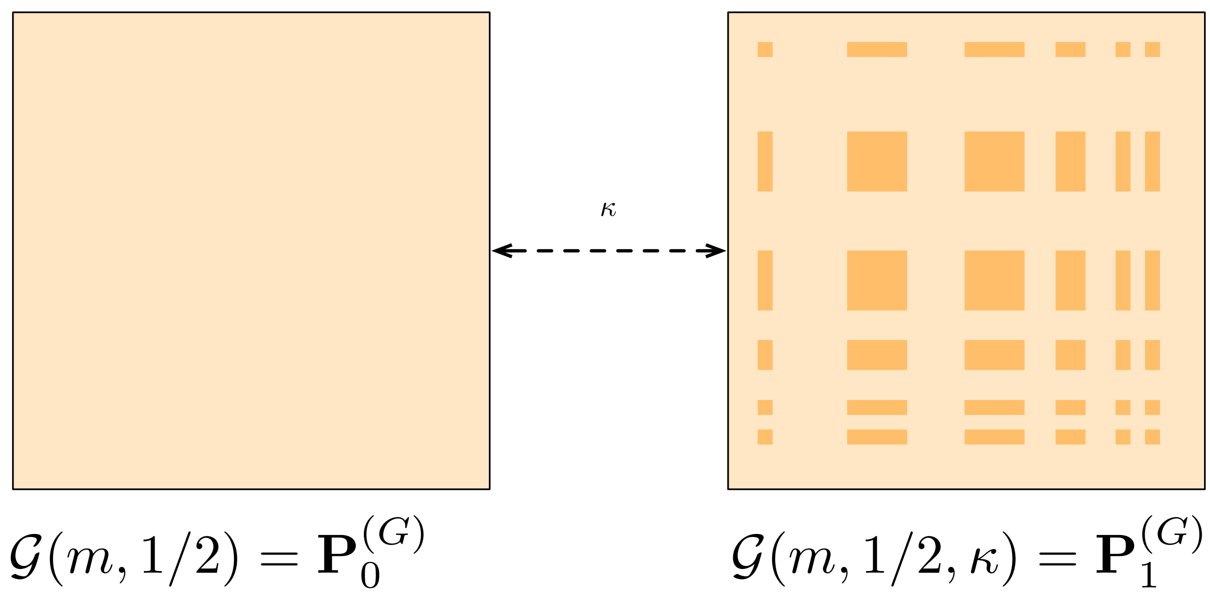
For any $a, b \in (0,1)$, all randomized $\mathcal{P}$-time tests ξ, there exists $\Gamma > 0$ such that

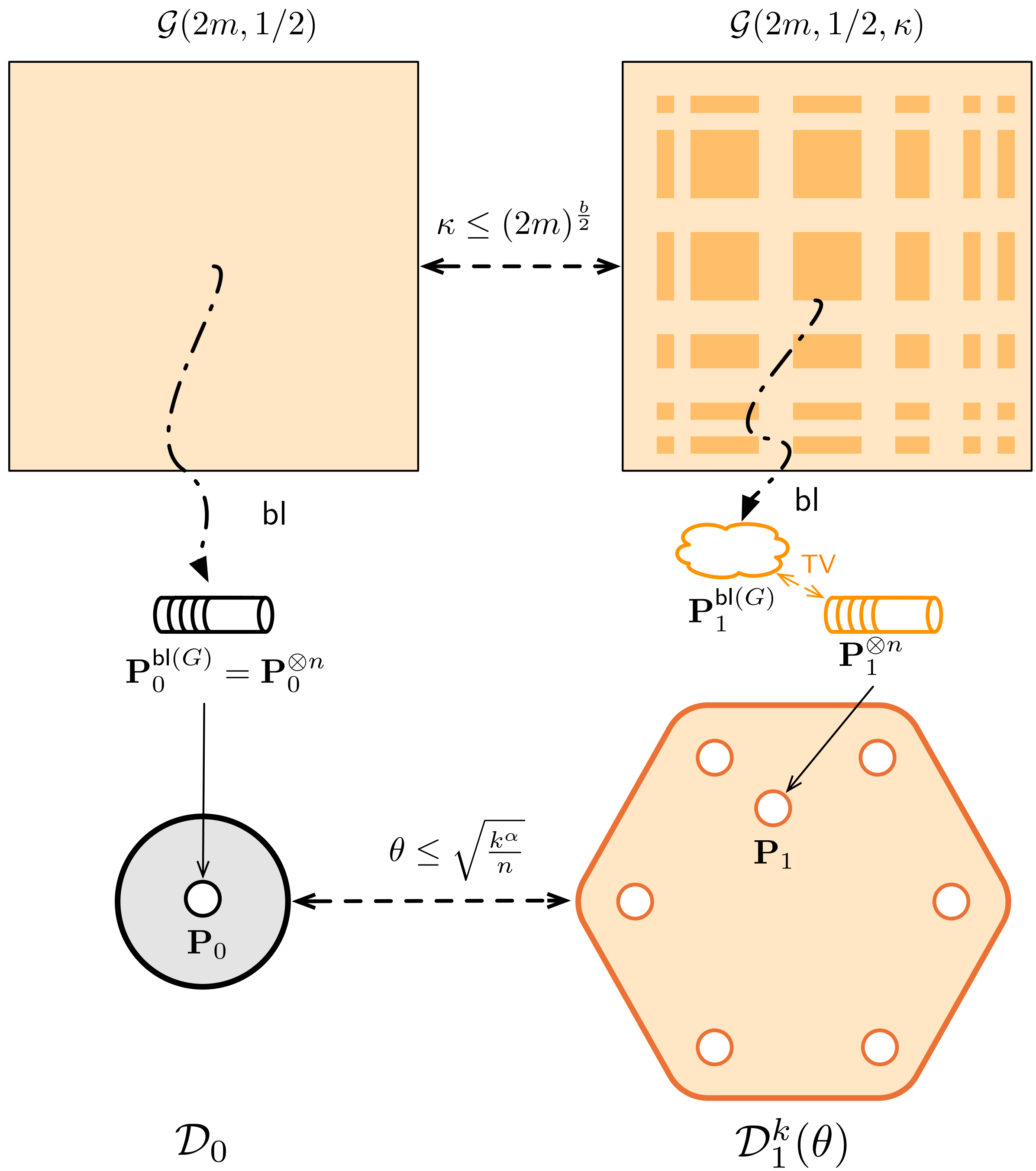
$$\mathbf{P}_0^{(G)}(\xi_{m,\kappa}(G) = 1) \vee \mathbf{P}_1^{(G)}(\xi_{m,\kappa}(G) = 0) \geq 1.2\delta\,, \quad \forall\, m^{\frac{a}{2}} < \Gamma\kappa < m^{\frac{b}{2}}\,.$$

Remarks

- Formalization of computational hardness of the statistical problem.
- Constant Γ can depend on ξ, a, b: Asymptotic nature of the class $\mathcal{P}$.

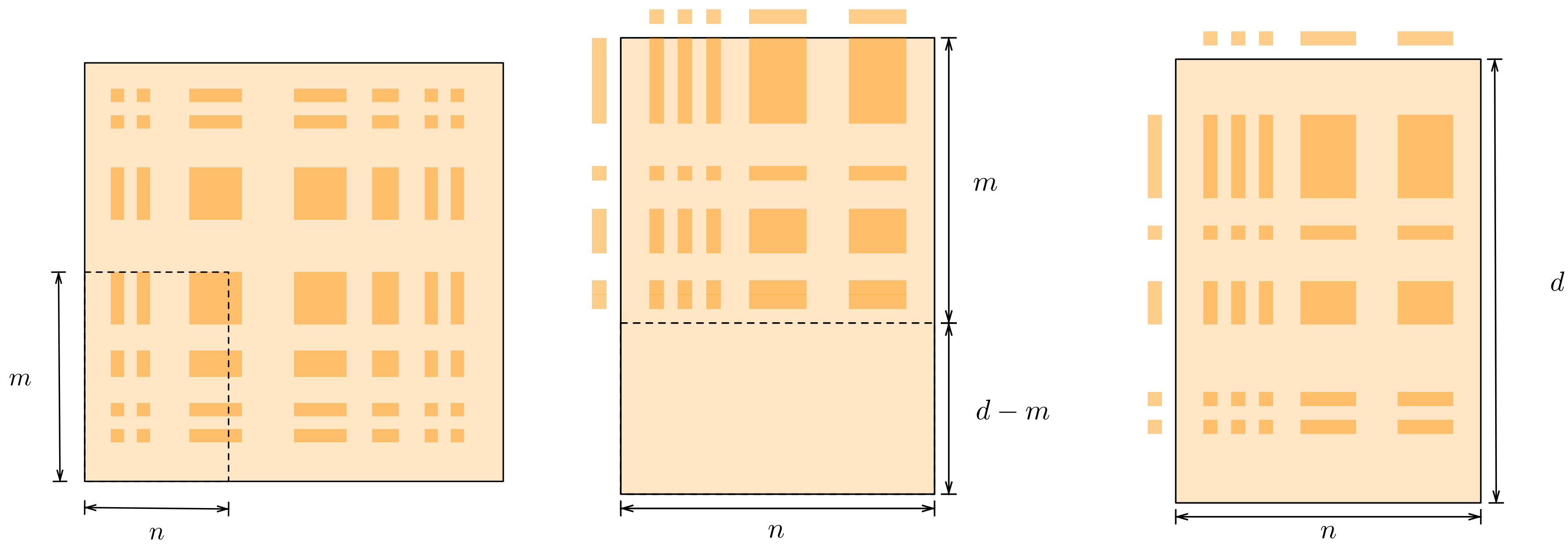
Create a randomized $\mathcal{P}$-time function from graphs to random vectors.




$\mathcal{G}(2m, 1/2)$
$\mathcal{G}(2m, 1/2, \kappa)$
$\kappa \leq (2m)^{\frac{b}{2}}$
bl
bl
TV
$\mathbf{P}_1^{\mathsf{bl}(G)}$
$\mathbf{P}_0^{\mathsf{bl}(G)} = \mathbf{P}_0^{\otimes n}$
$\mathbf{P}_1^{\otimes n}$
$\theta \leq \sqrt{\frac{k^\alpha}{n}}$
$\mathbf{P}_0$
$\mathbf{P}_1$
$\mathcal{D}_0$
$\mathcal{D}_1^k(\theta)$

Reduction description

Function based on the bottom-left corner of the adjacency matrix.



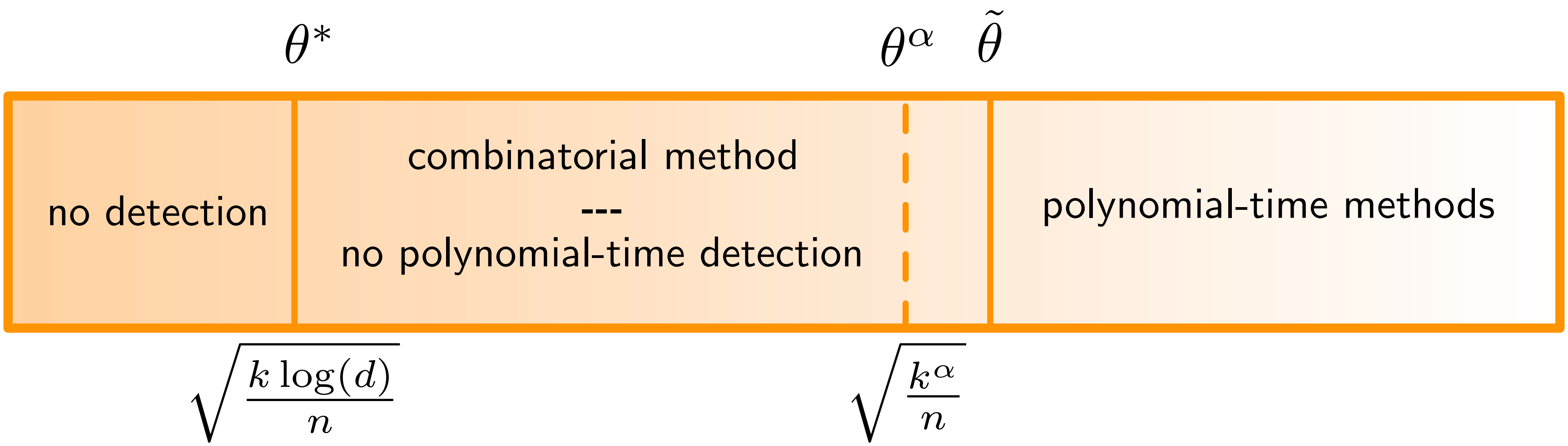
bottom-left $\qquad$ n vectors of $\mathbf{R}^d$ $\qquad$ shuffle and $\times\{-1, 1\}$

- $\mathcal{G}(2m, 1/2) \to n$ vectors i.i.d. with independent $\{-1, 1\}$ coefficients $\in \mathcal{D}_0$.
- $\mathcal{G}(2m, 1/2, \kappa_b) \to n$ vectors with distribution close (in TV) to i.i.d. $\in \mathcal{D}_1^k(\theta^\alpha)$.

Key observation: Sampling without replacement close to with replacement

Rates

- Detection at the rate θ^α would contradict Hypothethis $\mathrm{A_{PC}}$.
- SDP, MDP optimal among $\mathtt{P}$-time methods.



Take-home message: gap of $\sqrt{k}$ for methods in P-time

Conclusion

In this work

- Theoretical formulation of computational lower bounds.
- Link between Sparse PCA and Planted Clique problem.
- Optimal rates in $\mathbf{P}$-time for SDP, MDP.

Future work

- Computational Lower bounds for other problems
- Strengthening the complexity assumption, random $\mathtt{3-SAT}$, $\mathbf{NP}$-hardness...

Reference

Computational Lower Bounds for Sparse PCA.
arxiv.org/abs/1304.0828

supported by CAREER award DMS-1053987.