Nonlinear Laplacians

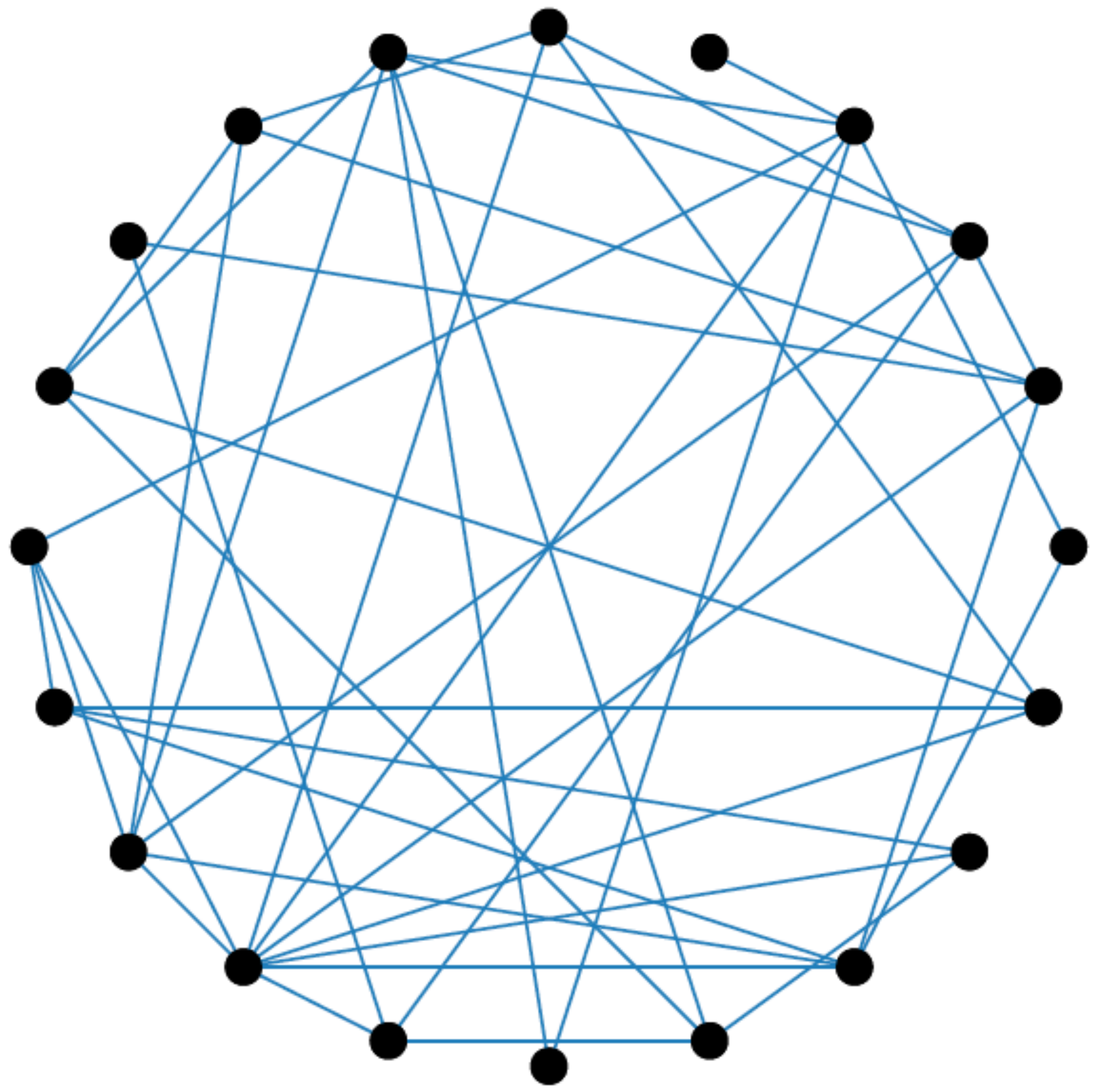
Tunable principal component analysis under directional prior information

Yuxin Ma, Dmitriy (Tim) Kunisky
Department of Applied Mathematics and Statistics, Johns Hopkins University
{yma93, kunisky}@jhu.edu

Graph Planted Clique Problem

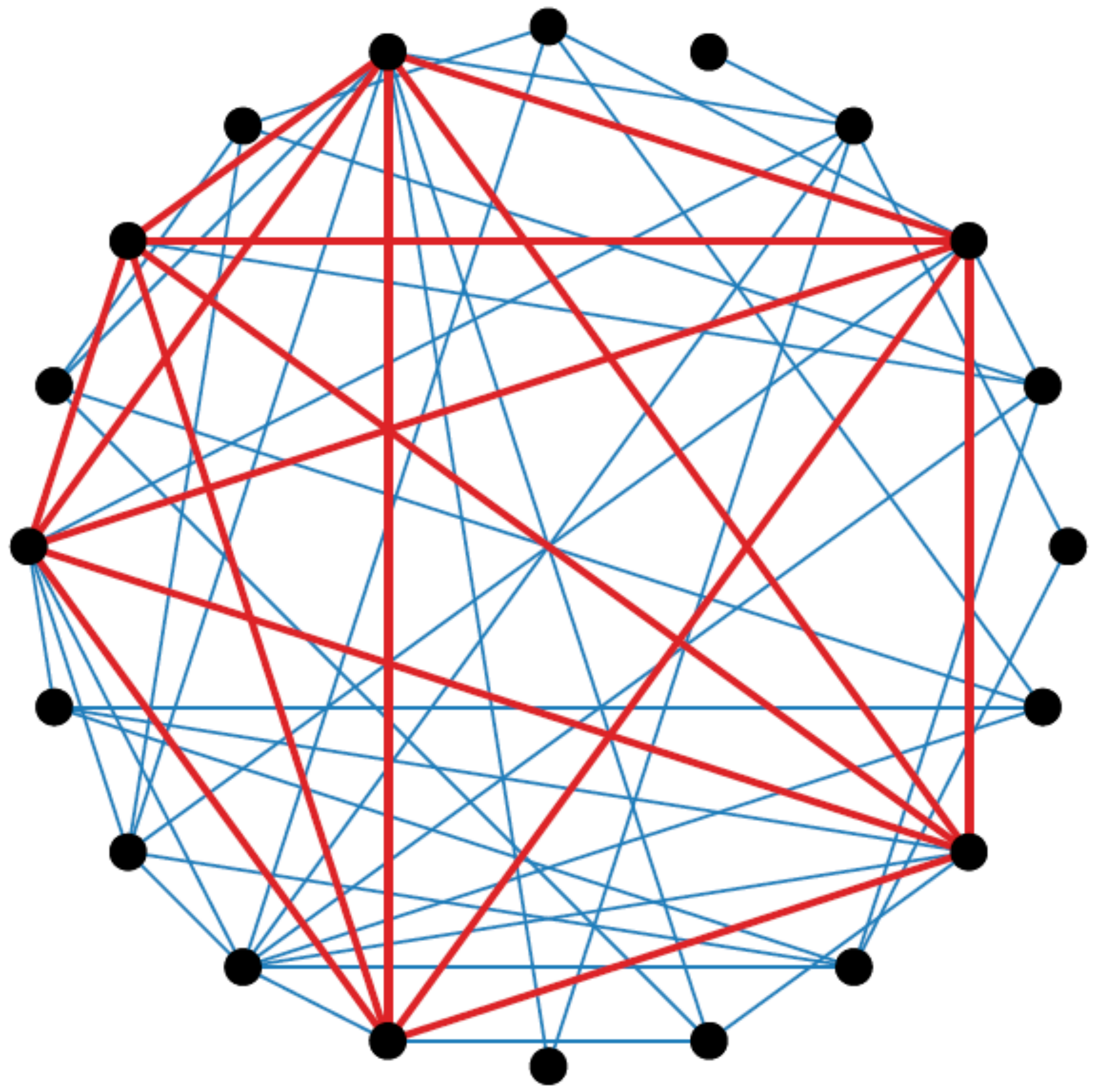
Graph Planted Clique Problem

- Draw G an Erdős–Rényi graph with n nodes. (Each edge independently with probability $1/2$).



Graph Planted Clique Problem

- Draw G an Erdős–Rényi graph with n nodes. (Each edge independently with probability $1/2$).
- Choose a subset of vertices S and add a clique (complete subgraph) on S.

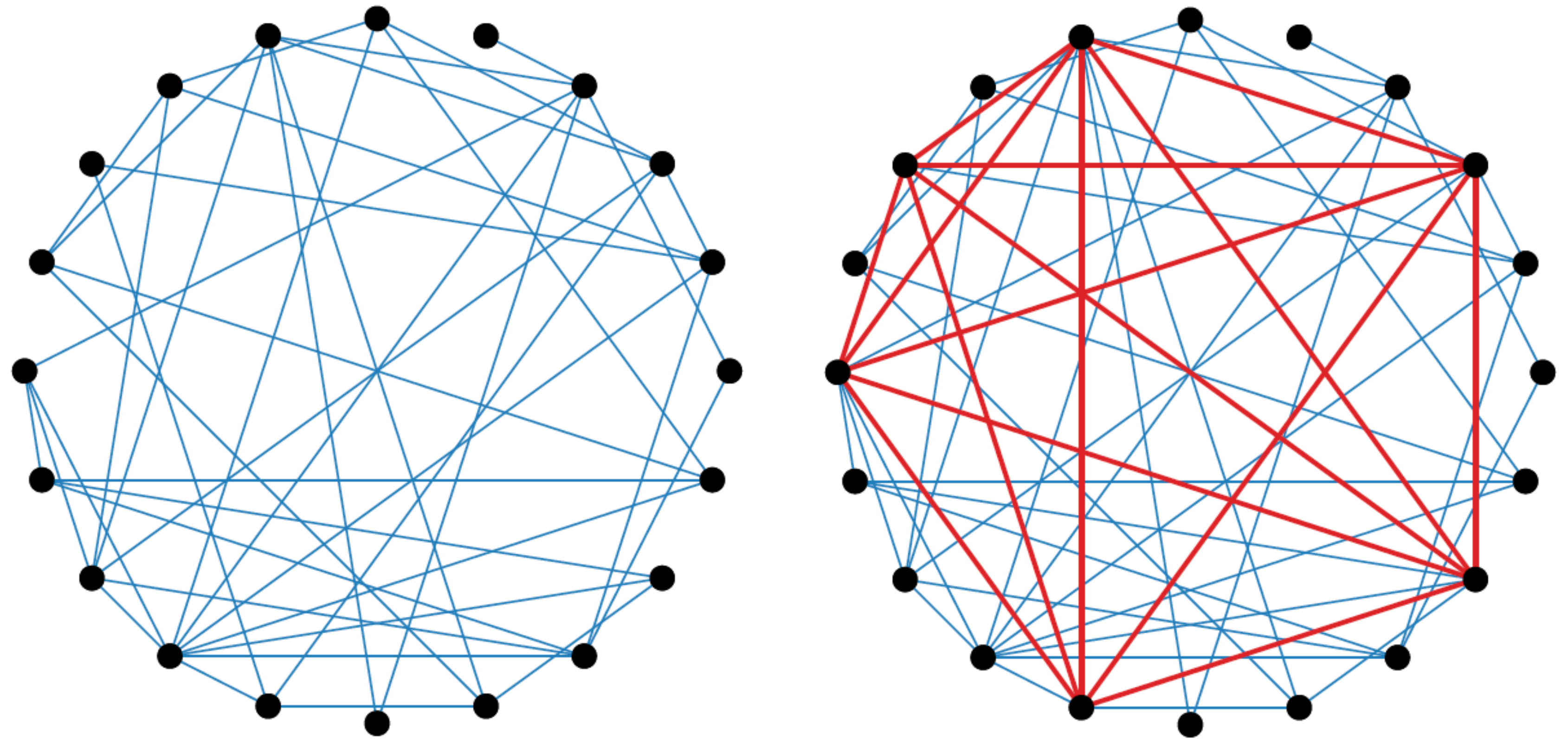


Graph Planted Clique Problem

- Draw G an Erdős–Rényi graph with n nodes. (Each edge independently with probability $1/2$).
- Choose a subset of vertices S and add a clique (complete subgraph) on S.

Want to either

- **Detection:** decide whether $S = \varnothing$. (Whether a clique is planted).
- **Estimation:** produce $\hat{S} \approx S$. (Recover the clique).



Naive spectral algorithm

Take $\mathbf{Y}$ the $\{\pm 1\}$-valued adjacency matrix of graph G, and perform PCA on it.

Naive spectral algorithm

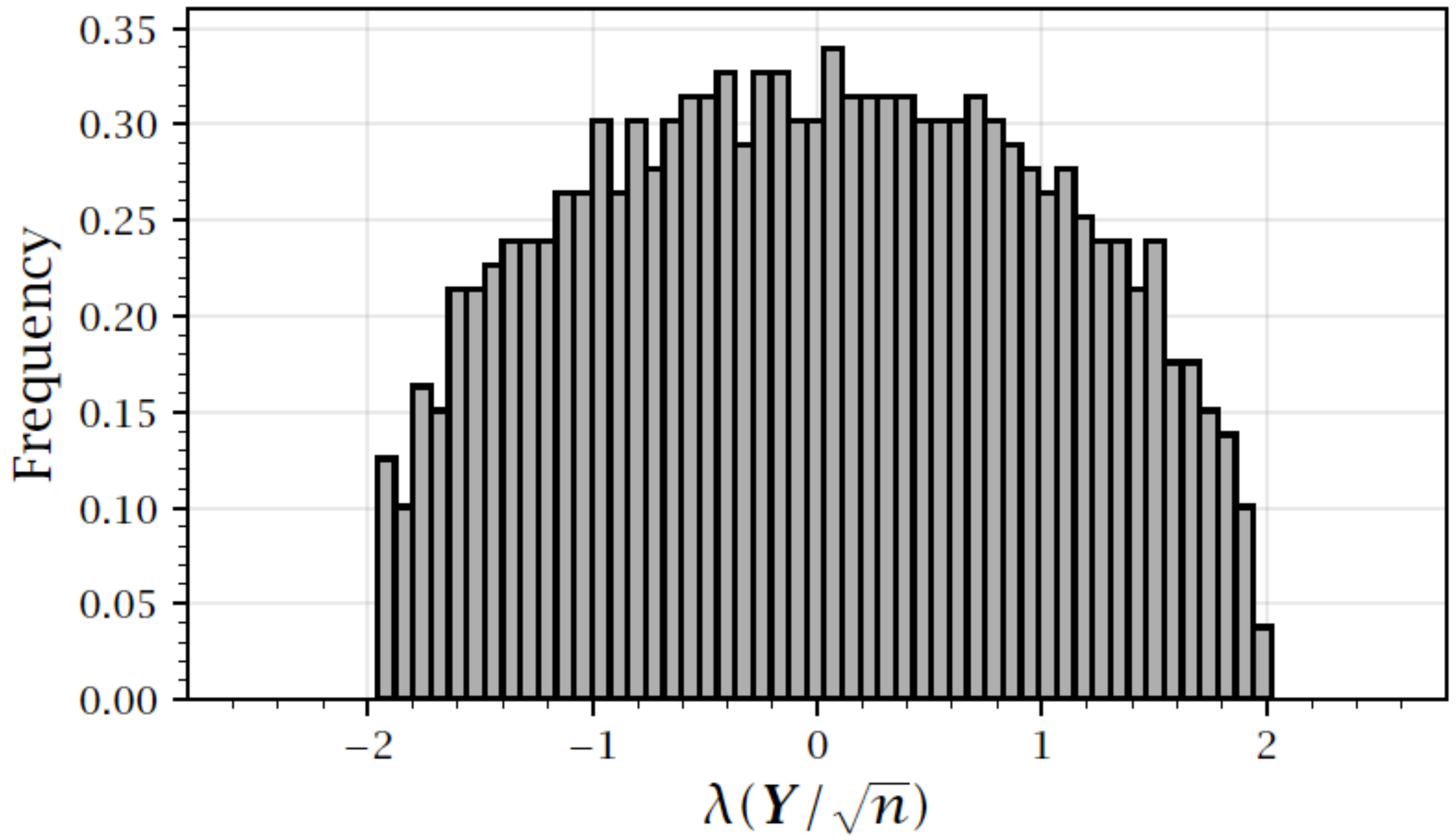
Take $\mathbf{Y}$ the $\{\pm 1\}$-valued adjacency matrix of graph G, and perform PCA on it.

- $\mathbf{Y} \approx \mathbf{1}_S \mathbf{1}_S^\top +$ (i.i.d. noise matrix)

Naive spectral algorithm

Take $\mathbf{Y}$ the $\{\pm 1\}$-valued adjacency matrix of graph G, and perform PCA on it.

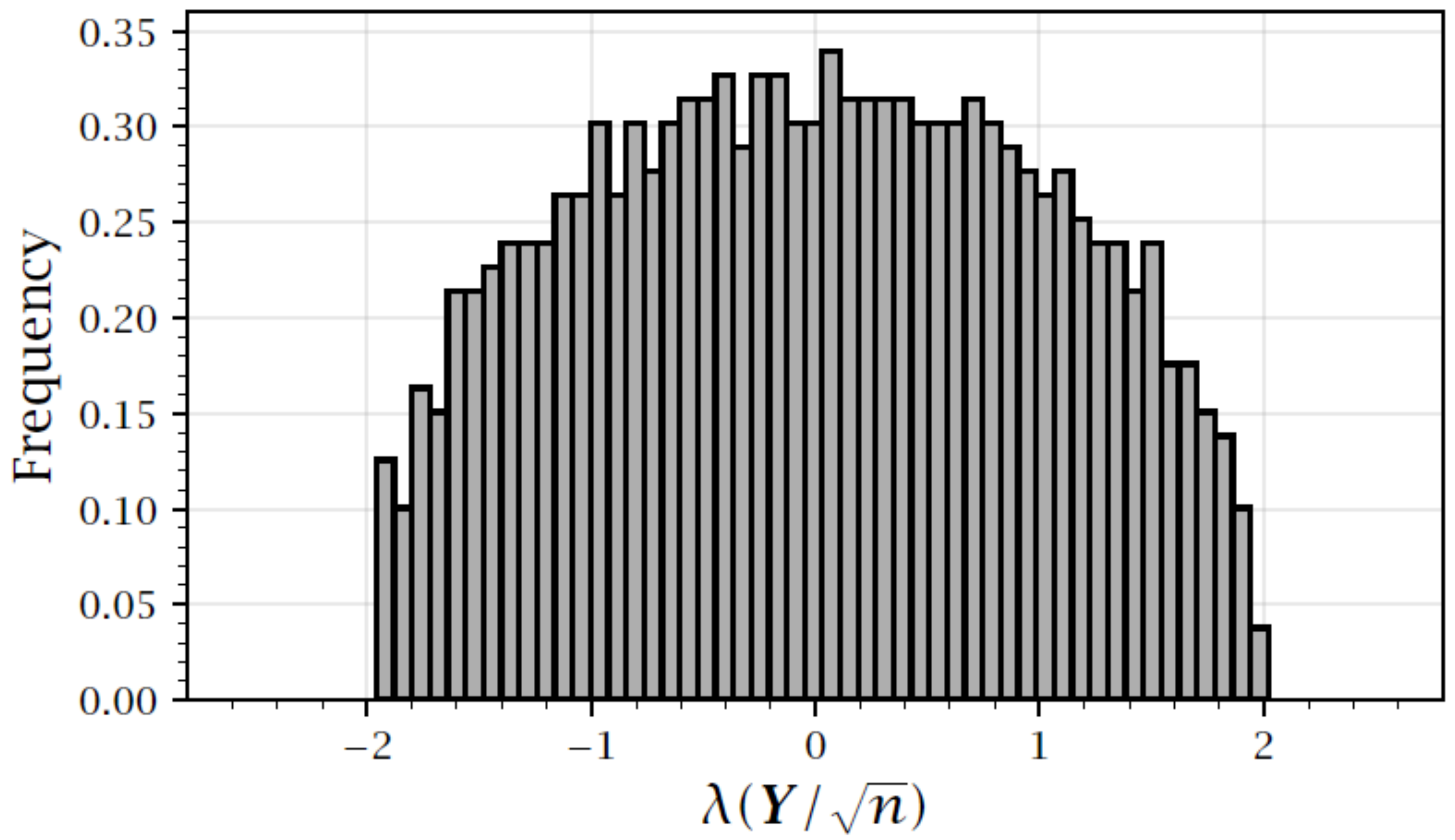
- $\mathbf{Y} \approx \mathbf{1}_S \mathbf{1}_S^\top +$ (i.i.d. noise matrix)

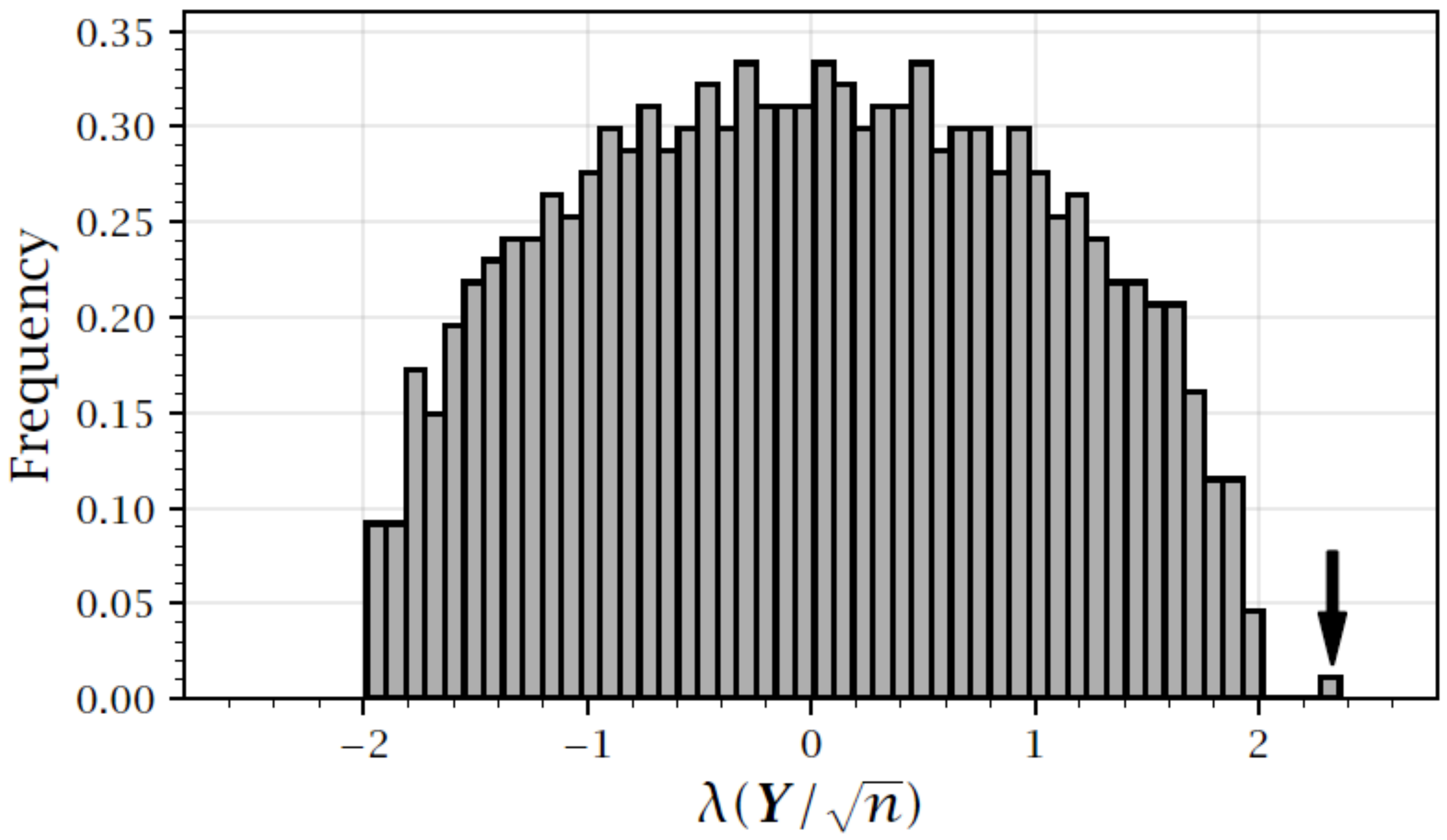


Naive spectral algorithm

Take $\mathbf{Y}$ the $\{\pm 1\}$-valued adjacency matrix of graph G, and perform PCA on it.

- $\mathbf{Y} \approx \mathbf{1}_S \mathbf{1}_S^\top +$ (i.i.d. noise matrix)

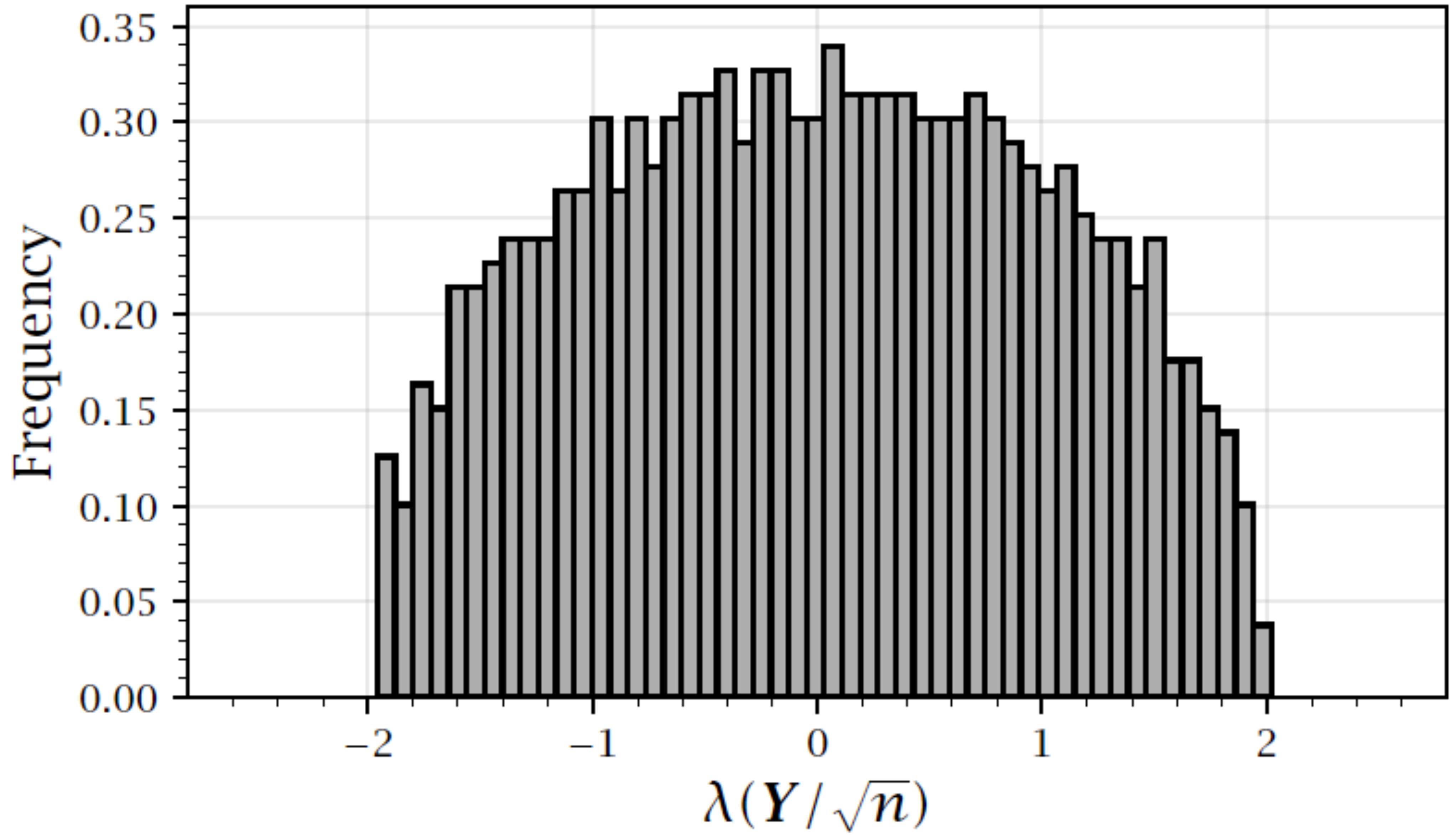


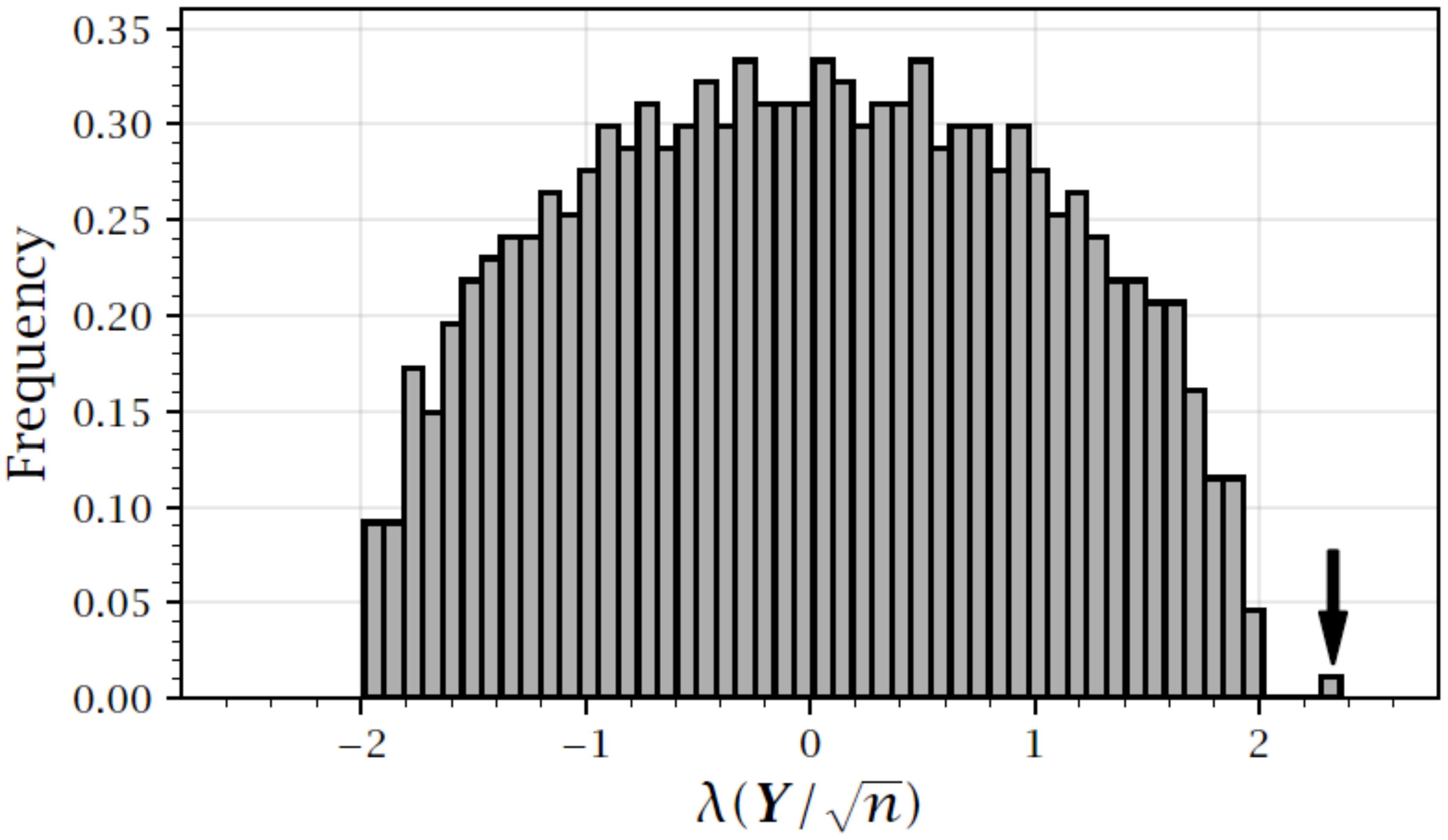


Naive spectral algorithm

Take $\mathbf{Y}$ the $\{\pm 1\}$-valued adjacency matrix of graph G, and perform PCA on it.

- $\mathbf{Y} \approx \mathbf{1}_S \mathbf{1}_S^\top +$ (i.i.d. noise matrix)
- Outlier eigenvalue $\iff$ $|S| > \sqrt{n}$





Naive spectral algorithm

- **Detection**: decide there's a planted clique if $\lambda_1\left(\frac{\mathbf{Y}}{\sqrt{n}}\right) > 2 + \varepsilon$
- **Estimation**: $v_1\left(\frac{\mathbf{Y}}{\sqrt{n}}\right) \approx \frac{\mathbf{1}_S}{\|\mathbf{1}_S\|}$

Non-trivial detection and estimation $\iff |S| > \sqrt{n}$

Our proposal: Nonlinear Laplacians

The naive spectral algorithm discards crucial prior information:
it works for any rank-one estimation $\mathbf{Y} = xx^\top +$ (i.i.d. noise matrix)

Our proposal: Nonlinear Laplacians

The naive spectral algorithm discards crucial prior information:
it works for any rank-one estimation $\mathbf{Y} = \boldsymbol{x}\boldsymbol{x}^\top + \text{(i.i.d. noise matrix)}$

- In Planted Clique, the signal $\mathbf{1}_S$ is **biased** towards $\mathbf{1}$.
- Nodes inside the clique have higher **degrees**.

Our proposal: Nonlinear Laplacians

The naive spectral algorithm discards crucial prior information:
it works for any rank-one estimation $\mathbf{Y} = \boldsymbol{x}\boldsymbol{x}^\top + \text{(i.i.d. noise matrix)}$

- In Planted Clique, the signal $\mathbf{1}_S$ is **biased** towards $\mathbf{1}$.
- Nodes inside the clique have higher **degrees**.

Idea: Use λ_1, v_1 of the Nonlinear Laplacian matrix

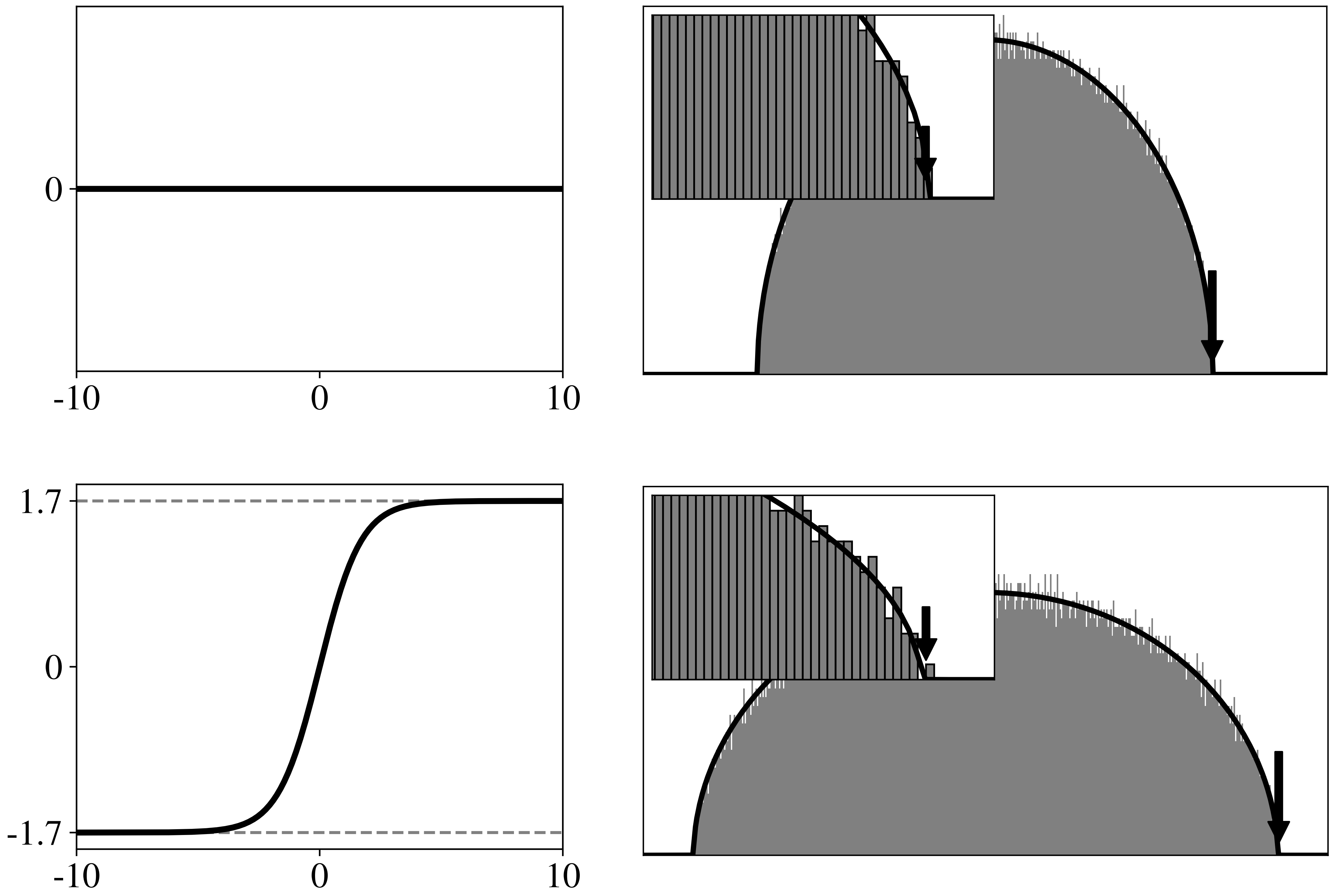
$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$

- $\hat{\mathbf{Y}}\mathbf{1}$ captures the degree information.
- $\sigma : \mathbb{R} \to \mathbb{R}$ bounded and monotone, applied entry-wise.

Our proposal: Nonlinear Laplacians

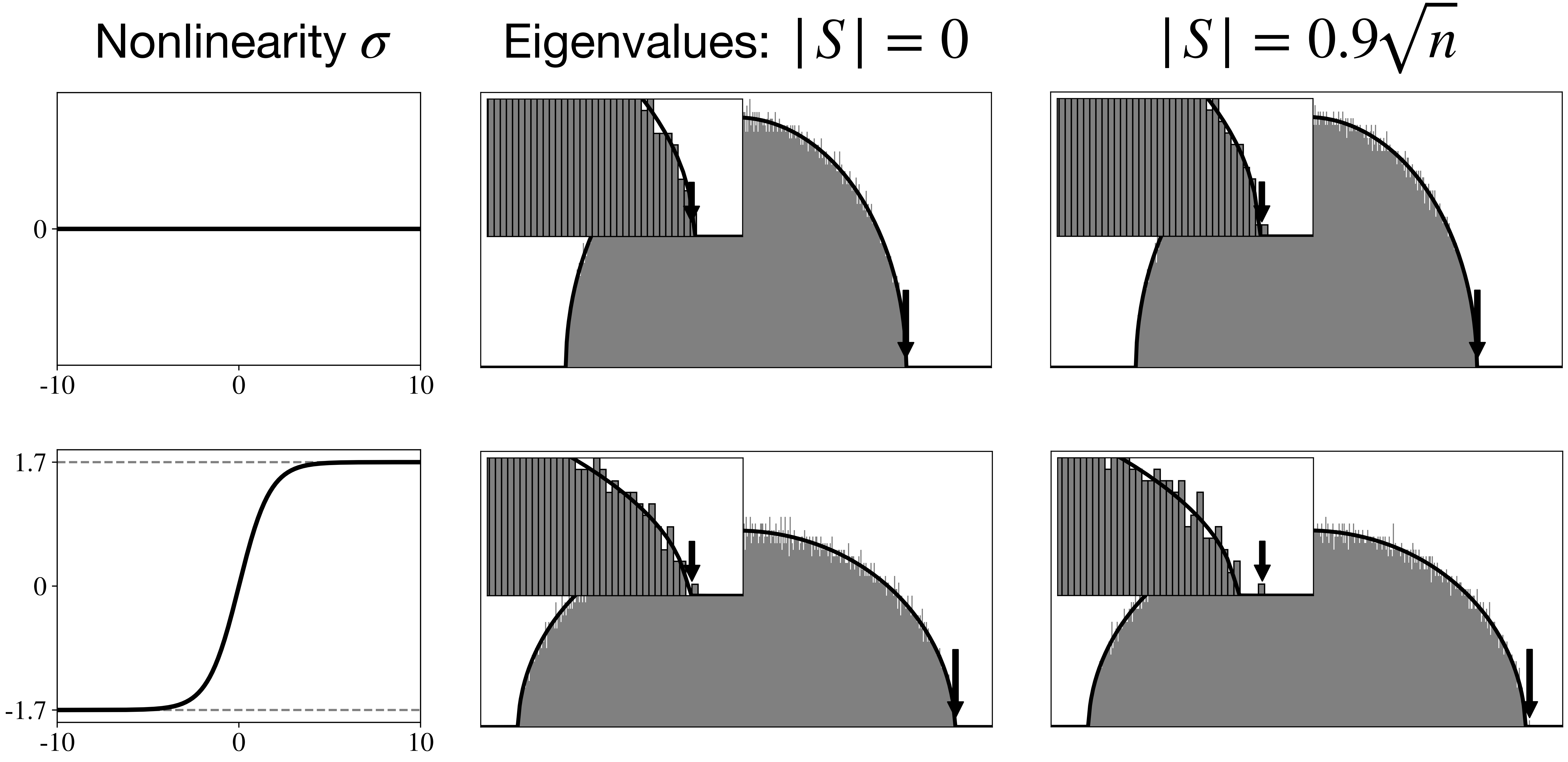
$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$

Nonlinearity σ Eigenvalues: $|S| = 0$



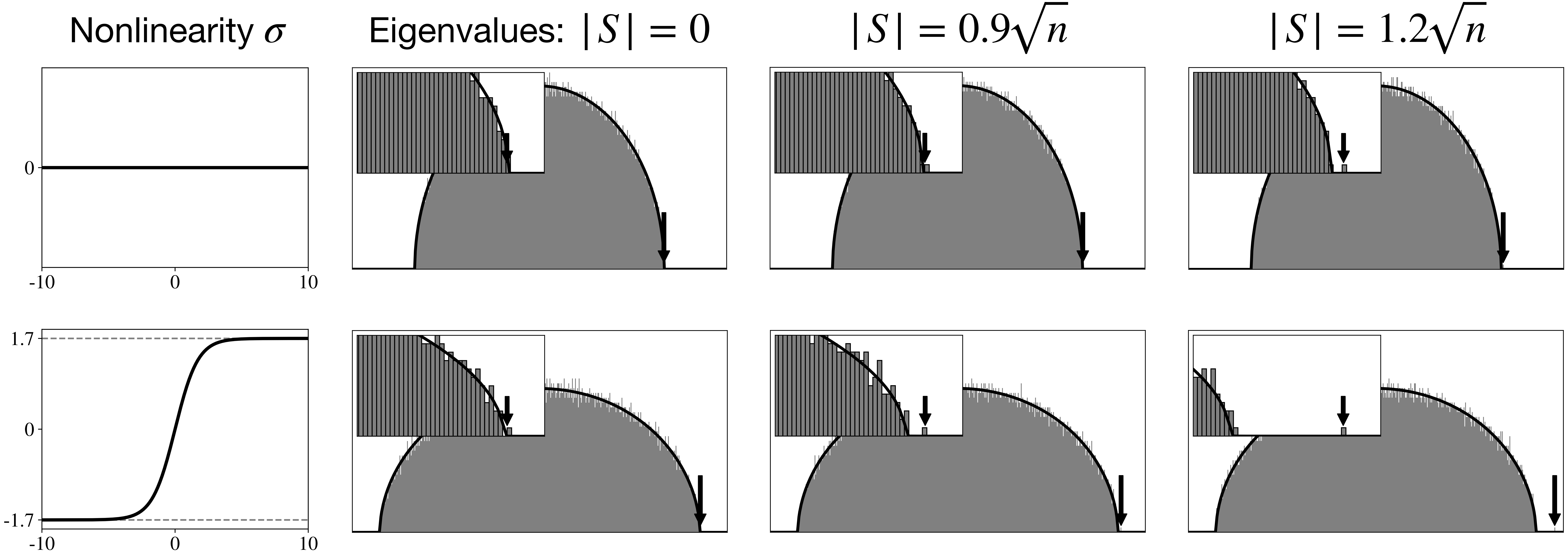
Our proposal: Nonlinear Laplacians

$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$



Our proposal: Nonlinear Laplacians

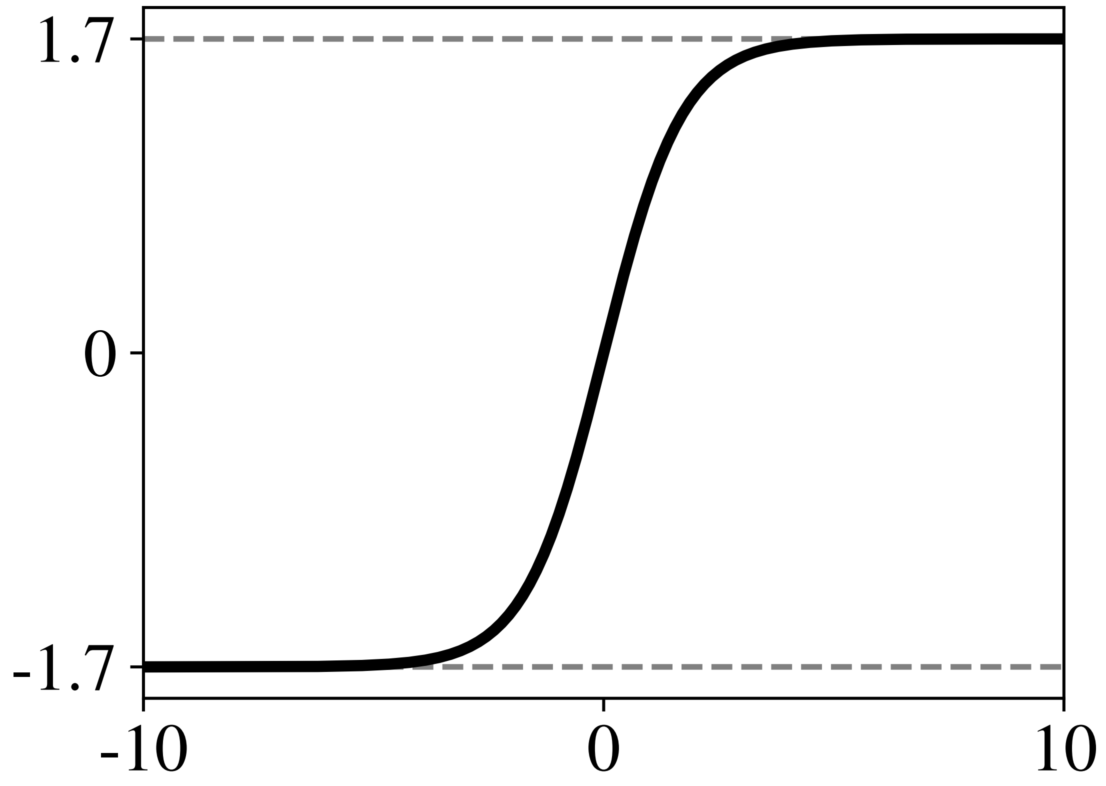
$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$



Our proposal: Nonlinear Laplacians

$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$

$\sigma =$



Non-trivial detection and estimation $\iff$ $|S| > 0.76\sqrt{n}$

Our proposal: Nonlinear Laplacians

$$\boxed{\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}}$$

A simple, tunable family of spectral algorithms that combines spectral and degree information.

Our proposal: Nonlinear Laplacians

$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$

A simple, tunable family of spectral algorithms that combines spectral and degree information.

- For any clique size $< \sqrt{n}$, neither the spectral information alone nor the degree information alone cannot detect.
- By combining them in this simple way, we achieve detection for any clique size $> 0.76\sqrt{n}$.

Our proposal: Nonlinear Laplacians

$$\mathbf{L}_\sigma(\mathbf{Y}) = \hat{\mathbf{Y}} + \mathrm{diag}(\sigma(\hat{\mathbf{Y}}\mathbf{1})), \quad \hat{\mathbf{Y}} = \frac{\mathbf{Y}}{\sqrt{n}}$$

A simple, tunable family of spectral algorithms that combines spectral and degree information.

- For any clique size $< \sqrt{n}$, neither the spectral information alone nor the degree information alone cannot detect.
- By combining them in this simple way, we achieve detection for any clique size $> 0.76\sqrt{n}$.

Poster on Thursday 4 Dec, 4:30 - 7:30 pm PST.