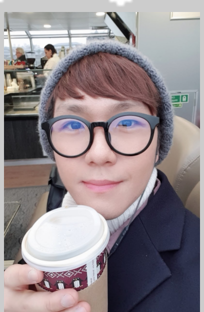
Adaptive Power Method: Eigenvector Estimation from Partial Observations

Seiyun Shin, Han Zhao, and Ilan Shomorony

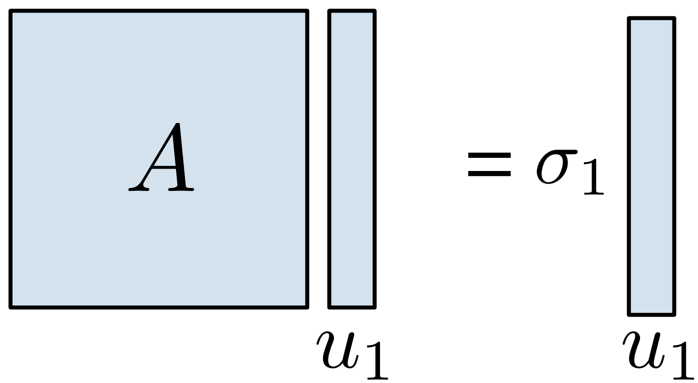
University of Illinois at Urbana-Champaign



ITA – February 14, 2023

Old problem: eigenvector computation

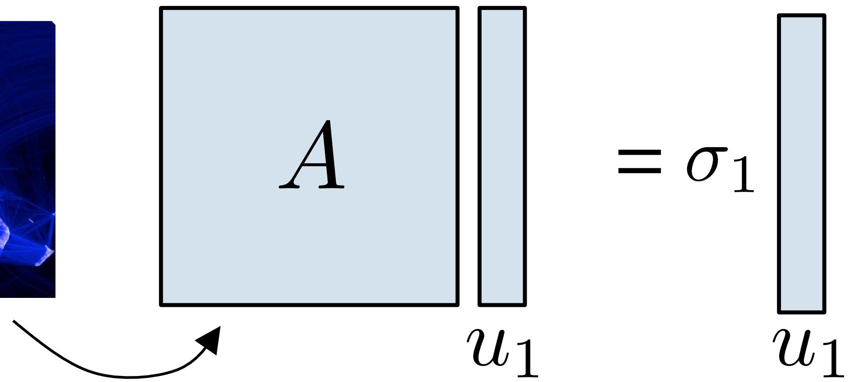
- Compute top eigenvector of a symmetric matrix A



- Many applications in data science:
  - PCA, low-rank approximation, spectral clustering
  - Network analysis: <u>eigenvector centrality</u>

Old problem: eigenvector computation

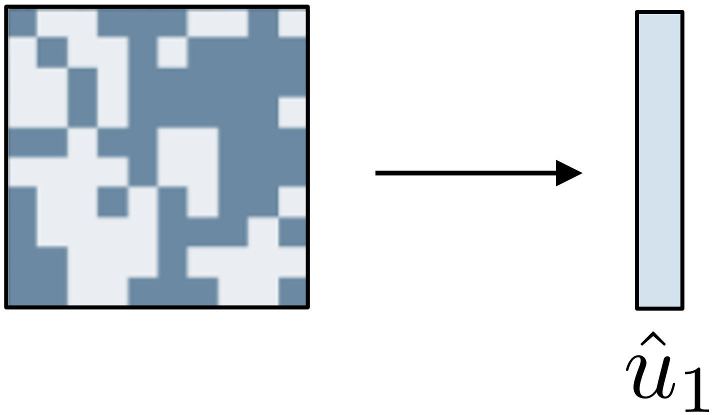
- Compute top eigenvector of a symmetric matrix A



- **Challenge:**
  - Matrix A could be large and stored in distributed manner
  - Fully observing A may be costly or infeasible
  - e.g., drug-interaction networks, biological networks

Eigenvector estimation from partial observations

- Symmetric matrix $A \in \{0,1\}^{n\times n}$



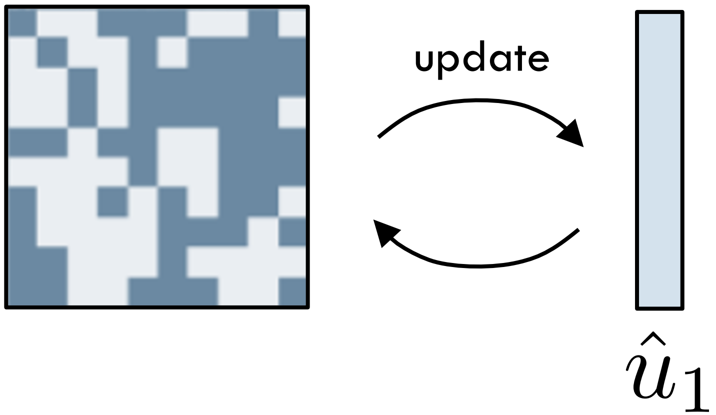
- Budget of B entry observations

Questions:

- How large does B have to be for $\hat{u}_1 \approx u_1$?
- Can adaptivity help?

Main result

- Adaptive Power Method (APM)



- Adaptively select new observations based on current $\hat{u}_1$

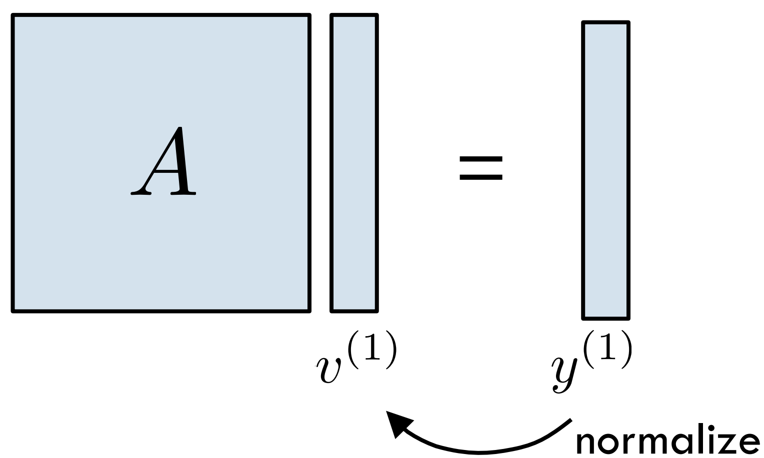
Theorem (informal): With $B = O\left(\frac{n}{\epsilon^2}\log^2\left(\frac{n}{\epsilon}\right)\right)$ observations,

$$\sin(\hat{u}_1, u_1) \leq \epsilon$$

with probability $1 - o(1)$.

Power Method (classical)

- Start with random unitary vector $v^{(0)}$



$$y^{(\ell)} = Av^{(\ell-1)}$$
$$v^{(\ell)} = y^{(\ell)} / \|y^{(\ell)}\|$$

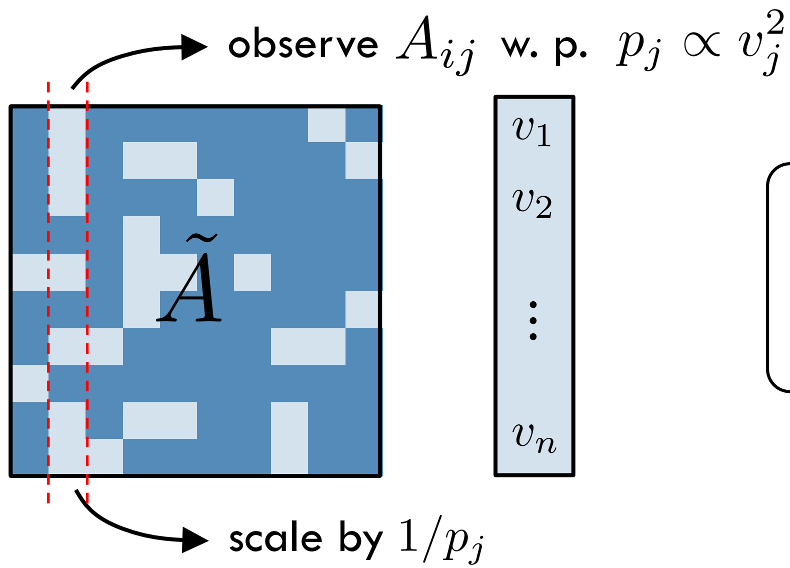
- After $L = O\left(\frac{\sigma_1}{\sigma_1 - \sigma_2} \log\left(\frac{n}{\epsilon}\right)\right)$ iterations,

$$\|u_1 - v^{(L)}\| \leq \epsilon$$

- What if we can only observe A partially?
  - We can estimate each $Av^{(\ell)}$

Adaptive Power Method

- Start with random unitary vector $v^{(0)}$



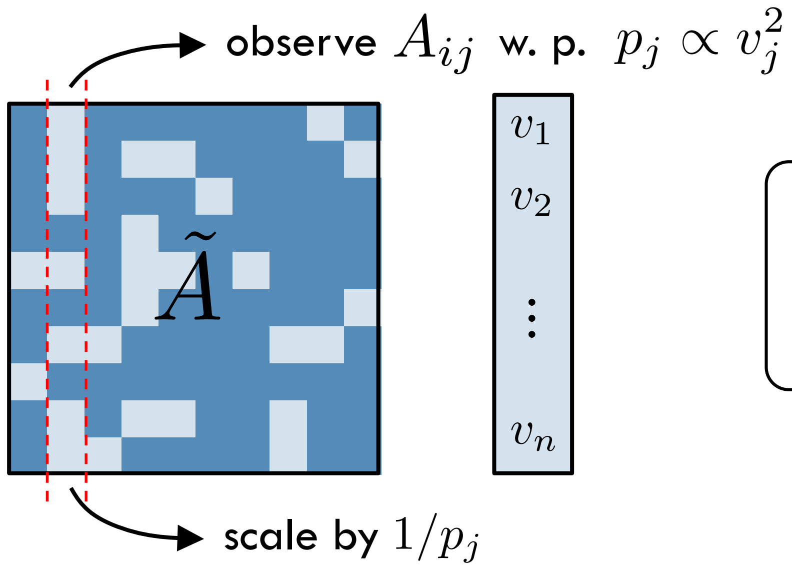
$$\boxed{\begin{aligned} y^{(\ell)} &= \tilde{A}^{(\ell)} v^{(\ell-1)} \\ v^{(\ell)} &= y^{(\ell)} / \|y^{(\ell)}\| \end{aligned}}$$

- With $O\left(\frac{n}{\epsilon^2}\log n\right)$ observations per iteration, w. p. $1-o(1)$,

$$\|\tilde{A}^{(\ell)} v^{(\ell-1)} - A v^{(\ell-1)}\| \leq \epsilon$$

- Proof is based on Bernstein's inequality
- Will $v^{(\ell)}$ converge to u_1 ?

Adaptive Power Method



$$
\begin{aligned}
y^{(\ell)} &= \tilde{A}^{(\ell)} v^{(\ell-1)} \\
v^{(\ell)} &= y^{(\ell)} / \|y^{(\ell)}\|
\end{aligned}
$$

- **Noisy Power Method**

  (Hardt and Price, 2014)

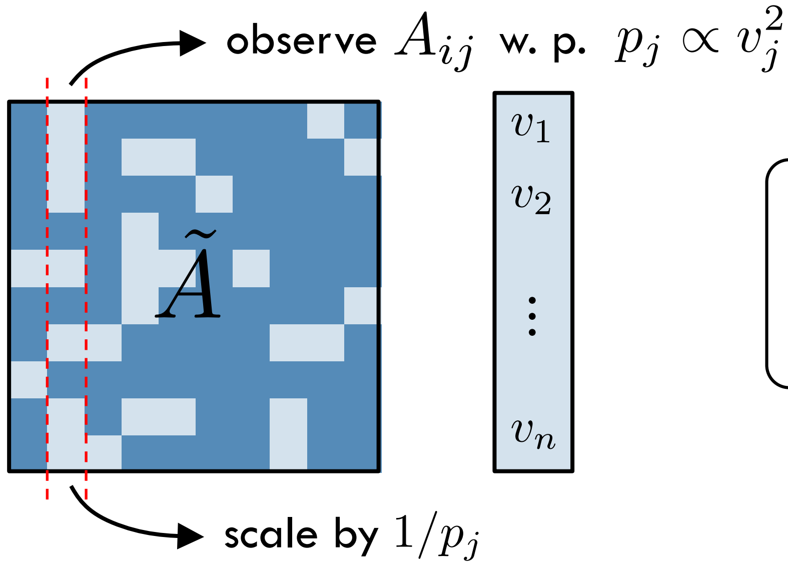
NPM:

$$
\begin{aligned}
y^{(\ell)} &= A v^{(\ell-1)} + n^{(\ell)} \\
v^{(\ell)} &= y^{(\ell)} / \|y^{(\ell)}\|
\end{aligned}
$$

Theorem: NPM converges to $u_1 + O(\epsilon)$ if all $n^{(\ell)}$ satisfy

1. $\|n^{(\ell)}\| \leq \frac{\epsilon}{5}(\sigma_1 - \sigma_2)$
2. $\|u_1^T n^{(\ell)}\| \leq \frac{1}{5\sqrt{n}}(\sigma_1 - \sigma_2)$

Adaptive Power Method

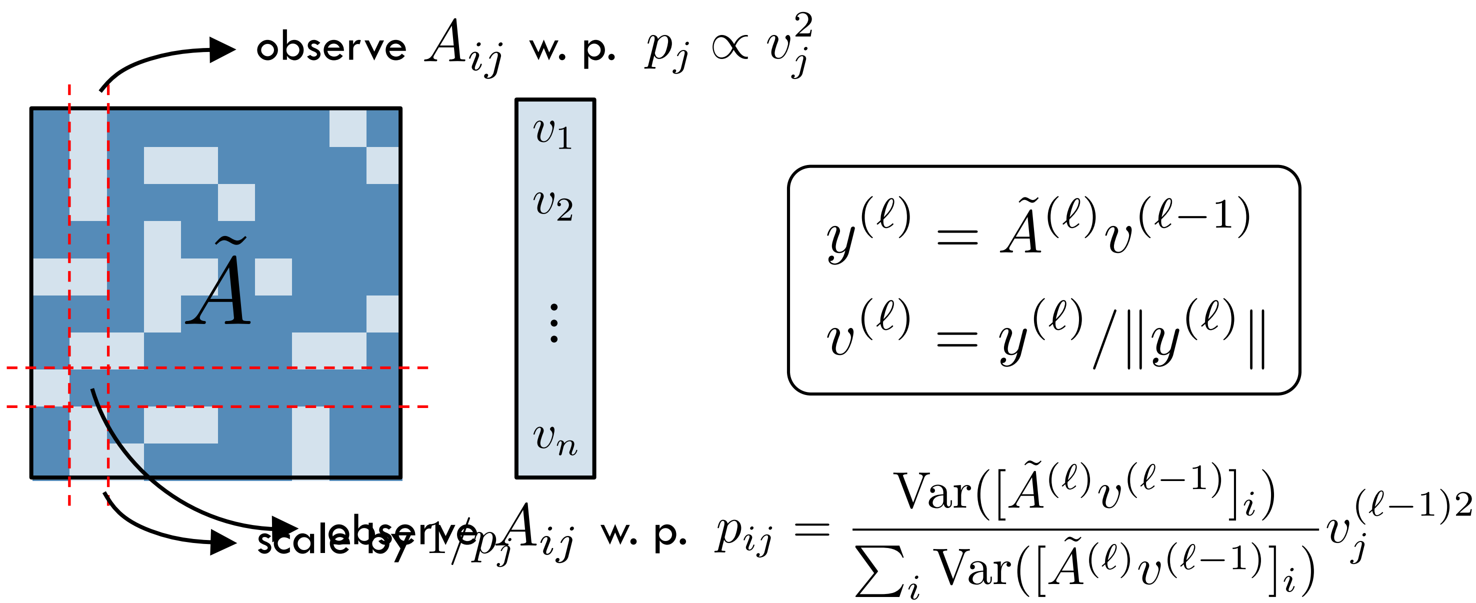


$$\begin{aligned} y^{(\ell)} &= \tilde{A}^{(\ell)} v^{(\ell-1)} \\ v^{(\ell)} &= y^{(\ell)} / \|y^{(\ell)}\| \end{aligned}$$

- View APM as a version of the Noisy Power Method

Theorem: Suppose $\sigma_1 - \sigma_2 = \Theta(n)$. Using $B = O\left(\frac{n}{\epsilon^2}\log^2\left(\frac{n}{\epsilon}\right)\right)$ observations and $L = O(\log(n/\epsilon))$ iterations, APM estimates the top eigenvector with $\sin(\hat{u}_1, u_1) \leq \epsilon$, with probability $1 - o(1)$.

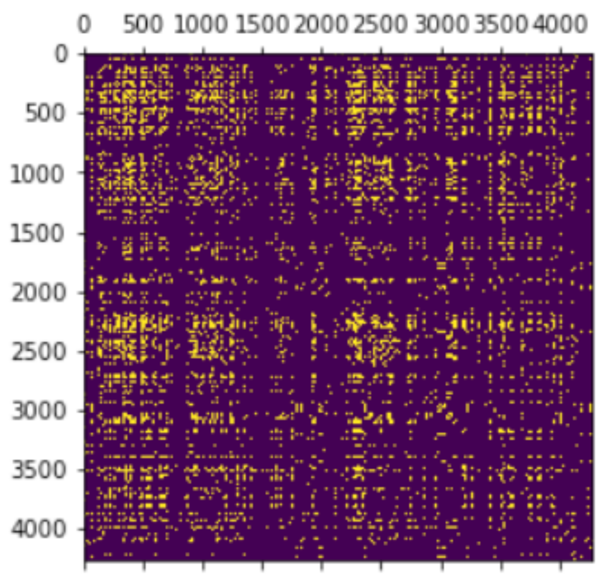
Adaptive Power Method Refinement

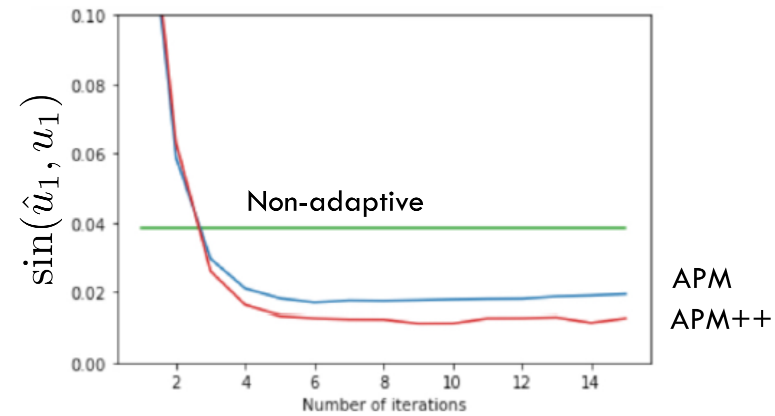


- This choice of p_{ij} minimizes variance of the estimator
- Non-uniform budget allocation; favors large eigenvector entries
- We call this APM++

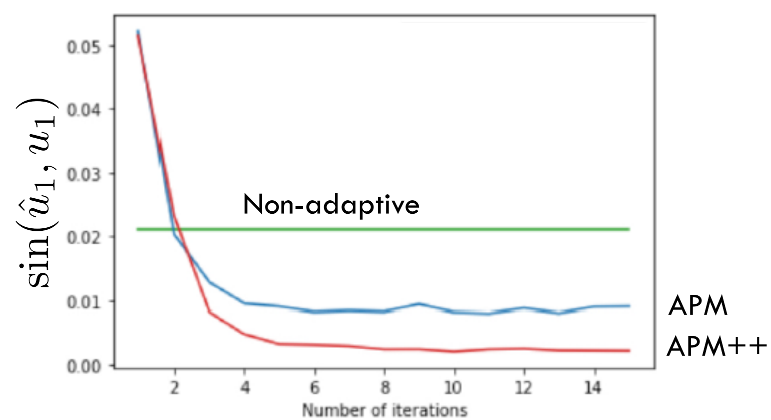
Numerical Results

- Drug-drug interaction network (Open Graph Benchmark)
  - 4k nodes and 2M edges





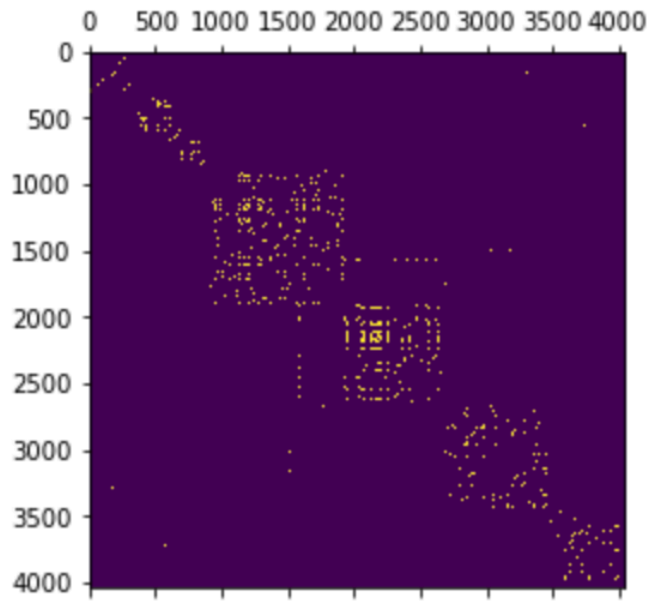
top-500 entries of u_1

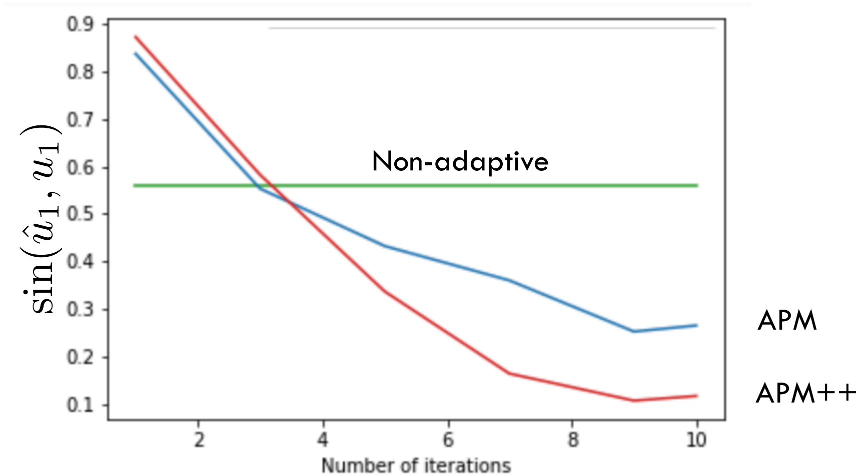


  - B = 50% of entries

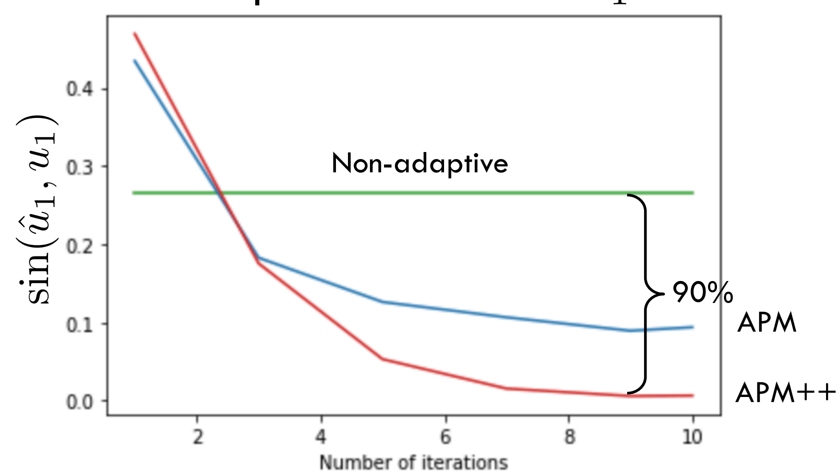
Numerical Results

- Stanford's "ego-facebook" network
  - 4k nodes and 90k edges





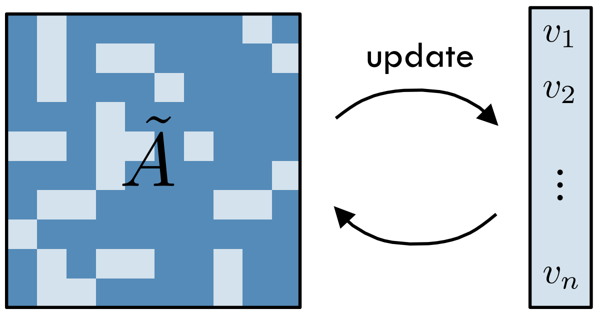
top-500 entries of u_1



  - B = 5% of entries

Concluding Remarks

- **APM:** adaptively estimate top eigenvector



Budget: $B = O\left(\frac{n}{\epsilon^2}\log^2\left(\frac{n}{\epsilon}\right)\right)$

Error: $\sin(\hat{u}_1, u_1) \leq \epsilon$

- Generalizations:
  - Non-binary A, top-k eigenvectors, asymmetric A ?
- Very sparse graphs? Exploit community structure?
- Partial graph observation in other settings: Graph NNs?

Thank you!