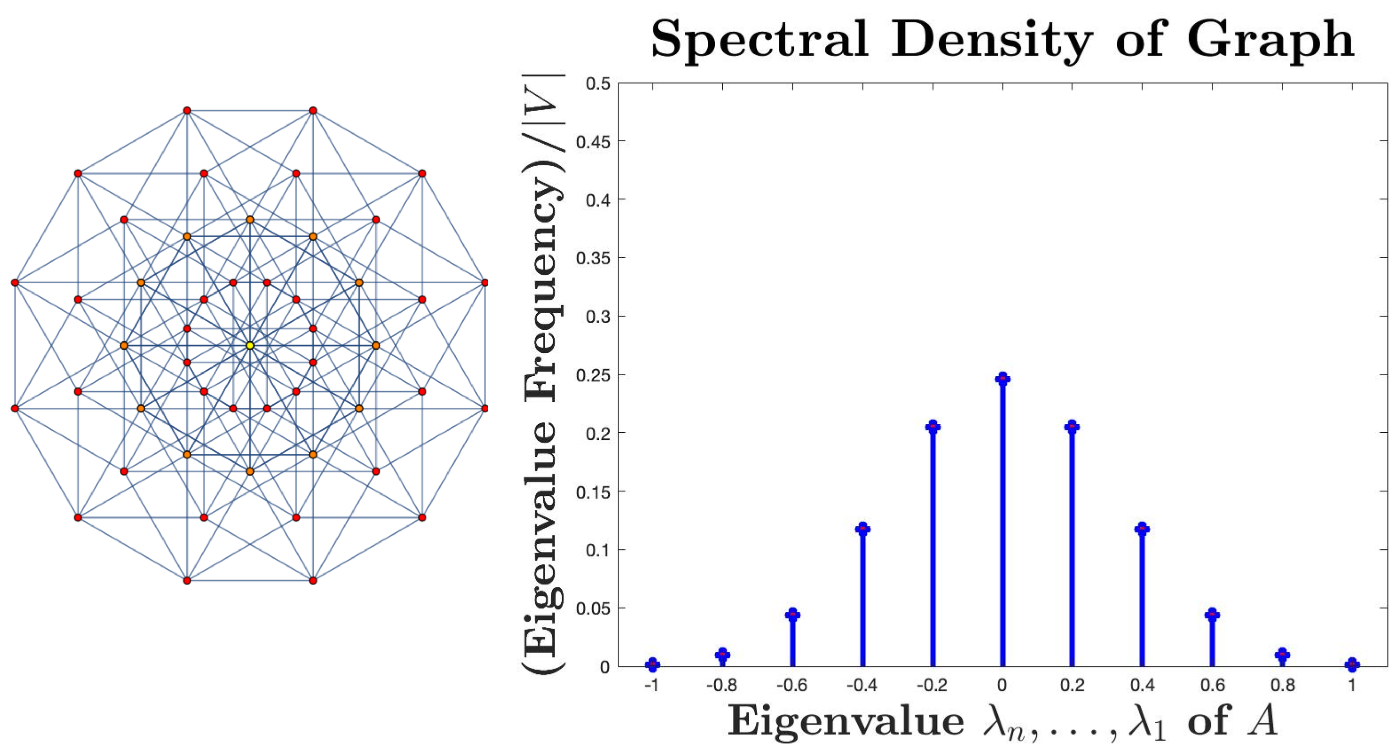
# Moments, Random Walks, and Limits for Spectrum Approximation

Yujia Jin (Stanford), Christopher Musco (NYU), Aaron Sidford (Stanford) and **Apoorv Vikram Singh** (NYU)

## Problem

**Spectral Density Estimation:** $G = (V, E)$, $A = D^{-\frac{1}{2}} G D^{-\frac{1}{2}}$.




**Problem:** Estimate spectral density to error $\varepsilon$ in *Wasserstein*-1 distance.
**Equivalently:** Output $\lambda'_1 \geq \cdots \geq \lambda'_n$, s.t. $\frac{1}{n}\sum_i |\lambda_i - \lambda'_i| \leq \varepsilon$.

## Existing results

- [CKSV'18] Given access to a random node and a random neighbour: Estimate spectral *density* to error $\leq \varepsilon$ in $W_1$ distance in time $2^{O(1/\varepsilon)}$.
- Idea: Estimate first $1/\varepsilon$ *moments* to accuracy $2^{O(-1/\varepsilon)}$, using *random-walks* for *spectrum approximation*, and return a distribution [KV'17].
- *Moments and Random Walks:* $j$-th moment $m_j = \frac{1}{n}\sum_{i=1}^n \lambda_i^j$.

$$m_j = \frac{1}{n}\text{Tr}(A^j) = \frac{1}{n}\sum_i \Pr[j\text{-step rw from } i \text{ to } i].$$

- Takes $2^{O(1/\varepsilon)}$-steps of random-walks to estimate the first $1/\varepsilon$ moments to accuracy $2^{O(-1/\varepsilon)}$.

- [BKM'22] Estimate spectral density to error $\leq \varepsilon$ in $W_1$ distance in time $n/\text{poly}(\varepsilon)$.

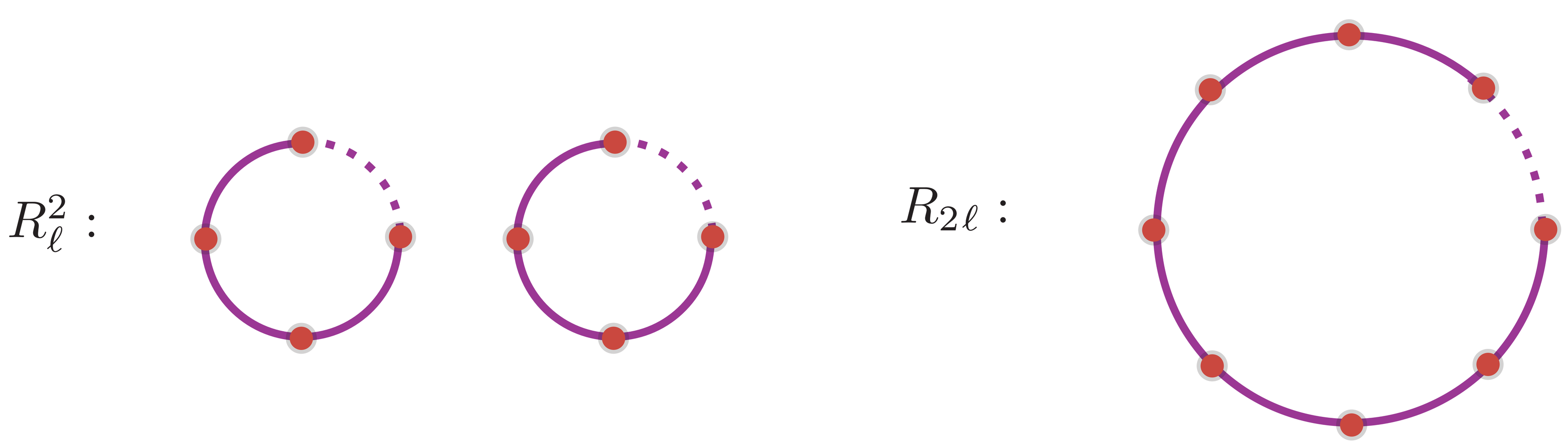
## Our results

Results of [KV'17] and [CKSV'18] are *tight*!!

1. $\exists$ distributions on $[-1, 1]$ that cannot be approximated to accuracy $\varepsilon$ in $W_1$ distance even if *all* of their moments are known to multiplicative accuracy $(1 \pm 2^{-\Omega(1/\varepsilon)})$.
2. No algo can compute an $\varepsilon$-accurate approximation to SDE with constant probability, even when given the transcript of $2^{\Omega(1/\varepsilon)}$ random walks of length $2^{\Omega(1/\varepsilon)}$.

## Initial idea

Generate 2 distributions such that moments are close and $W_1$ distance is far.

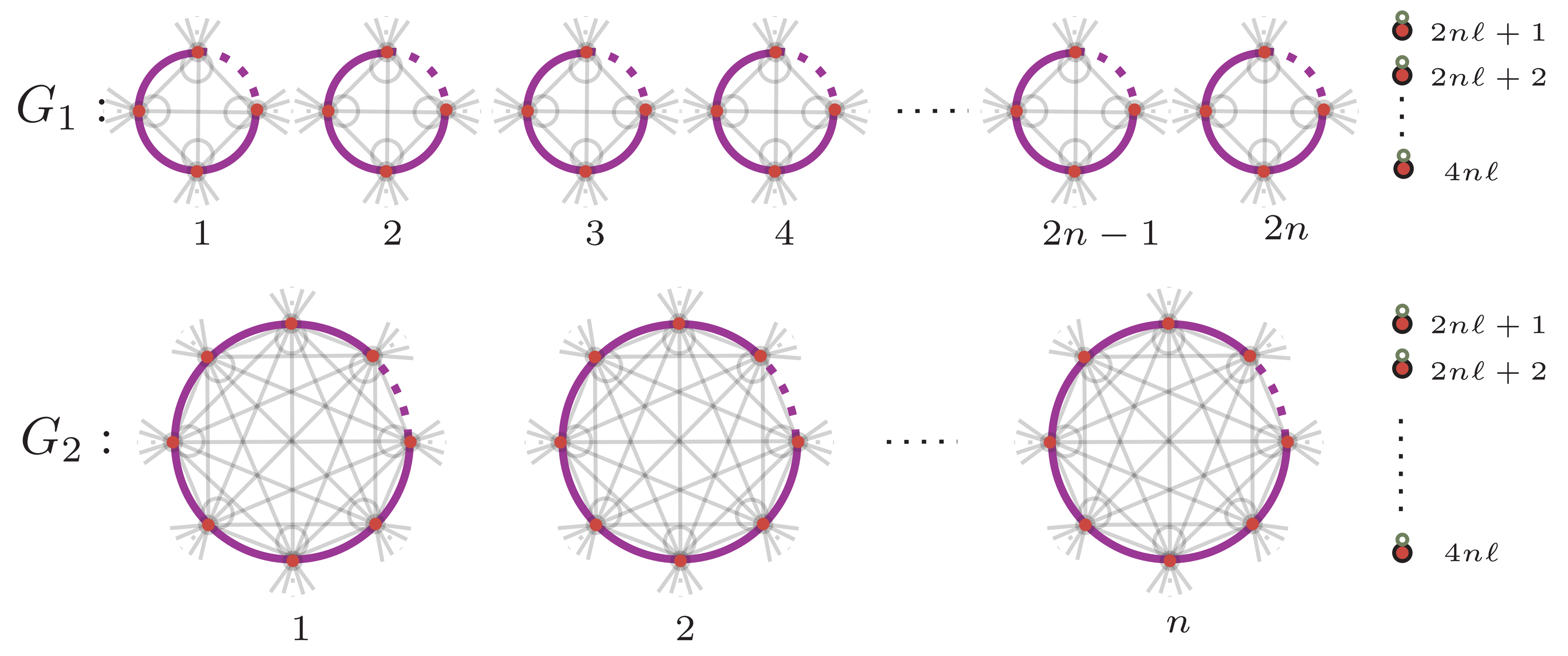



- $W_1(R_{2\ell}, R_\ell^2) = 1/\ell$. Choose $\ell = 1/\varepsilon$.
- First $\ell - 1$ moments are the same.
- Difference of $\ell$-th moment is $2^{-\ell}$.
- However, difference of $1/\varepsilon^2$-th moment is $\Omega(\varepsilon)$-away!

Paper: `https://savs95.github.io/rw_sde.pdf`

## Moment lower bound

Fix the issue by adding a *light* complete graph.

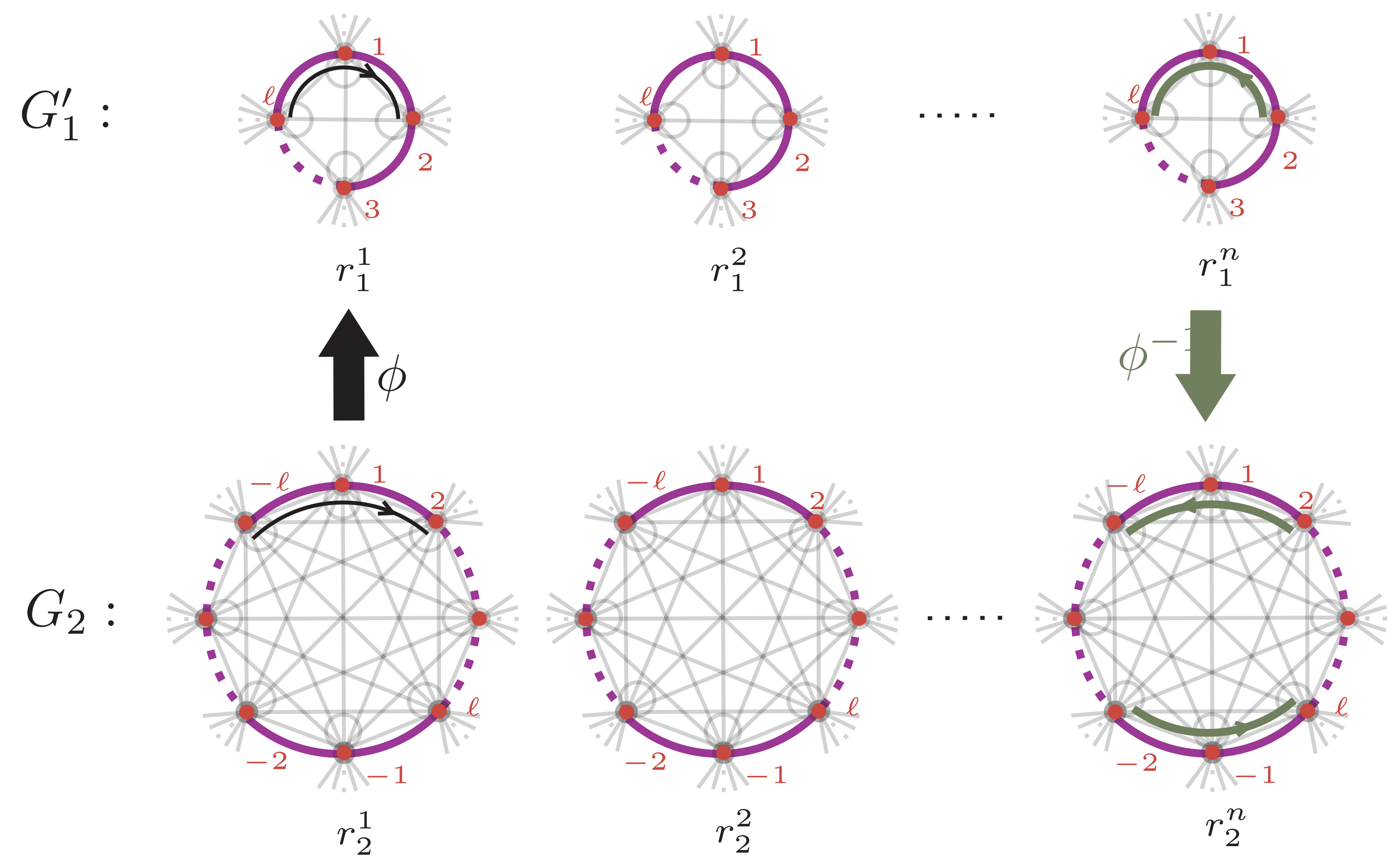



- $W_1(G_1, G_2) = 1/(4\ell)$. Choose $\ell = 1/(4\varepsilon)$.
- With probability $1/2$, the random walk stays inside the cycle.
- With probability $1/2$, the random walk goes to a random node.

$$\implies |m_j(G_1) - m_j(G_2)| = \begin{cases} 0, & \text{for } j < \ell \\ 2^{-\ell+1}, & \text{for } j \geq \ell \end{cases}.$$

## Random-walk lower bound

Random-walks started from random nodes.

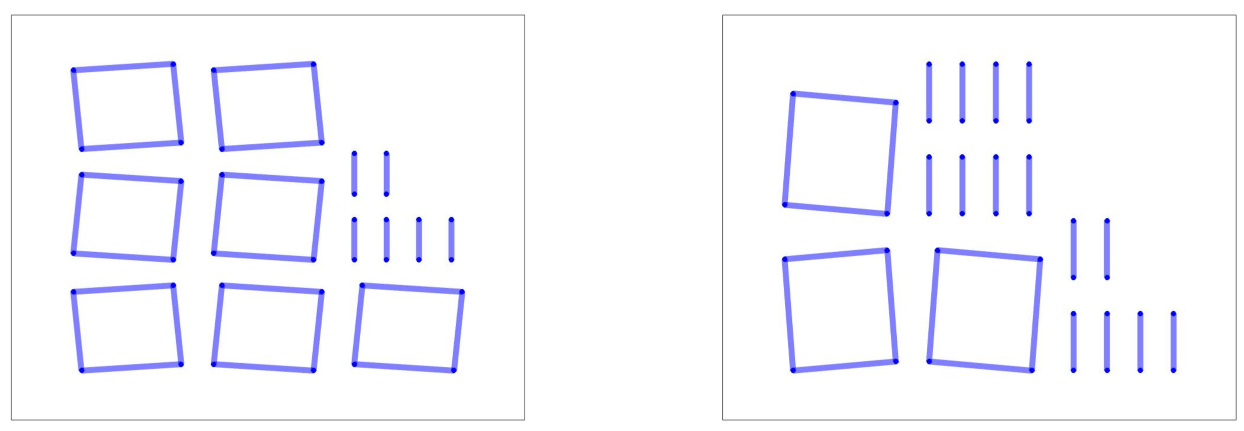



- $W_1(G'_1, G_2) > 1/2\ell$. Choose $\ell = 1/(2\varepsilon)$.
- Define a *coupling* between the random-walks using $\phi$ and $\phi^{-1}$.
- $S$: $2^{\Omega(1/\varepsilon)}$-sized transcript of random-walks in $G'_1$, or $G_2$.

$$d_{\text{TV}}(\mathbb{P}_{G'_1}(S), \mathbb{P}_{G_2}(S)) \leq \frac{1}{2}.$$

## Adaptive random-walks lower bound

$\frac{1}{2} + \varepsilon$ fraction of 4-cycles and $\frac{1}{2} - \varepsilon$ fraction of 2-cycles.



- Need $\Omega(1/\varepsilon^2)$-random-walk queries to distinguish between the two graphs.
- **Open Problem 1:** Improve lower bound?
- **Open Problem 2:** Improve upper bound?

## References

[BKM'22] Vladimir Braverman, Aditya Krishnan, and Christopher Musco. Sublinear Time Spectral Density Estimation *STOC*, 2022.

[CKSV'18] David Cohen-Steiner, Weihao Kong, Christian Sohler, and Gregory Valiant. Approximating the Spectrum of a Graph *KDD*, 2018.

[KV'17] Weihao Kong and Gregory Valiant. Spectrum Estimation from Samples. *The Annals of Statistics*, 2017.