

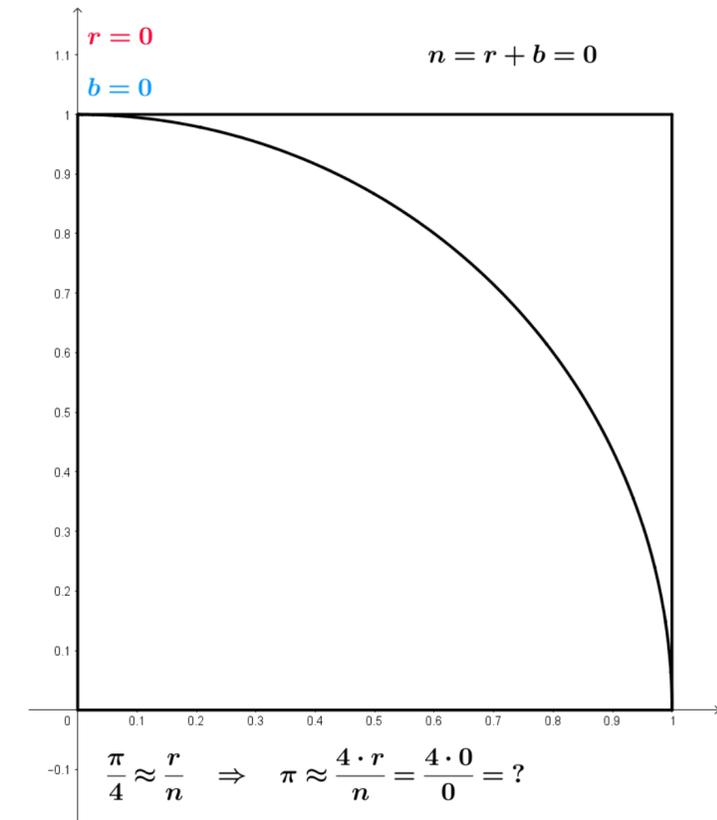
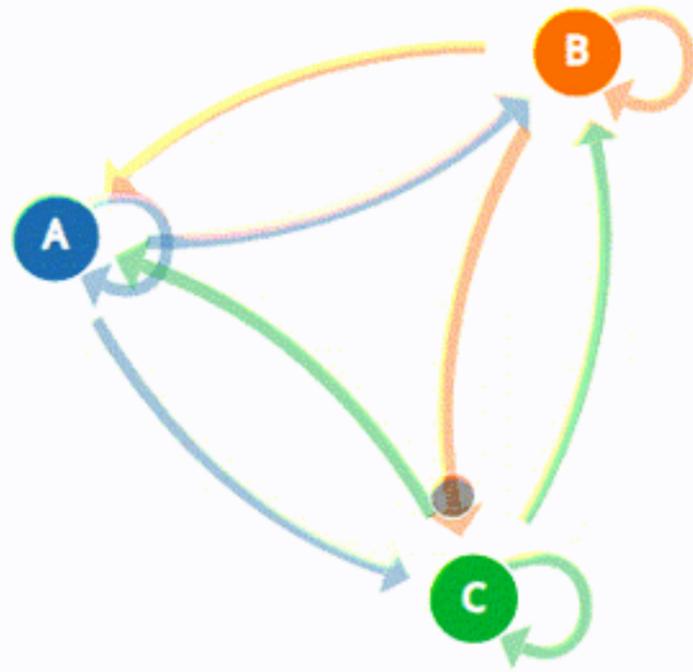
# Warm starts in Classical and Quantum computation

Shrinidhi Teganahally Sridhara

Equipe ICQ, LaBRI

Based on ongoing works with Yassine Hamoudi, Yvan Le Borgne and Adrian Tanasa

# The Markov chain Monte Carlo method (MCMC)

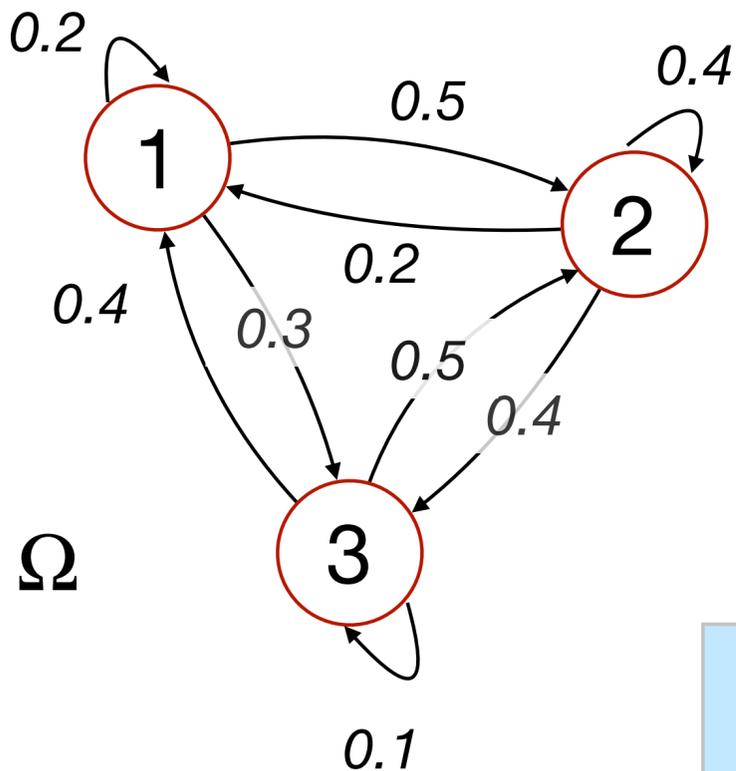


**Markov chains**

**MC + MC**

**Monte Carlo**

# Markov chains



$$P = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.4 & 0.5 & 0.1 \end{bmatrix}_{|\Omega| \times |\Omega|}$$

Transition matrix

## Mixing time

$t_{mix}(\epsilon) := t$  s.t.  $\mu_0 P^t$  is  $\epsilon$  **close** to  $\pi$  in TV distance

MC has mixed

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.26 \\ 0.49 \\ 0.29 \end{bmatrix} \xrightarrow{P} \dots \xrightarrow{P} \begin{bmatrix} 0.258 \\ 0.455 \\ 0.287 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.2574 \\ 0.4545 \\ 0.2881 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.2574 \\ 0.4545 \\ 0.2881 \end{bmatrix}$$

$\mu_0 \quad \mu_0 P \quad \mu_0 P^2 \quad \dots \quad \mu_0 P^t = \pi \quad \pi P = \pi$

## Perron-Frobenius Theorem:

If  $P$  satisfies **certain** conditions, then starting from **any** distribution  $\mu_0$ ,  $\exists$  unique  $\pi$  s.t.,  $\lim_{t \rightarrow \infty} \mu_0 P^t = \pi$

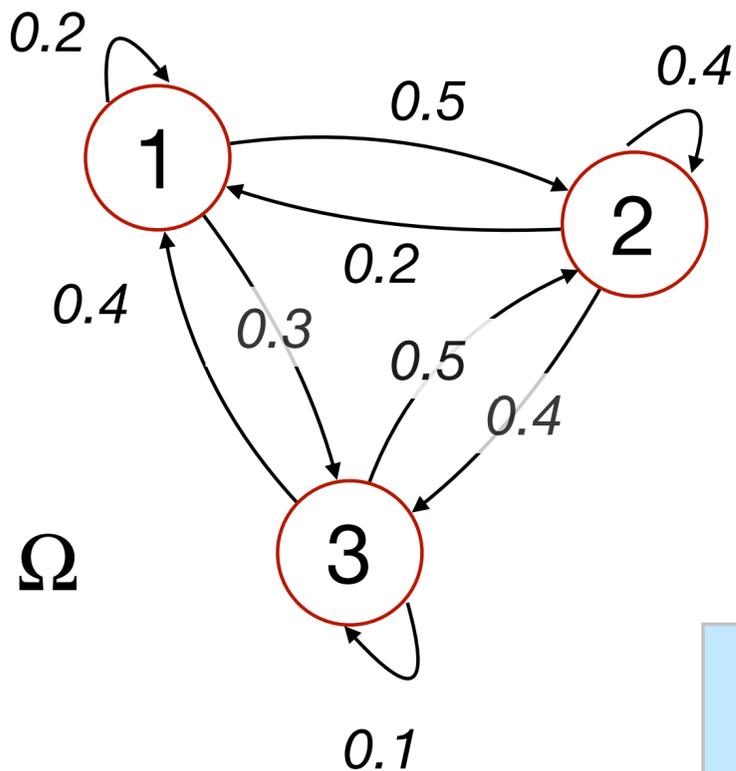
$\pi$  - unique left eigenvector of  $P$  and with the largest eigenvalue 1 normalized by the  $L_1$  norm

**Theorem (informal):**  $t_{mix}(\epsilon) = O\left(\frac{1}{1 - \lambda_2} \cdot \ln\left(\frac{\text{warmness}(\mu_0 \rightarrow \pi)}{\epsilon}\right)\right)$

Warmness - how close is  $\mu_0$  to  $\pi$ ?

$\mu_0$  is a warm start  $\longrightarrow$   $\mu_0$  close to  $\pi$   $\longrightarrow$  **Faster mixing**

# Markov chains



$$P = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.2 & 0.4 & 0.4 \\ 0.4 & 0.5 & 0.1 \end{bmatrix}_{|\Omega| \times |\Omega|}$$

Transition matrix

## Mixing time

$t_{mix}(\epsilon) := t$  s.t.  $\mu_0 P^t$  is  $\epsilon$  close to  $\pi$  in TV distance

MC has mixed

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.26 \\ 0.49 \\ 0.29 \end{bmatrix} \xrightarrow{P} \dots \xrightarrow{P} \begin{bmatrix} 0.258 \\ 0.455 \\ 0.287 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.2574 \\ 0.4545 \\ 0.2881 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} 0.2574 \\ 0.4545 \\ 0.2881 \end{bmatrix}$$

$\mu_0 \quad \mu_0 P \quad \mu_0 P^2 \quad \dots \quad \mu_0 P^t = \pi \quad \pi P = \pi$

## Perron-Frobenius Theorem:

If  $P$  satisfies **certain** conditions, then starting from **any** distribution  $\mu_0$ ,  $\exists$  unique  $\pi$  s.t.,  $\lim_{t \rightarrow \infty} \mu_0 P^t = \pi$

$\pi$  - unique left eigenvector of  $P$  and with the largest eigenvalue 1 normalized by the  $L_1$  norm

**Theorem (informal):**  $t_{mix}(\epsilon) = O \left( \frac{1}{1 - \lambda_2} \cdot \ln \left( \frac{\left\| \frac{\mu_0}{\pi} - 1 \right\|_{2,\pi}}{\epsilon} \right) \right)$

Warmness - how close is  $\mu_0$  to  $\pi$ ?

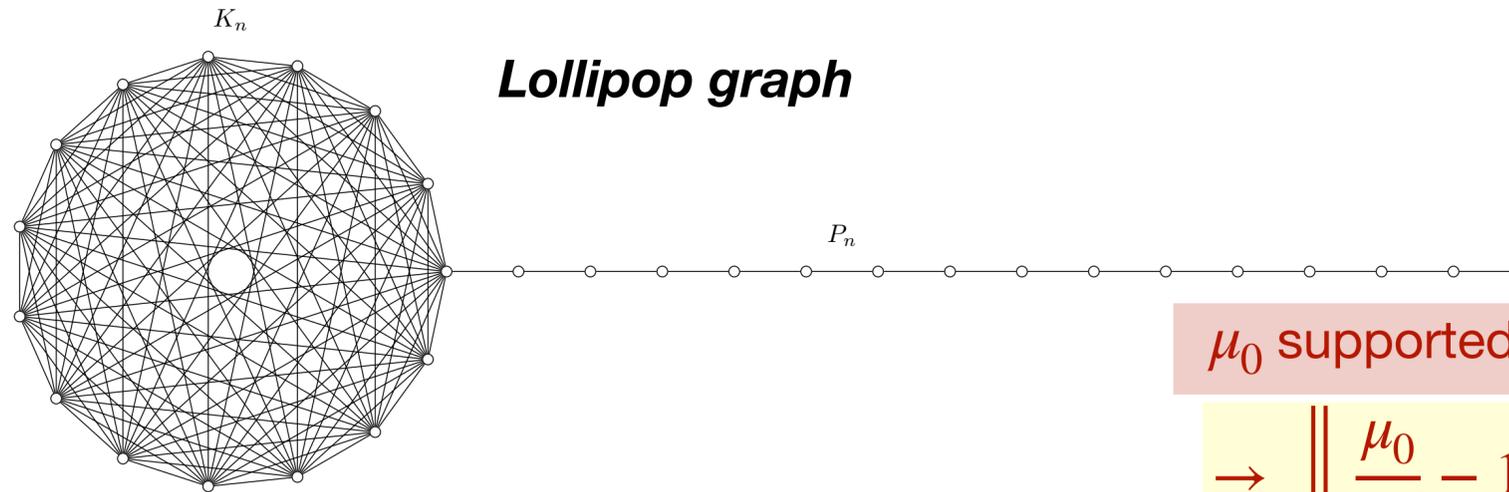
Usually measured by  $\left\| \frac{\mu_0}{\pi} - 1 \right\|_{2,\pi}^2 = \sum_{x \in \Omega} \pi(x) \left( \frac{\mu_0(x)}{\pi(x)} - 1 \right)^2$

$\mu_0$  is a warm start  $\longrightarrow$   $\mu_0$  close to  $\pi$   $\longrightarrow$  **Faster mixing**

# Warm starts advantage: A simple example

$\mu_0$  supported only on vertices of  $K_n$

$$\rightarrow \left\| \frac{\mu_0}{\pi} - 1 \right\|_{2,\pi} = \Theta(1)$$



$\mu_0$  supported only on vertices of  $P_n$

$$\rightarrow \left\| \frac{\mu_0}{\pi} - 1 \right\|_{2,\pi} = \Theta(\sqrt{n})$$

Directly implies mixing time improvement of  $O(\ln(n))$  \*

Usually in MC of interest,  $|\Omega| = 2^n$ ,  $\mu_0$  could be a cold start with  $\left\| \frac{\mu_0}{\pi} - 1 \right\|_{2,\pi} = \Theta(2^{-n})$

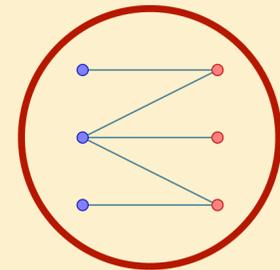
Thus, warm starts which are  $\Theta(1)$  close to  $\pi$  lead to  $poly(n)$  improvement in mixing time!

# Warm starts advantage: Permanent of 0-1 matrix

$$\text{perm}(A_{n \times n}) = \sum_{\sigma \in S_n} \prod_{i=1}^n A_{i, \sigma(i)}$$

When  $A$  is binary matrix,  $\text{perm}(A) = \mathcal{M} = \#\text{Perfect matchings of } G(A)$

$$\text{perm} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \#\text{Perfect matchings of}$$



But counting the number of perfect matchings **exactly** is  $\#P$ -Hard

**Monte Carlo**

But can we approximately count it?



Yes, if we can approximately sample a uniformly random perfect matching!

$$(1 - \epsilon) \cdot \text{perm}(A) \leq \widetilde{\text{perm}} \leq (1 + \epsilon) \cdot \text{perm}(A)$$

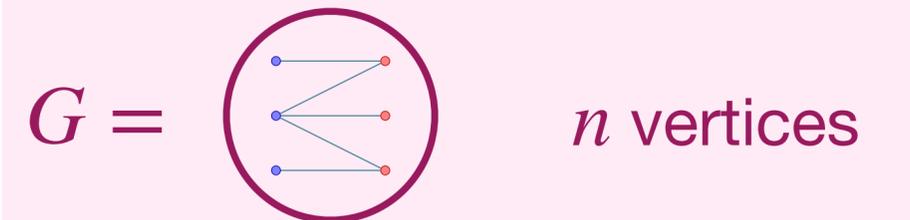
How to sample a uniformly random perfect matching?

A Markov chain which mixes fast and whose stationary distribution allows us to sample perfect matchings uniformly at random!

**Markov chain**

# Warm starts advantage: Permanent of 0-1 matrix

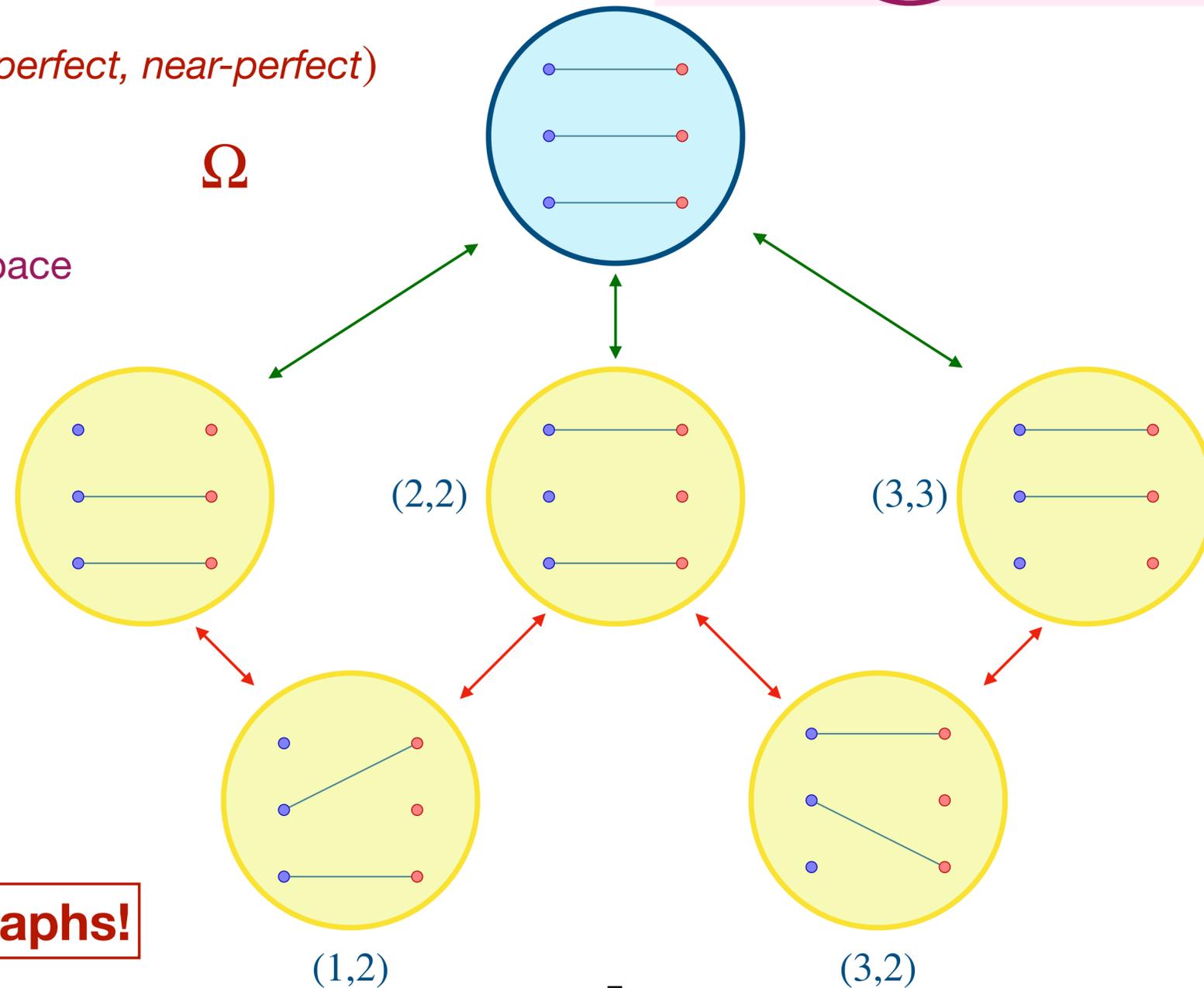
**Jerrum-Sinclair-Vigoda (JSV) chain:**



Stationary distribution  $\pi_G \sim \text{Unif}(\text{perfect, near-perfect})$

Usually  $2^{\Omega(n)}$  vertices in state space

Hole pattern  
(1,1)



Perfect matching

Addition/Deletions

Near-perfect matchings

Rotations

Near-perfect matchings

But this mixes in time  $\text{poly}(n)$  only if

$$\frac{\mathcal{M}}{\mathcal{M}_{u,v}} \geq \frac{1}{\text{poly}(n)}$$

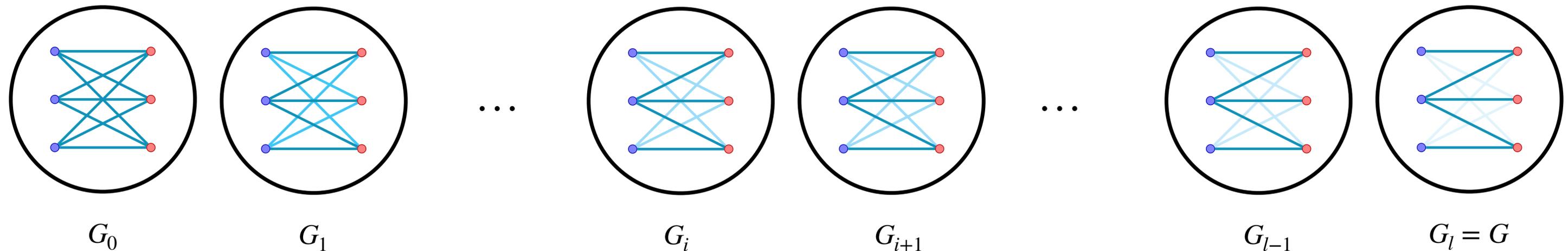
**Not true for all bipartite graphs!**

# Warm starts advantage: Permanent of 0-1 matrix

Work around: assign **weights** to every matching to reflect the **ratios** so that all matchings in the end have same weight

But knowing the ratio is as hard as finding the permanent!

Solution: Sequence of graphs  $G_i$  such that the ratios changes slowly from  $G_i$  to  $G_{i+1}$  and  $G_0$  is the complete bipartite graph



Slowly reduce activities of non-existing edges

$$\frac{\text{weights}_i}{\text{weights}_{i+1}} \text{ is small}$$

$\pi_i$  is a warm start for  $\pi_{i+1}$

JSV mixes in  $poly(n)$  steps on  $G_i, \forall i$

Thick edges - more activity

Thin edges - less activity

$\pi_l = \pi_G$  can be prepared in  $poly(n)$  time\*

Jerrum, Sinclair, Vigoda 2001

# Generic framework of using warm starts

## Simulated Annealing

Sequence of Markov chains  $\mathcal{M}(\beta)$  parameterized by  $\beta$

It is often easy to sample from  $\pi_{\beta_1}$  but difficult to sample from  $\pi_{\beta_n}$

$$\beta_1 > \beta_2 > \dots > \beta_i > \dots > \beta_{n-1} > \beta_n$$

$$\pi_{\beta_1} \xrightarrow{\beta_2} \pi_{\beta_2} \xrightarrow{\beta_3} \dots \xrightarrow{\beta_i} \pi_{\beta_i} \xrightarrow{\beta_{i+1}} \dots \xrightarrow{\beta_{l-1}} \pi_{\beta_{l-1}} \xrightarrow{\beta_n} \pi_{\beta_l}$$

*Inspired by physics  
- stable  
configuration of a  
system at low  
temperature*

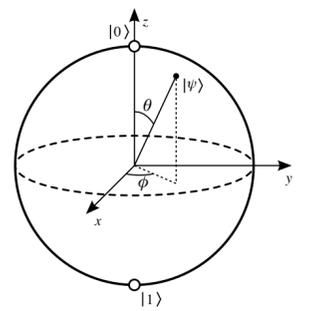
Sample from  $\pi_{\beta_i}$  acts as a **warm start** for  $\pi_{\beta_{i+1}}$

Hence, we can sample from the target  $\pi_{\beta_l}$  quickly!

*Simulated annealing of TSP solutions*

$\beta =$  length of tour

# Quantum Computation



**Classical**

**Quantum**

Probability vectors  $v \in \mathbb{R}^n$   
- normalized by  $L_1$ -norm

Eigenvalues of both  $P$  and  $U$  are related

Quantum states - vectors  $|v\rangle \in \mathbb{C}^n$   
- normalized by  $L_2$ -norm

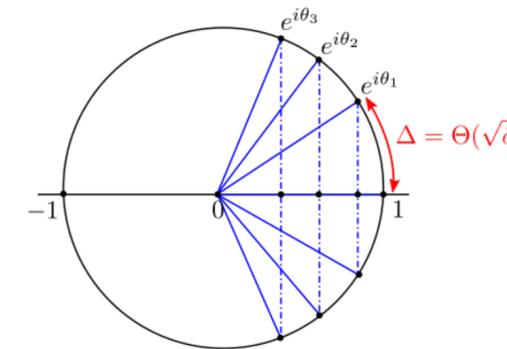
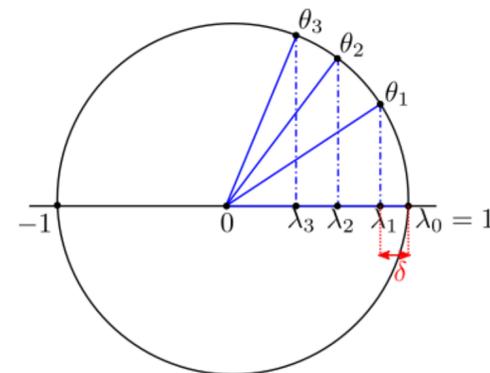
Only stochastic operators

Only unitary operators

Transition matrix  $P$

$\Pi = \text{diag}(\pi(1), \pi(2), \dots, \pi(n))$

Unitary  $U = \begin{pmatrix} \Pi^{1/2} P \Pi^{-1/2} & * \\ * & * \end{pmatrix}$



$\pi$  is top eigenvector of  $P$

$$\theta_i = \cos^{-1}(\lambda_i)$$

$$U|\pi\rangle = |\pi\rangle$$

**Quadratic speedup w.r.t spectral gap!**

**Theorem (informal):**  $quantum - t_{mix}(\epsilon) = O\left(\frac{1}{\sqrt{1-\lambda_2}} \cdot \left(\frac{\text{warmness}(\mu_0 \rightarrow \pi)}{\epsilon}\right)\right)$

**Warmness is more important for quantum!**

Usually measured by  $F(\mu_0, \pi) = \sum_{x \in \Omega} \sqrt{\mu_0(x)\pi(x)}$   $L_2$  - warmness

# Polynomial speedup for permanent

Current best classical

**Theorem (Bezáková et al. 2007):** There is a randomized algorithm that approximates the permanent of a binary  $n \times n$  matrix in time  $\tilde{O}(n^7)$ .

**Theorem (informal):**  $quantum - t_{mix}(\epsilon) = O\left(\frac{1}{\sqrt{1-\lambda_2}} \cdot \left(\frac{\text{warmness}(\mu_0 \rightarrow \pi)}{\epsilon}\right)\right)$



$O(n^2)$  speedup!

**Our theorem:** There is a quantum algorithm that approximates the permanent of a binary  $n \times n$  matrix in time  $\tilde{O}(n^5)$ .

**Techniques used:** Gap amplification, Quantum walks, Multidimensional Amplitude estimation, Non-destructive phase estimation and others

# Is the speedup polynomial always?

$H =$  adjacency matrix  $A$  of  $G$  with  $2^n$  vertices

$v$  - top eigenvector of  $H$  normalized by  $L_2$ -norm

**Quantum task:**

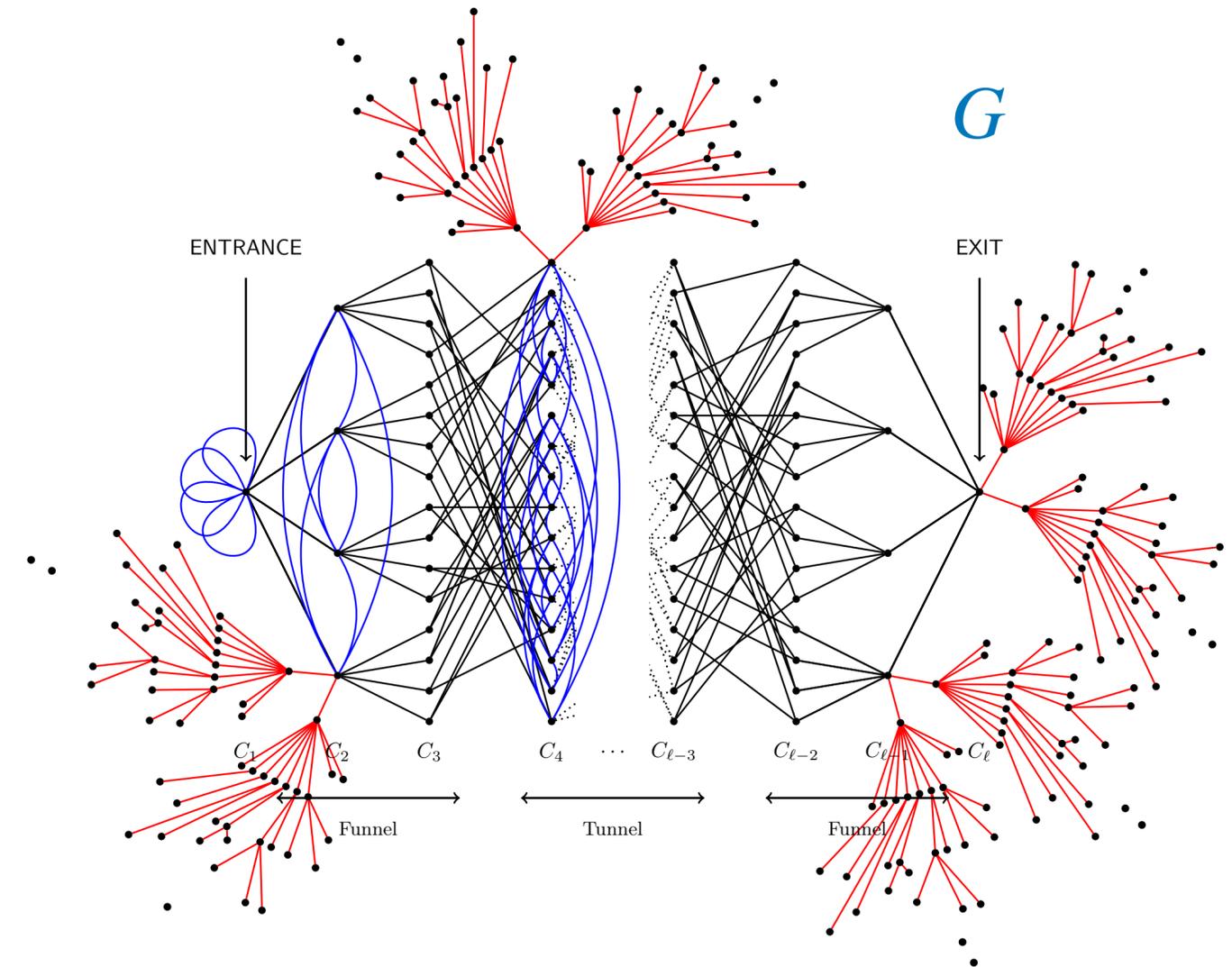
Re-align a  $L_2$ -warm start to  $v$

$O(\text{poly}(n))$  time

**Classical task:**

Re-align a  $L_1$ -warm start to  $v$

$\Omega\left(2^{n^{o(1)}}\right)$  time



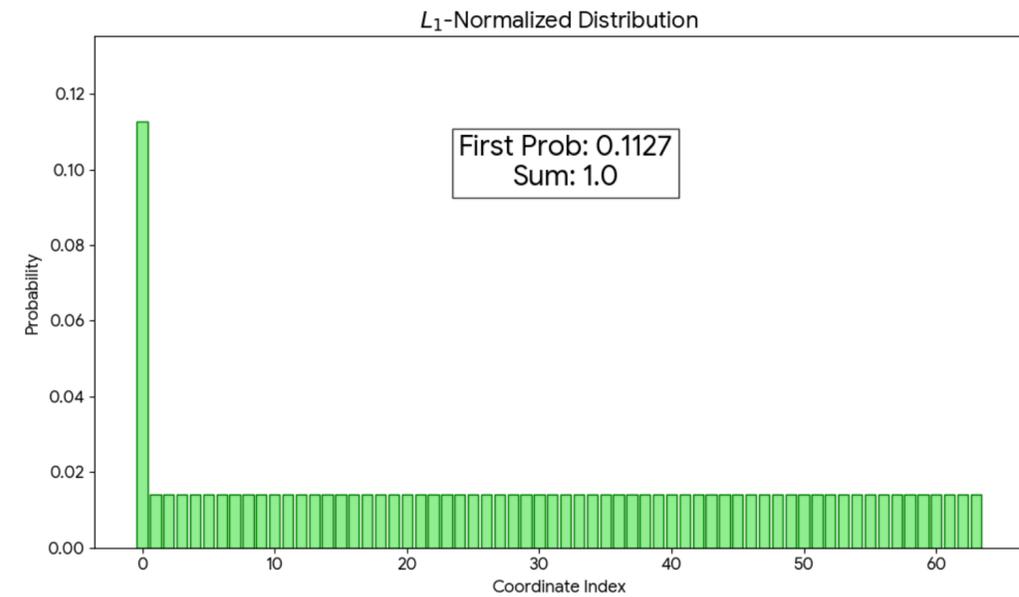
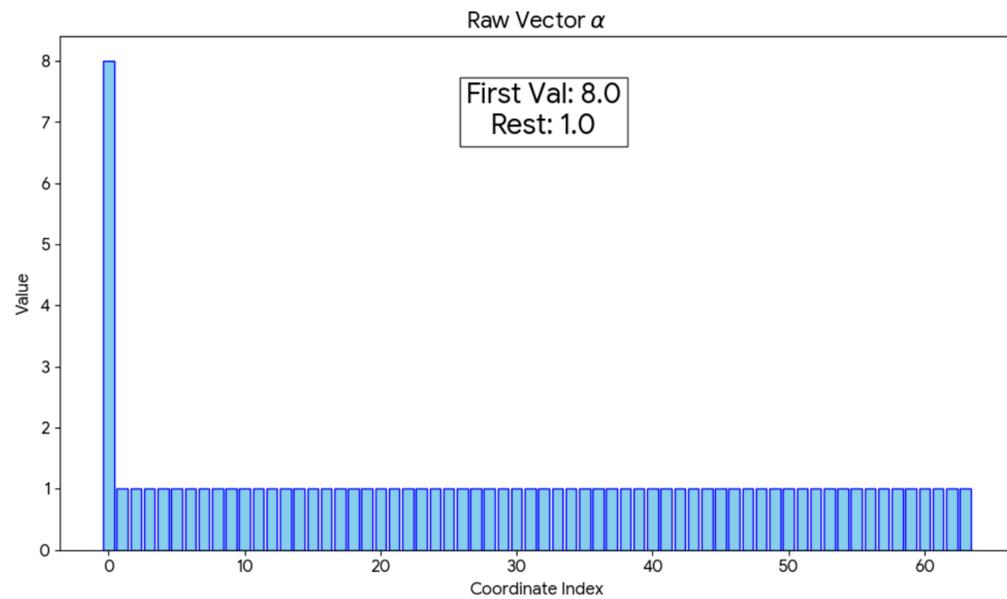
Gilyén, Hastings, Vazirani STOC 2021

Super-polynomial separation between Quantum and classical!

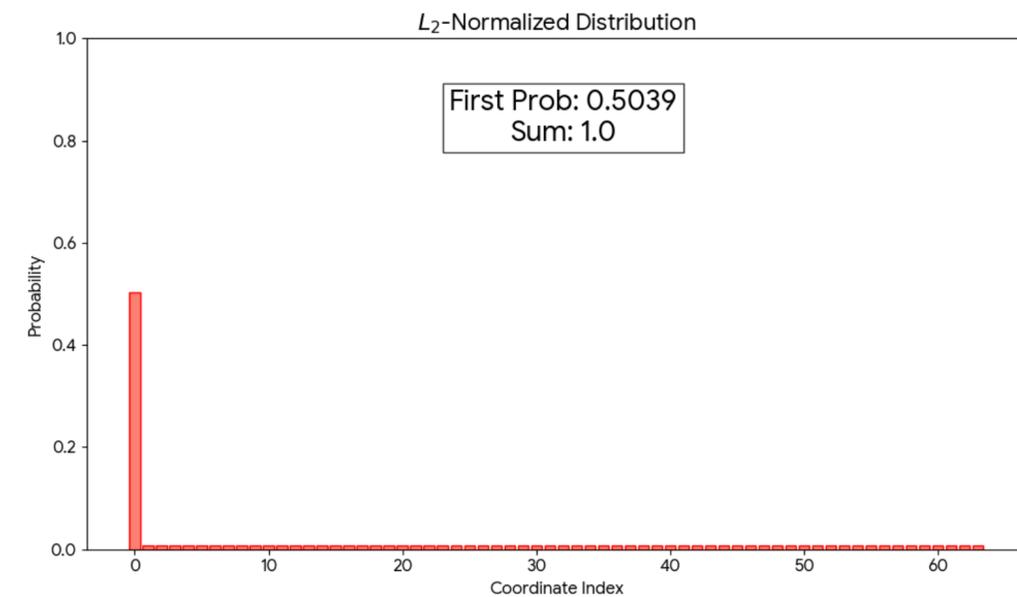
# $L_1$ vs $L_2$ normalization

$L_1$ -normalized distribution - almost uniform, with the first coordinate having weight  $1/(2^{n/2} + 1 - 2^{-n/2}) \approx 1/2^n$ .

Let  $\alpha = (2^{n/2}, 1, \dots, 1)$  of size  $2^n$



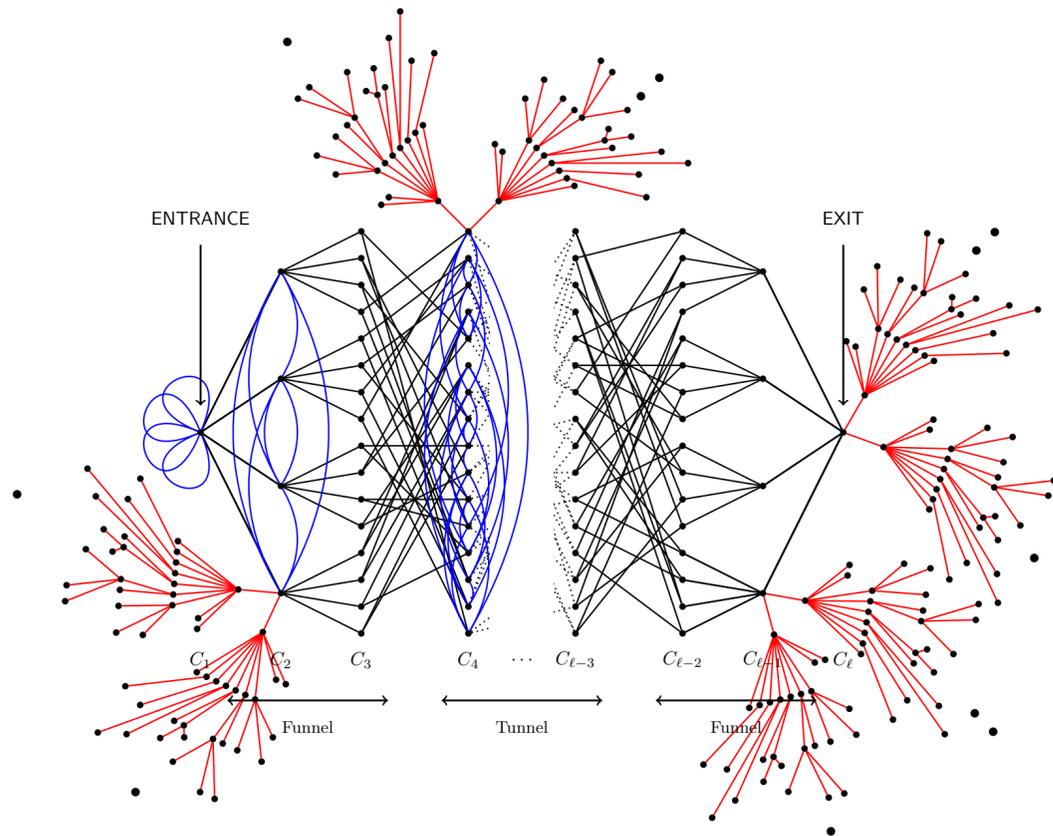
$L_2$ -normalized distribution - heavily biased toward its first coordinate, whose weight is  $1/(2 - 2^{-n}) \approx 1/2$ .



# Why is it not easy for classical always?

$v$  - top eigenvector of  $H$   
normalized by  $L_2$ -norm

$H =$  adjacency matrix  $A$  of  $G$  with  $2^n$  vertices



Same only when graph is regular!

Classical random walk on  $G$  - samples  
from the top eigenvector of  
 $P = \text{diag}(d(1), d(2), \dots, d(2^n))^{-1}A$

**This is different from  $v$ !**

In the top eigenvector of a tree,  $L_1$   
weight is more on leaves whereas  $L_2$   
weight is heavily concentrated on the  
root

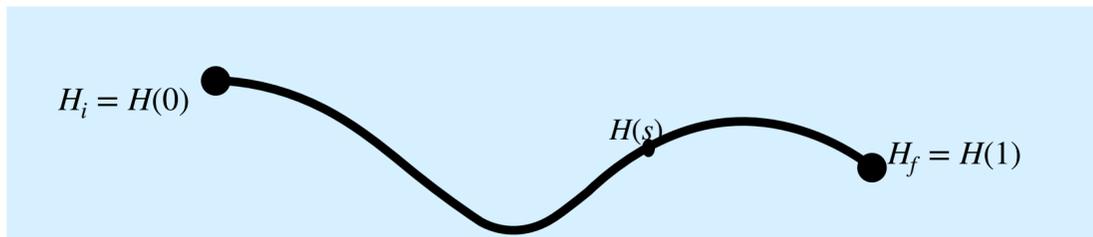


**Classical random walks** tend to move to the **trees**

**Quantum algorithm** mostly stays in the original graph!

# Quantum Adiabatic Algorithm

## Classical Optimization problems



$$h : \{0,1\}^n \rightarrow \mathbb{R}$$

Ex: MAX-SAT  
 $(x_0 \vee x_1) \wedge (x_0 \vee \neg x_1) \wedge (x_3 \vee x_2) \wedge (x_2 \vee \neg x_1)$

$$h(x) = \# \text{ Unsatisfied clauses}$$

$$H_f = \begin{bmatrix} h(0) & & & & \\ & h(1) & & & \\ & & \ddots & & \\ & & & h(i) & \\ & & & & \ddots \\ & & & & & h(2^n) \end{bmatrix} \quad H_f = \sum_{z \in \{0,1\}^n} h(z) |z\rangle\langle z|$$

$$H_i = - \sum_i X_i$$

$$|\psi(0)\rangle = |+\rangle^{\otimes n} = \frac{1}{2^n} \sum_{z \in \{0,1\}^n} |z\rangle$$

*Adiabatic evolution*

$$H_f = \sum_{z \in \{0,1\}^n} h(z) |z\rangle\langle z|$$

$$|\psi(1)\rangle \rightarrow \text{MAX-SAT solution}$$

*But doesn't mean all classical optimization problems can be solved very fast....*

### Approximately Adiabatic Evolution

The total simulation time required for adiabaticity depends on the spectrum of the Hamiltonian

$$\text{Gap: } \delta(s) = \lambda_1(s) - \lambda_0(s) \text{ and } \delta = \min_{s \in [0,1]} \delta(s)$$

Theorem [Aharonov et al]. Let the spectral gap of  $H(s)$  be at least  $\Delta$  for all  $0 \leq s \leq 1$ . For any  $\epsilon > 0$ , there exists  $T = \text{poly} \left( \frac{1}{\Delta}, \frac{1}{\epsilon}, \|H_i\| + \|H_f\| \right)$  such that  $|\psi(T)\rangle$  will be  $\epsilon$ -close to  $|\lambda_0(1)\rangle$ .

# Open questions

- *What is the quantum speedup to **approximately** sample a uniformly random spanning tree?*
- *Are there some **natural distributions** for which sampling/preparing the  $L_2$ -**normalized** vector exhibits **super polynomial separation** between quantum and classical randomized algorithms?*
- *Can we obtain an **exponential** separation?*
- ***Lower bounds** for classical **MCMC** and quantum analogs are rare, can we obtain classical and quantum lower bounds for approximating the **0-1 permanent**?*

**Thank you!**