# LECTURE 4
# ABOUT CLASSICAL AND QUANTUM RANDOM WALKS

Advanced Quantum Information and Computing

Thomas Debris-Alazard

Inria, École Polytechnique

# THE OBJECTIVE OF THE DAY

- To present the usefulness of Markov chains $\Big($via the birthday paradox$\Big)$
- To give you an introduction to the theory of Markov chains
- Ultimately to show you how quantum computing can increase the “efficiency” of Markov chains

# COURSE OUTLINE

1. Usefulness of Random Walks
2. Markov Chains
3. Random Walks on Graphs
4. Quantum Random Walks
5. Application: Finding Collisions

# YOU SAID RANDOM WALKS?

# COMPLEX NETWORK ANALYSIS

**Complex Networks:**

- Web network:
  - $\geq 10^{10}$ pages
  - average number 38 hyperlinks per page
- Bluesky:
  - $\approx 10^{7}$ users
  - a user follows about 100 other users

$\longrightarrow$ These networks involve a complex analysis:

unknown and changing topology, crawling the entire network is slow

Naive question: how to count the number of nodes in these networks?

$\big($for instance with sub-linear complexity in the number of nodes$\big)$

# COUNTING THE NUMBER OF NODES

**Assumption:**

It is possible to efficiently sample uniformly among a set

**Claim:**

Under the above assumption, we can estimate **the size** $n$ of the set with only $O\left(\sqrt{n}\right)$ samples!

**Idea: birthday paradox**

Assume that a list of persons are ready to enter a room one by one. Each person is let in and declares her birthday. How many people have to enter before two of them have the same birthday? A birthday collision is likely to happen at time $t \approx \sqrt{365} \approx 24$

$\longrightarrow$ By sampling elements uniformly at random in a set of size $n$ we expect $\sqrt{n}$ drawing for a collision to happen: a collision happen at time $t$, we deduce that the set has size $\sqrt{t}$

# BIRTHDAY PARADOX: RIGOROUS ANALYSIS

## Estimating the size of a set:

Pick elements uniformly at random in the set of size $n$ and mark them. Stop at the moment you fin a marked elements, *i.e.*, you found a collision

Let $T$ be the time of the first collision and $\mathsf{X}_i$ be the picked element during the drawing. We have

$$T = \inf_{i \geq 1} \left\{ \mathsf{X}_i \in \{\mathsf{X}_1, \ldots, \mathsf{X}_{i-1}\} \right\}$$

Notice that,

$$\mathbb{P}\left(T > i\right) = \frac{n(n-1)\ldots(n-i+1)}{n^i} = \frac{i!}{n!} \frac{d^i}{dz^i} \left[1+z\right]^n (0) = i! \frac{d^i}{dz^i} \left[1 + \frac{z}{n}\right]^n (0)$$

Now we have,

$$\mathbb{E}\left(T\right) = \sum_{i=0}^{+\infty} \mathbb{P}\left(T > i\right) = 1 + \sum_{i=1}^{+\infty} i! \cdot \frac{d^i}{dz^i} \left[1 + \frac{z}{n}\right]^n (0)$$

Using now that $i! = \int_0^{+\infty} t^i \mathrm{e}^{-t} dt$ $\Big($Gamma function$\Big)$ and $(1 + \frac{z}{n})^n = \sum_{i \geq 0} \frac{d^i}{dz^i} \left[1 + \frac{z}{n}\right]^n (0) \cdot z^i$,

$$\mathbb{E}\left(T\right) = 1 + \int_0^{+\infty} \left(1 + \frac{t}{n}\right)^n \mathrm{e}^{-t} dt = \sqrt{\frac{\pi n}{2}} + \frac{2}{3} + O\left(\frac{1}{\sqrt{n}}\right)$$

Our estimator $\widehat{n}$ of the size is therefore given by

$$\widehat{n} = \frac{2 \cdot (T - 2/3)^2}{\pi}$$

# UNIFORM SAMPLING?

But how to efficiently sample uniformly at random an element in a set?

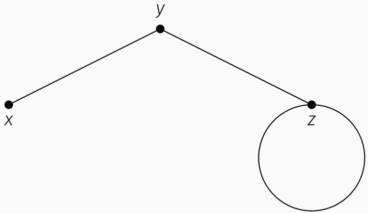
$\Big($remember that we don't even know the size of the set$\Big)$

# FUNDAMENTAL IDEA: RANDOM WALKS

**Fundamental idea: random walks**

Start from a point in the set and walk a certain amount of time to a “neighbour” with some probability

$\longrightarrow$ According to the “structure” of the walk $\Big($how are distributed neighbours and how we walk from one to the other$\Big)$ we can efficiently, *e.g.* with sub-linear walks, sample a uniform point!



# MARKOV CHAINS

# STOCHASTIC MATRIX

**Stochastic matrix:**

Given a finite set $\mathcal{X}$, a matrix $\mathbf{P} = (p(x,y))_{x,y\in\mathcal{X}}$ is said to be stochastic if

- $p(x,y) \geq 0$ for all $x, y \in \mathcal{X}$
- $\sum_{y\in\mathcal{X}} p(x,y) = 1$

**Fundamental fact**

If $\mathbf{x} = (q(x))_{x\in\mathcal{X}}$ is a distribution[a] and $\mathbf{P}$ is a stochastic matrix. Then, $\mathbf{x}^\top \mathbf{P}$ is a distribution

[a] for all $x \in \mathcal{X}$, $q(x) \geq 0$ and $\sum_{x\in\mathcal{X}} q(x) = 1$

- Let $(r(y))_{y\in\mathcal{X}}$ be the distribution defined as $\mathbf{x}^\top \mathbf{P}$. We have,

$$r(y) = \sum_{x\in\mathcal{X}} q(x)p(x,y)$$

- $r$ defines the distribution: pick $x$ according to $q$ and then pick $y$ with probability $p(x,y)$

# MARKOV CHAINS

*Markov chains give a rule to walk from one point to the other independently of the path we followed in the past*

## Markov chain:

Let $\mathcal{X}$ be a finite set, $(q(x))_{x\in\mathcal{X}}$ be a distribution and $\mathsf{P} = (p(x,y))_{x,y\in\mathcal{X}}$ be a stochastic matrix.

A $\Big($homogenous$\Big)$ Markov chain with state space $\mathcal{X}$, initial distribution $q$ and transition matrix $\mathsf{P}$ is a sequence of random variables $\mathsf{X}_0, \ldots, \mathsf{X}_t, \ldots$ such that

$$\mathbb{P}\left(\mathsf{X}_0 = x_0\right) = q(x_0) \quad \text{and} \quad \mathbb{P}\left(\mathsf{X}_{t+1} = x_{t+1} \mid \mathsf{X}_t = x_t, \ldots, \mathsf{X}_0 = x_0\right) = p(x_t, x_{t+1})$$

for all $t \in \mathbb{N}$ and $x_0, \ldots, x_{t+1} \in \mathcal{X}$ such that $\mathbb{P}\left(\mathsf{X}_0 = x_0, \ldots, \mathsf{X}_t = x_t\right) > 0$

## Remark:

The homogenous term refers to the fact that for each $t$ the transition matrix is the same

# WALK ACCORDING TO THE MARKOV CHAIN

**Proposition:**

Given a Markov chain $(\mathsf{X}_t)_t$ with initial distribution $(q(x))_{x\in\mathcal{X}}$, transition matrix $\mathbf{P} = (p(x,y))_{x,y\in\mathcal{X}}$,

$$\mathbb{P}\left(\mathsf{X}_t = x\right) = q^{(t)}(x) \quad \text{where} \quad \left(q^{(t)}(x)\right)_{x\in\mathcal{X}} \overset{\text{def}}{=} \left(q(x)\right)_{x\in\mathcal{X}}^{\top} \mathbf{P}^{t}$$

and,

$$\mathbb{P}\left(\mathsf{X}_{t+1} = x_{t+1} \mid \mathsf{X}_t = x_t\right) = p(x_t, x_{t+1})$$

$\Big(p(x,y)$: rule for moving from $x$ to $y$, we read from left to right$\Big)$

**Proof:**

Exercise!

**Notation:**

Given $\mathbf{P} = \left(p(x,y)\right)_{x,y\in\mathcal{X}}$, we denote $\mathbf{P}^{t} = \left(p^{(t)}(x,y)\right)_{x,y\in\mathcal{X}}$

# STATIONARY DISTRIBUTION

Starting from the distribution $\mathbf{x} = (q(x))_{x \in \mathcal{X}}$ and after $t$ walks we are distributed as $\mathbf{x}^\top \mathbf{P}^t$

**Stationary distribution:**

Let $\mathbf{P}$ be a stochastic matrix. A stationary distribution for $\mathbf{P}$ is a distribution $\pi$ such that

$$\pi^\top = \pi^\top \mathbf{P}$$

$\longrightarrow$ Starting from the stationary distribution and applying the walk keeps invariant the distribution!

# ERGODICITY AND CONVERGENCE TO EQUILIBRIUM

$$\left(\text{given } \mathbf{P} = \left(p(x,y)\right)_{x,y\in\mathcal{X}}, \text{ we denote } \mathbf{P}^{t} = \left(p^{(t)}(x,y)\right)_{x,y\in\mathcal{X}}\right)$$

**Ergodicity:**

A stochastic matrix **P** is said ergodic if there exists $t_0 \in \mathbb{N}$ such that

$$\forall x, y \in \mathcal{X}, \quad p^{(t_0)}(x,y) > 0$$

**Theorem:**

A stochastic matrix **P** is ergodic if and only if there exists a strict probability distribution[a] $\pi$ on $\mathcal{X}$ such that

$$\forall x, y \in \mathcal{X}, \quad p^{(t)}(x,y) \xrightarrow[t\to+\infty]{} \pi(y)$$

Furthermore, when **P** is ergodic, the above distribution $\pi$ is the unique stationary distribution for **P**

[a] $\pi(x) > 0$ for any $x \in \mathcal{X}$

# PROOF(I)

**Proof:**

Suppose that **P** is ergodic and $\varepsilon \overset{\text{def}}{=} \min\limits_{x,y\in\mathcal{X}} p^{(t_0)}(x,y) \in (0,1)$,

$$M^{(t)}(y) \overset{\text{def}}{=} \max_{x\in\mathcal{X}} p^{(t)}(x,y) \quad m^{(t)}(y) \overset{\text{def}}{=} \min_{x\in\mathcal{X}} p^{(t)}(x,y)$$

We have,

$$m^{(t)}(x,y) \leq \sum_z p(x,z)m^{(t)}(x,y) \leq \sum_z p(x,z)p^{(t)}(z,y) = p^{(t+1)}(x,y) \leq \sum_z p(x,z)M^{(t)}(y) = M^{(t)}(y)$$

We deduce that $t \mapsto M^{(t)}(x,y)$ and $t \mapsto m^{(t)}(x,y)$ are decreasing and increasing. Therefore they convergence as belonging to $(0,1)$. Call $\pi_1(y)$ and $\pi_2(y)$ their limits. For any $r \geq 0$ we have:

$$\begin{aligned}
p^{(t_0+r)}(x,y) &= \sum_z p^{(t_0)}(x,z)p^{(r)}(z,y) \\
&= \sum_z \left(p^{(t_0)}(x,z) - \varepsilon p^{(r)}(y,z)\right) p^{(r)}(z,y) + \varepsilon \cdot \sum_z p^r(y,z)p^{(r)}(z,y) \\
&\geq m^{(r)}(y) \sum_z \left(p^{(t_0)}(x,z) - \varepsilon p^{(r)}(y,z)\right) + \varepsilon \cdot p^{(2r)}(y,y) \\
&= (1-\varepsilon)\cdot m^{(r)}(y) + \varepsilon \cdot p^{(2r)}(y,y) \geq (1-\varepsilon)m^{(r)}(y) + \varepsilon \cdot m^{(2r)}(y,y)
\end{aligned}$$

where the inequality follows from the fact that $\left(\text{as } \varepsilon \geq p^{(t_0)}(x,z)\right)$,

$$p^{(t_0)}(x,z) - \varepsilon p^{(r)}(y,z) \geq p^{(t_0)}(x,z)\left(1 - p^{(r)}(y,z)\right) \geq 0$$

# PROOF(II)

**Proof:**

Similarly $M^{(n_0+r)}(y) \leq (1-\varepsilon)M^{(r)}(y) + \varepsilon \cdot M^{(2r)}(y,y)$. We deduce that for any $k$,

$$M^{(kn_0+r)}(y) - m^{(kn_0+r)}(y) \leq (1-\varepsilon)^k \left(M^{(r)}(y) - m^{(r)}(y)\right) \xrightarrow[k\to+\infty]{} 0$$

Therefore $\pi(y) \overset{\text{def}}{=} \pi_1(y) = \pi_2(y)$ and from above with the fact that $M^{(t)}$ and $m^{(t)}$ are decreasing and increasing, for $t = kn_0 + r$ where $0 \leq r \leq n_0$,

$$\left|p^{(t)}(x,y) - \pi(y)\right| \leq M^{(t)}(y) - m^{(t)}(y) \leq (1-\varepsilon)^{n/n_0}$$

and therefore $p^{(t)}(x,y) \xrightarrow[t\to+\infty]{} \pi(y)$. Furthermore,

$$p^{(t+1)}(x,y) \sum_z p^{(t)}(x,z)p(z,y)$$

we get with $t \to +\infty$,

$$\pi(y) = \sum_z \pi(z)p(z,y)$$

which shows that $\pi$ is a stationary distribution $\Big($it is a distribution as $\sum_z p^{(t)}(x,z) = 1$ and $p^{(t)}(x,z) \geq 0\Big)$. It is strict as $m^{(t)}(y) \geq \varepsilon > 0$

Conversely, suppose that $p^{(t)}(x,y) \xrightarrow[t\to+\infty]{} \pi(y) > 0$. We deduce easily that $\mathbf{P}$ is ergodic $\Big($a finite number of $y\Big)$. To prove uniqueness let $\pi'$ be another stationary distribution,

$$\pi'(y) = \sum_x \pi'(x)p^{(t)}(x,y) \xrightarrow[t\to+\infty]{} \sum_x \pi'(x)\pi(y) = \pi(y)$$

# COUNTING WITH MARKOV CHAINS

How to enumerate the size of a set $\mathcal{X}$ with Markov chains?

- Choose a transition matrix $\mathbf{P}$ $\Big($walking rules over $\mathcal{X}\Big)$ ergodic such that its stationary distribution $\pi_{\text{unif}}$ is the uniform distribution $\Big($no reason to be true$\Big)$
- Choose any initial distribution $q$, *e.g.* start on a fix point $x_0$ by choosing $q(x) = \begin{cases} 1 \text{ if } x = x_0 \\ 0 \text{ if } x \neq x_0 \end{cases}$

Start from any point according to the initial distribution $q$ and apply $t$ $\Big($large enough$\Big)$ walks according to $\mathbf{P}$

As $p^{(t)}(x, y) \xrightarrow[t \to +\infty]{} \pi_{\text{unif}}(y)$: we will sample a point on $\mathcal{X}$ according to $\pi_{\text{unif}}$ and then we apply the reasoning of the beginning of the lecture

But how to measure the efficiency of Markov chain to sample uniformly?

# SPEED OF CONVERGENCE

How many walks are necessary to be close enough to the stationary distribution?

$\longrightarrow$ We want to estimate the smallest $t$ such that $p^{(t)}(x, y) \approx \pi(y)$

Issue:

Recall that $p^{(t)}(x, y)$ is defined as $\mathbf{P}^t = \left(p^{(t)}(x, y)\right)_{x,y \in \mathcal{X}}$

$\longrightarrow$ Coefficients $p^{(t)}(x, y)$ involve a convoluted analysis

How could we reasonably overcome this issue?

# SPEED OF CONVERGENCE

How many walks are necessary to be close enough to the stationary distribution?

$\longrightarrow$ We want to estimate the smallest $t$ such that $p^{(t)}(x, y) \approx \pi(y)$

**Issue:**

Recall that $p^{(t)}(x, y)$ is defined as $\mathbf{P}^t = \left(p^{(t)}(x, y)\right)_{x,y \in \mathcal{X}}$

$\longrightarrow$ Coefficients $p^{(t)}(x, y)$ involve a convoluted analysis

How could we reasonably overcome this issue?

$\longrightarrow$ Spectral analysis: diagonalize the transition matrix $\mathbf{P}$

# SPEED OF CONVERGENCE AND SPECTRAL ANALYSIS (I)

*First of all we need to interpret transition matrices as linear operators*

Let $L(\mathcal{X}) = \{f : \mathcal{X} \to \mathbb{C}\}$ and define a linear operator $\mathbf{A} = (a(x,y))_{x,y\in\mathcal{X}} : L(\mathcal{X}) \to L(\mathcal{X})$ as,

$$\forall f \in L(\mathcal{X}), \quad (\mathbf{A}f) : x \in \mathcal{X} \longmapsto \sum_{y\in\mathcal{X}} f(y)a(x,y)$$

$\Big($ $\mathbf{A}f$ is nothing else than $\mathbf{A}\,(f(x))_{x\in\mathcal{X}}$, *i.e.*, columns space instead of the rows space$\Big)$

**Proposition:**

Given a transition matrix, 1 is always an eigenvalue of $\mathbf{P}$. Furthermore, if $\lambda$ is another eigenvalue we have $|\lambda| \leq 1$

**Proof:**

We clearly have $\mathbf{P1} = \mathbf{1}$ as $\mathbf{P}$ is stochastic. Furthermore, let $f \neq 0$ such that $\mathbf{P}f = \lambda f$. Let $x_0 \in \mathcal{X}$ such that $|f(y)| \leq |f(x_0)|$ for all $y \in \mathcal{X}$. Then,

$$|\lambda f(x_0)| = \left|\sum_{y\in\mathcal{X}} f(y)p(x_0,y)\right| \leq \sum_{y\in\mathcal{X}} |f(y)|p(x_0,y) \leq |f(x_0)|$$

as $\sum_{y\in\mathcal{X}} p(x_0,y) = 1$ and $|f(y)| \leq |f(x_0)|$

# SPEED OF CONVERGENCE AND SPECTRAL ANALYSIS (II)

Suppose that $\mathbf{P}$ is diagonalizable. Therefore $\mathbf{P}^\top$ is diagonalizable with the same eigenvalues

$$\begin{aligned}
\mathbf{x} &\overset{\text{def}}{=} \text{initial distribution} \\
\pi &\overset{\text{def}}{=} \text{eigenvector of } \mathbf{P}^\top \text{ corresponding to eigenvalue 1 with norm 1} \\
\mathbf{u}_i &\overset{\text{def}}{=} \text{eigenvector of } \mathbf{P}^\top \text{ corresponding to eigenvalue } \lambda_i \text{ with norm 1}
\end{aligned}$$

We have $\mathbf{x} = \alpha \cdot \pi + \sum_i \alpha_i \cdot \mathbf{u}_i$ and

$$\mathbf{x}^\top \mathbf{P}^t = \alpha \cdot \pi + \sum_i \lambda_i^t \alpha_i \mathbf{u}_i$$

Suppose now that $|\lambda_i| < 1$ for all $i$ $\left(\text{otherwise we only have that } |\lambda_i| \leq 1\right)$

We deduce that $\mathbf{x}^\top \mathbf{P}^t \xrightarrow[t \to +\infty]{} \alpha \cdot \pi$ as $|\lambda_{i_0}|^t$ where $|\lambda_{i_0}| = \max_i (|\lambda_i|)$

**Conclusion:**

Suppose $(i)$ $\mathbf{P}$ is diagonalizable, $(ii)$ 1 is eigenvalue with multiplicity one and $(iii)$ $-1$ is not an eigenvalue, then $\mathbf{x}^\top \mathbf{P}^t$ converges to equilibrium exponential fast as $|\lambda_{i_0}|^t$ where $|\lambda_{i_0}|$ be the second greatest absolute values of eigenvalues

# A PARTICULAR CASE: SYMMETRIC TRANSITION MATRICES

Suppose that $\mathbf{P}$ is symmetric

$\longrightarrow$ Eigenvectors of $\mathbf{P}$ and $\mathbf{P}^\top$ are the same. Furthermore $\mathbf{P}$ is diagonalizable

**In particular:**

$\pi = \frac{1}{\sharp\mathcal{X}}(1, \ldots, 1)$ $\big($uniform distribution$\big)$ is the eigenvector of norm 1 with eigenvalue 1 of $\mathbf{P}$ and $\mathbf{P}^\top$. We deduce that if 1 has multiplicity one and $-1$ is not an eigenvalue, then $\mathbf{x}^\top\mathbf{P}$ converges to the uniform distribution

*It would be a pity to only consider Markov chains with symmetric transition matrices for the spectral analysis. . .*

# REVERSIBLE MARKOV CHAINS

*A natural extension of symmetric transition matrices...*

**Reversible Markov chain:**

Given a stochastic matrix $\mathbf{P} = (p(x,y))_{x,y\in\mathcal{X}}$, we say that $\mathbf{P}$ is reversible if there exists a strict probability measure $\pi$ on $\mathcal{X}$ such that

$$\forall x, y \in \mathcal{X}, \quad \pi(x)p(x,y) = \pi(y)p(y,x)$$

We say that $\mathbf{P}$ and $\pi$ are in detailed balance

⟶ Starting from $x$ with probability $\pi(x)$ and walking to $y$ is the same than starting from $y$ with probability $\pi(y)$ and walking to $x$

**Fact:**

If $(\mathbf{P}, \pi)$ are in detailed balance, then $\pi$ is a stationary distribution $\left(\pi^\top \mathbf{P} = \pi^\top\right)$, *i.e.*,

$$\sum_{x\in\mathcal{X}} p(x,y)\pi(x) = \sum_{x\in\mathcal{X}} p(y,x)\pi(y) = \pi(y)$$

⟶ Very useful to define a Markov chain with prescribed stationary distribution $\pi$

## REVERSIBLE MARKOV CHAINS AND SPECTRAL ANALYSIS

Given $f_1, f_2 \in L(\mathcal{X})$,

$$\langle f_1, f_2 \rangle_\pi \overset{\text{def}}{=} \sum_{x \in \mathcal{X}} f_1(x) \overline{f_2(x)} \pi(x) \quad \Big(\text{scalar product}\Big)$$

**Proposition:**

$(\mathbf{P}, \pi)$ detailed balance if and only if $\mathbf{P}$ is self-adjoint with respect to the scalar product $\langle \cdot, \cdot \rangle_\pi$

**Consequence:**

If $(\mathbf{P}, \pi)$ are in detailed balance, then $\mathbf{P}$ is diagonalizable over the reals: it exists $\{\lambda_x : \ x \in \mathcal{X}\} \subseteq \mathbb{R}$ and $\mathbf{U} \in \mathbb{R}^{\sharp\mathcal{X} \times \sharp\mathcal{X}}$ such that

$$\begin{cases} \mathbf{PU} = \mathbf{U\Delta} \\ \mathbf{U}^\top \mathbf{DU} = \mathbf{Id} \end{cases} \qquad \text{where} \quad \mathbf{\Delta} = \mathbf{Diag}(\lambda_x)_{x \in \mathcal{X}} \quad \text{and} \quad \mathbf{D} = \mathbf{Dia}(\pi(x))_{x \in \mathcal{X}}$$

$\longrightarrow$ The columns of $\mathbf{U}$ are eigenvectors of $\mathbf{P}$ and $\mathbf{U}$ is unitary with respect to $\pi$

# PROOF

Proof:

- Suppose that **P** is self-adjoint with respect to $\langle\cdot,\cdot\rangle_\pi$ . Let $\delta_x$ be the Kronecker symbol. We have,

$$\pi(x)p(x,y)\langle \mathbf{P}\delta_y,\delta_x\rangle_\pi = \langle\delta_y,\mathbf{P}\delta_x\rangle_\pi = \pi(y)p(y,x)$$

- Suppose that $(\mathbf{P},\pi)$ are in detailed balance. Let $f_1, f_2 \in L(\mathcal{X})$, we have

$$\begin{aligned}\langle \mathbf{P}f_1, f_2\rangle_\pi &= \sum_{x\in\mathcal{X}}\left(\sum_{y\in\mathcal{X}} p(x,y)f_1(y)\right)\overline{f_2(x)}\pi(x)\\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} \pi(x)p(x,y)f_1(y)\overline{f_2(x)}\\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{X}} \pi(y)p(y,x)f_1(y)\overline{f_2(x)}\\ &= \langle f_1, \mathbf{P}f_2\rangle_\pi\end{aligned}$$

# DIAGONALIZABLE DECOMPOSITION

**Consequence:**

If $(\mathbf{P}, \pi)$ are in detailed balance, then $\mathbf{P}$ is diagonalizable over the reals: it exists $\{\lambda_x : \ x \in \mathcal{X}\} \subseteq \mathbb{R}$ and $\mathbf{U} \in \mathbb{R}^{\sharp\mathcal{X} \times \sharp\mathcal{X}}$ such that

$$\begin{cases} \mathbf{PU} = \mathbf{U\Delta} \\ \mathbf{U}^\top \mathbf{DU} = \mathbf{Id} \end{cases} \quad \text{where} \ \ \mathbf{\Delta} = \mathbf{Diag}(\lambda_x)_{x \in \mathcal{X}} \ \ \text{and} \ \ \mathbf{D} = \mathbf{Dia}(\pi(x))_{x \in \mathcal{X}}$$

$\longrightarrow$ The above decomposition is powerful to determine $p^{(t)}(x,y)$ as function of eigenvalues of $\mathbf{P}$

**Proposition:**

Using above notation, $\mathbf{P}^t = \mathbf{U\Delta}^t\mathbf{U}^\top\mathbf{D}$, *i.e.*, $p^{(t)}(x,y) = \pi(y) \cdot \sum_{z \in \mathcal{X}} \lambda_z^t \cdot u(x,z) \cdot u(y,z)$

**Proof:**

We just need to notice that $\mathbf{U}^\top\mathbf{D}$ is the inverse of $\mathbf{U}$

# SPECTRAL GAP AND CONVERGENCE TO EQUILIBRIUM

## Spectral gap:

Given a Markov chain with transition matrix **P** admitting $(\lambda_i)_i$ as eigenvalues, its spectral gap is defined as

$$\delta \overset{\text{def}}{=} 1 - \max_i \left( |\lambda_i| : \ 0 < |\lambda_i| < 1 \right)$$

## Theorem:

Suppose that $(\mathbf{P}, \pi)$ are in detailed balance, 1 is eigenvalue of **P** with multiplicity 1 and $-1$ is not an eigenvalue. Let $q^{(t)}$ be the distribution after $t$ walks $\mathbf{x}^\top \mathbf{P}^t$ where **x** be the initial distribution starting at any fixed point. Then,

$$\|q^{(t)} - \pi\|_{1/\pi} = \sqrt{\sum_{y \in \mathcal{X}} (q^{(t)}(y) - \pi(y))^2 \cdot \frac{1}{\pi(y)}} \leq (1-\delta)^t \cdot \sum_z u(x,z)^2 \xrightarrow[t \to +\infty]{} 0$$

where $\delta \in (0,1)$ is the spectral gap of **P**

In particular, the Markov chain with transition matrix **P** is ergodic and $\pi$ is its stationary distribution

## Fundamental consequence:

If anyone wants $\|\mathbf{x}^\top \mathbf{P}^t - \pi\| \leq \varepsilon$, then $t = \ln(\varepsilon)/\ln(1-\delta) \approx 1/\delta$ is enough: sub-exponential number of walks is enough to reach equilibrium $\Big($but be careful on the spectral gap$\Big)$

# PROOF

**Proof:**

Let $\lambda_z$ be the eigenvalues of $\mathbf{P}$ with associate eigenvectors $\mathbf{u}_z$. By assumption it exists a unique $z_0$ such that $|\lambda_z| < 1$ for $z \neq z_0$ and $\lambda_{z_0} = 1$. Therefore,

$$p^{(t)}(x,y) = \pi(y) \sum_{z \in \mathcal{X}} \lambda_z^t \cdot u(x,z) \cdot u(y,z) = \pi(y) \cdot u(x,z_0) \cdot u(y,z_0) + \pi(y) \sum_{z \neq z_0} \lambda_z^t \cdot u(x,z) \cdot u(y,z)$$

Furthermore, $u(\alpha, z_0) = 1$ for all $\alpha$ as $\mathbf{1}$ is the eigenvector of $\mathbf{P}$ associated to eigenvalue 1.

Therefore,

$$\left(p^{(t)}(x,y) - \pi(y)\right)^2 = \pi(y)^2 \sum_{z_1, z_2 \neq z_0} \lambda_{z_1}^t \cdot \lambda_{z_2}^t \cdot u(x,z_1) \cdot u(x,z_2) \cdot u(y,z_1) \cdot u(y,z_2)$$

Notice now that $\mathbf{x}^\top \mathbf{P}^t$ is distributed as $\left(p^{(t)}(x_0,y)\right)_{y \in \mathcal{X}}$. We deduce,

$$\begin{aligned} \|\mathbf{x}^\top \mathbf{P}^t - \pi\|_{1/\pi}^2 &= \sum_y \left(p^{(t)}(x_0,y) - \pi(y)\right)^2 \frac{1}{\pi(y)} \\ &= \sum_{z_1, z_2 \neq z_0} \lambda_{z_1}^t \cdot \lambda_{z_2}^t \cdot u(x,z_1) \cdot u(x,z_2) \cdot \sum_y u(y,z_1) \cdot u(y,z_2) \pi(y) \\ &= \sum_{z \neq z_0} u(x,z)^2 \cdot \lambda_z^{2t} \end{aligned}$$

where in the last equality we used that $\mathbf{U}^\top \mathbf{D} \mathbf{U} = \mathbf{Id}$. It concludes the proof

# A PARTICULAR CASE: SYMMETRIC TRANSITION MATRICES

Suppose that $\mathsf{P}$ is symmetric

$\longrightarrow$ $(\mathsf{P}, \pi_{\text{unif}})$ are in detailed balance where $\pi_{\text{unif}}$ is the uniform distribution

$$\tfrac{1}{n} \cdot p(x, y) = \tfrac{1}{n} \cdot p(y, x)$$

- If the spectral gap $\delta$ is $< 1$ $\Big($1 eigenvalue with multiplicity one and $-1$ is not an eigenvalue$\Big)$, applying $t$ walks is converging with $t$ to $\pi_{\text{unif}}$ which is the unique stationary distribution
- With our notation $\sqrt{\sharp\mathcal{X}} \cdot \mathsf{U}$ are unity matrices and

  $$\|q^{(t)} - \pi\|_2 \leq (1 - \delta)^t$$

  where $\|\cdot\|_2$ is the Euclidean norm over $L(\mathcal{X})$

# TO TAKE AWAY

- Markov chain: set of rules for walking from $x$ to $y$

  $\longrightarrow$ Rules are given by a transition matrix $\mathsf{P} = (p(x,y))_{x,y\in\mathcal{X}}$

- Invariant distribution: initial distribution such that applying walking rules does not change the distribution

- Rules for $t$ walks are given by $\underbrace{\mathsf{P}\ldots\mathsf{P}}_{t \text{ times}} = \left(p^{(t)}(x,y)\right)_{x,y\in\mathcal{X}}$

- Ergodic Markov chain: $\exists t > 0$ such that $p^{(t)}(x,y) > 0$

- Ergodic Markov chain and convergence to the equilibrium: applying $\Big($whatever is the starting distribution$\Big)$ $t$ walks for $t \to +\infty$ amounts to sample according to the unique invariant distribution

- To determine the speed of convergence: we restrict ourself to spectral analysis

  $\longrightarrow$ $\mathsf{P}$ is usually “symmetric” with respect to some distribution $\pi$: $\pi(x)p(x,y) = \pi(y)p(y,x)$

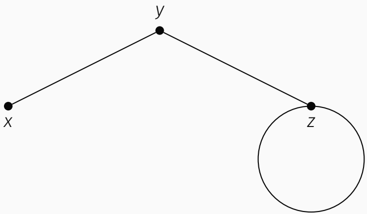
- The spectral gap $\delta$ $\Big($under the good hypothesis with 1 and $-1$ as eigenvalues$\Big)$ determines the speed of convergence as $\approx \frac{1}{\delta}$

# RANDOM WALKS ON GRAPHS

# VISUALISATION OF MARKOV CHAINS

*Some of you will no doubt be complaining that Markov chains are not very "visual"*

⟶ But they are! Transition rules define a graph, and reciprocally graphs define Markov chains. . .

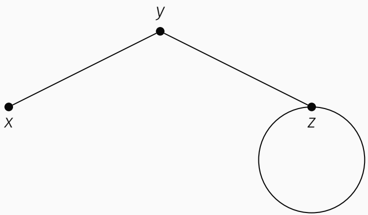


*We will restrict ourself to the following kind of graphs, in particular* ***undirected***

**(Undirected) Graph:**

A graph is a couple $\mathcal{G} = (V, E)$ where $V$ is a set of vertices and $E$ is a set of edges which are undirected pairs $\{x, y\}$ of $V$ possibly collapsing to a singleton

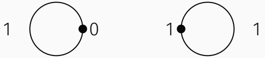
Notation $x \sim y$ denotes any edge $\{x, y\} \in E$



$\longrightarrow V = \{x, y, z\}$ and $E = \{\{x, y\}, \{y, z\}, \{z\}\}$

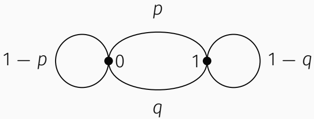
Any Markov chain defines a graph structure

$$\mathsf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



$$V = \{0, 1\}, E = \{\{0\}, \{1\}\}$$

$$\mathsf{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$



$$V = \{0, 1\}, E = \{\{0\}, \{0, 1\}, \{1\}\}$$

# WEIGHTED GRAPHS

**Weighted graph:**

Let $(V, E)$ be a graph. A weight on this graph is $w : V \times V \longrightarrow \mathbb{R}_{\geq 0}$ such that

$$\begin{cases} w(x,y) = w(y,x) & \big(\text{symmetry}\big) \\ w(x,y) > 0 & \text{if and only if } x \sim y, \textit{ i.e.}, \{x,y\} \in E \end{cases}$$

**From weighted graphs to Markov chains:**

With a weight $w$ on $(V, E)$ we associate a stochastic matrix $\mathbf{P} = (p(x,y))_{x,y \in V}$ by setting,

$$p(x,y) = \frac{w(x,y)}{W(x)} \quad \text{where} \;\; W(x) \overset{\text{def}}{=} \sum_{z \in V} w(x,z)$$

The corresponding Markov chain is called a random walk on $(V, E)$

This Markov chain is in detailed balance with the distribution

$$\pi(x) = \frac{W(x)}{W} \quad \text{where} \;\; W = \sum_{z \in V} W(z)$$

⟶ Weighted graphs are reversible Markov chains. Reciprocally reversible Markov chains define weighted graphs by setting $w(x,y) = \pi(x)p(x,y)$ and $x \sim y$ when $w(x,y) > 0$

# AN IMPORTANT EXAMPLE: SIMPLE RANDOM WALKS

Given $(V, E)$ and $x \in V$, we define $\mathbf{deg}\, x = \sharp\{y \in V : \ x \sim y\}$ $\big($number of neighbours of $x\big)$ and

$w(x, y) = 1$ if and only if $x \sim y$

$\longrightarrow$ It defines the simple random walk: from $x$, we walk to a neighbor $y$ with probability $\frac{1}{\mathbf{deg}\, x}$

- If $(V, E)$ is $d$-regular, *i.e.*, $\mathbf{deg}\, x = d$ for all $x \in V$. Then the associate transition matrix is symmetric. In particular, the equilibrium is given by the uniform distribution

# CONNECTED GRAPHS

**Connected graph:**

A graph $(V, E)$ is said to be connected is for any $x, y \in V$ it exists a path between them, *i.e.*,

$$\forall x, y \in V, \ \exists z_1, \dots, z_t \in V : \ x \sim z_1 \sim z_2 \cdots \sim z_t \sim y$$

# CONNECTED GRAPHS AND SPECTRAL ANALYSIS

Proposition $\left(\text{admitted}\right)$:

Let $(V, E, w)$ be a weighted graph. Let $\mathbf{P}$ be the transition matrix of its associated random walk. The multiplicity of the eigenvalue $1$ of $\mathbf{P}$ is one if and only if $(V, E)$ is a connected graph

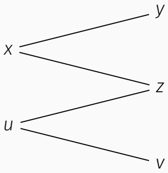
*Note that connectedness is a structural property and therefore the multiplicity of* $1$ *does not depend on the weight* $w$

# BIPARTITE GRAPHS

**Bipartite graph:**

A graph $(V, E)$ is said bipartite if it exists a non-trivial partition of vertices $V = V_1 \bigsqcup V_2$ such that $E \subseteq \left\{\{x_1, x_2\} : \; x_1 \in V_1 \text{ and } x_2 \in V_2\right\}$

$$V = \{x, y, z, u, v\} \text{ and } E = \left\{\{x, y\}, \{x, z\}, \{u, z\}, \{u, v\}\right\}$$



# BIPARTITE GRAPHS AND SPECTRAL ANALYSIS

**Proposition** $\big($admitted$\big)$:

Let $(V, E, w)$ be a connected weighted graph. Let $\mathsf{P}$ be the transition matrix of its associated random walk. Then the following are equivalent:

$(i)$ $(V, E)$ is bipartite

$(ii)$ The spectrum $\sigma(\mathsf{P})$ of $\mathsf{P}$ is symmetric, *i.e.*, $\lambda \in \sigma(\mathsf{P}) \iff -\lambda \in \sigma(\mathsf{P})$

$(iii)$ $-1$ is an eigenvalue, *i.e.*, $-1 \in \sigma(\mathsf{P})$

*Another example of a structural $\big($geometrical$\big)$ property that reflects on the spectral theory of the graph independently on the weight function.*

# CONNECTED AND BIPARTITE GRAPHS VERSUS ERGODICITY

**Theorem:**

Let $(V, E, w)$ be a connected not bipartite weighted graph. Let $q^{(t)}$ be the distribution of the associated random walk detailed balance with $\pi$ starting from a fixed point after $t$ walks. Then,

$$\|q^{(t)} - \pi\|_{1/\pi} = O\left((1-\delta)^t\right) \xrightarrow[t\to+\infty]{} 0$$

where $\delta \in (0,1)$ is the spectral gap of $\mathbf{P}$ and $\pi$ is its unique stationary distribution defined as $\pi(x) = W(x)/W$ where $W(x) = \sum_{y\in V} w(x,y)$ and $W = \sum_{x\in V} W(x)$

**Proof:**

We just combine Propositions on Slides 39 and 37 with Theorem on Slide 26

$\longrightarrow t = \ln(\varepsilon)/\ln(1-\delta) \approx 1/\delta$ steps of the random walk are enough to pick an element with distribution $\pi$

# Application: finding a marked vertex (i)

**Problem: find a marked vertex**

- **Input:** A connected non-bipartite graph $(V, E)$ and $f : V \to \{0, 1\}$ with $f(v) = 1$ if and only if $v$ is "marked"
- **Output:** A marked vertex, *i.e.*, $v \in V$ such that $f(v) = 1$

Suppose that the proportion of marked vertices is $\varepsilon$, *i.e.*, $\varepsilon = \frac{\sharp\{v\in V:\ f(v)=1\}}{\sharp V}$ and we can define a random walk with the uniform distribution as stationary distribution

- An algorithm to solve this problem: iterate
  - (*i*) Perform $1/\delta$ steps $\Big($where $\delta$ spectral gap$\Big)$ of the random walk
  - (*ii*) Output the current vertex if it is marked otherwise return to Step (*i*)

# APPLICATION: FINDING A MARKED VERTEX (II)

Cost of this approach:

- $S$ **setup cost:** the cost to set up the initial probability distribution
- $U$ **update cost:** the cost to perform one step of the random walk
- $C$ **check cost:** the cost to check if a vertex is marked, *i.e.*, to compute $f(v)$
- $\varepsilon$ proportion of marked vertices and $\delta$ spectral gap

$$\text{Complexity for finding a marked vertex} = S + \frac{1}{\varepsilon} \cdot \left( C + \frac{1}{\delta} \cdot U \right)$$

# QUANTUM RANDOM WALKS

# OBJECTIVE

To decrease the cost of the random walk by in particular “decreasing the number of walks”

| Classical approach | Quantum approach |
|---|---|
| $S + \frac{1}{\varepsilon} \cdot \left(C + \frac{1}{\delta} \cdot U\right)$ | $S + \frac{1}{\sqrt{\varepsilon}} \cdot \left(C + \frac{1}{\sqrt{\delta}} \cdot U\right)$ |

Table 1: Cost of classical and quantum approaches

Quantum computing enables surprisingly to increase the speed of convergence from $1/\delta$ to $1/\sqrt{\delta}$

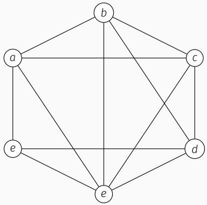
$\Big($the factor $1/\sqrt{\varepsilon}$ was expected via a Grover like approach$\Big)$

# ASSUMPTION

**Assumption:**

We will only consider graphs which are $d$-regular, *i.e.*, each vertex $x \in V$ has exactly $d$ neighbours

$\longrightarrow$ The 4-regular graph:



**Consequence:**

By using the weight $w(x, y) = 1/d$, it defines a random walk with symmetric transition matrix and the uniform as stationary distribution.

$\longrightarrow$ From one vertex we walk to one of its $d$ neighbours with probability $\frac{1}{d}$

# FIRST (QUANTUM) TRY

*As in the classical case: walk from one vertex to one of its d neighbours with probability $\frac{1}{d}$,*

*i.e., uniform choice over the neighbours*

$$|j\rangle \xrightarrow{U?} |\delta_j\rangle = \frac{1}{\sqrt{d}} \sum_{k \in V:\ (j,k) \in E} |k\rangle$$

**Issue:**

$|\delta_j\rangle$ and $|\delta_k\rangle$ may not be orthogonal while the $|j\rangle$ for $j \in V$ are

⟶ It can be fixed by moving to a larger Hilbert space

# QUANTUM APPROACH

**Quantum random walk framework:**

We will not only keep track of the current vertex but also of the neighbours $\Big($or predecessors$\Big)$

- **Quantum states:** superposition of elements $|x\rangle\,|y\rangle$ where $x \in V$ is the current vertex and $y \sim x$ a neighbour of $x$

# NOTATION

- **State space:** Span$\left\{ |x\rangle\,|y\rangle \;:\; x, y \in V \right\}$ where the $|x\rangle$ are orthonormal for $x \in V$
- **Quantum states:**

$$\begin{aligned} \mathcal{M} &= \text{marked vertices} \\ M &= \sharp\mathcal{M} \text{ number of marked vertices} \\ N &= \sharp V \text{ number of vertices} \\ |\psi_j\rangle &= \frac{1}{\sqrt{d}} \sum_{k:(j,k)\in E} |k\rangle \end{aligned}$$

**Starting idea:**

Build in the state space the superposition of $|j\rangle\,|\psi_j\rangle$ for the marked vertices, *i.e.*, $j \in \mathcal{M}$

- **Good and bad state:**

$$\begin{aligned} |G\rangle &= \frac{1}{\sqrt{M}} \sum_{j\in\mathcal{M}} |j\rangle\,|\psi_j\rangle \\ |B\rangle &= \frac{1}{\sqrt{N-M}} \sum_{j\notin\mathcal{M}} |j\rangle\,|\psi_j\rangle \\ |U\rangle &= \frac{1}{\sqrt{N}} \sum_{j\in V} |j\rangle\,|\psi_j\rangle \end{aligned}$$

# NOTATION

$$\begin{aligned}
\mathcal{M} &= \text{ marked vertices} \\
M &= \sharp\mathcal{M} \text{ number of marked vertices} \\
N &= \sharp V \text{ number of vertices} \\
|\psi_j\rangle &= \frac{1}{\sqrt{d}} \sum_{k:(j,k)\in E} |k\rangle \\
|G\rangle &= \frac{1}{\sqrt{M}} \sum_{j\in\mathcal{M}} |j\rangle\, |\psi_j\rangle \\
|B\rangle &= \frac{1}{\sqrt{N-M}} \sum_{j\notin\mathcal{M}} |j\rangle\, |\psi_j\rangle \\
|U\rangle &= \frac{1}{\sqrt{N}} \sum_{j\in V} |j\rangle\, |\psi_j\rangle
\end{aligned}$$

**Fact:**

$$|U\rangle = \sqrt{\varepsilon}\, |G\rangle + \sqrt{1-\varepsilon}\, |B\rangle \quad \Big(\text{we used that } \varepsilon = M/N \text{ proportion of marked vertices}\Big)$$

$\longrightarrow$ Our goal is to make $|U\rangle$ close to $|G\rangle$ $\Big($then by measuring we obtain a solution$\Big)$

Does it remind you something?

# GROVER APPROACH

$$|U\rangle = \sin\theta\,|G\rangle + \cos\theta\,|B\rangle \qquad \text{where } \sin\theta = \sqrt{\varepsilon}$$

$$\longrightarrow \theta \approx \sqrt{\varepsilon}\ \left(\text{we suppose } \varepsilon \text{ small}\right)$$

**Grover's algorithm:**

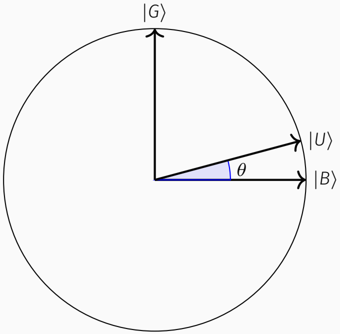
1. Build $|U\rangle$
2. Repeat $O\left(\frac{1}{\sqrt{\varepsilon}}\right)$ times:
   - ($i$) Reflection trough $|B\rangle$
   - ($ii$) Reflection trough $|U\rangle$
3. Measure the first register and check that $j \in V$ is marked, *i.e.*, $f(j) = 1$

⟶ For the justification that it succeeds with probability ≈ 1: see the following lecture

# PICTURING THE ALGORITHM

Supposing that we can make reflections over a quantum state

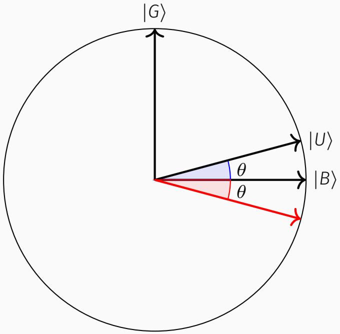
*We start by building $|U\rangle$*



# PICTURING THE ALGORITHM

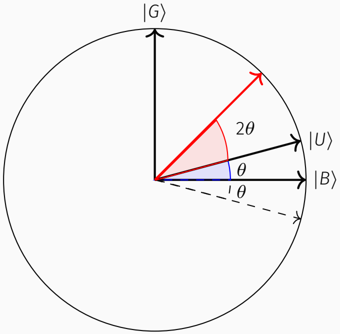
Supposing that we can make reflections over a quantum state

*Reflection over $|B\rangle$*



Supposing that we can make reflections over a quantum state

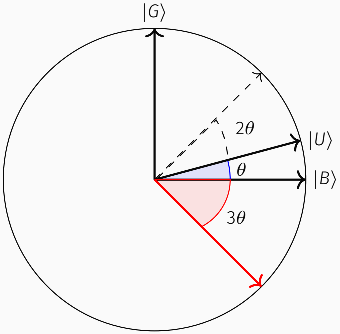
*Reflection over $|U\rangle$*



# PICTURING THE ALGORITHM

Supposing that we can make reflections over a quantum state

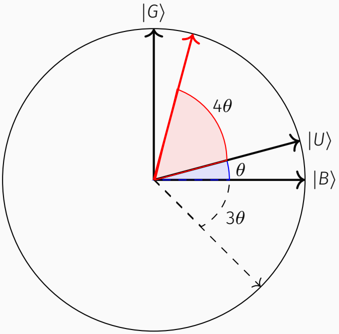
*Reflection over $|B\rangle$*



# PICTURING THE ALGORITHM

Supposing that we can make reflections over a quantum state

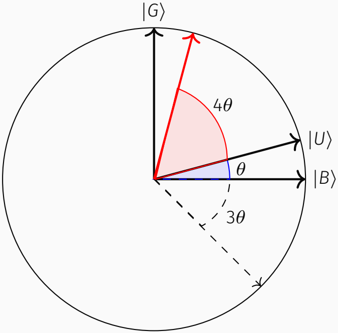
*Reflection over $|U\rangle$*



# PICTURING THE ALGORITHM

Supposing that we can make reflections over a quantum state

*and so on up to $\pi/2\ldots$*



**Number $k$ of iterations to reach $|G\rangle$: $\theta \longrightarrow (2k+1)\theta$**

Choose the number $k$ of iterations $\Big($reflections over $|B\rangle$ and $|U\rangle\Big)$ such that

$$(2k+1)\theta = \frac{\pi}{2} \iff k = \frac{\pi}{4\theta} - \frac{1}{2} \approx \frac{\pi}{4\sqrt{\varepsilon}}$$

# YOU SAID REFLECTIONS?

But how to build both reflections?

## Reflection over $|B\rangle$:

$$|B\rangle = \sin\theta \sum_{j:\ f(j)=0} |j\rangle \left|\psi_j\right\rangle = \sin\theta \sum_{j:\ f(j)=0} \ \sum_{k:\ (j,k)\in E} |j\rangle\,|k\rangle$$

- The reflection $\mathbf{R}_{|B\rangle}$ over $|B\rangle$ is nothing else than:

$$\mathbf{R}_{|B\rangle} : |j\rangle\,|k\rangle \longmapsto (-1)^{f(j)}\,|j\rangle\,|k\rangle$$

It can be performed in quantum time $O\left(\text{Cost(f)}\right)$ where Cost($f$) is the classical running time to run $f$ $\Big($see the followiong Exercise Session$\Big)$ which will be $C$ $\Big($check cost$\Big)$

## Reflection over $|U\rangle$: first approach

We can run the reflection $\mathbf{R}_{|U\rangle}$ in quantum time $O\left(\text{Cost}(f) + \text{Cost}(|U\rangle)\right)$ where Cost($|U\rangle$) is the quantum cost to build $|U\rangle$, *i.e.*, to run the unitary $\mathbf{U}\,|\mathbf{0}\rangle = |U\rangle$ $\Big($see the following Exercise Session$\Big)$

$\longrightarrow$ Using this approach is nothing else than Grover's algorithm. . .

The interest of quant. random walks is to take advantage of structure to run $\mathbf{R}_{|U\rangle}$ more efficiently!

$$|U\rangle = \frac{1}{\sqrt{N}} \sum_{j \in V} |j\rangle\, |\psi_j\rangle = \frac{1}{\sqrt{Nd}} \sum_{j \in V} \sum_{k:\ (j,k) \in E} |j\rangle\, |k\rangle = \frac{1}{\sqrt{Nd}} \sum_{k \in V} \sum_{j:\ (j,k) \in E} |j\rangle\, |k\rangle = \frac{1}{\sqrt{N}} \sum_{j \in V} |\psi_j\rangle\, |j\rangle$$

$$\begin{aligned} \mathcal{A} &= \mathrm{Span}\left\{\, |j\rangle\, |\psi_j\rangle \,\right\} \\ \mathcal{B} &= \mathrm{Span}\left\{\, |\psi_j\rangle\, |j\rangle \,\right\} \\ \mathsf{P} &= \text{transition matrix of the underlying random walk} \end{aligned}$$

**Fundamental idea:**

Given the reflection $\mathsf{R}_{\mathcal{A}}$ and $\mathsf{R}_{\mathcal{B}}$ over $\mathcal{A}$ and $\mathcal{B}$,

$$\mathsf{W}(\mathsf{P}) \overset{\text{def}}{=} \mathsf{R}_{\mathcal{B}} \mathsf{R}_{\mathcal{A}}$$

$\longrightarrow$ By the above decomposition $\mathsf{W}(\mathsf{P})\, |U\rangle = |U\rangle$ as $|U\rangle \in \mathcal{A} \cap \mathcal{B}$:

$|U\rangle$ eigenvector of $\mathsf{W}(\mathsf{P})$ with eigenvalue $1$

**Remark:**

One may wonder why we added the dependence in $\mathsf{P}$ by writing $\mathsf{W}(\mathsf{P})$. It turns out that $\mathsf{W}(\mathsf{P})$ can be interpreted as four steps of the underlying classical random walk $\Big($see later$\Big)$

# EIGENVALUES AND EIGENVECTORS OF THE QUANTUM WALK

$|U\rangle$ eigenvector of $\mathbf{W(P)}$ with eigenvalue 1: $\mathbf{W(P)}\,|U\rangle = |U\rangle$

What are the others eigenvalues of $\mathbf{W(P)}$?

$\longrightarrow$ They are basically eigenvalues of $\mathbf{P}$

**Theorem** $\Big($admitted$\Big)$:

The eigenvalues of $\mathbf{P}$ are the $\mathrm{e}^{2i\pi\theta_j}$ where $\cos \pi\theta_j = \lambda_j$ with the $\lambda_j$ being eigenvalues of $\mathbf{P}$

$\Big($remember, all eigenvalues $\lambda$ of $\mathbf{P}$ verify $|\lambda| \leq 1\Big)$

**Fundamental consequence:**

$\pi\theta_j \bmod \pi/2 \geq \sqrt{2\delta}$ as $|\lambda_i| \leq 1 - \delta$ where $\delta$ is the spectral gap!

# PHASE GAP

**Phase gap:**

The phase gap $\Delta(\mathsf{P})$ of $\mathsf{P}$ is the value of $\theta$ where $\pi\theta$ is the smallest angle in $(0, 2\pi)$ such that $\cos \pi\theta$ is an eigenvalue $\neq 1$ of $\mathsf{P}$

**Proposition:**

$$\Delta(\mathsf{P}) \geq \frac{1}{\pi} \cdot \sqrt{2\delta} \quad \text{where} \;\; \delta \text{ spectral gap of } \mathsf{P}$$

**Proof:**

Let $\theta$ achieving the maximum of $\cos 2\pi\theta$ for $\Delta(\mathsf{P})$, we have $1 - \cos 2\pi\theta = \delta$. But,

$$1 - \delta = \cos \pi\theta \geq 1 - (\pi\theta)^2 / 2$$

# FUNDAMENTAL REMARK: PHASE ESTIMATION PROBLEM

**Phase estimation:**

- **Input:** $n \in \mathbb{N}$, a unitary $\mathbf{U}$ and an eigenstate $|u\rangle$:
$$\mathbf{U}\,|u\rangle = \mathrm{e}^{2i\pi\varphi}\,|u\rangle$$
- **Output:** $\widetilde{\varphi} \in [0,1)$ such that $|\varphi - \widetilde{\varphi}| < 2^{-n}$, *i.e.*, the knowledge of the eigenvalue with precision $n$

$\longrightarrow$ There is a quantum algorithm $\mathbf{U}_{\mathrm{PE}}$ solving this problem

**Proposition:**

The phase estimation $\mathbf{U}_{\mathrm{PE}}\Big($before the last step measuring in the computational basis$\Big)$ computes,

$$\left|0^t\right\rangle |u\rangle \mapsto |\psi_u\rangle\,|u\rangle$$

such that $|\psi_u\rangle$ is an approximation of $\varphi$, *i.e.*, when measuring the first register we obtain $\widetilde{\varphi} \in \{0,1\}^t$ admitting the same first $n$ bits than $\varphi$ with probability $\geq 1-\varepsilon$ if $t$ is chosen as

$$t = n + \left\lceil \log\left(2 + \frac{1}{2\varepsilon}\right)\right\rceil$$

Furthermore, the algorithm uses $O(t^2)$ elementary gates and $t$ calls to controlled-$\mathbf{U}^{2^j}$ for $0 \leq j < t$ which has a cost $O\Big(2^t \cdot \mathrm{Cost}(\mathbf{U})\Big)$

# REFLECTION OVER $|U\rangle$

$$\left(\text{given } \mathbf{x} = (x_1, \ldots, x_t) \in \{0,1\}^t,\ |\mathbf{x}| = \textstyle\sum_{i=0}^t x_i 2^i\right)$$

### Reflection over $|U\rangle$:

Given a quantum state $\left|0^t\right\rangle |\psi\rangle$ where $|\psi\rangle$ is an eigenvector of $\mathbf{W(P)}$ with eigenvalue $\mathrm{e}^{2i\pi\theta_j}$

1. Apply $\mathbf{U}_{\mathrm{PE}}$ for $\mathbf{W(P)}$ with precision $n = -\log_2 \frac{1}{2\pi} \cdot \sqrt{\delta}$
2. Apply on the first $n$ registers $\mathbf{Z}_\delta : |\mathbf{x}\rangle = \begin{cases} |\mathbf{x}\rangle \text{ if } |\mathbf{x}| < \frac{1}{2\pi} \cdot \sqrt{\delta} \\ -|\mathbf{x}\rangle \text{ if } |\mathbf{x}| > \frac{1}{2\pi} \cdot \sqrt{\delta} \end{cases}$
3. Apply $\mathbf{U}_{\mathrm{PE}}^{-1}$

$\longrightarrow$ **Claim**: this algorithm with $n = -\log_2 \frac{1}{2\pi} \cdot \sqrt{\delta}$ performs $\mathbf{R}_{|U\rangle}$

### Analysis:

Suppose that $\mathbf{U}_{\mathrm{PE}} \left|0^t\right\rangle |\psi\rangle = |\mathbf{x}\rangle\, |u\rangle$ where $\mathbf{x} \in \{0,1\}^t$ and its first $n$ bits are those of $\theta_j$

- Given the eigenvector $|\psi\rangle$ with eigenvalue $\mathrm{e}^{2i\pi\theta_j}$ where $\theta_j \neq 0$: we have $\theta_j \geq \frac{1}{\pi} \cdot \sqrt{2\delta}$ and $|\mathbf{x} - \theta_j| \leq 2^{-n} = \sqrt{\pi\delta}$ therefore $|\mathbf{x}| \geq \frac{1}{\pi} \cdot \sqrt{2\delta} - \frac{1}{2\pi} \cdot \sqrt{\delta} > \frac{1}{2\pi} \cdot \sqrt{\delta}$. The algorithm returns
$$-\left|0^t\right\rangle |\psi\rangle$$
- Given the eigenvector $|U\rangle$ with eigenvalue 1: we have $|x| < 2^n = \sqrt{\pi\delta}$. The algorithm returns
$$\left|0^t\right\rangle |\psi\rangle$$

Therefore, as eigenvectors of $\mathbf{U}$ for an orthonormal basis and $|U\rangle$, our algorithm computes $\mathbf{R}_{|U\rangle}$

# REFLECTION OVER $|U\rangle$: COST

**Reflection over $|U\rangle$:**

Given a quantum state $\left|0^t\right\rangle|\psi\rangle$ where $|\psi\rangle$ is an eigenvector of $\mathbf{W(P)}$ with eigenvalue $\mathrm{e}^{2i\pi\theta_j}$

1. Apply $\mathbf{U}_{\text{PE}}$ for $\mathbf{W(P)}$ with precision $n = -\log_2 \frac{1}{2\pi}\cdot\sqrt{\delta}$
2. Apply on the first $n$ registers $\mathbf{Z}_\delta : |\mathbf{x}\rangle = \begin{cases} |\mathbf{x}\rangle \text{ if } |\mathbf{x}| < \frac{1}{2\pi}\cdot\sqrt{\delta} \\ -|\mathbf{x}\rangle \text{ if } |\mathbf{x}| > \frac{1}{2\pi}\cdot\sqrt{\delta} \end{cases}$
3. Apply $\mathbf{U}_{\text{PE}}^{-1}$

The cost of $\mathbf{U}_{\text{PE}}$ is $O\left(2^t\cdot \text{Cost}\left(\mathbf{W(P)}\right)\right)$ with

$$t = n + \left\lceil \log\left(2+\frac{1}{2\varepsilon}\right)\right\rceil \quad \text{where } n = -\log_2\frac{1}{2\pi}\cdot\sqrt{\delta}$$

Here $1-\varepsilon$ is the probability after measuring that $\mathbf{U}_{\text{PE}}$ computes an approximation over the first $n$ bits. We choose $\varepsilon$ as a constant small enough. The whole cost of this algorithm is

$$O\left(\frac{1}{\sqrt{\delta}}\cdot\text{Cost}\left(\mathbf{W(P)}\right)\right)$$

**Be careful:**

In our analysis we supposed that $\mathbf{U}_{\text{PE}}\left|0^t\right\rangle|\psi\rangle = |\mathbf{x}\rangle|u\rangle$ where $\mathbf{x}\in\{0,1\}^t$ gives the first $n$ bits of $\theta_j$. It is not true but we can show that we have a sufficiently well approximation of $\mathbf{R}_{|U\rangle}$

# RUNNING TIME OF THE QUANTUM WALK

We need to provide an implementation of W(P)

$$\begin{aligned}\mathcal{A} &= \mathsf{Span}\Big\{\, |j\rangle \left|\psi_j\right\rangle \Big\} \\ \mathcal{B} &= \mathsf{Span}\Big\{\, \left|\psi_j\right\rangle |j\rangle \Big\}\end{aligned}$$

$\mathsf{W}(\mathsf{P}) \overset{\text{def}}{=} \mathsf{R}_{\mathcal{B}}\mathsf{R}_{\mathcal{A}}$ where $\mathsf{R}_{\mathcal{B}}, \mathsf{R}_{\mathcal{A}}$ are reflections over $\mathcal{A}$ and $\mathcal{B}$

$$(1)\ \ |j\rangle\,|0\rangle \mapsto |j\rangle \left|\psi_j\right\rangle, \quad (2)\ \ |0\rangle\,|j\rangle \mapsto \left|\psi_j\right\rangle |j\rangle$$

Both operations are the quantum versions of the classical operation "starting from $j \in V$ walks uniformly to one of its neighbours" which is run with our notation in time $U$ $\Big($see Slide 42$\Big)$

$\longrightarrow$ Operations (1) and (2) can be implemented in quantum time $O(U)$

- Implementing $\mathsf{R}_{\mathcal{A}}$ $\Big($*resp.* $\mathsf{R}_{\mathcal{B}}\Big)$: use the inverse of (1) $\Big($*resp.* (2)$\Big)$ put a $-$ in front if the second $\Big($*resp.* first$\Big)$ register and apply (1) $\Big($*resp.* (2)$\Big)$

W(P) can be run in quantum time $O(U)$ $\Big($via the quantization of four steps of random walks$\Big)$

# QUANTUM WALK: RECAP

- **Setup cost** $S$: we need to build $|U\rangle$ to feed him to the phase estimation. Notice that during the phase estimation $\mathbf{U}_{\mathrm{PE}}$ we do not modify $|U\rangle$. Therefore we can build it once even if we apply many times $\mathbf{U}_{\mathrm{PE}}$. It basically costs to classically setup on vertex and to call this procedure quantumly in superposition
- We repeat $O\left(\frac{1}{\sqrt{\varepsilon}}\right)$ times reflections $\mathbf{R}_{|B\rangle}$ and $\mathbf{R}_{|U\rangle}$
- **Check-up cost** $C$: we run $\mathbf{R}_{|B\rangle}$ in time $O(C)$ where $C$ is the running time of $f(x) = 0$ or $1$ $\Big($time to classically check if a solution is marked$\Big)$
- **Update cost** $U$: we run $\mathbf{R}_{|U\rangle}$ in time $O\left(\frac{1}{\sqrt{\delta}} \cdot U\right)$ where $U$ is the running time to classically update the Markov chain $\Big($walking uniformly from one vertex to one of its neighbours$\Big)$

**Quantum walk running time:**

$$S + \frac{1}{\sqrt{\varepsilon}}\left(C + \frac{1}{\sqrt{\delta}} \cdot U\right)$$

# COMPARISON OF ALL STRATEGIES

- Setup cost $S$
- Check-up cost $C$
- Update cost $U$

| Standard Search | Random Walk | Grover Algorithm | Quantum Random Walk |
|---|---|---|---|
| $\frac{1}{\varepsilon} \cdot (S + C)$ | $S + \frac{1}{\varepsilon} \cdot \left(C + \frac{1}{\delta} \cdot U\right)$ | $\frac{1}{\sqrt{\varepsilon}} \cdot (C + S)$ | $S + \frac{1}{\sqrt{\varepsilon}} \cdot \left(C + \frac{1}{\sqrt{\delta}} \cdot U\right)$ |

Table 2: Cost of classical and quantum approaches

**Remark:**

In Grover's algorithm we run $\mathsf{R}_{|U\rangle}$ by using the standard technique, *i.e.*, the one in which we need to build $|U\rangle$ which has a cost $S$

# APPLICATION: FINDING COLLISION

# OUR COLLISION PROBLEM

**Golden collision finding:**

- **Input:** a function $f : \{0,1\}^n \longrightarrow \{0,1\}^n$ with the promise that it exists a unique pair $(x_0, x_1)$ such that $f(x_0) = f(x_1)$ and $x_0 \neq x_1$
- **Output:** $(x_0, x_1)$

**Our goal:**

We want to solve this problem by minimizing the number of queries to $\mathbf{U}_f : |x\rangle\,|y\rangle \mapsto |x\rangle\,|y + f(x)\rangle$

$\longrightarrow$ We can solve this problem via Grover's algorithm over pairs with $O\left(\sqrt{2^{2n}}\right) = O\left(2^n\right)$ queries to $\mathbf{U}_f$

Can we do with less queries?

$\longrightarrow$ Yes with $O\left(2^{2n/3}\right)$ queries! By using the quantum random walk approach. . .

# OUR APPROACH: FINDING A MARKED VERTEX IN A PARTICULAR GRAPH

We first index $\{0,1\}^n$ as $\{y_1,\ldots,y_{2^n}\}$

**Our considered graph: for some parameter $r$**

- Vertices $V$: given $R \subseteq \{0,\ldots,2^n\}$ we define

$$v_R \stackrel{\text{def}}{=} \Big(v_R(\text{in}), v_R(\text{out})\Big) \quad \text{where} \quad v_R(\text{in}) \stackrel{\text{def}}{=} (y_i)_{i\in R} \text{ and } v_R(\text{out}) \stackrel{\text{def}}{=} (f(y_i))_{i\in R}$$

  Then,

$$V \stackrel{\text{def}}{=} \Big\{v_R \; : \; \sharp R = r\Big\}$$

- Edges $E$: we have $v_R \sim v_{R'}$ if $R$ and $R'$ only differs by one element, *i.e.*, $v_R(\text{in})$ and $v_R(\text{in})$ differ by one element
- Marked vertices: $v_R$ is marked if two elements in $v_R(\text{out})$ are equal, *i.e.*, $v_R(\text{in})$ contains $(x_0, x_1)$

$\longrightarrow$ This graph is known as the **Johnson graph $J(2^n, r)$!**

**Remark:**

If $R$ is too large $\Big(\text{think } 2^n\Big)$ then just the cost to build one vertex is prohibitive

# JOHNSON GRAPH

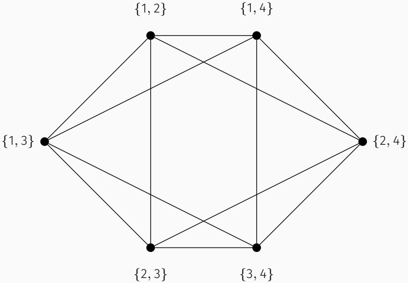


Table 3: Johnson graph $J(4, 2)$

**Johnson graph property:**

$J(2^n, r)$ has $\binom{2^n}{r}$ vertices, it is $r\left(2^n - r\right)$ regular and its spectral gap is given by

$$\delta = \frac{2^n}{r(2^n - r)}$$

$\longrightarrow$ In particular, $\delta \approx \frac{1}{r}$ when $r \ll 2^n$

# QUANTUM RANDOM WALK TO FIND COLLISIONS

**Our considered graph: for some parameter $r$**

- Vertices $V$: given $R \subseteq \{0, \ldots, 2^n\}$ we define

$$v_S \overset{\text{def}}{=} \left(v_R(\text{in}), v_R(\text{out}\right) \quad \text{where } \; v_R(\text{in}) \overset{\text{def}}{=} (y_i)_{i\in R} \text{ and } v_R(\text{out}) \overset{\text{def}}{=} (f(y_i))_{i\in R}$$

  Then,

$$V \overset{\text{def}}{=} \left\{v_R \, : \, \sharp R = r\right\}$$

- Edges $E$: we have $v_R \sim v_{R'}$ if $R$ and $R'$ only differs by one element, *i.e.*, $v_R(\text{in})$ and $v_R(\text{in})$ differ by one element
- Marked vertices: $v_R$ is marked if two elements in $v_R(\text{out})$ are equal, *i.e.*, $v_R(\text{in})$ contains $(x_0, x_1)$

$\longrightarrow$ It is the graph $J(2^n, r)$

- **Setup** $S$: to setup a vertex requires $r$ queries to $f$ and to build a superposition $r$ queries to $\mathbf{U}_f$
- **Update** $U$: two queries to $f$, therefore two queries to $\mathbf{U}_f$
- **Check-up** $C$: zero query to $f$ as all information is in $v_S(\text{out})$
- The proportion of marked vertices is $\varepsilon = \frac{r(r-1)}{2^n\left(2^n-1\right)} \approx \frac{r^2}{2^{2n}}$

# QUANTUM QUERY COST TO SOLVE GOLDEN COLLISION PROBLEM

- **Setup** $S$: to setup a vertex requires $r$ queries to $f$ and to build a superposition $r$ queries to $\mathsf{U}_f$
- **Update** $U$: two queries to $f$, therefore two queries to $\mathsf{U}_f$
- **Check-up** $C$: zero query to $f$ as all information is in $v_S(\text{out})$
- The proportion of marked vertices is $\varepsilon = \frac{r(r-1)}{2^n\left(2^n-1\right)} \approx \frac{r^2}{2^{2n}}$

Final cost to find the collision in terms of queries to $f$:

$$S + \frac{1}{\sqrt{\varepsilon}} \cdot \left(C + \frac{1}{\sqrt{\delta}} \cdot U\right) = r + \frac{\sqrt{2^n}}{r} \cdot \sqrt{r} = r + \frac{2^n}{\sqrt{r}}$$

$\longrightarrow$ It is minimized to $r = 2^{2n/3}$ and is equal to $2^{2n/3}$ $\Big($claiming query complexity$\Big)$

# EXERCISE SESSION