Quantum Walks: What's Missing?

Yevgeniy Kovchegov
Oregon State University

Quantum Fourier Transforms

Discrete Fourier transform over $\mathbb{C}^N$ is an $N \times N$ unitary matrix. Thus, there is an algorithmic advantage of implementing it on quantum computers.

- Classical Fourier transform uses $O(N^2)$ gates.

- Let $N = 2^n$. Fast Fourier transform (FFT) uses $O(n2^n) = O(N \log N)$ gates.

- Let $N = 2^n$. Quantum Fourier transform (QFT) uses $O(n^2) = O(\log^2 N)$ gates.

The QFT is a quantum implementation of the FFT via Hadamard gates and controlled phase shift gates.

Applied mathematician Gilbert Strang described the FFT as *the most important numerical algorithm of our lifetime*.

Randomized algorithms.

Applied mathematician Gilbert Strang described the FFT as *the most important numerical algorithm of our lifetime*.

The QFT is equally revered in quantum computation. Notable applications being the phase estimation algorithm and Shor's algorithm.

Another successful approach in classical computing: randomized algorithms would often achieve algorithmic speedup, and in some cases, randomized algorithms are the only practical means of solving a problem.

Similarly to Grover's algorithm, randomized algorithm produces an optimal or close to optimal solution with probability $1 - \epsilon$.

Goal: A quantum approach that replicates the success of randomized algorithms.

Metropolis-Hastings algorithm.

Goal: simulating an S-valued random variable distributed according to a given probability distribution $\pi(z)$, given a complex nature of large discrete space S.

MCMC: generating a Markov chain $\{X_t\}$ over S, with distribution $\mu_t(z) = P(X_t = z)$ converging rapidly to its unique stationary distribution, $\pi(z)$.

Metropolis-Hastings algorithm: Consider a connected neighborhood network with points in S. Suppose we know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any two neighbor points z and z' on the network.

Let for z and z' connected by an edge of the network, the transition probability be set to

$$p(z, z') = \frac{1}{M} \min\left\{1, \frac{\pi(z')}{\pi(z)}\right\} \quad \text{for } M \text{ large enough.}$$

Metropolis-Hastings algorithm.

Consider a connected neighborhood network with points in S.

Suppose we know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any two neighbor points z and z' on the network.

Let for z and z' connected by an edge of the network, the transition probability be set to

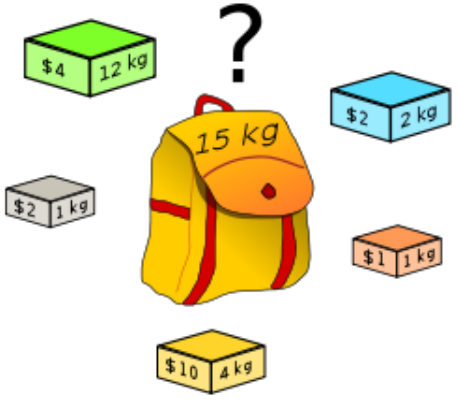
$$p(z,z') = \frac{1}{M} \min\left\{1, \frac{\pi(z')}{\pi(z)}\right\} \quad \text{for } M \text{ large enough.}$$

Specifically, M can be any number greater than the maximal degree in the neighborhood network.

Let $p(z,z)$ absorb the rest of the probabilities, i.e.

$$p(z,z) = 1 - \sum_{z':\ z\sim z'} p(z,z')$$

Knapsack problem. The **knapsack problem** is a problem in combinatorial optimization: Given a set of items, each with a mass and a value, determine the number of each item to include in a collection so that the total weight is less than or equal to a given limit and the total value is as large as possible. Knapsack problem is NP complete.



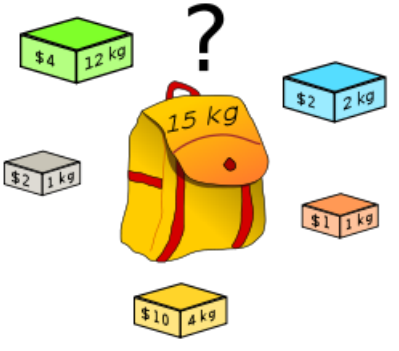
Source: Wikipedia.org

Knapsack problem. Given m items of various weights w_j and value v_j, and a knapsack with a weight limit R. Assuming the volume and shape do not matter, find the most valuable subset of items that can be carried in the knapsack.

Mathematically: we need $z = (z_1, \ldots, z_m)$ in

$$S = \left\{ z \in \{0,1\}^m \ : \ \sum_{j=1}^{m} w_j z_j \leq R \right\}$$

maximizing $U(z) = \sum_{j=1}^{m} v_j z_j$.



Source: Wikipedia.org

Knapsack problem. Find $z = (z_1, \ldots, z_m)$ in

$$S = \left\{ z \in \{0,1\}^m \ : \ \sum_{j=1}^{m} w_j z_j \leq R \right\} \text{ maximizing } \ U(z) = \sum_{j=1}^{m} v_j z_j .$$

• **MCMC approach:** Assign weights $\pi(z) = \frac{1}{Z_\beta} \exp\left\{\beta\, U(z)\right\}$ to each $z \in S$ with $\beta = \frac{1}{T}$, where

$$Z_\beta = \sum_{z \in S} \exp\left\{\beta\ U(z)\right\}$$

is called partition function. Next, for each $z \in S$ consider a **clique** $\mathcal{C}_z$ of neighbor points in S. Consider a Markov chain over S that jumps from z to a neighbor $z' \in \mathcal{C}_z$ with probability

$$p(z, z') = \frac{1}{m} \ \min\left\{1, \ \frac{\pi(z')}{\pi(z)}\right\}.$$

Knapsack problem. Assign weights $\pi(z) = \frac{1}{Z_\beta} \exp\{\beta\ U(z)\}$ to each $z \in S$ with $\beta = \frac{1}{T}$, where

$$Z_\beta = \sum_{z \in S} \exp\{\beta\ U(z)\}$$

is called partition function. Next, for each $z \in S$ consider a **clique** $\mathcal{C}_z$ of neighbor points in S. Consider a Markov chain over S that jumps from z to a neighbor $z' \in \mathcal{C}_z$ with probability

$$p(z, z') = \frac{1}{m}\ \min\left\{1,\ \frac{\pi(z')}{\pi(z)}\right\}.$$

Observe that

$$\frac{\pi(z')}{\pi(z)} = \exp\{\beta\ \big(U(z') - U(z)\big) = \exp\{\beta\ \big(v \cdot (z' - z)\big)\},$$

where $v = (v_1, \ldots, v_m)$ is the values vector.

Knapsack and other optimization problems.

- **Issues:**

(i) Running time?

Analyzing mixing time is challenging in MCMC for real-life optimization problems such as knapsack problem. With few exceptions – no firm foundation exists, and no performance guaranteed.

(ii) Optimal T?

T is usually chosen using empirical observations, trial and error, or certain heuristic.

Often, simulated annealing approach is used.

Simulated annealing.

Usually, we let $\pi(z) = \frac{1}{Z_\beta} \exp\left\{\beta\ U(z)\right\}$ to each $z \in S$ with $\beta = \frac{1}{T}$, and $p(z, z') = \frac{1}{M}\ \min\left\{1,\ \frac{\pi(z')}{\pi(z)}\right\}$.

• **Idea:** What if we let temperature T change with time t, i.e. $T = T(t)$? When T is large, the Markov chain is more diffusive; as T gets smaller, the value X_t stabilizes around the maxima.

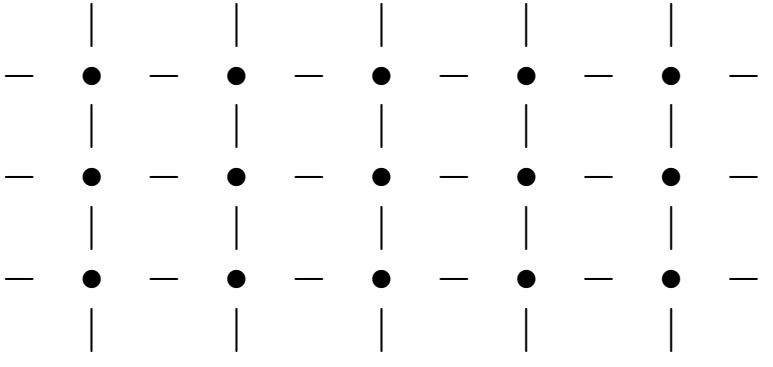
The method was independently devised by S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi in 1983, and by V. Černý in 1985.

Name comes from *annealing in metallurgy*, a technique involving heating and controlled cooling.

Gibbs Sampling: Ising Model.

Every vertex v of $G = (V, E)$ is assigned a spin $\sigma(v) \in \{-1, +1\}$. The probability of a configuration $\sigma \in \{-1, +1\}^V$ is

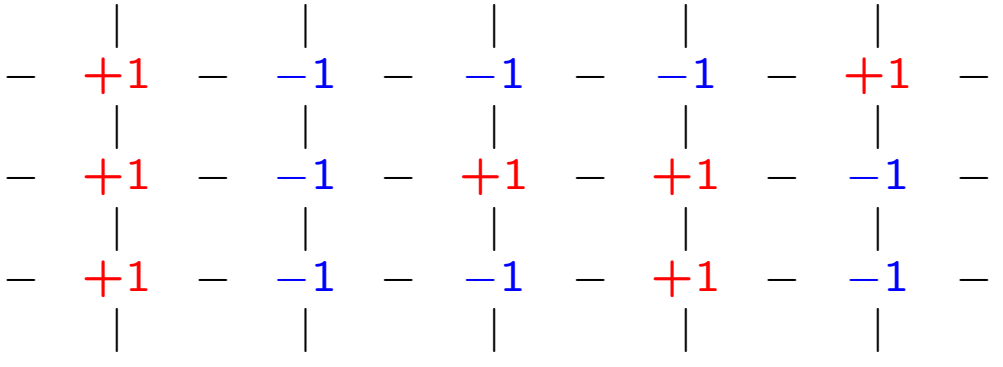
$$\pi(\sigma) = \frac{e^{-\beta \mathcal{H}(\sigma)}}{Z(\beta)}, \quad \text{where} \quad \beta = \frac{1}{T}$$



Gibbs Sampling: Ising Model.

Every vertex v of $G = (V, E)$ is assigned a spin $\sigma(v) \in \{-1, +1\}$. The probability of a configuration $\sigma \in \{-1, +1\}^V$ is

$$\pi(\sigma) = \frac{e^{-\beta \mathcal{H}(\sigma)}}{Z(\beta)}, \quad \text{where} \quad \beta = \frac{1}{T}$$



Gibbs Sampling: Ising Model.

$\forall \sigma \in \{-1,+1\}^V$, the Hamiltonian (energy function)

$$\mathcal{H}(\sigma) = -\frac{1}{2} \sum_{u,v:\, u \sim v} \sigma(u)\sigma(v) = \quad - \sum_{edges \;\; e=[u,v]} \sigma(u)\sigma(v)$$

and probability of a configuration $\sigma \in \{-1,+1\}^V$ is

$$\pi(\sigma) = \frac{e^{-\beta \mathcal{H}(\sigma)}}{Z(\beta)}, \quad \text{where} \;\; \beta = \frac{1}{T}$$

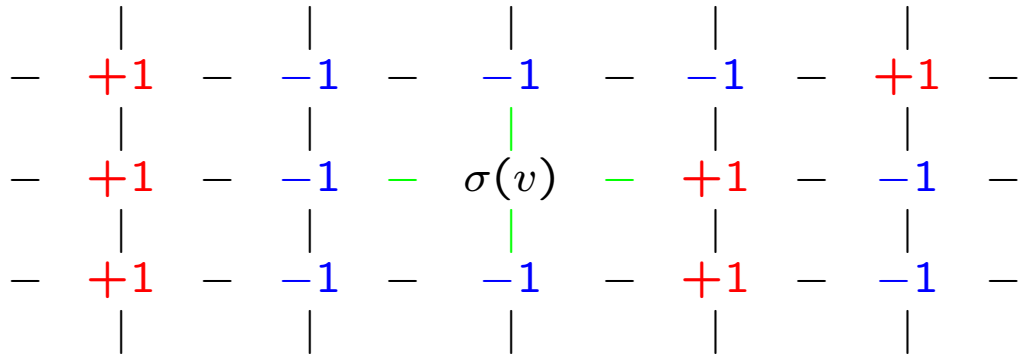
$Z(\beta) = \sum_{\sigma \in \{-1,+1\}^V} e^{-\beta \mathcal{H}(\sigma)}$ - normalizing factor.

The local Hamiltonian $\mathcal{H}_{local}(\sigma, v) = -\sum_{u:\, u \sim v} \sigma(u)\sigma(v)$.

Conditional probability for $\sigma(v)$ is expressed via $\mathcal{H}_{local}(\sigma, v)$:

$$\mathcal{H}(\sigma) = \mathcal{H}_{local}(\sigma, v) - \sum_{e=[u_1,u_2]:\, u_1,u_2 \neq v} \sigma(u_1)\sigma(u_2)$$

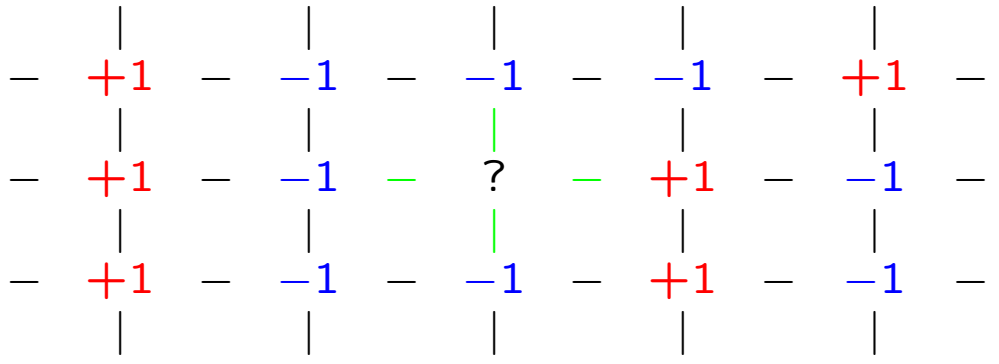
Gibbs Sampling: Ising Model via Glauber dynamics.



Conditional probability for $\sigma(v)$ is expressed via $\mathcal{H}_{local}(\sigma, v)$:

$$\mathcal{H}(\sigma) = \mathcal{H}_{local}(\sigma, v) - \sum_{e=[u_1,u_2]:\, u_1,u_2 \neq v} \sigma(u_1)\sigma(u_2)$$

Gibbs Sampling: Ising Model via Glauber dynamics.



Randomly pick $v \in V$ with probability $\frac{deg(v)}{2|E|}$. Next, erase the spin $\sigma(v)$, and replace it with $+1$ or -1 with probabilities

$$P(\sigma \to \sigma_+) = \frac{e^{-\beta \mathcal{H}(\sigma_+)}}{e^{-\beta \mathcal{H}(\sigma_-)} + e^{-\beta \mathcal{H}(\sigma_+)}} = \frac{e^{-\beta \mathcal{H}_{local}(\sigma_+, v)}}{e^{-\beta \mathcal{H}_{local}(\sigma_-, v)} + e^{-\beta \mathcal{H}_{local}(\sigma_+, v)}} = \frac{e^{-2\beta\sigma(v)}}{e^{-2\beta\sigma(v)} + e^{2\beta\sigma(v)}}$$

$$\text{and} \quad P(\sigma \to \sigma_-) = \frac{e^{2\beta\sigma(v)}}{e^{-2\beta\sigma(v)} + e^{2\beta\sigma(v)}}.$$

Gibbs Sampling: Ising Model via Glauber dynamics.
Randomly pick $v \in V$ with probability $\frac{deg(v)}{2|E|}$. Next, erase the spin $\sigma(v)$, and replace it with $+1$ or -1 with probabilities

$$P(\sigma \to \sigma_+) = \frac{e^{-\beta \mathcal{H}(\sigma_+)}}{e^{-\beta \mathcal{H}(\sigma_-)} + e^{-\beta \mathcal{H}(\sigma_+)}} = \frac{e^{-\beta \mathcal{H}_{local}(\sigma_+, v)}}{e^{-\beta \mathcal{H}_{local}(\sigma_-, v)} + e^{-\beta \mathcal{H}_{local}(\sigma_+, v)}} = \frac{e^{-2\beta\sigma(v)}}{e^{-2\beta\sigma(v)} + e^{2\beta\sigma(v)}}$$

$$\text{and} \quad P(\sigma \to \sigma_-) = \frac{e^{2\beta\sigma(v)}}{e^{-2\beta\sigma(v)} + e^{2\beta\sigma(v)}}.$$

Here, $\sigma_{+,v} = \sigma_+$ and $\sigma_{-,v} = \sigma_-$ are given by

$$\sigma_+(u) = \begin{cases} \sigma(u) & \text{if } u \neq v, \\ +1 & \text{if } u = v \end{cases} \quad \text{and} \quad \sigma_-(u) = \begin{cases} \sigma(u) & \text{if } u \neq v, \\ -1 & \text{if } u = v \end{cases}$$

So, the transition probabilities are

$$p(\sigma, \sigma_{+,v}) = \frac{deg(v)}{2|E|} \frac{e^{2\beta\sigma(v)}}{e^{-2\beta\sigma(v)} + e^{2\beta\sigma(v)}} \text{ and } p(\sigma, \sigma_{-,v}) = \frac{deg(v)}{2|E|} \frac{e^{2\beta\sigma(v)}}{e^{2\beta\sigma(v)} + e^{2\beta\sigma(v)}}.$$

Glauber dynamics: Rapid mixing.

Glauber dynamics - a random walk on state space S (here $\{-1,+1\}^V$) s.t. needed π is stationary w.r.t. Glauber dynamics.

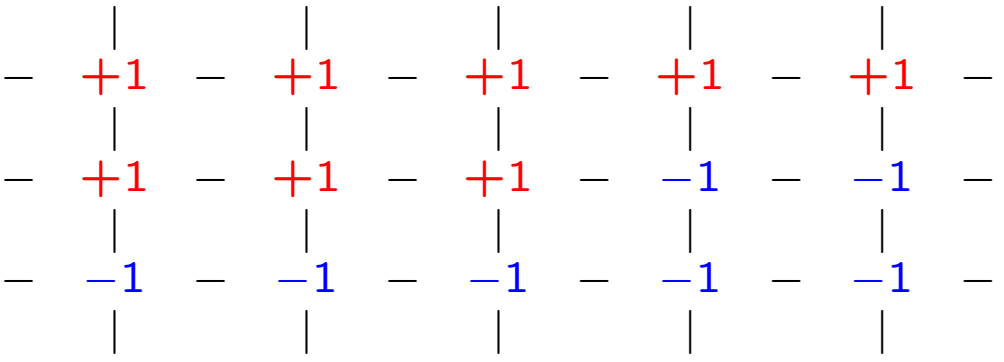
In high temperatures (i.e. $\beta = \frac{1}{T}$ small enough) it takes $O(n \log n)$ iterations to get "ε-close" to π. Here $|V| = n$.

Need: $\boxed{\max_{v \in V} deg(v) \cdot \tanh(\beta) < 1}$

Thus the Glauber dynamics is a fast way to generate π. It is an important example of Gibbs sampling.

Close enough distribution and mixing time.

What is "ε-close" to π? Start with σ_0:



If $P_t(\sigma)$ is the probability distribution after t iterations, the total variation distance

$$\|P_t - \pi\|_{TV} = \frac{1}{2} \sum_{\sigma \in \{-1,+1\}^V} |P_t(\sigma) - \pi(\sigma)| \leq \varepsilon \ .$$

Close enough distribution and mixing time.

Total variation distance:

$$\|\mu-\nu\|_{TV} := \frac{1}{2}\sum_{x\in S}|\mu(x)-\nu(x)| = \sup_{A\subset S}|\mu(A)-\nu(A)|.$$

Mixing time: let $\mu_t = \mu_0 P^t$, then

$$t_{mix}(\varepsilon) := \inf\{t \,:\, \|\mu_t - \pi\|_{TV} \le \varepsilon, \quad \text{all } \mu_0\}.$$

In high temperature, $t_{mix}(\varepsilon) = O(n \log n)$.

Quantum walks: the interchange framework.

In
Framework for discrete-time quantum walks and a symmetric walk on a binary tree (Phys. Rev. A, 2011) by Z. Dimcovic, D. Rockwell, I. Milligan, R. Burton, T. Nguyen, and Y. Kovchegov,
a general framework for discrete-time quantum walks is constructed as a quantum analog of memory-2 Markov chains (walks).

Consider a simply connected graph (network) $G = (V, E)$. Let $N = |V|$.

We construct a quantum walk $|\psi(t)\rangle \in \mathbb{C}^N \times \mathbb{C}^N$ via the interchange framework.

Let $|i\rangle$ and $|j\rangle$ represent two neighboring vertices in $G = (V, E)$. Then,

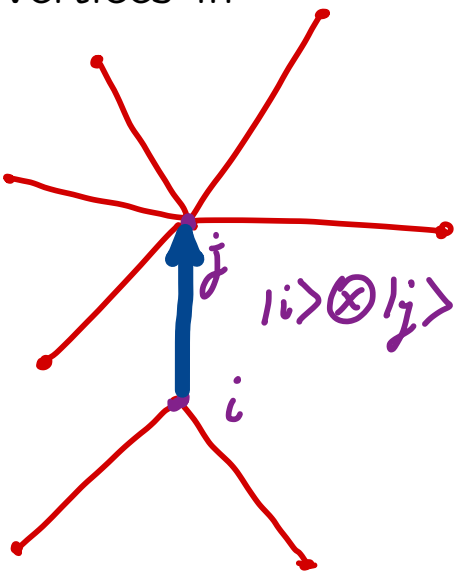
$$|\psi(t)\rangle = \sum_{i,j} c_{ij}(t)\, |i\rangle \otimes |j\rangle$$

Quantum walks: the interchange framework.

We construct a quantum walk $|\psi(t)\rangle \in \mathbb{C}^N \times \mathbb{C}^N$ via the interchange framework.

Let $|i\rangle$ and $|j\rangle$ represent two neighboring vertices in $G = (V, E)$. Then,

$$|\psi(t)\rangle = \sum_{i,j} c_{ij}(t)\, |i\rangle \otimes |j\rangle$$



evolves according to

$$|\psi(t+1)\rangle = UX\, |\psi(t)\rangle,$$

where

$$X : \quad |i\rangle \otimes |j\rangle \to |j\rangle \otimes |i\rangle$$

is the interchange operator, and

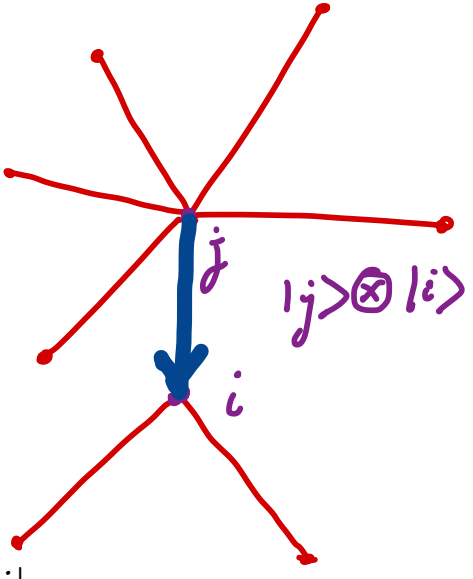
$$U = \sum_j \Pi_j \otimes U_j, \quad \text{where} \quad \Pi_j = |j\rangle\langle j|\,.$$

Quantum walks: the interchange framework.

We construct a quantum walk $|\psi(t)\rangle \in \mathbb{C}^N \times \mathbb{C}^N$ via the interchange framework.

Let $|i\rangle$ and $|j\rangle$ represent two neighboring vertices in $G = (V, E)$. Then,

$$|\psi(t)\rangle = \sum_{i,j} c_{ij}(t)\, |i\rangle \otimes |j\rangle$$



evolves according to

$$|\psi(t+1)\rangle = UX\, |\psi(t)\rangle,$$

where

$$X : \quad |i\rangle \otimes |j\rangle \to |j\rangle \otimes |i\rangle$$

is the interchange operator, and

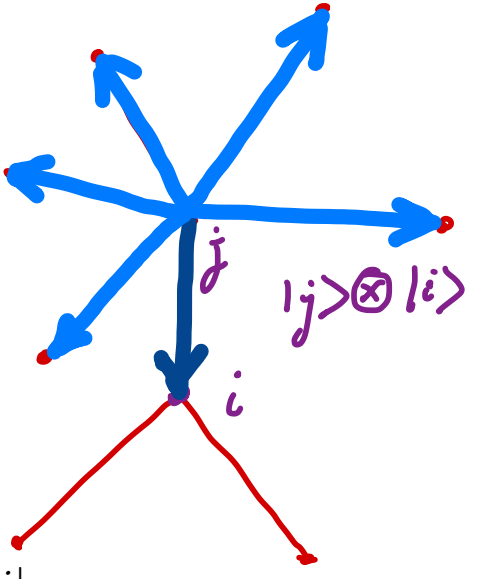
$$U = \sum_j \Pi_j \otimes U_j, \quad \text{where} \quad \Pi_j = |j\rangle\langle j|\,.$$

Quantum walks: the interchange framework.

We construct a quantum walk $|\psi(t)\rangle \in \mathbb{C}^N \times \mathbb{C}^N$ via the interchange framework.

Let $|i\rangle$ and $|j\rangle$ represent two neighboring vertices in $G = (V, E)$. Then,

$$|\psi(t)\rangle = \sum_{i,j} c_{ij}(t)\, |i\rangle \otimes |j\rangle$$



evolves according to

$$|\psi(t+1)\rangle = UX\, |\psi(t)\rangle,$$

where

$$X : \quad |i\rangle \otimes |j\rangle \to |j\rangle \otimes |i\rangle$$

is the interchange operator, and

$$U = \sum_j \Pi_j \otimes U_j, \quad \text{where} \quad \Pi_j = |j\rangle\langle j|\,.$$

Quantum walks: the interchange framework.

Framework for discrete-time quantum walks and a symmetric walk on a binary tree (Phys. Rev. A, 2011) by Z. Dimcovic, D. Rockwell, I. Milligan, R. Burton, T. Nguyen, and Y. Kovchegov,
studied a DTQW on a binary tree. Here, for a non-root vertex j,

$$U_j = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & U_j^{loc} \end{array}\right], \quad \text{where } U_j^{loc} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & e^{2\pi i/3} & e^{2\pi i/3} \\ e^{2\pi i/3} & 1 & e^{2\pi i/3} \\ e^{2\pi i/3} & e^{2\pi i/3} & 1 \end{bmatrix}.$$

neighbor points

DTQW evolution: $|\psi(t+1)\rangle = UX\,|\psi(t)\rangle$ with

$X: \ |i\rangle\otimes|j\rangle \to |j\rangle\otimes|i\rangle$ and $U = \sum_j \Pi_j \otimes U_j$, where $\Pi_j = |j\rangle\langle j|$.

Quantum walks: what's missing?

- We are supposed to know the ratios of $\frac{\pi(z')}{\pi(z)}$ for any pair of neighbors z and z' on the network.

Approach: local unitary operators U_j in the interchange framework.

- Not knowing the running time.

Approach: Using adiabatic analog of simulated annealing.

On Alternating Quantum Walks (Physica A, 2017) by J. Rousseva and Y. Kovchegov

Stable adiabatic times for Markov chains (Stochastics, 2016) by K. Bradford, Y. Kovchegov, and T. Nguyen

Adiabatic times for Markov chains and applications (J. Stat. Phys., 2011) by K. Bradford and Y. Kovchegov

A note on adiabatic theorem for Markov chains (Stat.& Prob. Letters, 2010) by Y. Kovchegov