Quantum Algorithm for Link Prediction in Complex Networks

Rishi Vora (MS21113)

IISER Mohali

2025-04-12

- Complex system — A system whose collective behavior is difficult to derive from a knowledge of the system's components.
- Network science studies complex systems represented as networks (graphs) of interconnected entities.
- How these networks form, evolve, and influence behaviors across domains like social networks, biological systems, etc.
- Link prediction — forecasts where new connections might form in a network.
- Approaches for link prediction include machine learning techniques, studying global perturbations.
- Paths based methods — Predict links based on path length between different nodes.

Link prediction methods take as input a graph $G(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of nodes with size $N = |\mathcal{V}|$ and $\mathcal{E}$ is the set of undirected links, often described by the adjacency matrix $A \in \{0, 1\}^{N \times N}$ where

$$A_{ij} = A_{ji} = \begin{cases} 1 \text{ if } (i, j) \in \mathcal{E} \\ 0 \text{ otherwise} \end{cases}$$

and output a matrix of predictions $P \in \mathbb{R}^{N \times N}$ where each entry p_{ij} is a score value quantifying the likelihood of a link existing between nodes i and j.

A matrix $S \in \mathbb{R}^{N \times N}$ quantifies the similarity between any two nodes.

2.2.1 TCP based

- Predictions are based on similarity between two nodes i.e. $P = S$.
- Similarity is quantified based on the number of shared connections, i.e., paths of even lengths $(S = A^{2n})$.
- Useful in social networks where the number of common friends is a good predictor of future connections.

2.2.2 L3 based

- Predictions are based on node i being similar to the existing neighbours of another node j, i.e. $P = AS$.
- Similarity is quantified based on number of similar neighbours, i.e., paths of odd lengths $(S = A^{2n+1})$.
- Useful in biological networks like protein pairs.

- Classical methods produce likelihood score p_{ij} for all possible (i, j)

- But practically, we only need the ones which cross a certain threshold.

- We use quantum walks to encode path-based link prediction [1].

- We encode the prediction scores in the amplitudes of a superposition and perform measurements. The predictions with the highest score will be naturally sampled with higher probability.

- Utilises both even and odd path lengths.
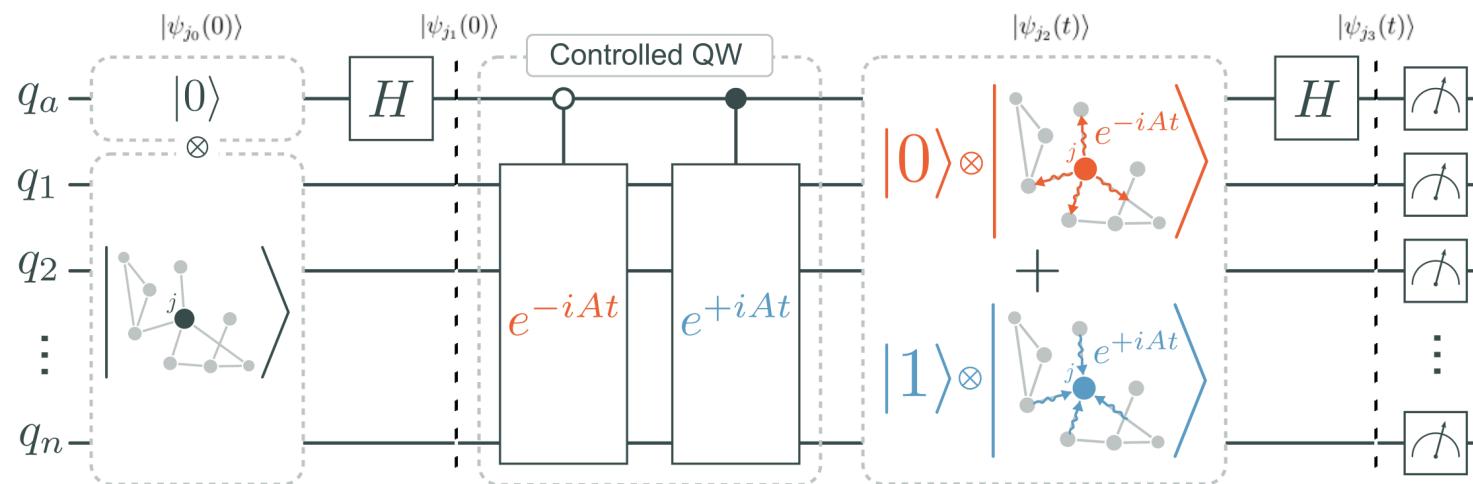
We use the continuous-time quantum walk (CTQW) formalism. The Hilbert space of the walk is defined by $\{|j\rangle\}$ where $j \in \mathcal{V}$. The Hamiltonian is the adjacency matrix A of the graph [2]. The walk itself is defined as the solution of the differential equation

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = A|\psi(0)\rangle$$

which is

$$|\psi(t)\rangle = e^{-iAt}|\psi(0)\rangle$$

3.2 The method

3.2.1 The circuit
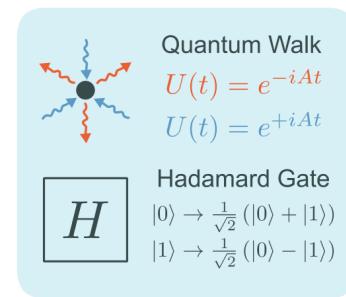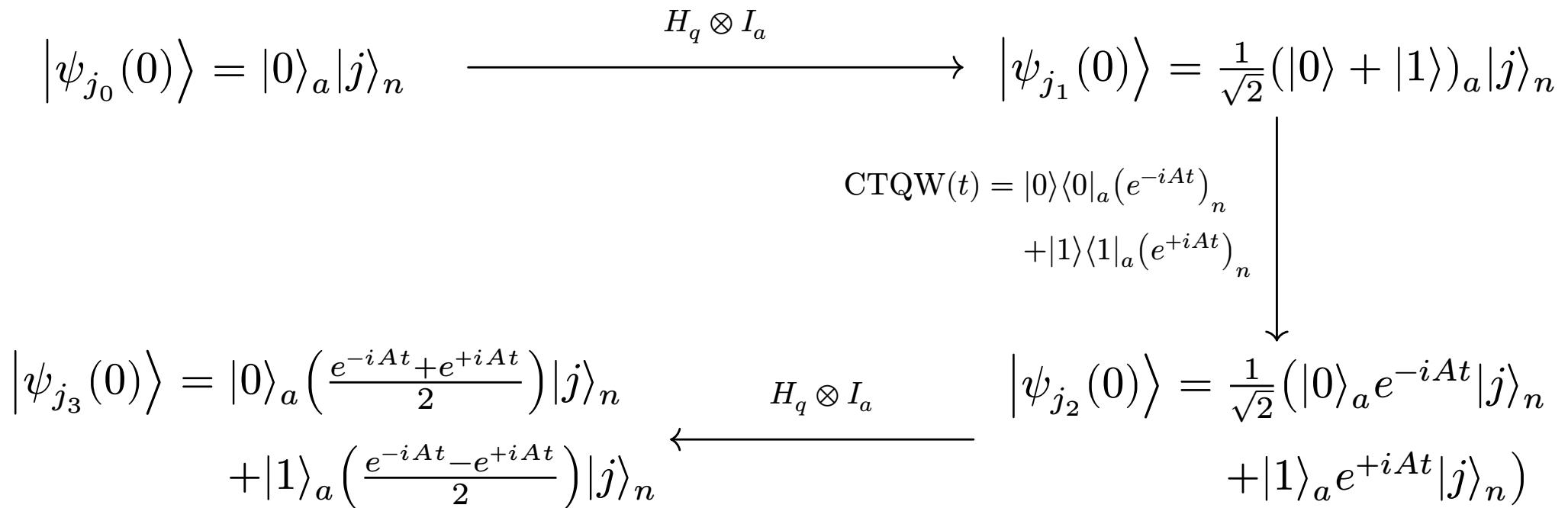
Figure 1: Circuit for Quantum Link Prediction [1]

3.2.2 The algorithm

- There are $n = \log_2 N$ qubits (q_n) to serve as binary label to each node N and an extra ancillary qubit q_a.
- The state evolves as:

$$\left|\psi_{j_0}(0)\right\rangle = |0\rangle_a |j\rangle_n \xrightarrow{\quad H_q \otimes I_a \quad} \left|\psi_{j_1}(0)\right\rangle = \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle)_a |j\rangle_n$$

$$\text{CTQW}(t) = |0\rangle\langle 0|_a \left(e^{-iAt}\right)_n + |1\rangle\langle 1|_a \left(e^{+iAt}\right)_n$$

$$\left|\psi_{j_3}(0)\right\rangle = |0\rangle_a \left(\tfrac{e^{-iAt}+e^{+iAt}}{2}\right)|j\rangle_n + |1\rangle_a \left(\tfrac{e^{-iAt}-e^{+iAt}}{2}\right)|j\rangle_n \xleftarrow{\quad H_q \otimes I_a \quad} \left|\psi_{j_2}(0)\right\rangle = \tfrac{1}{\sqrt{2}}\left(|0\rangle_a e^{-iAt}|j\rangle_n + |1\rangle_a e^{+iAt}|j\rangle_n\right)$$

The final state before measurement can be resolved as

$$|0\rangle_a \left(\sum_{k=0}^{\infty} c_{\text{even}}(k,t) A^{2k} \right) |j\rangle_n + i|1\rangle_a \left(\sum_{k=0}^{\infty} c_{\text{odd}}(k,t) A^{2k+1} \right) |j\rangle_n$$

where

$$c_{\text{even}}(k,t) = \frac{(-1)^k t^{2k}}{(2k)!}$$

$$c_{\text{odd}}(k,t) = \frac{(-1)^{k+1} t^{2k+1}}{(2k+1)!}$$

3.2.3 Obtaining results

To obtain the relevant predictions from this state, we repeatedly sample as follows: First, we measure q_1 yielding $|0\rangle$ or $|1\rangle$, collapsing the rest of the qubits to either

$$|\psi(t)\rangle_n^{\text{even}} \propto \left(\sum_{k=0}^{\infty} c_{\text{even}}(k,t) A^{2k} \right) |j\rangle \quad \text{OR} \quad |\psi(t)\rangle_n^{\text{odd}} \propto \left(\sum_{k=0}^{\infty} c_{\text{odd}}(k,t) A^{2k+1} \right) |j\rangle$$

Now when q_n is measured, the probability of yielding binary string corresponding to node i is

$$p_{ij}^{\text{even}} \propto \left\langle i \left| \sum_{k=0}^{\infty} c_{\text{even}}(k,t) A^{2k} \right| j \right\rangle \quad \text{OR} \quad p_{ij}^{\text{odd}} \propto \left\langle i \left| \sum_{k=0}^{\infty} c_{\text{odd}}(k,t) A^{2k+1} \right| j \right\rangle$$

3.2.4 Disadvantages

The immediate cons of this approach are:

- A sample may correspond to an even or odd prediction, which can only be known after measurement of q_a. So this may result in a sampling overhead (based on the requirements) and unwanted predictions will have to be discarded.
- The sample may correspond to initial node itself or an already existing link, which also needs to be discarded, resulting in more overhead.

These overhead may not be significant depending upon the degree distribution of the network and the samples taken per node.

Bibliography

[1] J. P. Moutinho, A. Melo, B. Coutinho, I. A. Kovács, and Y. Omar, "Quantum link prediction in complex networks," *Physical Review A*, vol. 107, 2023, doi: 10.1103/physreva.107.032605.

[2] R. Portugal, *Quantum Walks and Search Algorithms.* Springer International Publishing, 2018. doi: 10.1007/978-3-319-97813-0.

Thank you!

Questions?