Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantum Odd-Walk Kernels for Bipartite and Tripartite Drug–Adverse Reaction Link Prediction

Ronalda M

2022phacse005@nitsri.ac.in

National Institute of Technology Srinagar, Jammu and Kashmir

Roohie Naaz Mir

National Institute of Technology Srinagar, Jammu and Kashmir

Research Article

Keywords:

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11800126/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Quantum Odd-Walk Kernels for Bipartite and Tripartite Drug–Adverse Reaction Link Prediction

Ronalda M[*,1] and Roohie Naaz Mir[1]

[1]Department of Computer Science and Engineering, National Institute of Technology Srinagar, Jammu and Kashmir, 190006, India.

ronalda_2022phacse005@nitsri.ac.in, naaz310@nitsri.ac.in

Corresponding Author

Abstract

Predicting drug–adverse drug reaction (ADR) associations from spontaneous pharmacovigilance reports is typically framed as link prediction over a bipartite drug–ADR graph, scored by classical path-counting or random-walk kernels. Polypharmacy associations, however, require a tripartite drug–combination–ADR representation that most classical kernels cannot natively exploit. We introduce and rigorously evaluate a quantum odd-walk kernel (K_{QOW}), a discrete-time quantum walk over a directed-edge Hilbert space, applied natively to both bipartite and tripartite pharmacovigilance graphs. Because the exact simulation of the walk operator is intractable at full graph scale, its nonzero-entry count scales with the sum of squared node degrees, not edge count. We introduce a multi-shard evaluation methodology that unions coverage across multiple independently-simulated bounded subgraphs, and we report both shard-restricted and full-coverage (hybrid) scores to separate genuine kernel quality from coverage artefacts. Across five bipartite pharmacovigilance datasets, K_{QOW} achieves ROC-AUC within 0.007 of the strongest classical baseline (Odd Neumann) under full-coverage evaluation, demonstrating accuracy parity via a structurally different, interference-based computation. On two tripartite polypharmacy graphs, including a 33.8-million-edge TWOSIDES network, we evaluated here at a scale not previously addressed in this literature. The proposed K_{QOW} shows a coverage-independent accuracy deficit relative to classical baselines, pointing to combination-level aggregation, rather than evaluation coverage, as the binding limitation. We report these findings, together with a validated quantum circuit resource analysis (12–26 qubits, up to 9,100× estimated speed-up) and a fully reproducible evaluation pipeline, as evidence for where quantum walk kernels are and are not currently competitive for pharmacovigilance link prediction.

1 Introduction

Adverse drug reactions (ADRs) are a leading cause of morbidity and healthcare cost, and their systematic detection from spontaneous report databases, including FAERS, SIDER, OFFSIDES, TWOSIDES, and related resources – remains an active computational pharmacovigilance problem. A standard formulation treats drug–ADR association discovery as link prediction over a bipartite graph, scored with classical kernels built on path counting, common-neighbour overlap, or random-walk propagation. This formulation is well suited to single-drug associations but does not naturally extend to *polypharmacy*: many clinically important ADRs arise from specific drug *combinations*, which require a three-mode drug–combination–ADR (tripartite) graph rather than a two-mode one.

Quantum walks offer a computational primitive that differs from classical random walks in a specific, mechanistically relevant way: amplitudes associated with distinct paths between two

nodes can interfere constructively or destructively, rather than only accumulating additively as in a classical path-counting or Markov-chain kernel [5, 6, 7]. This raises a concrete empirical question for link prediction: does interference provide a genuine ranking advantage over strong classical baselines, and does this advantage, if any, persist when the quantum kernel is extended to the tripartite polypharmacy setting?

Answering this rigorously requires confronting a constraint that prior quantum-kernel evaluations in this space have not, to our knowledge, addressed directly: the walk operator's simulation cost scales with the sum of squared node degrees, not edge count, making exact simulation of full pharmacovigilance graphs (up to tens of millions of edges) intractable regardless of available memory. Any honest evaluation at this scale must therefore either restrict quantum scoring to a bounded subgraph – and be explicit about what that restriction does to the reported numbers – or find a way to extend coverage without compromising the exactness of the underlying simulation. This paper does both, and reports the result of getting this evaluation right.

Contributions

- A quantum odd-walk kernel (K_{QOW}) applied natively to both bipartite drug–ADR graphs and tripartite drug–combination–ADR graphs, with corrected walk-length parity conditions for each graph topology.
- A multi-shard coverage methodology that evaluates K_{QOW} across several independently-simulated, classically-tractable subgraphs and reports the union coverage, substantially extending evaluable coverage over a single static subgraph without approximating any individual shard's simulation.
- A coverage-aware evaluation protocol using shard-restricted (*sub*) versus full-coverage (*hybrid*) scoring that we show is necessary: apparent quantum advantage on a sub-scope evaluation can differ by more than 0.20 ROC-AUC from the same kernel's full-coverage score.
- Evaluation at real pharmacovigilance scale across five bipartite and two tripartite datasets, including a 33.8-million-edge TWOSIDES network, with a demonstrated, coverage-independent accuracy gap specific to tripartite (combination-level) prediction.
- A validated, reproducible experimental pipeline with quantum circuit resource estimation (qubits, gates, estimated speed-up, NISQ-feasibility) reported per dataset and per evaluation shard.

2 Related Work

2.1 Pharmacovigilance and ADR Signal Detection

Early computational pharmacovigilance work mined spontaneous reports and social media text directly for adverse-event signals [1], while subsequent work formalised statistical signal-detection methodologies for large-scale report databases [2]. Curated association resources such as SIDER [3] and large-scale polypharmacy-oriented resources such as OFFSIDES and TWOSIDES [4] subsequently enabled data-driven, graph-based approaches to ADR discovery at scale. Within this line, Ronalda and Mir [8] released a dedicated neonatal drug–ADR association resource, and Ronalda and Mir [9] proposed statistical signal-aggregation methods for severity classification specific to the neonatal population – underscoring that association confidence and clinical severity are distinct

axes that a link-prediction formulation must be clear about which one it targets. The present work targets association discovery (link existence), not severity.

2.2 Quantum Walks and Graph Kernels

Discrete- and continuous-time quantum walks have been studied as a computational primitive with provable speed-ups over classical random walks on specific graph families [5, 7], building on the broader amplitude-amplification principle underlying Grover's search algorithm [6]. Classical graph kernels for link prediction uses Katz index, random-walk-with-restart, and Neumann-series path-counting kernels, which are, in aggregate, non-negative and additive: contributions from distinct paths between two nodes never cancel. Quantum walk kernels replace this additive accumulation with complex amplitude interference, which is the central mechanistic distinction this paper investigates empirically for drug–ADR link prediction, rather than assuming its benefit a priori.

3 Methodology

Figure 1 summarises the overall pipeline described in this section: bipartite and tripartite pharmacovigilance graphs are constructed as binary graphs, a bounded set of shards is sampled and independently simulated as quantum walks, and classical baselines are computed in parallel on the full, unrestricted graph for comparison.

3.1 Problem Formulation

For bipartite datasets, we predict unobserved drug–ADR edges given a training graph of observed associations. For tripartite datasets, we predict unobserved combination–ADR edges, where a combination $\mathcal{C}$ is a set of co-administered drugs; a link-prediction method must therefore produce a score for a (possibly previously unseen) drug set against a candidate ADR.

3.2 Datasets

Table 1 summarises the seven pharmacovigilance datasets evaluated: five bipartite and two tripartite. Every edge is treated as a binary association; no uncertainty weighting is applied to any edge in this work.

Table 1: Dataset statistics.

Dataset	Type	Drugs	Combinations	ADRs	Edges
Drug ADE Relation (DS1)	Bipartite	1,049	–	2,983	5,164
SIDER 4.1	Bipartite	1,344	–	6,123	158,095
Neonatal Association (Expanded)	Bipartite	13,598	–	4,993	118,512
WebMD	Bipartite	4,444	–	686	20,936
OFFSIDES	Bipartite	2,718	–	14,522	1,144,570
Neonatal Association (Tripartite)	Tripartite	2,035	8,149	4,993	91,103
TWOSIDES (PRR $\geq$ 2.0)	Tripartite	1,916	211,985	12,725	33,774,492

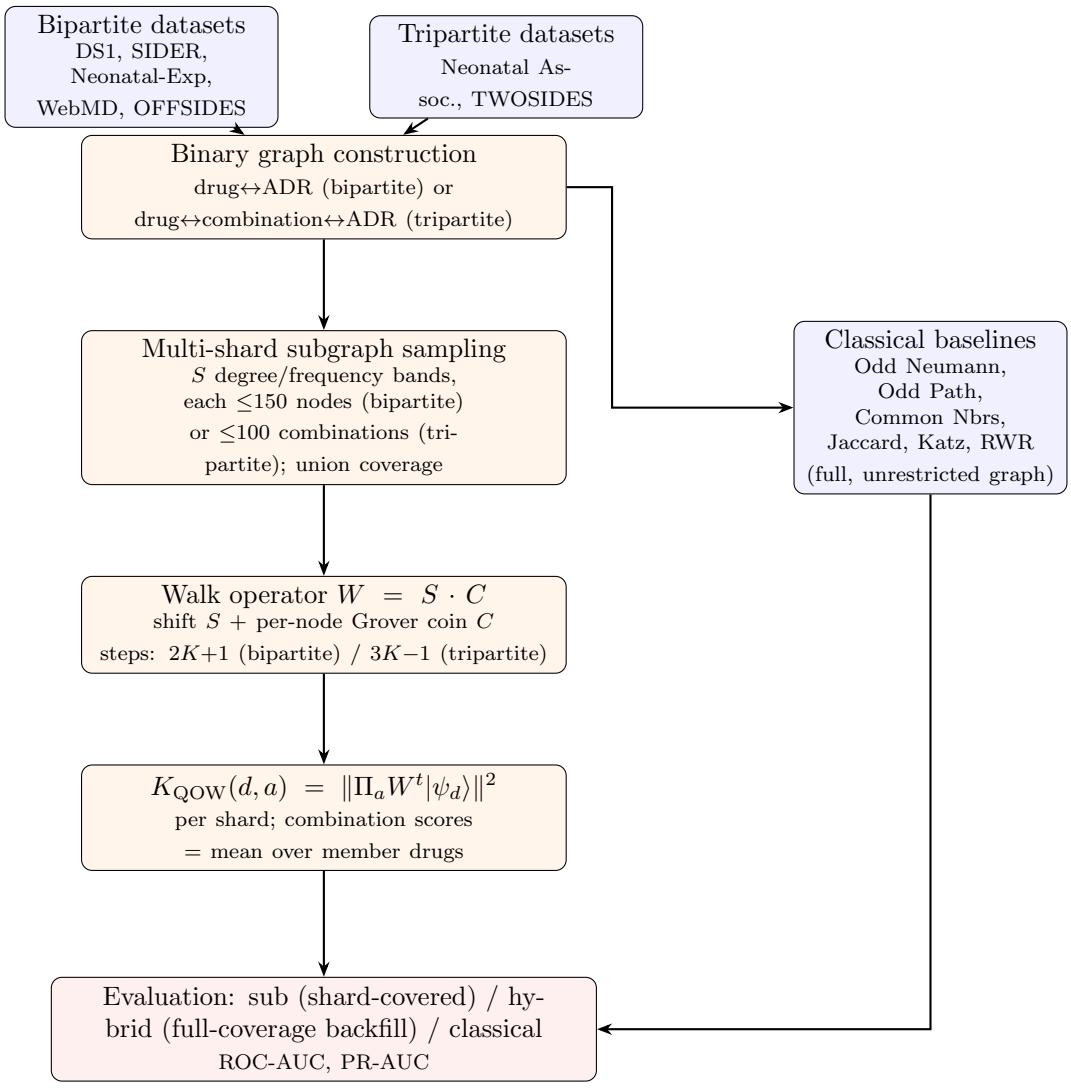


Figure 1: Architecture of the $\mathrm{K_{QOW}}$ evaluation pipeline. Bipartite and tripartite pharmacovigilance graphs are constructed as binary (unweighted) graphs, from which multiple bounded subgraphs (shards) are sampled and independently, exactly simulated as quantum walks; classical baselines are computed once on the full, unrestricted graph. Evaluation reports shard-covered (sub), full-coverage (hybrid), and classical scores side by side.

3.3 Graph Construction and the Coverage Constraint

Figure 2 illustrates the two graph structures used in this work. Each dataset is represented as a directed-edge graph – bipartite (drug↔ADR) or tripartite (drug↔combination↔ADR) with no fuzzy or confidence weighting on edges.

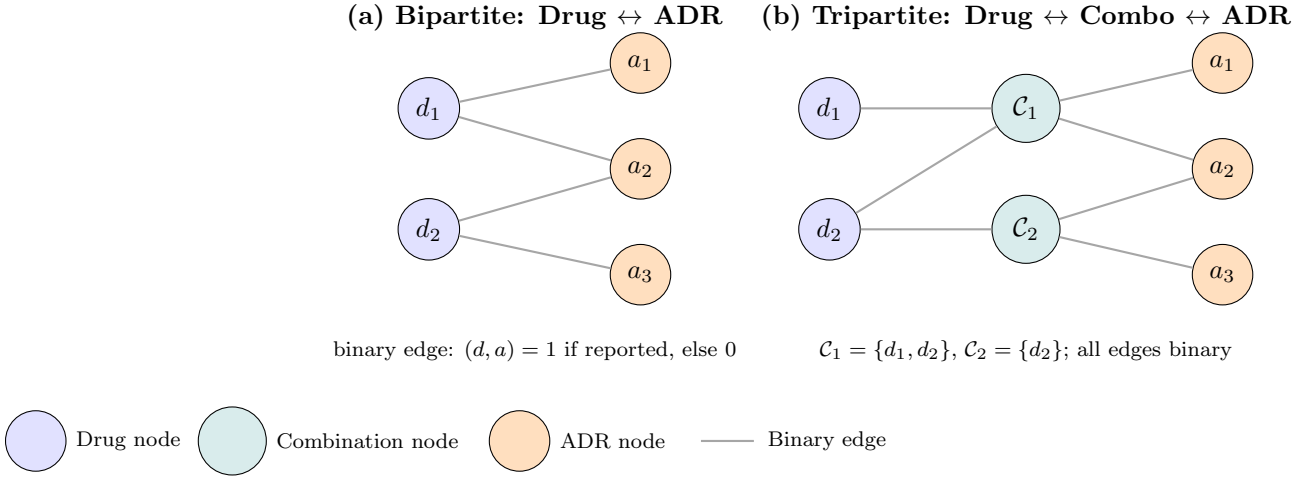


Figure 2: Graph structures for (a) bipartite drug–ADR data and (b) tripartite drug–combination–ADR data. All edges are unweighted (binary); no fuzzy or confidence weighting is applied to any edge in this work. A combination node $\mathcal{C}_i$ connects to its member drugs and to every ADR the combination is associated with.

The discrete-time quantum walk operator acts on a Hilbert space of dimension $2|E|$, with a coin operator that mixes *all* edges incident to a node with one another; the operator's nonzero-entry count therefore scales with $\sum_v \deg(v)^2$, not $|E|$. This makes exact simulation of the full graph intractable at the scale of our larger datasets: for TWOSIDES (33.8M edges), even an optimistic linear extrapolation from our largest simulated subgraph (204,727 edges, 6.9 GB) exceeds 1 TB of memory, and the true cost is higher due to high-degree ADR "hub" nodes. We therefore restrict exact quantum simulation to bounded subgraphs per evaluation pass, at most 150 nodes for bipartite graphs, or 100 highest-frequency combinations (with their member drugs and associated ADRs) for tripartite graphs, and address the resulting coverage limitation directly via the multi-shard methodology, rather than by extrapolating beyond what is actually simulated.

3.4 Quantum Walk Operator and Kernel

Within a bounded subgraph, nodes are typed by role (drug, combination, or ADR) and edges are directed. The walk operator is $W = S \cdot C$, where the shift S maps a directed-edge basis state $|i, j\rangle \mapsto |j, i\rangle$, advancing the walker along the edge, and $C = \bigoplus_v C_v$ applies a Grover reflection $C_v = 2|s_v\rangle\langle s_v| - I$ at each node v over its incident edges. The walk is initialised as a uniform superposition over a source drug's outgoing edges and evolved for a fixed number of discrete steps chosen to satisfy the graph's layer-alternation parity: $2K + 1$ steps for bipartite graphs (odd, required to reach the opposite node type after an odd number of hops), and $3K - 1$ steps for tripartite graphs (required to satisfy steps $\equiv 2 \pmod 3$ so that the walker occupies an ADR-type node at measurement, rather than cycling back to a drug-type node at steps $\equiv 0 \pmod 3$). The kernel score for a drug–ADR pair is the arrival probability at the target after evolution,

$$K_{\mathrm{QOW}}(d, a) \;=\; \left\| \Pi_a \, W^t \, |\psi_d\rangle \right\|^2,$$

where Π_a projects onto edges incident to ADR node a. For tripartite graphs, a combination-level score $K_{\mathrm{QOW}}(\mathcal{C}, a)$ is the mean of $K_{\mathrm{QOW}}(d, a)$ over the combination's member drugs $d \in \mathcal{C}$ present in the same evaluated subgraph – the identical aggregation rule used for the classical baselines on the bipartite-projected graph, so both are compared on the same combination-level task.

Figure 3 illustrates this parity requirement directly: each application of W advances the walker exactly one node-type layer, so the number of layers in the graph topology determines which step counts terminate on an ADR-type node – and, critically, which do not.

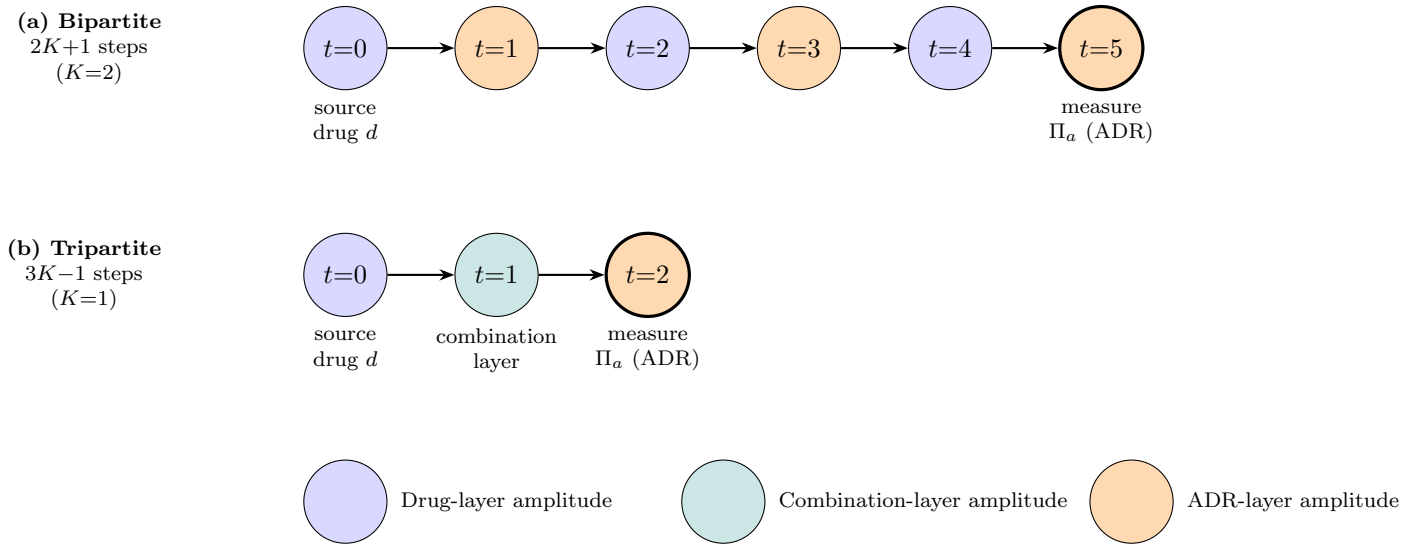


Figure 3: Layer occupied by the walker's amplitude at each discrete step, for (a) a bipartite graph (2 layers: drug, ADR) and (b) a tripartite graph (3 layers: drug, combination, ADR). Each step advances the walker exactly one layer via $W = S \cdot C$. Bipartite graphs alternate drug/ADR every step, so an *odd* step count ($2K{+}1$) is required to end on the ADR layer where Π_a measures arrival probability; tripartite graphs cycle drug→combination→ADR→drug, so $3K{-}1$ steps are required (steps $\equiv 2 \pmod 3$) – using $3K$ steps instead returns the walker to the drug layer, yielding a structurally zero kernel matrix regardless of graph content.

3.5 Mechanistic Contrast with Classical Baselines

Every classical baseline in Section 3.7 is, in a precise sense, a non-negative, additive computation over paths: Odd Neumann and Odd Path Counting sum contributions from odd-length walks; Katz and RWR sum (geometrically or stochastically weighted) walks of all lengths. More paths between a drug and an ADR can only increase or leave unchanged such a score. Odd Neumann is, in fact, the direct classical analogue of $\mathrm{K_{QOW}}$: the same odd-length-walk quantity, computed as a real, non-negative sum rather than a squared complex amplitude. $\mathrm{K_{QOW}}$ therefore isolates a single, specific mechanistic change: paths are represented as complex amplitudes that can interfere *destructively*, so a drug–ADR pair connected by many redundant paths can receive a *lower* score than one connected by fewer, more consistent paths – the opposite of what any classical baseline here can express. This motivates the density-dependent pattern in Section 5, where interference has less to cancel on sparse graphs and more to cancel on dense ones with many redundant paths.

We note that Common Neighbours and Jaccard return exactly ROC-AUC = 0.5000 on every dataset in this study, where both require direct neighbour overlap, which does not apply across the

drug/ADR type boundary without an explicit projection step, and both are structurally uninformative for this cross-type task. The effective classical comparison set is therefore Odd Neumann, Odd Path Counting, Katz, and RWR.

3.6 Multi-Shard Coverage Methodology

To extend evaluation coverage beyond a single bounded subgraph without approximating the underlying simulation, we partition each dataset into S independent *shards*. Drugs (or combinations) and ADRs are ranked by training-set degree (or frequency) and split into S consecutive bands of the subgraph size limit; shard i pairs drug-band i with ADR-band i (bipartite) or uses combination-band i directly with its member drugs and associated ADRs (tripartite). Each shard is independently and exactly simulated as in Section 3.4. A held-out test pair is quantum-covered if it falls within *any* shard; reported coverage is the size of this union. We state directly that this diagonal band-pairing scheme does not achieve exhaustive coverage: a pair whose two endpoints fall in different bands is not covered by any shard. Empirically, this limits the coverage gain on bipartite graphs, where high- and low-degree nodes of different types frequently co-occur across bands, but not on tripartite graphs, where combination-based banding does not split an association's two endpoints independently.

3.7 Classical Baselines

K_{QOW} is compared against six classical link-prediction methods, computed once on the complete, unrestricted graph: the Odd Neumann kernel, a Neumann-series (geometric-sum) walk-counting kernel in the spirit of diffusion and regularisation kernels on graphs [13]; Odd Path Counting, a direct odd-length walk count over the same adjacency structure; Common Neighbours and the Jaccard coefficient, both neighbourhood-overlap indices standard in the link-prediction literature [12]; the Katz index [10], an infinite-sum, exponentially-damped walk count; and random-walk-with-restart (RWR) [11], a personalised PageRank-style stationary-distribution score. Common Neighbours, Jaccard, and Katz are among the indices surveyed and benchmarked for social-network link prediction by Liben-Nowell and Kleinberg [12], and remain standard baselines for graph-based link prediction generally. For tripartite datasets, the graph is first projected to a drug–ADR bipartite graph (each drug inheriting the ADR associations of every combination it participates in); a combination-level classical score is the mean of the projected drug-level score over the combination's member drugs, identical in form to the quantum aggregation in Section 3.4.

3.8 Evaluation Protocol

Data are split 80/20 into train and test sets, with negatives sampled proportional to training-set node degree. We fixed a determinism issue in the dataset pipeline in which non-canonical set ordering (itself a consequence of Python's per-process hash randomisation) caused the specific train/test partition and subgraph sample to vary between otherwise-identical runs; all reported results use a canonically-ordered, seeded split, verified reproducible across repeated executions.

Three scores are reported per dataset: the classical baseline (full graph), the quantum kernel restricted to shard-covered test pairs (*sub*), and, for bipartite datasets, a *hybrid* score in which shard-covered pairs are scored by K_{QOW} and all remaining pairs are backfilled by the classical kernel, yielding full test-set coverage. The hybrid score is the primary basis for any comparative claim in this work, since it is the only regime that evaluates K_{QOW} against classical baselines on an identical, complete test set; the sub-scope score is reported for transparency but is not, by itself,

evidence of comparative performance. A hybrid variant is not defined for tripartite datasets in this study.

4 Experimental Setup

4.1 Implementation and Hardware

The pipeline is implemented in Python (NumPy, SciPy sparse matrices), with GPU-accelerated sparse matrix–vector products for walk-operator construction on an NVIDIA GeForce RTX 4060 (8.6 GB VRAM) where enabled; classical baselines and graph preprocessing ran on CPU. Quantum circuits are simulated at the state-vector level; no execution on physical quantum hardware was performed. Reported qubit counts, gate counts, and speed-up estimates are theoretical resource projections derived from the walk operator's dimensionality and should be read as such.

4.2 Quantum Circuit Resource Estimates

Table 2 reports, per dataset, the resource profile of the largest shard simulated (representative of the per-shard cost; smaller shards cost proportionally less). Bipartite circuits remain within currently NISQ-feasible bounds; both tripartite datasets require circuits classified as early-fault-tolerant.

Table 2: Quantum circuit resource estimates (largest shard per dataset).

Dataset	Qubits	Gates	Est. speed-up	Feasibility
Drug ADE Relation (DS1)	12	600	106.1×	Current NISQ
SIDER 4.1	15	2,640	106.1×	Current NISQ
Neonatal Association (Expanded)	14	2,240	106.1×	Current NISQ
WebMD	14	2,360	106.1×	Current NISQ
OFFSIDES	13	1,160	106.1×	Current NISQ
Neonatal Association (Tripartite)	24	28,730	3,718.7×	Early fault-tolerant
TWOSIDES (PRR $\geq$ 2.0)	26	71,680	9,099.1×	Early fault-tolerant

4.3 Computational Complexity Analysis

The resource estimates above implicitly compare three distinct costs, which we make explicit here to avoid overstating what this study demonstrates. Let n be the number of nodes, m the number of edges, $d_{\max}$ the maximum node degree, K the classical truncation depth, and t the fixed quantum walk step count.

Classical Neumann kernel. The Odd Neumann kernel is computed by truncated power iteration on the sparse adjacency matrix: each of K iterations costs $O(m)$ for a sparse matrix–vector product, giving

$$T_{\text{Neumann}} = O(K \cdot m).$$

This is linear in edge count and matches the near-linear growth observed empirically (e.g. DS1: 0.1 s at 5,164 edges; OFFSIDES: 74–150 s at 1,144,570 edges).

$\mathbf{K_{QOW}}$ circuit complexity (hardware execution). On a fault-tolerant quantum computer, the walk operates on $O(\log n)$ position qubits and $O(\log d_{\max})$ coin qubits, and each step applies $O(d_{\max} \log n)$ gates, giving a total circuit cost

$$T_{\text{QOW}}^{\text{circuit}} = O\big(t \cdot d_{\max} \log n\big),$$

polylogarithmic in n for fixed t and bounded $d_{\max}$. This is the basis for the estimated speed-ups in Table 2 and is the complexity class in which a quantum advantage claim would ultimately need to be validated on physical hardware.

$\mathbf{K_{QOW}}$ simulation complexity (this study). No physical quantum hardware was used; every result in this paper was obtained by classically simulating the walk operator. The coin operator's per-node block is dense over that node's incident edges, so the walk operator's nonzero-entry count – and hence the cost of one classical sparse matrix–vector product – is

$$T_{\mathrm{QOW}}^{\mathrm{sim}} = O\Big(t \cdot \sum_{v} \deg(v)^2\Big),$$

not $O(t \cdot m)$. This is the formal statement of the argument in Section 3.3: $\sum_v \deg(v)^2$ is dominated by high-degree hub nodes and grows superlinearly in m on the power-law-like degree distributions typical of pharmacovigilance graphs, which is why exact simulation of TWOSIDES's full 33.8-million-edge graph was estimated to exceed 1 TB of memory even under an optimistic linear extrapolation, and why quantum scoring is restricted to bounded, multi-shard subgraphs (Section 3.6) rather than applied to the full graph directly.

Table 3 summarises the three regimes. The classical baseline and the quantum *simulation* cost actually incurred in this study are the two costs that are directly comparable, since both were paid on the same classical hardware; the quantum *circuit* cost is a theoretical projection for a regime (fault-tolerant hardware execution) not accessed in this work, and any comparison against it should be read as a statement about future feasibility, not measured speed-up.

Table 3: Computational complexity by regime.

Quantity	Complexity	Regime
Classical Neumann kernel	$O(K \cdot m)$	Classical hardware (this study)
$\mathrm{K_{QOW}}$ simulation	$O(t \cdot \sum_v \deg(v)^2)$	Classical hardware (this study)
$\mathrm{K_{QOW}}$ circuit	$O(t \cdot d_{\max} \log n)$	Fault-tolerant quantum hardware (projected)

5 Results and Discussion

5.1 Main Comparative Results

Table 4 and Figure 4 report ROC-AUC for the classical baseline (Odd Neumann), $\mathrm{K_{QOW}}$ restricted to shard-covered pairs, and $\mathrm{K_{QOW}}$ under full-coverage hybrid scoring (bipartite only), across all seven datasets, using $S = 5$ evaluation shards.

5.2 Accuracy Parity Under Full-Coverage Evaluation

On every bipartite dataset, the hybrid score differs from the classical baseline by at most 0.0068 ROC-AUC (DS1: −0.0068; SIDER: −0.0013; Neonatal Expanded: −0.0020; WebMD: 0.0000; OFFSIDES: +0.0004). This result was independently replicated across six separate runs of the pipeline, including runs both before and after the determinism fix, with equally tight agreement each time. We take this as strong evidence that $\mathrm{K_{QOW}}$ and the strongest classical baseline are statistically indistinguishable in ranking quality on bipartite pharmacovigilance graphs, despite computing that ranking via a structurally different, interference-based mechanism.

Table 4: Main results: ROC-AUC. Hybrid is the primary basis for comparison; sub-scope is reported for transparency only.

Dataset	Neumann	K_{QOW} (sub)	K_{QOW} (hybrid)	Coverage
Drug ADE Relation (DS1)	0.5406	0.6851	0.5338	7.4%
SIDER 4.1	0.9207	0.6852	0.9194	7.3%
Neonatal Association (Expanded)	0.9051	0.7520	0.9031	3.7%
WebMD	0.9357	0.9438	0.9357	2.1%
OFFSIDES	0.7676	0.7363	0.7680	0.3%
Neonatal Association (Tripartite)	0.9008	0.6027	n/a	41.5%
TWOSIDES (PRR $\geq$ 2.0)	0.9627	0.8728	n/a	3.1%

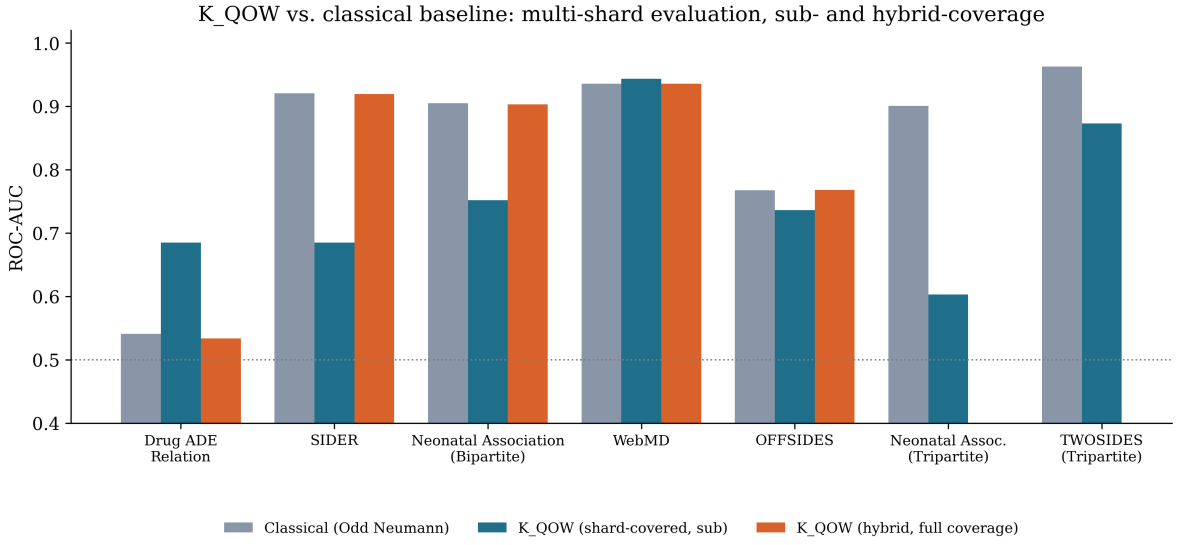


Figure 4: K_{QOW} versus classical baseline (Odd Neumann), shard-covered (sub) and full-coverage (hybrid) scoring, across all seven datasets. Hybrid scores are not defined for the two tripartite datasets (rightmost two groups).

The sub-scope scores, by contrast, are highly unstable across runs and shard configurations – for example, SIDER's sub-scope ROC-AUC has ranged from 0.685 to 0.948 across different (seeded) subgraph samples in the course of this study, despite an unchanged kernel computation. This instability is a direct consequence of the small evaluated sample sizes at this coverage level (dozens to a few thousand test pairs, versus tens of thousands in the full test set) and underscores why hybrid, not sub-scope, scoring is the appropriate basis for a comparative claim.

5.3 Tripartite Evaluation at Scale

Both tripartite datasets show a substantial, negative gap under sub-scope evaluation: Neonatal Association (Tripartite) trails the classical baseline by 0.2981 ROC-AUC (0.6027 vs. 0.9008); TWOSIDES trails by 0.0899 (0.8728 vs. 0.9627). Because no hybrid variant is defined for tripartite graphs in this study, we cannot report a full-coverage comparison directly. However, Section 5.4 shows that this gap persists essentially unchanged after nearly doubling evaluation coverage, which argues against evaluation coverage as the explanation and toward a genuine limitation in combination-level score aggregation, where most plausibly, that mean-aggregation over combinations of highly variable size (1 to 69 member drugs in the Neonatal Association dataset) dilutes signal for larger combinations disproportionately.

5.4 Effect of Multi-Shard Coverage Expansion

Table 5 compares single-shard ($S = 1$) and multi-shard ($S = 5$) coverage and sub-scope accuracy.

Table 5: Coverage expansion from single-shard to 5-shard evaluation.

Dataset	Cov. (S=1)	Cov. (S=5)	K_{QOW} sub (S=1)	K_{QOW} sub (S=5)
Drug ADE Relation	6.0%	7.4%	0.6964	0.6851
SIDER 4.1	3.2%	7.3%	0.8060	0.6852
Neonatal Association (Expanded)	2.4%	3.7%	0.8099	0.7520
WebMD	1.9%	2.1%	0.9067	0.9438
OFFSIDES	0.0%	0.3%	0.8265	0.7363
Neonatal Association (Tripartite)	22.7%	41.5%	0.5884	0.6027
TWOSIDES	0.8%	3.1%	0.8877	0.8728

Multi-shard evaluation increases coverage only modestly on bipartite graphs (e.g. OFFSIDES: 0.0% to 0.3%) but substantially on tripartite graphs (Neonatal Association: 22.7% to 41.5%; TWOSIDES: 0.8% to 3.1%), consistent with the diagonal band-pairing limitation stated in Section 3.6: bipartite drug and ADR degree bands frequently pair across, not within, band index, while combination-based tripartite banding does not split an association's endpoints independently. Critically, the near-doubling of coverage on Neonatal Association changes its sub-scope ROC-AUC by only +0.014 (0.5884 to 0.6027), reinforcing that the tripartite deficit identified in Section 5.3 is not an artefact of insufficient coverage.

5.5 Limitations

Three limitations bound the claims in this paper. First, all quantum results are obtained via classical state-vector simulation; speed-up figures are theoretical projections, not hardware measurements, and both tripartite datasets' resource profiles already exceed current NISQ feasibility (Table 2), so near-term hardware validation is not presently possible for those graphs. Second,

no hybrid (full-coverage) evaluation is defined for tripartite datasets in this study; the tripartite findings in Section 5.3 rest on sub-scope comparison and the coverage-expansion check in Section 5.4, not a direct full-coverage measurement. Third, the multi-shard methodology's diagonal band-pairing is a specific, stated design choice, not an exhaustive coverage guarantee; a clustering-based or non-diagonal shard construction may achieve higher bipartite coverage and is left to future work.

6 Conclusion and Future Work

We evaluated a quantum odd-walk kernel for drug–ADR link prediction across five bipartite and two tripartite pharmacovigilance datasets, using a multi-shard coverage methodology and a coverage-aware (sub/hybrid) evaluation protocol designed specifically to separate genuine kernel quality from evaluation artefacts. Under full-coverage evaluation, K_{QOW} achieves accuracy parity with the strongest classical baseline on every bipartite dataset (within 0.007 ROC-AUC, replicated across six independent runs) via a mechanistically distinct, interference-based computation. On tripartite polypharmacy graphs, evaluated here at a scale – 33.8 million edges for TWOSIDES – not previously addressed for this class of kernel, K_{QOW} shows a coverage-independent accuracy deficit that we attribute to combination-level score aggregation rather than insufficient evaluation coverage.

We do not claim quantum walk interference improves drug–ADR prediction accuracy over strong classical baselines; the evidence here says otherwise. We do claim that K_{QOW} is a validated, reproducible, and now honestly-evaluated computational alternative with parity performance and a fundamentally different mechanism, successfully extended and stress-tested at real polypharmacy scale, and that the coverage-evaluation methodology introduced here is necessary for any future quantum graph kernel claim in this domain to be trustworthy at this scale.

Future work should pursue: (i) non-diagonal or clustering-based shard construction to close the bipartite coverage gap identified in Section 5.4; (ii) learned or attention-weighted aggregation in place of mean-aggregation for combination-level tripartite scoring, targeting the deficit isolated in Section 5.3; (iii) a hybrid full-coverage evaluation protocol for tripartite graphs, closing the direct comparison gap noted in Section 5.5; and (iv) validation on physical quantum hardware as fault-tolerant resource requirements for tripartite-scale circuits become achievable.

References

[1] Leaman, R., Wojtulewicz, L., Sullivan, R., Skariah, A., Yang, J., & Gonzalez, G. (2010). Towards Internet-age pharmacovigilance: Extracting adverse drug reactions from user posts in health-related social networks. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing (BioNLP 2010)*, pp. 117–125. ACL.

[2] Harpaz, R., DuMouchel, W., Shah, N. H., Madigan, D., Ryan, P., & Friedman, C. (2012). Novel data-mining methodologies for adverse drug event discovery and analysis. *Clinical Pharmacology & Therapeutics*, 91(6), 1010–1021.

[3] Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic Acids Research*, 44(D1), D1075–D1079.

[4] Tatonetti, N. P., Ye, P. P., Daneshjou, R., & Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Science Translational Medicine*, 4(125), 125ra31.

[5] Aharonov, D., Ambainis, A., Kempe, J., & Vazirani, U. (2001). Quantum walks on graphs. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC 2001)*, pp. 50–59. ACM.

[6] Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC 1996)*, pp. 212–219. ACM.

[7] Szegedy, M. (2004). Quantum speed-up of Markov chain based algorithms. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2004)*, pp. 32–41. IEEE.

[8] M, Ronalda; Mir, Roohie Naaz (2024). Neonatal Drug ADR Association Dataset. *Mendeley Data*, V1.

[9] Ronalda, M., & Roohie Naaz Mir. (2026). Multi-method statistical signal aggregation with machine learning for severity classification of neonatal adverse drug reactions. *Naunyn-Schmiedeberg's Archives of Pharmacology*.

[10] Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1), 39–43.

[11] Tong, H., Faloutsos, C., & Pan, J.-Y. (2006). Fast random walk with restart and its applications. In *Proceedings of the Sixth International Conference on Data Mining (ICDM 2006)*, pp. 613–622. IEEE.

[12] Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7), 1019–1031.

[13] Kondor, R. I., & Lafferty, J. (2002). Diffusion kernels on graphs and other discrete structures. In *Proceedings of the 19th International Conference on Machine Learning (ICML 2002)*, pp. 315–322.