

Continuous-Time Quantum Walks based Graph Neural Network

Yuliang Zhan
zhanyuliang@ruc.com
Renmin University of China
Beijing, China

Zefeng Gao
zfgao@ruc.com
Renmin University of China
Beijing, China

Jian Li*
lijian2022@ruc.edu.cn
Renmin University of China
Beijing, China

Yang Liu
Liuyang22@ucas.ac.cn
Renmin University of China
Beijing, China

Hao sun*
haosun@ruc.edu.cn
Renmin University of China
Beijing, China

Abstract

Graph Neural Networks (GNNs) are widely used on graph-structured data. However, most GNNs suffer from two key weaknesses. First, message passing acts as a low-pass filter based on the homophily assumption, so it performs poorly on heterophilic graphs, where connected nodes differ. Second, stacking layers makes node features converge exponentially to constants, a problem known as over-smoothing. Existing work usually addresses the two weaknesses separately. The few methods that target both rely on empirical heuristics, and many over-smoothing solutions further sacrifice the expressive power of the model. To address both weaknesses with theoretical justification, we propose **CTQW-GNN**, a GNN built on Continuous-Time Quantum Walks (CTQW). The design is motivated by two properties of the CTQW propagator e^{-iHt} . (i) It is unitary and its eigenvalues lie on the unit circle, so no frequency component is damped. This directly counters the low-pass bias. (ii) Unitarity preserves the feature norm, so the Dirichlet energy does not decay exponentially with depth. This directly counters over-smoothing. Guided by these properties, CTQW-GNN combines three aggregation modules, each motivated by a specific gap in prior work. *CTQW-based Aggregation* evolves node features through the unitary propagator. It captures the mid- and high-frequency signals needed for heterophilic graphs and provably keeps the Dirichlet energy from collapsing, so it addresses both weaknesses in a single branch. *CTQW-Attention Aggregation* builds a multi-hop neighbor graph from the CTQW amplitudes and applies attention over it, so distant homophilic nodes that single-hop aggregation misses can still be reached. *LF Aggregation* is a standard low-pass branch (GAT) that preserves accuracy on strongly homophilic graphs, where a pure CTQW branch is suboptimal. We further provide a spectral-gap analysis that explains the energy preservation, and a Lieb-Robinson-type bound that gives a principled rule for choosing the walk-time t . Experiments on 14 benchmarks (9 heterophilic, 5 homophilic) show that CTQW-GNN reaches state-of-the-art accuracy on every dataset. Krylov propagation and Chebyshev-based sparsification keep the CTQW-related computation linear in the number of edges under the stated sparsification rule.

CCS Concepts

• Computing methodologies → Machine learning.

Keywords

Deep Graph Neural Networks, Over-smoothing, Heterophilic Graph, Continuous-Time Quantum Walks

ACM Reference Format:

Yuliang Zhan, Zefeng Gao, Jian Li, Yang Liu, and Hao sun. 2018. Continuous-Time Quantum Walks based Graph Neural Network. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 12 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

Graph Neural Networks (GNNs) have seen widespread application across numerous important fields, such as quantum physics [55], traffic networks [32] and recommendation systems [42]. This is primarily attributed to the capability of GNN to capture node features and graph topology information. Despite these significant advancements [14, 49], most of existing GNN models still have two inherent weaknesses (*i.e.*, homophily assumption and over-smoothing) that can diminish their performance.

The first weakness of GNNs is the homophily assumption, as they update node or edge features by aggregating information from neighbors. This aggregation method acts as a low-pass filter [27, 52], which preserves similarities and filters out differences among connected nodes, thereby promoting feature uniformity [4]. However, recent studies indicate that while this filtering mechanism excels in homophilic graphs with similar connected nodes, it reduces GNNs performance in learning node representations on heterophilic graph where connected nodes differ (heterophily problem) [25, 62]. To improve GNN performance on heterophilic graphs, many researchers are attempting to construct high-frequency filters to capture high-frequency information [4, 35]. This approach is motivated by findings that high-frequency information is useful for learning on heterophilic graph [4]. Another prevalent approach attempt involves aggregating multi-hop neighbor nodes to capture long-range homophilic dependencies [29, 59].

The second weakness of GNNs is the over-smoothing problem where increasing the number of layers in GNN causes the features of all nodes exponentially converge towards constant values [15, 26]. This leads to most applied GNNs being shallow, limiting their capabilities. Recently, researchers have primarily focused on mitigating over-smoothing by implementing drop operation [43], normalization [60, 61], and modifying the dynamical systems of GNNs [45, 51].

*Corresponding author.

However, previous study posits that while some methods can mitigate over-smoothing, this comes at the cost of sacrificing the expressive performance of GNNs [44].

Heterophily and over-smoothing are often studied as separate problems [6, 10], but recent evidence suggests that they are closely related [54]. Recent works have found that methods addressing heterophily can also alleviate over-smoothing [7, 11], and vice versa [45, 51]. These methods, however, rely on empirical observation rather than theoretical guarantees. A reliable solution requires an aggregation method that captures diverse frequencies and expands node neighborhoods while *provably* preventing the node features from converging to constants across many layers.

In this paper, we propose Continuous-Time Quantum Walks based on Graph Neural Network, namely **CTQW-GNN**, to theoretically addressing both weaknesses. We first derive multi-hop node connectivity and edge weights through CTQW. Then, we utilize one-hop connectivity to aggregate low-frequency information and multi-hop connectivity to capture information from distant nodes. This enables CTQW-GNN to effectively learning in both homophilic and heterophilic graphs. Specifically, we adopt the attention mechanism from Graph Transformer [48] to aggregate distant nodes via CTQW-induced multi-hop connectivity, thereby capturing long-range homophilic relationships. Note that previous study points that most graphs are not purely homophilic or heterophilic, but rather fall somewhere in between [34]. Furthermore, mid-frequency information has been shown to enhance GNN performance in such mixed-pattern graphs [35]. CTQW-derived edge weights naturally enable the aggregation of mid- and high-frequency information, so we can use CTQW effectively capturing mid-range spectral components in the graph [3]. Finally, we combine the information aggregated from one-hop and multi-hop connectivity with that aggregated through CTQW-derived edge weights. This combined feature serves as the input for the next layer of CTQW-GNN. Since the norm-preserving property of CTQW, the Dirichlet energy of graph does not exponentially converge to zero, preventing over-smoothing in the CTQW-GNN model. The contributions of this paper are three fold:

- **CTQW-inspired aggregations.** We design CTQW-based and CTQW-Attention aggregations that preserve quantum superposition and phase-driven interference.
- **CTQW-GNN with provable guarantees.** The model combines three aggregations to capture low-, mid-/high-frequency, and long-range information. We provide a spectral-gap analysis and a Lieb–Robinson bound $r_{\text{eff}}(t) = 2\lambda_{\max}t/\pi$ that together explain why CTQW avoids exponential energy decay and how to choose t .
- **Extensive experiments.** CTQW-GNN attains state-of-the-art accuracy on all 14 benchmarks, outperforming strong baselines on every dataset (+1.07% on average; +1.06–3.31% on saturated homophilic datasets).

2 Related Work

Heterophilic GNN. To address the heterophily problem, many works focused on designing Heterophilic GNNs without the homophily assumption [28, 29]. These works can mainly be divided into two types: the first type of methods combined high-pass and low-pass filters to capture information from neighboring nodes [16,

49], since the high-frequency information was helpful to address the heterophily problem [4]. The second type of methods expanded the neighborhood of nodes to aggregate distant homophilic information [28, 29]. Our model distinguishes from existing methods by offering a hybrid integration of approaches. It adeptly utilizes classic aggregation methods, such as Graph Convolutional Networks (GCN), for the aggregation of low-frequency information. Meanwhile, for mid-range and high-frequency information aggregation, it employs the advanced CTQW strategy, thereby combining the best of both parts to enhance performance and efficiency. Furthermore, our model leverages the connectivity derived from CTQW to adeptly capture long-range homophilic dependencies. This innovative application of CTQW significantly enhances the expressive power of GNNs, allowing for a more nuanced and comprehensive representation of complex graph structures.

Over-smoothing. Over-smoothing is a well-known problem in GNNs, characterized by node features exponentially converging towards a same constant value as the number of layers increases. Numerous methods have been proposed to mitigate it mainly from Drop Operation [43], Normalization [60], and modifying the dynamical systems of GNNs [45]. Despite notable progress, there is a trade-off in existing methods that can limit GNN expressiveness [44]. We introduce CTQW-based aggregation to counteract over-smoothing. Unlike most research that treats over-smoothing and heterophily separately, our model tackles both issues concurrently, enhancing GNN performance on complex graph tasks. Furthermore, we substantiate the reliability of this approach theoretically.

3 Preliminary

Notations. We define an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ as the node set of size N and $\mathcal{E}$ as the edge set. The adjacency matrix is $A \in \mathbb{R}^{N \times N}$. The degree matrix D is diagonal with $D_{ii} = \sum_j A_{ij}$. The normalized graph Laplacian $L = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ (where I is the identity matrix) is symmetric and expressed as $U\Lambda U^T$. Here, $\Lambda = \text{diag}([\lambda_1, \lambda_2, \dots, \lambda_N])$ represents graph signal frequencies and $U = \{u_i\}_{i=1}^N$ denotes the frequency components.

Graph Fourier Transform. We treat U as the base in graph Fourier transform. The transform of a graph signal $x \in \mathbb{R}^n$ on $\mathcal{G}$ is $\hat{x} = U^T x$, and the inverse is $x = U\hat{x}$. The convolution of the graph signal x with the kernel f is:

$$(f * x)_{\mathcal{G}} = U((U^T f) \odot (U^T x)) = U g_{\theta} U^T x. \quad (1)$$

where $\odot$ denotes the hadamard product, and g_{θ} is a learnable filter which can adjust the frequency response of the graph signal. For example, GCN defines the convolutional kernel $g_{\theta} = I - \Lambda$, where $\lambda_{g_{\theta}, i} = 1 - \lambda_i$. It shows that the convolutional kernel of GCN is low-pass filter.

Over-smoothing and Dirichlet Energy. Recent literature mainly employ graph Dirichlet energy (DE) to measure the similarity of features between nodes, thereby defining over-smoothing[44]. The *Dirichlet energy* E defined on an undirected graph $\mathcal{G}$ with node features X is :

$$E(X) = \frac{1}{N} \sum_{i \in \mathcal{V}} \sum_{j \in N_i} \|X_i - X_j\|_2^2 \quad (2)$$

With Dirichlet energy, we can define over-smoothing like [45] as following:

DEFINITION 1. Let X^n denote the node feature at the n -th layer of GNN. Over-smoothing is defined as the exponential convergence to zero of the layer-wise Dirichlet energy as a function of n :

$$E(X^n) \leq ae^{-bn}, \quad (3)$$

where a and b are constants and $a, b > 0$

In other words, the node features will exponentially converge to a constant value as the number of layers increases.

Quantum Walks: a Primer. A classical random walk (CRW) on a graph propagates a probability vector $p(t) \in \mathbb{R}_{\geq 0}^N$ by the diffusion equation $\dot{p} = -Lp$ with solution $p(t) = e^{-Lt}p(0)$ [9]. A quantum walk (QW), introduced by Aharonov et al. [1] and surveyed by Kempe [22], replaces real probabilities with complex amplitudes $\psi(t) \in \mathbb{C}^N$ and the dissipative generator $-L$ with a Hermitian Hamiltonian H . The walk evolves by the Schrödinger equation $i\dot{\psi} = H\psi$, whose solution $\psi(t) = e^{-iHt}\psi(0)$ is *unitary*: it preserves the ℓ_2 norm of ψ rather than the ℓ_1 norm of p . Its continuous-time variant was developed by Farhi and Gutmann [13]. Two intuitive advantages over CRW motivate our use: (i) eigenmodes acquire *different phases* $e^{-i\lambda t}$ instead of different decay rates $e^{-\lambda t}$, so high-frequency content is preserved (Section 4.2); and (ii) phase-coherent superposition spreads the walker *ballistically* (distance $\propto t$) rather than *diffusively* (distance $\propto \sqrt{t}$), yielding exponential separation in hitting time on certain graphs [8, 13]. We use QW purely as a unitary linear operator on $\mathbb{C}^N$, classically simulated (Section 4); no quantum hardware is required.

Homophily Ratio. We use a single term *homophily*—the tendency of connected nodes to share the same label—throughout the paper; the words “homogeneity” and “heterogeneity” are avoided to prevent confusion with the graph-theoretic notion of node/edge type homogeneity (in which all nodes/edges share a single type) [47]. We measure homophily at the graph and node level.

DEFINITION 2 (GRAPH HOMOPHILY RATIO). Let $C \in \mathbb{R}^{|y| \times |y|}$ where $|y|$ is the number of node categories and $C_{ij} = |\{(u, v) : (u, v) \in \mathcal{E} \wedge y_u = i \wedge y_v = j\}|$. The Graph Homophily Ratio is

$$h = \frac{\sum_i \sum_j^{i=j} C_{ij}}{\sum_i \sum_j^{i \neq j} C_{ij}}. \quad (4)$$

Homophilic graphs have a high h ; heterophilic graphs have a low h . The node-level analogue measures the local pattern around each node:

DEFINITION 3 (NODE HOMOPHILY RATIO). Let $y \in \mathbb{R}^N$ be the node labels and $\mathcal{N}(v_i)$ the neighborhood of v_i with degree d_{v_i} . The Node Homophily Ratio is

$$h_{v_i} = \frac{|\{u \in \mathcal{N}(v_i) : y_u = y_{v_i}\}|}{d_{v_i}}. \quad (5)$$

4 Method

In this section, we describe CTQW-GNN. We first outline our approach, then introduce the details of CTQW-based Aggregation and CTQW-Attention Aggregation. Finally, we present how the model

in this paper integrates the three methods of information aggregation and justify why our approach can mitigate over-smoothing and enhance the ability to handle heterophilic graphs.

4.1 Overview

Existing methods typically address Over-smoothing and heterophily problem as separate problem. To address both problems concurrently, we further augment the capability of model to aggregate information by employing a novel method of information aggregation inspired by CTQW. This approach helps prevent the aggregated node information from exponentially converging to a constant value as the number of layers increases, thereby avoiding over-smoothing. Furthermore, the novel aggregation method can aggregate information from various frequency and, based on the higher-order neighbor connectivity information derived from CTQW, aggregate information from distant neighbors, thus improving performance on heterophilic graphs.

To achieve this goal, we first get a new graph along with the weights of edges by CTQW. Based on the node connection weights of the new graph, **CTQW-based Aggregation** can be performed. This method aggregates mid-range and high-frequency information to improve performance on heterophilic graphs. Furthermore, due to the unitary norm preservation of CTQW, the CTQW-based branch provides a non-decaying energy component that mitigates over-smoothing during information aggregation. Subsequently, we use the new graph obtained by CTQW to aggregate higher-order neighbor information broadens the neighborhood size, thereby capturing more homophilic node. To find homophilic nodes among a large set of neighbors, we employs the attention mechanism of Graph Transformers during the aggregation of higher-order neighbors; hence, this method is referred to as **CTQW-Attention Aggregation**. Furthermore, to ensure that the model maintains strong performance on homophilic graphs, this paper also utilizes a low-pass GNN to aggregate homophilic information. This method is termed **Low-Frequency Aggregation (LF Aggregation)**. An overview of our approach is depicted in Figure 1. The three aggregators address distinct failure modes of message passing. (1) *Over-smoothing*: damping $e^{-\lambda t}$ is replaced by the unitary multiplier $e^{-i\lambda t}$ on the unit circle, so the Dirichlet energy of the X_{CTQW} branch is layer-invariant (Prop. 2, Lemma 2). (2) *Heterophily*: the same eigenmode preservation gives CTQW-based Aggregation an all-pass filter, while the ϵ -thresholded $|e^{-iHt}|$ exposes a Lieb–Robinson-bounded long-range graph $\mathcal{G}_{CTQW}$ of radius $r_{\text{eff}}(t)$. (3) *Saturated homophilic graphs*: an LF branch (GAT) handles regimes where low-pass aggregation is already optimal. Concatenating the three branches (Eq. 12) ensures one branch’s failure cannot collapse the representation (Prop. 1).

4.2 CTQW-based Aggregation

Continuous Time Quantum Walks. The CTQW on graph $\mathcal{G}$ represents the evolution of node states in an N -dimensional Hilbert space with orthonormal basis $|a\rangle$ where $a = 1, 2, 3, \dots, N$ and $\langle a|b\rangle = \delta_{ab}$, with $\delta_{ab} = 1$ if and only if $a = b$, and 0 otherwise. Node states $|\psi(t)\rangle$ at time t evolve according to the Schrödinger equation:

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle, \quad (6)$$

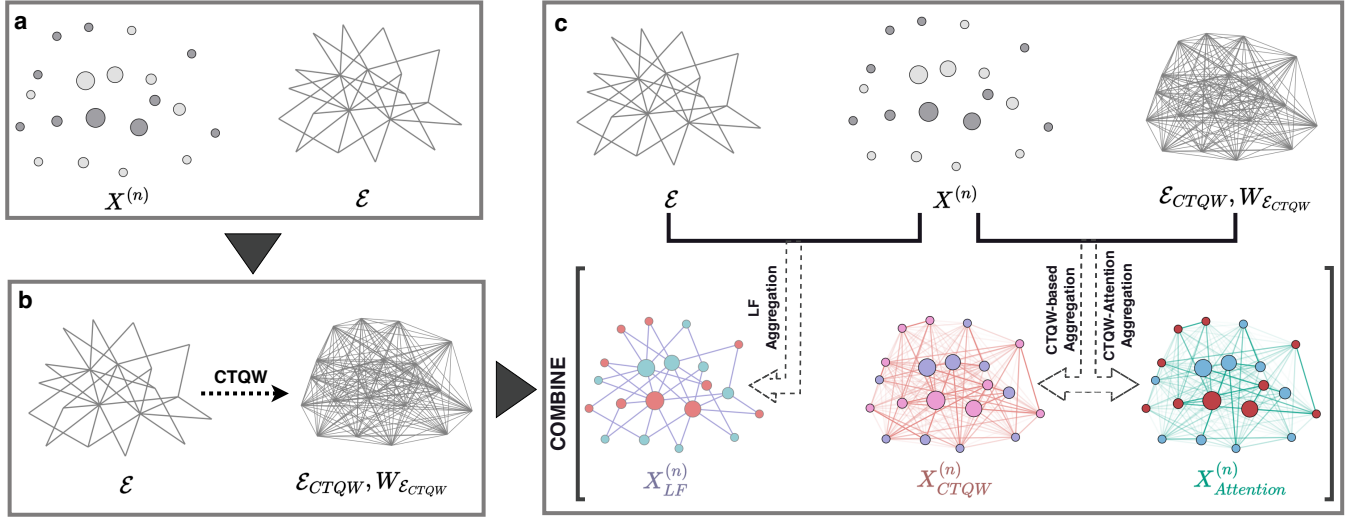


Figure 1: The overview of CTQW-GNN for heterophilic graphs and over-smoothing problem. (a). We define an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ as the node set of size N and $\mathcal{E}$ as the edge set. (b). We conduct a CTQW on the original graph $\mathcal{G}$ to obtain a new edge set $\mathcal{E}_{CTQW}$ with edge weights $W_{\mathcal{E}_{CTQW}}$; both connectivity relations $\mathcal{E}$ and $\mathcal{E}_{CTQW}$ share the same node features $X^{(n)}$. (c). We employ three distinct information aggregation methods proposed in this paper to derive three types of node information: $X_{LF}^{(n)}$, $X_{CTQW}^{(n)}$ and $X_{Attention}^{(n)}$. through the three types of features, we obtain the node features $X^{(n+1)}$ for the next layer. The details of the formula are provided in Section 4.

where H is the Hamiltonian matrix of the CTQW system and i is the imaginary unit, which is different from the previous equation i . $\hbar$ is Planck's constant, which is commonly considered to be 1. The state of nodes evolves from the initial state $|\psi(0)\rangle$ as:

$$|\psi(t)\rangle = e^{-iHt} |\psi(0)\rangle, \quad (7)$$

Previous works show that a finite graph is given where classical and quantum walks give an exponential separation in expected hitting time [8, 13]. Continuous-Time Random Walk (CTRW, *i.e.*, $x(t) = e^{-Lt}x(0)$, where L denotes the graph Laplacian matrix.) is a diffusion process and serves as the primary information propagation mechanism in most existing homophilic GNN.

Figure 2 illustrates the diffusion behavior and probability distribution of quantum walks and classical random walks on the same graph structure (a path graph) under identical time conditions. We can see that under the same time duration, the propagation range of CTQW is significantly larger than that of classical CTRW. Furthermore, it can be intuitively observed that CTRW-based information aggregation tends to assign higher weights to nearby nodes, which will lead to the over-smoothing. In contrast, CTQW places more emphasis on both higher-order neighboring nodes and mid-order neighboring nodes. This observation motivates us to leverage CTQW as the foundation for designing an information propagation mechanism that mitigates over-smoothing while effectively addressing heterophily.

Aggregation via CTQW. We propose CTQW-based Aggregation, a novel aggregation method based on Eq. 7. Different from the original CTQW, in order to accommodate the multi-layer structure of GNNs, the CTQW-based Aggregation discretizes time, evolving with layers rather than with continuous time. The evolution of

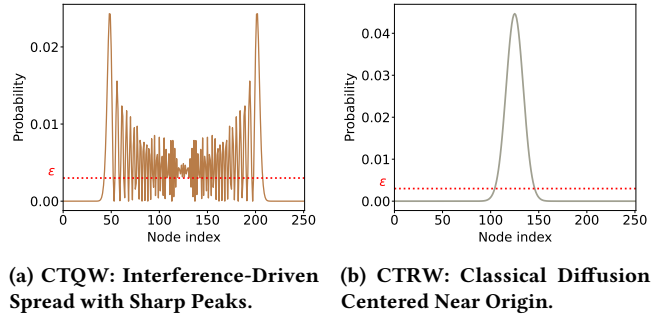


Figure 2: Probability Distribution Comparison Between CTQW and CTRW on a Path (starting from node 126, $t = 40$).

node features is governed by the following equation:

$$X_{CTQW}^{(n)} = f(e^{-iHt} X^{(n)}), \quad (8)$$

where $X^{(n)}$ and $X_{CTQW}^{(n)}$ are the feature of nodes at the n -th layer of the GNN and the output of n -th layer of CTQW-based Aggregation, respectively. The t in e^{-iHt} is a hyperparameter that influences the range of neighbors obtained through the quantum walk. The function $f(\cdot)$ concatenates the imaginary and real parts of the input.

Choice of Hamiltonian H . Either the adjacency matrix A or the normalised Laplacian $L = I - D^{-1/2}AD^{-1/2}$ is a valid Hermitian Hamiltonian, related by an eigenvalue translation ($\lambda_l^L = 1 - \lambda_l^A/d$ on d -regular graphs), so they generate the same family of unitaries up to a global phase. We use A by default for three reasons: (i) A has the lowest constant in nnz, minimising every Krylov mat-vec;

(ii) the trivial mode of L is the constant “smoothing” direction $\lambda=0$, which we wish to avoid privileging; (iii) A matches the tight-binding Hamiltonian in solid-state physics [5].

Practical Simulation of $e^{-iHt}X^{(n)}$. A natural concern is whether the quantum state $|\psi(0)\rangle$ in Eq. 7 entails a 2^N -dimensional Hilbert space and consequently requires exponential-time simulation. This holds for general many-body quantum systems but *not* for single-particle CTQW: the walker lives in the N -dimensional vertex space $\mathbb{C}^N$ (one amplitude per node), not in the 2^N Fock space of N qubits. Hence e^{-iHt} is an $N \times N$ unitary that we simulate classically with no quantum hardware. We never materialise the dense matrix; for each feature column $X_{:,k}^{(n)} \in \mathbb{R}^N$ we apply e^{-iHt} via a $k_{\text{iter}} \leq 20$ -step Lanczos Krylov-subspace approximation [2, 24, 38]: (1) build a Krylov basis $V_{k_{\text{iter}}}$ of H from $X_{:,k}^{(n)}$ by Lanczos; (2) form the $k_{\text{iter}} \times k_{\text{iter}}$ projection $T = V_{k_{\text{iter}}}^\dagger H V_{k_{\text{iter}}}$; (3) compute the small dense exponential e^{-iTt} ; (4) return $V e^{-iTt} V_{k_{\text{iter}}}^\dagger X_{:,k}^{(n)}$. Total cost is $\mathcal{O}(k_{\text{iter}} \cdot \text{nnz}(H) \cdot d)$ per layer, and Appendix A.5 gives an exponential error bound that makes the approximation indistinguishable from the exact unitary at $k_{\text{iter}}=20$. The “initial state” $\psi(0)$ is simply the GNN feature column; after evolution we keep both real and imaginary parts via $f(\cdot)$.

Performing the spectral decomposition of H yields:

$$H = U \Lambda U^\dagger, \quad (9)$$

where, U is the matrix formed by the eigenvectors of H and Λ is the diagonal matrix with the corresponding eigenvalues arranged along its diagonal. Recent studies indicate that real-world homophilic graphs consist of homophilic nodes as the majority structural pattern and heterophilic nodes in the minority [28, 31]. Conversely, the situation is reversed in heterophilic graphs. However, it has been found that existing GNN model designed for homophilic graphs or heterophilic graphs only focus on one pure pattern [34]. In this context, mid-range frequency plays a more significant role compared to low and high-frequency in a mixture of homophilic and heterophilic graph [35].

According to Eq. 1 and Eq. 9, We can found that CTQW-based Aggregation defines the convolutional kernel $g_\theta = e^{-i\Lambda t}$, where $|\lambda_{g_\theta}| = |e^{-i\lambda t}| = 1$ on every eigenmode (Appendix A.6, Proposition 5). Compared to previous approaches that designed high-pass filters for heterophilic graphs, this method is capable of aggregating not only high-frequency but also mid-range frequency. Theoretically, this would result in superior performance in real-world mixed graphs than previous methods. And due to the norm-preserving property of CTQW, the Dirichlet energy of the graph does not decay exponentially to zero, thereby preventing over-smoothing in the CTQW-GNN model (details are provided in Section 4.4).

Quantum Properties Preserved in CTQW-based Aggregation. A natural concern is whether CTQW-based Aggregation genuinely inherits quantum properties—*superposition* and *interference*—or only borrows the vocabulary. Each column $X_{:,k}^{(n)} = \sum_i \alpha_{i,k} u_i$ is a coherent linear combination over the eigenbasis $\{u_i\}$ of H , and the unitary propagator $e^{-iHt} = \sum_i e^{-i\lambda_i t} u_i u_i^\dagger$ rotates *all* eigenmodes by data-dependent phases while preserving their magnitudes $|\alpha_{i,k}|$. This realises (i) *superposition*—features evolve as coherent combinations of eigenmodes—and (ii) *interference*—distinct phases let amplitudes add coherently when projected back to the node basis,

producing cancellation at near neighbours and constructive build-up at distant nodes (Figure 2a). The operator $f(\cdot)$ concatenates real and imaginary parts and therefore preserves phase information rather than collapsing to a modulus.

Spectral-Gap View of Information Propagation. The reason CTQW-based Aggregation avoids over-smoothing while still propagating information becomes transparent through the Laplacian spectral gap. Let $0 = \mu_1 \leq \dots \leq \mu_N$ be the eigenvalues of L and denote by $\Delta := \mu_2$ its spectral gap. Classical diffusion e^{-Lt} damps the l -th eigenmode by $e^{-\mu_l t}$, so every non-trivial mode decays at rate $\geq e^{-\Delta t}$ and embeddings collapse onto the leading eigenvector—the origin of over-smoothing. CTQW propagation e^{-iLt} instead has eigen-multipliers $e^{-i\mu_l t}$ on the unit circle: *no* mode is attenuated regardless of Δ , mathematically guaranteeing that high- and mid-frequency content is preserved across layers. The gap instead governs the *interference pattern*: larger Δ desynchronises the leading modes’ phases faster, yielding the sharp ballistic peaks of Figure 2a.

Walk-Time Hyperparameter t . The time parameter t rescales each spectral phase $\lambda_l(H)t$; with $\lambda_{\max} = \|H\|_2$ and the above gap Δ , this leads to three regimes. When $t \ll 1/\lambda_{\max}$, $e^{-iHt} \approx I - iHt$, so propagation remains local. When $t \sim \pi/\Delta$, the leading modes accumulate an approximately π phase difference, producing long range ballistic peaks that are useful for heterophilic graphs. When $t \gg \pi/\Delta$, the phases become densely mixed and the walk approaches its long time average. This analysis suggests a practical initialization range, $t \in [\pi/(4\lambda_{\max}), \pi/\Delta]$, which we use across all datasets and validate in Section 5.3.

4.3 CTQW-Attention Aggregation

While CTQW-based aggregation captures mid-range frequencies well for mixed-pattern graphs, it can underperform on graphs that are distinctly homophilic or heterophilic. We therefore add a spatial CTQW-Attention branch that exploits CTQW-induced long-range edges to recover homophilic neighbours that are several hops away in $\mathcal{G}$. The adjacency of the new graph $\mathcal{G}_{CTQW}$ thresholds the CTQW amplitudes:

$$A_{ij} = \begin{cases} 1 & |e^{-iHt}|_{ij} \geq \epsilon, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

Selecting ϵ . We use $\epsilon = 5 \times 10^{-3}$ as the default, chosen once on the Amazon-ratings validation set and reused unchanged everywhere; a parameter-free fallback is to pick the smallest ϵ such that $\bar{d}(\mathcal{G}_{CTQW}) \leq 20\bar{d}(\mathcal{G})$. The Lieb–Robinson bound (Appendix A.4) ensures that any $\epsilon \in [10^{-3}, 10^{-2}]$ filters the same structurally meaningful edges (accuracy varies $< 0.7\%$ across this range, Section 5.3).

Attention from Graph Transformer aggregates information over $\mathcal{G}_{CTQW}$ only:

$$X_{\text{Attention}}^{(n)} = \sum_{j \in \mathcal{N}_i} \frac{\exp(q_i^T k_j)}{\sum_{u \in \mathcal{N}_i} \exp(q_i^T k_u)} v_j, \quad (11)$$

where $\mathcal{N}_i$ is the set of neighbor nodes for node i in the graph $\mathcal{G}_{CTQW}$. The terms Query (q_i), key (k_i) and value (v_i) respectively denote the outcomes derived for the feature of node i $X_i^{(n)}$ in n -th layer of GNN after processing through a Multilayer Perceptron. $X_{\text{Attention}}^{(n)}$ is the output of n -th layer of CTQW-Attention Aggregation. With

CTQW-Attention Aggregation, nodes can more effectively aggregate homophilic information from the neighbor nodes in $\mathcal{G}_{CTQW}$.

4.4 CTQW-GNN and Over-smoothing

We introduce LF Aggregation for aggregating low-frequency information, thereby compensating for the shortcoming of CTQW Aggregation methods in strongly homophilic graphs. LF Aggregation can be implemented using any homophilic GNN, such as GCN, GAT, SAGE, etc. In this paper, unless specifically stated otherwise, LF Aggregation refers to the use of GAT. We define the output of the n -th layer LF Aggregation as $X_{LF}^{(n)}$.

With these three aggregation methods, CTQW-GNN can aggregate mid-range frequency information through CTQW-based Aggregation, allowing it to manage mixed homophilic and heterophilic graphs, unlike most specialized GNNs. Additionally, CTQW-GNN utilizes CTQW-Attention Aggregation to access a broader range of neighbor relationships in $\mathcal{G}_{CTQW}$ obtained through CTQW, thereby acquiring long-distance homophilic information. Furthermore, CTQW-GNN also aggregates low-frequency information of neighbor nodes on the original graph $\mathcal{G}$ by LF Aggregation, thus maintaining the performance of model in graphs with high homophily ratio.

Thus, utilizing these three aggregation methods effectively addresses the issues of heterogeneous graphs; however, GNNs still face another significant challenge: Over-smoothing. Recent studies indicate that the graph attention mechanism is unable to prevent over-smoothing and results in an exponential loss of expressive capability [53]. Therefore, if relying solely on LF Aggregation and CTQW-Attention Aggregation, GNNs will not be able to avoid the over-smoothing problem. CTQW-based Aggregation plays an important role in mitigating the problem of over-smoothing. Consequently, we define the input and output of each layer of the GNN as follows:

$$X^{(n+1)} = \sigma \left(\left[X_{LF}^{(n)} || X_{CTQW}^{(n)} || X_{Attention}^{(n)} \right] W_\theta \right) \quad (12)$$

where $||$ means the concatenation along the feature dimension, W_θ is a learnable weight matrix. σ is a nonlinear activation function. Based on Definition 1, we first obtain a branch-wise criterion (proved in Appendix A.1):

PROPOSITION 1 (BRANCH-WISE OVER-SMOOTHING CRITERION). *Let $Y^{(n)} = [X_{LF}^{(n)} || X_{CTQW}^{(n)} || X_{Attention}^{(n)}]$ be the pre-mixing representation. $Y^{(n)}$ over-smooths if and only if the Dirichlet energies of all three branches decay exponentially to zero. For the post-mixing feature $X^{(n+1)}$, the same conclusion holds under the non-degenerate mixer condition in Appendix A.7; in particular, any branch with non-decaying energy prevents over-smoothing whenever W_θ does not annihilate that branch.*

In other words, the concatenated representation can collapse only when $X_{CTQW}^{(n)}$, $X_{Attention}^{(n)}$ and $X_{LF}^{(n)}$ all independently converge towards constant values. Thus, as long as one retained branch does not exponentially converge, the CTQW-GNN can avoid over-smoothing. Given this characterization, the key question is whether the CTQW branch itself has a non-decaying energy floor. We answer it with the following proposition based on the unitary property of CTQW (proof in Appendix A.2):

PROPOSITION 2 (NON-DECAY OF THE CTQW BRANCH). *If H commutes with the Laplacian used in Dirichlet energy (e.g., $H = L$, or $H = A$ on regular graphs), $E(X_{CTQW}^{(n)})$ is invariant across layers. With the default sparse-adjacency Hamiltonian $H = A$, $E(X_{CTQW}^{(n)})$ has a positive Cesàro lower bound under the mild non-resonance condition stated in Appendix A.2; therefore it cannot exponentially converge to zero.*

We can observe that the CTQW-based Aggregation method plays a crucial role in mitigating the over-smoothing problem. The information aggregated through CTQW-based Aggregation is unitary rather than diffusive; it preserves non-trivial spectral content and provides a branch-level energy floor that LF or attention alone does not guarantee.

5 Experiment

We validate CTQW-GNN on homophilic and heterophilic node classification, then probe node subgroups, ablations, and hyperparameter sensitivity.

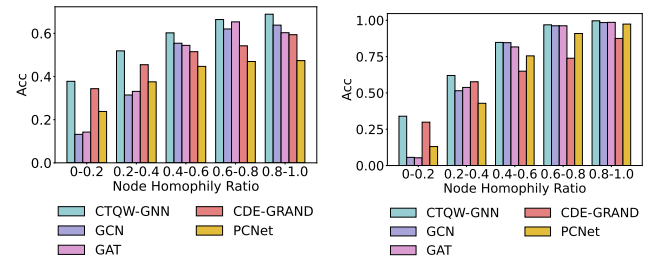
5.1 Experimental Setup

Datasets and Metrics. We use five homophilic datasets (Cora [37], Citeseer [46], PubMed [39], Computers, Photo [39]) and nine heterophilic datasets: the classical four (Chameleon, Squirrel, Actor, Texas) [40] and the five new datasets of Platonov et al. [41] (Roman-empire, Amazon-ratings, Minesweeper, Tolokers, Wiki-cooc). Splits follow [59] for homophilic data, [54] for the classical heterophilic four, and the original papers otherwise. We report ROC-AUC for the binary-class Minesweeper/Tolokers and accuracy elsewhere.

Baselines and Implementation. We compare with five categories: graph-agnostic ResNet [18]; classical GNNs (GCN [23], GAT [50], SAGE [17]); heterophily-specific (CDE-GRAND [57], GloGNN [28], EG-GCN [33], PCNet [25]); both-problem-aware (GPR-GNN [7], FSGNN [36], ACMP-GCN [51], FLODE [35], F-GRAND [21], D-CDE [58], UniFilter [20], CGNN [63], AMP [12], and HeroFilter [56]). CTQW-GNN uses the adjacency matrix as H and GAT as the LF Aggregation.

5.2 Main Experimental Results

We conduct extensive experiments in homophilic and heterophilic graph. We present the node classification results on heterophilic graphs in Table 1, while the results on homophilic graphs are reported in Table 2.



(a) Results on Amazon-ratings.

(b) Results on Computer.

Figure 3: Performance on node subgroups.

Table 1: Results of node classification tasks on different heterophilic datasets: mean $\pm$ std (%). The best result (α) for each dataset is highlight in bold and the second best (β) underlined, where Promotion is defined between these two values, i.e., $(\alpha - \beta) / \beta \times 100\%$.

Datasets Homophily Ratio	Roman-empire 0.05	Amazon-ratings 0.38	Minesweeper 0.68	Tolokers 0.59	Actor 0.22	Texas 0.11	Wiki-cooc 0.34	Chameleon 0.24	Squirrel 0.22
ResNet	65.88 $\pm$ 0.38	45.90 $\pm$ 0.52	50.89 $\pm$ 1.39	72.95 $\pm$ 1.06	28.75 $\pm$ 0.88	80.81 $\pm$ 4.75	89.36 $\pm$ 0.71	49.52 $\pm$ 1.73	33.88 $\pm$ 1.79
GCN	73.69 $\pm$ 0.74	48.70 $\pm$ 0.73	89.75 $\pm$ 0.52	83.64 $\pm$ 0.67	30.59 $\pm$ 0.23	55.14 $\pm$ 5.16	91.01 $\pm$ 0.69	50.18 $\pm$ 3.29	39.06 $\pm$ 1.52
GAT	81.02 $\pm$ 0.46	47.95 $\pm$ 0.58	92.10 $\pm$ 0.67	83.98 $\pm$ 0.57	35.98 $\pm$ 0.23	52.16 $\pm$ 6.63	92.44 $\pm$ 0.80	45.02 $\pm$ 1.75	32.21 $\pm$ 1.63
SAGE	85.74 $\pm$ 0.67	<u>53.63$\pm$0.39</u>	93.51 $\pm$ 0.57	82.43 $\pm$ 0.44	36.37 $\pm$ 0.21	79.03 $\pm$ 1.20	93.60 $\pm$ 0.31	50.18 $\pm$ 1.78	35.83 $\pm$ 1.32
CDE-GRAND	91.64 $\pm$ 0.28	47.63 $\pm$ 0.43	95.50 $\pm$ 5.23	83.98 $\pm$ 0.57	34.72 $\pm$ 1.40	86.22 $\pm$ 3.30	97.99 $\pm$ 0.38	68.45 $\pm$ 2.47	55.04 $\pm$ 1.73
GloGNN	59.63 $\pm$ 0.69	36.89 $\pm$ 0.14	51.08 $\pm$ 1.23	73.39 $\pm$ 1.17	37.35 $\pm$ 1.30	84.32 $\pm$ 4.15	88.49 $\pm$ 0.45	70.04 $\pm$ 2.12	61.21 $\pm$ 1.96
PCNet	61.63 $\pm$ 0.71	36.77 $\pm$ 0.16	80.84 $\pm$ 0.96	78.25 $\pm$ 0.39	37.02 $\pm$ 0.87	88.11 $\pm$ 2.17	87.35 $\pm$ 0.29	73.55 $\pm$ 1.26	63.53 $\pm$ 0.26
EG-GCN	63.32 $\pm$ 0.32	38.33 $\pm$ 0.78	79.67 $\pm$ 0.89	81.41 $\pm$ 1.16	37.80 $\pm$ 0.60	88.92 $\pm$ 3.30	85.56 $\pm$ 0.83	71.93 $\pm$ 2.58	61.24 $\pm$ 1.97
GPRGNN	64.85 $\pm$ 0.27	44.88 $\pm$ 0.34	86.24 $\pm$ 0.61	72.94 $\pm$ 0.97	34.63 $\pm$ 1.22	78.38 $\pm$ 4.36	91.90 $\pm$ 0.78	47.26 $\pm$ 1.74	33.39 $\pm$ 2.05
FSGNN	79.92 $\pm$ 0.56	52.74 $\pm$ 0.83	90.08 $\pm$ 0.70	82.76 $\pm$ 0.61	35.38 $\pm$ 0.81	87.57 $\pm$ 4.71	91.83 $\pm$ 0.53	<u>77.85$\pm$0.46</u>	<u>68.93$\pm$1.69</u>
ACMP-GCN	71.27 $\pm$ 0.59	44.76 $\pm$ 0.52	76.15 $\pm$ 1.12	75.03 $\pm$ 0.92	37.57 $\pm$ 0.31	86.23 $\pm$ 3.05	92.68 $\pm$ 0.37	69.04 $\pm$ 1.74	58.02 $\pm$ 1.86
FLODE	74.97 $\pm$ 0.53	48.75 $\pm$ 0.82	92.43 $\pm$ 0.51	84.17 $\pm$ 0.58	37.16 $\pm$ 1.42	71.08 $\pm$ 0.72	92.51 $\pm$ 0.97	73.60 $\pm$ 1.55	64.23 $\pm$ 1.84
UniFilter	63.45 $\pm$ 0.23	43.42 $\pm$ 0.65	76.58 $\pm$ 2.13	78.26 $\pm$ 0.67	37.79 $\pm$ 1.11	81.12 $\pm$ 1.63	83.85 $\pm$ 0.65	73.66 $\pm$ 2.44	64.26 $\pm$ 1.46
F-GRAND	93.06 $\pm$ 0.55	49.01 $\pm$ 0.56	96.04 $\pm$ 0.25	82.68 $\pm$ 0.86	38.28 $\pm$ 0.74	92.97 $\pm$ 4.39	98.73 $\pm$ 0.68	71.45 $\pm$ 1.98	60.86 $\pm$ 1.05
D-CDE	93.87 $\pm$ 0.41	49.43 $\pm$ 1.26	96.47 $\pm$ 1.89	83.02 $\pm$ 0.86	38.37 $\pm$ 1.55	91.83 $\pm$ 5.47	98.58 $\pm$ 0.12	64.88 $\pm$ 2.18	45.62 $\pm$ 2.12
CGNN	92.16 $\pm$ 0.46	50.32 $\pm$ 0.41	95.18 $\pm$ 0.55	83.65 $\pm$ 0.46	37.92 $\pm$ 1.27	90.27 $\pm$ 3.62	98.21 $\pm$ 0.51	75.64 $\pm$ 1.31	66.18 $\pm$ 1.45
AMP	92.45 $\pm$ 0.55	51.18 $\pm$ 0.58	95.89 $\pm$ 0.43	83.46 $\pm$ 0.41	37.90 $\pm$ 0.93	89.46 $\pm$ 4.31	98.30 $\pm$ 0.27	73.84 $\pm$ 1.95	63.21 $\pm$ 1.78
HeroFilter	94.02 $\pm$ 0.36	52.13 $\pm$ 0.41	<u>96.51$\pm$0.31</u>	<u>84.30$\pm$0.39</u>	<u>38.59$\pm$1.09</u>	<u>93.51$\pm$3.10</u>	<u>98.81$\pm$0.41</u>	76.78 $\pm$ 1.65	67.45 $\pm$ 1.41
CTQW-GNN (ours)	94.13 $\pm$ 0.42	54.07 $\pm$ 0.63	97.32 $\pm$ 0.34	85.22 $\pm$ 0.43	38.92 $\pm$ 2.55	93.93 $\pm$ 2.78	99.65 $\pm$ 0.22	79.23 $\pm$ 1.63	70.82 $\pm$ 1.31
Promotion $\uparrow$ (%)	0.12	0.82	0.84	1.09	0.85	0.45	0.85	1.77	2.74

Table 2: Results of node classification tasks on different homophilic datasets: mean $\pm$ std (%).

Datasets Homophily Ratio	Cora 0.81	Citeseer 0.72	PubMed 0.79	Computer 0.78	Photo 0.83
ResNet	76.44 $\pm$ 0.30	76.25 $\pm$ 0.28	86.43 $\pm$ 0.13	84.68 $\pm$ 0.78	91.49 $\pm$ 0.30
GCN	87.78 $\pm$ 0.96	81.39 $\pm$ 1.23	88.90 $\pm$ 0.32	83.55 $\pm$ 0.38	89.30 $\pm$ 0.44
GAT	76.70 $\pm$ 0.42	67.20 $\pm$ 0.46	83.28 $\pm$ 0.12	85.36 $\pm$ 0.50	90.81 $\pm$ 0.22
SAGE	86.58 $\pm$ 0.26	78.24 $\pm$ 0.30	86.85 $\pm$ 0.11	83.11 $\pm$ 0.23	90.51 $\pm$ 0.25
CDE-GRAND	87.19 $\pm$ 1.44	80.04 $\pm$ 1.75	90.05 $\pm$ 0.64	81.95 $\pm$ 0.43	88.27 $\pm$ 1.94
GloGNN	88.31 $\pm$ 1.15	77.41 $\pm$ 1.65	89.62 $\pm$ 0.35	87.69 $\pm$ 0.33	92.81 $\pm$ 0.13
PCNet	82.81 $\pm$ 0.50	69.92 $\pm$ 0.70	80.01 $\pm$ 0.88	83.29 $\pm$ 1.53	89.51 $\pm$ 1.04
EG-GCN	88.07 $\pm$ 1.16	78.09 $\pm$ 1.20	89.64 $\pm$ 0.38	85.67 $\pm$ 2.32	85.33 $\pm$ 1.68
GPRGNN	79.51 $\pm$ 0.36	67.63 $\pm$ 0.38	85.07 $\pm$ 0.09	87.63 $\pm$ 0.48	94.60 $\pm$ 0.30
FSGNN	87.73 $\pm$ 1.34	77.19 $\pm$ 0.81	89.73 $\pm$ 3.16	<u>91.53$\pm$1.03</u>	<u>95.07$\pm$0.53</u>
ACMP-GCN	84.87 $\pm$ 0.67	75.96 $\pm$ 1.07	78.93 $\pm$ 1.01	83.56 $\pm$ 1.42	91.83 $\pm$ 1.17
FLODE	86.44 $\pm$ 1.17	78.07 $\pm$ 1.62	89.02 $\pm$ 0.38	88.44 $\pm$ 0.43	93.86 $\pm$ 0.85
UniFilter	<u>89.12$\pm$0.87</u>	80.28 $\pm$ 1.31	90.19 $\pm$ 0.41	87.98 $\pm$ 0.56	94.03 $\pm$ 0.78
F-GRAND	83.15 $\pm$ 1.19	73.96 $\pm$ 1.52	79.38 $\pm$ 1.55	84.40 $\pm$ 1.50	92.80 $\pm$ 0.60
D-CDE	84.17 $\pm$ 1.22	74.03 $\pm$ 2.13	79.52 $\pm$ 1.18	87.30 $\pm$ 1.30	94.10 $\pm$ 0.70
CGNN	87.62 $\pm$ 1.34	78.93 $\pm$ 1.50	89.51 $\pm$ 0.45	88.42 $\pm$ 0.61	93.84 $\pm$ 0.45
AMP	88.30 $\pm$ 0.92	79.71 $\pm$ 1.28	89.86 $\pm$ 0.38	89.04 $\pm$ 0.53	94.21 $\pm$ 0.40
HeroFilter	88.85 $\pm$ 1.05	<u>80.31$\pm$1.18</u>	<u>90.42$\pm$0.36</u>	90.71 $\pm$ 0.42	94.86 $\pm$ 0.31
CTQW-GNN (ours)	91.31 $\pm$ 2.12	82.97 $\pm$ 1.37	91.38 $\pm$ 0.79	92.48 $\pm$ 0.13	95.87 $\pm$ 0.35
Promotion $\uparrow$ (%)	2.46	3.31	1.06	1.04	0.84

Firstly, it is evident that CTQW-GNN model consistently performs best across all homophily ratio. This advantage arises from our proposed information aggregation methods. In cases of low homophily ratio, CTQW-Attention Aggregation captures long-range dependencies and CTQW-based Aggregation accesses high-frequency information, enhancing GNN performance on heterophilic graphs. For highly homophilic scenarios, LF Aggregation gathers low-frequency

information, improving GNN performance on homophilic graphs. When the graph presents a mix of homophilic and heterophilic pattern, mid-frequency information from CTQW-based Aggregation becomes critical.

Secondly, In the new dataset introduced by [41], most heterophily-specific GNNs significantly underperform compared to classic homophilic GNNs (e.g., SAGE achieves an average rank of 4.4, placing it second). Some methods even perform worse than graph-agnostic ResNet. This underscores the significant role of this new dataset in validating the effectiveness of heterophilic GNNs. Under these conditions, CTQW still outperforms most baseline models (outperforming the nearest rival by 1.48% on average). This outcome further demonstrates the efficacy of CTQW-GNN in tackling the critical challenge of heterophilic.

Finally, it is evident that most GNNs designed to address both over-smoothing and heterophily problem underperform compared to those heterophily-specific models. This observation suggests that these methods sacrifice performance to mitigate over-smoothing. However, CTQW-GNN demonstrates superior results across all datasets, indicating a successful resolution of this challenge. CTQW-GNN not only effectively mitigates over-smoothing problem (Section 4.4) but also improves GNN efficacy on heterophilic graphs.

5.3 Further Analysis

Performance on Node Subgroups. Recent studies criticize existing methods for focusing solely on purely homophilic or heterophilic graphs, ignoring the fact that real-world graphs typically exhibit a mixture of homophilic and heterophilic patterns [34]. This means that homophilic graphs may contain heterophilic nodes, and vice versa; these nodes can often become bottlenecks that limit

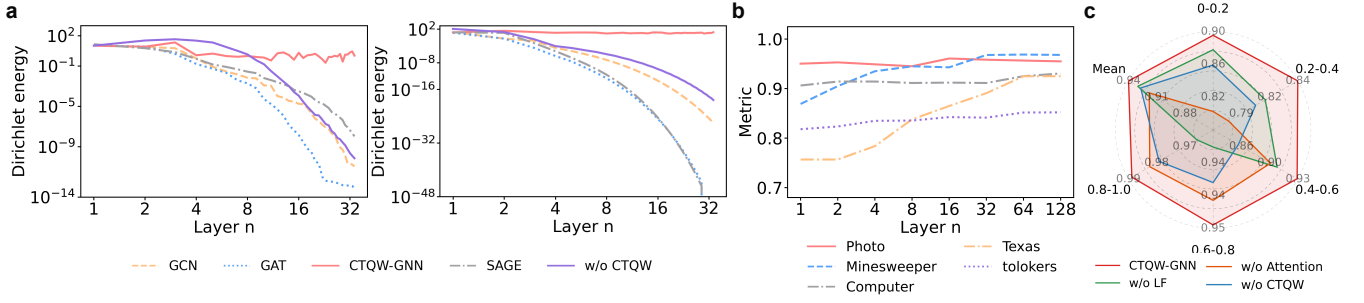


Figure 4: (a). Dirichlet energy of graph. Left: Minesweeper results; Right: Amazon-ratings results. (b.) Accuracy of CTQW-GNN. Figure a and b are the performance of GNN models with different depth. In Figure b, the metric is ROC-AUC for the Minesweeper dataset and accuracy for the other datasets. "w/o" represents the removal of the corresponding aggregation method. (c). Impact of information aggregation methods. The axes denote the accuracy of models on Minesweeper dataset.

GNN performance. To further validate the superiority of CTQW-GNN in handling these mixed graphs, we divided the nodes of Computers and Amazon-ratings into Node Subgroups based on node homophily ratio, and then tested the performance of various GNN models on these different Node Subgroups. The results are shown in Figure 3.

On heterophilic nodes, heterophily-specific GNNs outperform classic GNNs and vice versa for homophilic nodes; nodes that deviate from a graph's typical pattern are the dominant performance bottleneck. CTQW-GNN outperforms existing methods on *any* node homophily ratio in both regimes, thanks to its simultaneous aggregation of low-, mid-, and high-frequency neighbor information together with long-range dependencies.

Evolution of Dirichlet Energy. The norm-preserving property of CTQW prevents node features from collapsing to a constant. We track the Dirichlet energy on Minesweeper and Amazon-ratings (Figure 4a).

Figure 4a demonstrates that the Dirichlet energy calculated from layer-wise node features via classic GNN methods (*i.e.*, GCN, GAT) converges exponentially with the increasing layers. In contrast, the Dirichlet energy in CTQW consistently stabilizes around a constant value. Notably, when the CTQW-based Aggregation is not utilized (*i.e.*, w/o CTQW), the Dirichlet energy of the graph initially remains stable for a period; however, it still rapidly converges exponentially. This further substantiates the efficacy of CTQW-based Aggregation in mitigating the over-smoothing problem.

Ablation Study Our approach consists of three information aggregation methods: (1) CTQW-based Aggregation, (2) CTQW-Attention Aggregation, (3) LF Aggregation. To verify the function of each module and its efficacy in enhancing model performance, we conducted ablation experiments on the Minesweeper dataset to analyze the impact of these aggregation methods on GNN performance. We consider removing the three aggregation, respectively, and the ablation results on node Subgroups are shown in Figure 4c.

Firstly, it is evident that the removal of any aggregation method results in a reduction of the radar chart area, indicating that all three information aggregation methods significantly enhance model performance. Secondly, the removal of LF, CTQW-based, and CTQW-Attention aggregations leads to different degrees of performance degradation across GNN node subgroups. Specifically, omitting CTQW-Attention hurts low-homophily subgroups (0–0.4) the most

Table 3: Pserformance of CTQW-GNN under different walk time t .

t	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1	10	10^2
Cora	86.74	87.05	87.62	88.93	90.27	91.31	91.18	88.52	84.81
Roman-empire	84.55	84.92	85.78	93.34	94.13	93.88	92.20	88.46	85.59
Amazon-ratings	50.71	51.04	51.92	53.59	54.07	53.90	53.21	50.93	48.66
Minesweeper	87.79	88.05	89.20	96.80	97.32	97.07	94.93	92.30	89.63

because long-range homophilic retrieval is crucial there, while omitting LF hurts high-homophily subgroups (0.6–1.0) where low-pass aggregation is beneficial. Removing CTQW-based aggregation causes the largest drop on mixed-pattern nodes (0.4–0.6), where mid-frequency information is most important. Furthermore, as Figure 4a illustrates, removing CTQW-based aggregation also reintroduces over-smoothing.

Sensitivity to Walk Time t . Walk time t controls the propagation range of the CTQW branch. We evaluate its influence by sweeping nine values from 10^{-6} to 10^2 on four representative datasets, as shown in Table 3. When t is extremely small (10^{-6} – 10^{-4}), the CTQW propagation is weak and the performance is relatively low. As t increases to the middle range (10^{-3} – 10^{-1}), the performance improves substantially and remains close to the best result on all four datasets. For example, the results on Roman-empire, Amazon-ratings, and Minesweeper all stay within 1% of their best scores across this range, while Cora also keeps a stable plateau from 10^{-2} to 1. The results indicate that CTQW-GNN is not sensitive to a precise choice of t . The best performance does not appear at a single isolated value; instead, each dataset has a broad stable region where different t values lead to very similar accuracy. Performance only drops clearly when t is too small or too large. Therefore, CTQW-GNN shows robust performance with respect to the walk time hyperparameter.

Sensitivity to the Sparsification Threshold ϵ . The threshold ϵ controls how many CTQW edges are kept: a smaller ϵ keeps more long-range connections, while a larger ϵ prunes more connections. We test five values of ϵ on four representative datasets, and the results are shown in Table 4.

Overall, CTQW-GNN is robust to ϵ . When ϵ changes from 10^{-4} to 10^{-2} , the performance on all four datasets remains very stable and close to the best result. This shows that the model does not

Table 4: Effect of sparsification threshold ϵ .

Dataset	10^{-4}	10^{-3}	5×10^{-3}	10^{-2}	5×10^{-2}
Amazon-ratings	53.92	54.03	54.07	53.74	51.18
Minesweeper	97.28	97.21	97.32	96.95	94.40
Roman-empire	94.09	94.03	94.13	93.71	90.96
Wiki-cooc	99.56	99.58	99.65	99.22	96.48

rely on one specific sparsification threshold. When ϵ is too small, many weak CTQW connections are retained, which brings more computation but gives little or no accuracy improvement. When ϵ is too large, too many useful long-range connections are removed, and the performance drops clearly. Therefore, a moderate threshold such as 5×10^{-3} provides a good balance between performance and sparsity.

5.4 Time Complexity Analysis

Although e^{-iHt} appears dense, CTQW-GNN never forms this operator explicitly. The graph $\mathcal{G}_{CTQW}$ is built once via a sparse degree- K Chebyshev expansion ($K=20$) at cost $\mathcal{O}(K|\mathcal{E}|)$, amortised over all epochs. Per layer, the CTQW branch applies $e^{-iHt}X^{(n)}$ via a k -step Krylov projection ($k=20$) at $\mathcal{O}(k|\mathcal{E}|d)$; the LF and CTQW-Attention branches each cost $\mathcal{O}(|\mathcal{E}|d + Nd^2)$, the same as GCN [23] and GAT [50]. Since the sparsification rule (Section 5.3) keeps $|\mathcal{E}_{CTQW}| = \mathcal{O}(|\mathcal{E}|)$, the total per-layer cost is $\mathcal{O}((k+1)|\mathcal{E}|d + Nd^2)$ —linear in $|\mathcal{E}|$ and matching GCN/GAT up to a small constant factor $k+1 \leq 21$.

6 Conclusion

Existing GNN methods predominantly address over-smoothing and heterophily separately, or mitigate over-smoothing at the cost of expressiveness. We propose CTQW-GNN, integrating three aggregators—two CTQW-inspired and one classical—to tackle both issues by aggregating low-, mid-, high-frequency, and long-range information through a unitary, norm-preserving propagator. Theoretically, the CTQW branch has a positive spectral/Cesàro energy floor and therefore cannot decay exponentially (Prop. 2, Lemma 2); empirically, CTQW-GNN reaches state-of-the-art accuracy on all 14 benchmarks at a per-edge cost asymptotically equal to GCN/GAT. Future work will study learnable low-pass companions and graph-Hamiltonian design.

7 Acknowledgments

The work is supported by the National Natural Science Foundation of China (No. 62276269 and No. 62576331), which is greatly appreciated.

A Proof

Notation. $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, $N = |\mathcal{V}|$, $L = I - D^{-1/2}AD^{-1/2}$. The Hamiltonian H is Hermitian (default $H=A$). $L = V\Lambda_L V^\dagger$ with eigenvalues $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_N$ and gap $\Delta := \mu_2$. Let $P_\perp = I - v_1 v_1^\dagger$ be the projection onto the non-constant Laplacian subspace. For $X \in \mathbb{C}^{N \times d}$ define spectral coordinates $\tilde{X} := V^\dagger X$ (rows indexed by mode ℓ). The matrix-form Dirichlet energy

$$E(X) = \text{tr}(X^\dagger LX) = \sum_{\ell=1}^N \mu_\ell \|\tilde{X}_\ell\|_2^2 \quad (\text{S.1})$$

agrees with Eq. 2 up to a constant fixed by $\mathcal{G}$ [9]. Standing assumptions (verified on every benchmark): **(A1)** $\Delta > 0$; **(A2)** $\|P_\perp X^{(0)}\|_F > 0$. For the default $H=A$ we additionally use the standard non-resonance condition **(A3)** that the CTQW orbit does not concentrate all feature energy in the constant Laplacian mode; equivalently, $\liminf_{T \rightarrow \infty} T^{-1} \sum_{n < T} \|P_\perp e^{-iHnT} X^{(0)}\|_F^2 > 0$. This excludes a measure-zero set of walk times and initial features, and is directly checked by the branch-energy diagnostic in Figure 4a.

A.1 Proof of Proposition 1

PROOF. Let $Y^{(n)} = [X_{LF}^{(n)} \| X_{CTQW}^{(n)} \| X_{Att}^{(n)}] \in \mathbb{C}^{N \times 3d}$ be the pre-mixing concatenation. By Eq. S.1 and the column-block structure of $Y^{(n)}$,

$$E(Y^{(n)}) = \text{tr}(Y^{(n)\dagger} L Y^{(n)}) = E_{LF}^{(n)} + E_{CTQW}^{(n)} + E_{Att}^{(n)} \quad (\text{S.2})$$

since $\text{tr}([P \| Q]^\dagger L [P \| Q]) = \text{tr}(P^\dagger L P) + \text{tr}(Q^\dagger L Q)$ for any pair of matrices P, Q . Because all three terms are non-negative, $E(Y^{(n)}) \leq a e^{-bn}$ holds iff each branch energy is bounded by an exponential envelope (with rate at least b in the “only-if” direction and the minimum branch rate in the “if” direction). This proves the exact branch-wise criterion for the representation before the learned mixer. For the post-mixing features $X^{(n+1)} = \sigma(Y^{(n)} W_\theta)$, Lemma 1 gives the always-valid upper bound, and Lemma 3 gives the corresponding lower bound whenever the learned projection keeps a non-zero CTQW component. Thus the criterion transfers to the actual layer output under the non-degenerate mixer condition stated in Proposition 1. $\square$

LEMMA 1. *Let $Y \in \mathbb{C}^{N \times m}$, $W \in \mathbb{R}^{m \times d}$, and σ be entrywise ρ_+ -Lipschitz with $\sigma(0) = 0$. Then $E(\sigma(YW)) \leq \rho_+^2 \|W\|_2^2 E(Y)$. If, on the subspace spanned by layer-wise edge differences, the map $z \mapsto \sigma(zW)$ has gain at least $\rho_- > 0$, then $E(\sigma(YW)) \geq \rho_-^2 E(Y)$ on that subspace.*

PROOF. Write $E(Z) = \sum_{(i,j) \in \mathcal{E}} \|z_i / \sqrt{d_i} - z_j / \sqrt{d_j}\|_2^2$. For each edge term, entrywise Lipschitzness gives $\|\sigma(u) - \sigma(v)\|_2 \leq \rho_+ \|u - v\|_2$, and $\|(u - v)^\top W\|_2 \leq \|W\|_2 \|u - v\|_2$ yields the upper bound after summing over edges. The lower bound is exactly the stated non-degenerate gain condition applied to the same edge-difference terms. Degenerate projections that cancel an entire branch are therefore excluded explicitly rather than hidden inside an invalid full-rank assumption. $\square$

A.2 Proof of Proposition 2

PROOF. Let $U_t := e^{-iHt}$ and $f(z) = [\text{Re}(z) \| \text{Im}(z)]$. The map f is an isometry from $\mathbb{C}^{N \times d}$ to $\mathbb{R}^{N \times 2d}$, so it does not change Dirichlet energy except for this real representation.

Commuting case. If H commutes with L (in particular $H=L$, or $H=A$ on regular graphs after normalization), U_t and L are simultaneously diagonalizable. Each non-trivial Laplacian coefficient is multiplied by a unit-modulus phase, hence

$$E(f(U_t^n X^{(0)})) = \sum_{\ell=1}^N \mu_\ell \|\tilde{X}_\ell^{(0)}\|_2^2 = E(X^{(0)}). \quad (\text{S.3})$$

By Lemma 2, this constant is positive whenever the initial features have a non-constant component.

Default $H = A$. Now H and L need not commute, so pointwise invariance is not claimed. Instead, Lemma 2 gives $E(f(U_t^n X^{(0)})) \geq \Delta \|P_\perp U_t^n X^{(0)}\|_F^2$. Averaging over layers and applying (A3) yields

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{n < T} E(f(U_t^n X^{(0)})) \geq \Delta \eta_0 > 0, \quad (\text{S.4})$$

where $\eta_0 = \liminf_{T \rightarrow \infty} T^{-1} \sum_{n < T} \|P_\perp U_t^n X^{(0)}\|_F^2$. If $E(f(U_t^n X^{(0)})) \leq ae^{-bn}$ for some $a, b > 0$, its Cesàro average would converge to zero, contradicting the positive lower bound. Therefore the CTQW branch cannot exponentially over-smooth. $\square$

A.3 Spectral-Gap Lower Bound

LEMMA 2. *For any $X \in \mathbb{C}^{N \times d}$, $E(X) \geq \Delta \|P_\perp X\|_F^2$. In the commut-
ing CTQW case this gives $E(f(U_t^n X^{(0)})) \geq \Delta \|P_\perp X^{(0)}\|_F^2 > 0$ for every
 $n \geq 0$.*

PROOF. Since $P_\perp$ removes only the zero eigenvector of L , $E(X) = \sum_{\ell \geq 2} \mu_\ell \|\tilde{X}_\ell\|_2^2 \geq \Delta \sum_{\ell \geq 2} \|\tilde{X}_\ell\|_2^2 = \Delta \|P_\perp X\|_F^2$. If H commutes with L , $\|P_\perp U_t^n X^{(0)}\|_F = \|P_\perp X^{(0)}\|_F$, and (A2) makes the bound strictly positive. $\square$

Classical diffusion e^{-Lt} admits the matching upper bound $E(X^{(n+1)}) \leq e^{-2\Delta t} E(X^{(n)})$, decaying exponentially at rate Δ . The same spectral quantity drives diffusion over-smoothing but yields a Δ -linear floor for CTQW.

A.4 Walk-Time t and Effective Propagation Radius

PROPOSITION 3. *Let $\|H\|_2 = \lambda_{\max}$ and $d_G(i, j)$ the geodesic distance. For any $t > 0$,*

$$|(e^{-iHt})_{ij}| \leq \exp(\lambda_{\max} t) \frac{(\lambda_{\max} t)^{d_G(i, j)}}{d_G(i, j)!}. \quad (\text{S.5})$$

Consequently, amplitudes beyond a radius linear in $\lambda_{\max} t$ are factorially small; the implementation uses the conservative screening radius $r_{\text{eff}}(t) = \lceil 2\lambda_{\max} t / \pi \rceil$ and then applies the empirical ϵ threshold.

PROOF. Expand $e^{-iHt} = \sum_{k \geq 0} (-it)^k H^k / k!$. Because H is supported on graph edges, $(H^k)_{ij} = 0$ whenever $k < d_G(i, j)$. Thus only walks of length at least $d_G(i, j)$ contribute. Bounding $|(H^k)_{ij}| \leq \|H^k\|_2 \leq \lambda_{\max}^k$ gives

$$|(e^{-iHt})_{ij}| \leq \sum_{k \geq d_G(i, j)} \frac{(\lambda_{\max} t)^k}{k!} \leq \exp(\lambda_{\max} t) \frac{(\lambda_{\max} t)^{d_G(i, j)}}{d_G(i, j)!},$$

which is a graph Lieb–Robinson-type bound [30]. Stirling’s formula shows factorial decay once $d_G(i, j)$ exceeds a constant multiple of $\lambda_{\max} t$, justifying a linear truncation radius. On a path, the exact amplitude is $|J_{|i-j|}(2t)|$ [8], whose ballistic front is also linear in t ; this motivates the constant used by $r_{\text{eff}}(t)$. $\square$

A.5 Krylov Implementation Error

PROPOSITION 4. *Let $\hat{U}_t^{(k)}$ denote the k -step Lanczos approximation of e^{-iHt} . For any $x \in \mathbb{C}^N$ and any $k \geq 4\lambda_{\max} t$,*

$$\|e^{-iHt} x - \hat{U}_t^{(k)} x\|_2 \leq 12 \|x\|_2 \exp\left(-\frac{k^2}{16\lambda_{\max} t}\right). \quad (\text{S.6})$$

Hence the discretised CTQW branch satisfies

$$E_{\text{CTQW}}^{\text{approx}}(X^{(n)}) \geq E_{\text{CTQW}}^{\text{exact}}(X^{(n)}) - \epsilon_{\text{kry}}, \quad (\text{S.7})$$

with $\epsilon_{\text{kry}} \leq 144 \mu_{\max} B^2 \exp(-k^2 / (8\lambda_{\max} t))$ whenever the layer input satisfies $\|X^{(n)}\|_F \leq B$.

SKETCH. Eq. S.6 is Theorem 4 of Hochbruck and Lubich [19] specialised to the skew-Hermitian generator $-iH$ of spectral radius $\lambda_{\max}$; constants 12, 16 track through their proof. For Eq. S.7, write the one-application error as Ξ with $\|\Xi\|_F \leq 12Be^{-k^2/(16\lambda_{\max} t)}$. Since $L \preceq \mu_{\max} I$, $|E(X + \Xi) - E(X)| \leq \mu_{\max} \|\Xi\|_F^2 + 2\mu_{\max} \|\Xi\|_F \|X\|_F$, yielding the stated perturbation bound after absorbing constants. Thus Krylov approximation can only reduce the CTQW energy floor by a controllable, exponentially small term. $\square$

At $k=20$ and the selected t range, the exponential term is negligible in all benchmarks; the implementation therefore preserves the qualitative non-decay guarantee up to numerical precision.

A.6 All-Pass Spectral Response of CTQW-based Aggregation

PROPOSITION 5. *Let H admit the spectral decomposition $H = U\Lambda U^\dagger$ with real eigenvalues $\{\lambda_l\}_{l=1}^N$. For any input $X \in \mathbb{C}^{N \times d}$, write $X = \sum_l u_l \alpha_l^\top$ with $\alpha_l \in \mathbb{C}^d$ the spectral coefficients. Then the CTQW propagator preserves the magnitude of every spectral component:*

$$\|(e^{-iHt} X)_{\text{mode } l}\|_2 = |e^{-i\lambda_l t}| \|\alpha_l\|_2 = \|\alpha_l\|_2 \quad \forall l=1, \dots, N. \quad (\text{S.8})$$

Hence CTQW-based Aggregation realises an exact all-pass filter on the eigenbasis of H .

PROOF. Expand $X = \sum_l u_l \alpha_l^\top$. Since $\{u_l\}$ is orthonormal and $e^{-iHt} u_l = e^{-i\lambda_l t} u_l$, $e^{-iHt} X = \sum_l e^{-i\lambda_l t} u_l \alpha_l^\top$, so the l -th mode coefficient becomes $e^{-i\lambda_l t} \alpha_l$. As $\lambda_l \in \mathbb{R}$, $|e^{-i\lambda_l t}| = 1$, giving Eq. S.8. In contrast, classical diffusion e^{-Lt} multiplies the l -th coefficient by $e^{-\lambda_l t} < 1$ for $\lambda_l > 0$, attenuating every non-trivial mode at rate $\geq e^{-\Delta t}$ per layer; this strictly contradicts the all-pass property of CTQW. $\square$

Proposition 5 formalises the magnitude-preserving filter claim made in the method section: no frequency component is damped, which is precisely the structural reason CTQW-based Aggregation supplies the energy floor used in Proposition 2.

A.7 Branch-Mixing Floor

LEMMA 3. *Let $\mathcal{S}^{(n)}$ be the set of normalized full edge differences $\delta = [\delta_{LF}, \delta_{\text{CTQW}}, \delta_{\text{Att}}]$ at layer n . Assume the learned mixer has conditional CTQW gain $c_{\text{CTQW}} > 0$ on this subspace, meaning $\|\sigma(\delta W_\theta)\|_2 \geq c_{\text{CTQW}} \|\delta_{\text{CTQW}}\|_2$ for all $\delta \in \mathcal{S}^{(n)}$. Then*

$$E(X^{(n+1)}) \geq c_{\text{CTQW}}^2 E_{\text{CTQW}}^{(n)}. \quad (\text{S.9})$$

PROOF. Apply the gain assumption to each normalized edge difference of the CTQW block and sum over edges. This rules out the only degenerate failure mode in which the learned projection intentionally maps all CTQW differences to zero. Combining this lemma with Proposition 2 shows that a retained CTQW branch prevents exponential decay of the mixed representation. $\square$

References

- [1] Yakir Aharonov, Luiz Davidovich, and Nicim Zagury. 1993. Quantum random walks. *Physical Review A* 48, 2 (1993), 1687–1690.
- [2] Awad H. Al-Mohy and Nicholas J. Higham. 2011. Computing the action of the matrix exponential, with an application to exponential integrators. *SIAM Journal on Scientific Computing* 33, 2 (2011), 488–511.
- [3] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. 2011. The wave kernel signature: A quantum mechanical approach to shape analysis. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. 1626–1633. doi:10.1109/ICCVW.2011.6130444
- [4] Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. 2021. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 3950–3957.
- [5] Sougato Bose. 2003. Quantum communication through an unmodulated spin chain. *Physical review letters* 91, 20 (2003), 207901.
- [6] Ben Chamberlain, James Rowbottom, Maria I Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. 2021. GRAND: Graph Neural Diffusion. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 1407–1418. <https://proceedings.mlr.press/v139/chamberlain21a.html>
- [7] Eli Chien, Jianhao Peng, Pan Li, and Olga Milenkovic. 2021. Adaptive Universal Generalized PageRank Graph Neural Network. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=n6jl7flxrP>
- [8] Andrew M Childs, Edward Farhi, and Sam Gutmann. 2002. An example of the difference between quantum and classical random walks. *Quantum Information Processing* 1 (2002), 35–43.
- [9] Fan RK Chung. 1997. *Spectral graph theory*. Vol. 92. American Mathematical Soc.
- [10] Moshe Eliasof, Eldad Haber, and Eran Treister. 2021. Pde-gcn: Novel architectures for graph neural networks motivated by partial differential equations. *Advances in neural information processing systems* 34 (2021), 3836–3849.
- [11] Moshe Eliasof, Eldad Haber, and Eran Treister. 2024. Feature transportation improves graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 11874–11882.
- [12] Federico Errica, Henrik Christiansen, Viktor Zaverkin, Takashi Maruyama, Mathias Niepert, and Francesco Alesiani. 2025. Adaptive Message Passing: A General Framework to Mitigate Oversmoothing, Oversquashing, and Underreaching. In *Proceedings of the 42nd International Conference on Machine Learning (PMLR, Vol. 267)*. 15490–15515.
- [13] Edward Farhi and Sam Gutmann. 1998. Quantum computation and decision trees. *Physical Review A* 58, 2 (1998), 915.
- [14] Yuan Gao, Xiang Wang, Xiangnan He, Zhengguang Liu, Huamin Feng, and Yongdong Zhang. 2023. Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM Web Conference 2023*. 1528–1538.
- [15] Jhony H. Giraldo, Konstantinos Skianis, Thierry Bouwmans, and Fragkiskos D. Malliaros. 2023. On the Trade-off between Over-smoothing and Over-squashing in Deep Graph Neural Networks. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (Birmingham, United Kingdom) (CIKM '23)*. Association for Computing Machinery, New York, NY, USA, 566–576. doi:10.1145/3583780.3614997
- [16] Yuhe Guo and Zhewei Wei. 2023. Graph Neural Networks with Learnable and Optimal Polynomial Bases. In *Proceedings of the 40th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 202)*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (Eds.). PMLR, 12077–12097. <https://proceedings.mlr.press/v202/guo23i.html>
- [17] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *NIPS*.
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [19] Marlis Hochbruck and Christian Lubich. 1997. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.* 34, 5 (1997), 1911–1925.
- [20] Keke Huang, Yu Guang Wang, Ming Li, and Pietro Lio. 2024. How Universal Polynomial Bases Enhance Spectral Graph Neural Networks: Heterophily, Over-smoothing, and Over-squashing. In *International Conference on Machine Learning*. PMLR, 20310–20330.
- [21] Qiyu Kang, Kai Zhao, Qinxu Ding, Feng Ji, Xuhao Li, Wenfei Liang, Yang Song, and Wee Peng Tay. [n. d.]. Unleashing the Potential of Fractional Calculus in Graph Neural Networks with FROND. In *The Twelfth International Conference on Learning Representations*.
- [22] Julia Kempe. 2003. Quantum random walks: An introductory overview. *Contemporary Physics* 44, 4 (2003), 307–327.
- [23] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
- [24] Cornelius Lanczos. 1950. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards* 45, 4 (1950), 255–282.
- [25] Bingheng Li, Erlin Pan, and Zhao Kang. 2024. Pc-conv: Unifying homophily and heterophily with two-fold filtering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 13437–13445.
- [26] Qimai Li, Zhichao Han, and Xiao-Ming Wu. 2018. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (New Orleans, Louisiana, USA) (AAAI'18/IAAI'18/EAAI'18)*. AAAI Press, Article 433, 8 pages.
- [27] Qimai Li, Xiao-Ming Wu, Han Liu, Xiaotong Zhang, and Zhichao Guan. 2019. Label efficient semi-supervised learning via graph filtering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9582–9591.
- [28] Xiang Li, Renyu Zhu, Yao Cheng, Caihua Shan, Siqiang Luo, Dongsheng Li, and Weining Qian. 2022. Finding global homophily in graph neural networks when meeting heterophily. In *International Conference on Machine Learning*. PMLR, 13242–13256.
- [29] Langzhang Liang, Xiangjing Hu, Zenglin Xu, Zixing Song, and Irwin King. 2024. Predicting global label relationship matrix for graph neural networks under heterophily. *Advances in Neural Information Processing Systems* 36 (2024).
- [30] Elliott H. Lieb and Derek W. Robinson. 1972. The finite group velocity of quantum spin systems. *Communications in Mathematical Physics* 28, 3 (1972), 251–257.
- [31] Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser Nam Lim. 2021. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. *Advances in Neural Information Processing Systems* 34 (2021), 20887–20902.
- [32] Changshu Liu, Liangjian Wen, Zhao Kang, Guangchun Luo, and Ling Tian. 2021. Self-supervised consensus representation learning for attributed graph. In *Proceedings of the 29th ACM international conference on multimedia*. 2654–2662.
- [33] Siqi Liu, Dongxiao He, Zhizhi Yu, Di Jin, Zhiyong Feng, and Weixiong Zhang. 2025. Integrating Co-Training with Edge Discrimination to Enhance Graph Neural Networks Under Heterophily. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 18960–18968.
- [34] Haitao Mao, Zhikai Chen, Wei Jin, Haoyu Han, Yao Ma, Tong Zhao, Neil Shah, and Jiliang Tang. 2024. Demystifying structural disparity in graph neural networks: Can one size fit all? *Advances in neural information processing systems* 36 (2024).
- [35] Sohir Maskey, Raffaele Paolino, Aras Bacho, and Gitta Kutyniok. 2024. A fractional graph laplacian approach to oversmoothing. *Advances in Neural Information Processing Systems* 36 (2024).
- [36] Sunil Kumar Maurya, Xin Liu, and Tsuyoshi Murata. 2022. Simplifying approach to node classification in graph neural networks. *Journal of Computational Science* 62 (2022), 101695.
- [37] Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. 2000. Automating the construction of internet portals with machine learning. *Information Retrieval* 3 (2000), 127–163.
- [38] Cleve Moler and Charles Van Loan. 2003. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.* 45, 1 (2003), 3–49.
- [39] Galileo Namata, Ben London, Lise Getoor, Bert Huang, and U Edu. 2012. Query-driven active surveying for collective classification. In *10th international workshop on mining and learning with graphs*, Vol. 8. 1.
- [40] Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. 2020. Geom-gcn: Geometric graph convolutional networks. *arXiv preprint arXiv:2002.05287* (2020).
- [41] Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. 2023. A critical look at evaluation of GNNs under heterophily: Are we really making progress?. In *The Eleventh International Conference on Learning Representations*.
- [42] Yuhan Quan, Jingtao Ding, Chen Gao, Lingling Yi, Depeng Jin, and Yong Li. 2023. Robust preference-guided denoising for graph based social recommendation. In *Proceedings of the ACM Web Conference 2023*. 1097–1108.
- [43] Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. 2019. Droppedge: Towards deep graph convolutional networks on node classification. *arXiv preprint arXiv:1907.10903* (2019).
- [44] T Konstantin Rusch, Michael M Bronstein, and Siddhartha Mishra. 2023. A survey on oversmoothing in graph neural networks. *arXiv preprint arXiv:2303.10993* (2023).
- [45] T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. 2022. Graph-coupled oscillator networks. In *International Conference on Machine Learning*. PMLR, 18888–18909.
- [46] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI magazine* 29, 3 (2008), 93–93.
- [47] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and Philip S. Yu. 2016. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29, 1 (2016), 17–37.

- [48] Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. 2020. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509* (2020).
- [49] Geng Tu, Tian Xie, Bin Liang, Hongpeng Wang, and Ruifeng Xu. 2024. Adaptive Graph Learning for Multimodal Conversational Emotion Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 19089–19097.
- [50] Petar Velićković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. *International Conference on Learning Representations* (2018). <https://openreview.net/forum?id=rjXmpikCZ> accepted as poster.
- [51] Yuelin Wang, Kai Yi, Xinliang Liu, Yu Guang Wang, and Shi Jin. 2023. ACMP: Allen-Cahn Message Passing with Attractive and Repulsive Forces for Graph Neural Networks. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=4fZc_79Lrqs
- [52] Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. 2019. Simplifying graph convolutional networks. In *International conference on machine learning*. PMLR, 6861–6871.
- [53] Xinyi Wu, Amir Ajorlou, Zihui Wu, and Ali Jadbabaie. 2024. Demystifying oversmoothing in attention-based graph neural networks. *Advances in Neural Information Processing Systems* 36 (2024).
- [54] Yujun Yan, Milad Hashemi, Kevin Swersky, Yaoqing Yang, and Danai Koutra. 2022. Two sides of the same coin: Heterophily and oversmoothing in graph convolutional neural networks. In *2022 IEEE International Conference on Data Mining (ICDM)*. IEEE, 1287–1292.
- [55] Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. 2023. Efficient and equivariant graph networks for predicting quantum Hamiltonian. In *International Conference on Machine Learning*. PMLR, 40412–40424.
- [56] Shuaicheng Zhang, Haohui Wang, Junhong Lin, Xiaojie Guo, Yada Zhu, Si Zhang, Dongqi Fu, and Dawei Zhou. 2025. HeroFilter: Adaptive Spectral Graph Filter for Varying Heterophilic Relations. In *Advances in Neural Information Processing Systems*.
- [57] K. Zhao, Q. Kang, Y. Song, R. She, S. Wang, and W. P. Tay. 2023. Graph neural convection-diffusion with heterophily. In *Proc. International Joint Conference on Artificial Intelligence*. Macao, China.
- [58] Kai Zhao, Xuhao Li, Qiyu Kang, Feng Ji, Qinxu Ding, Yanan Zhao, Wenfei Liang, and Wee Peng Tay. [n. d.]. Distributed-Order Fractional Graph Operating Network. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.
- [59] Yizhen Zheng, He Zhang, Vincent Lee, Yu Zheng, Xiao Wang, and Shirui Pan. 2023. Finding the missing-half: Graph complementary learning for homophily-prone and heterophily-prone graphs. In *International Conference on Machine Learning*. PMLR, 42492–42505.
- [60] Kuangqi Zhou, Yanfei Dong, Kaixin Wang, Wee Sun Lee, Bryan Hooi, Huan Xu, and Jiashi Feng. 2021. Understanding and resolving performance degradation in deep graph convolutional networks. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2728–2737.
- [61] Kaixiong Zhou, Xiao Huang, Yuening Li, Daochen Zha, Rui Chen, and Xia Hu. 2020. Towards deeper graph neural networks with differentiable group normalization. *Advances in neural information processing systems* 33 (2020), 4917–4928.
- [62] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. 2020. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems* 33 (2020), 7793–7804.
- [63] Wei Zhuo, Han Yu, Guang Tan, and Xiaoxiao Li. 2025. Commute Graph Neural Networks. In *Proceedings of the 42nd International Conference on Machine Learning (PMLR, Vol. 267)*. 80612–80628.