Supplementary Material for MGKAN

Anonymous submission

Performance Comparison

We evaluate MGKAN on the specific-direction (Task 1) and asymmetric-direction (Task 2) prediction tasks on the DS 1 and DS 2 datasets, as reported in Tables 1 and 2. To assess efficiency, we include a runtime efficiency analysis in **Supplementary Material**, which shows that MGKAN maintains competitive runtime performance.

Task 1: specific-direction prediction. As shown in Table 1, MGKAN consistently outperforms all baselines on Task 1. While directed graph learning methods exhibit strong structural modeling capabilities, they achieve lower AUROC and F1, indicating that structural information alone is insufficient for accurate asymmetric DDI prediction. On the other hand, existing asymmetric DDI baselines primarily rely on the DDI network, limiting their ability to capture complex interaction patterns. MGKAN addresses these limitations by integrating multiple views—including co-interaction and similarity networks—which provide complementary topological and semantic information. For instance, drugs without direct interactions can still be indirectly connected through co-interaction networks, while the similarity view introduces auxiliary biochemical features. This multi-view architecture enables MGKAN to generate more expressive and discriminative embeddings than models restricted to a single view.

Task 2: asymmetric-direction prediction. In contrast, Task 2 is more challenging, as it requires precise identification of the interaction direction. As shown in Table 2, MGKAN achieves the best performance on both datasets, improving accuracy by 2.02% on DS 1 and 3.85% on DS 2 over the strongest baseline, DGAT-DDI. While baselines such as DGAT-DDI and DRGATAN incorporate directional message passing, MGKAN enhances directional modeling by integrating KANs into both the message passing and fusion stages. Compared to conventional MLPs used in other GNN baselines, the learnable basis functions in KANs allow MGKAN to capture nonlinear directional patterns more effectively. In addition, its complementary fusion module integrates multi-view embeddings, further enriching the learned representations. These designs collectively contribute to MGKAN's superior performance in asymmetric DDI prediction.

Ablation Study

We conduct an ablation study to evaluate the contribution of each module in MGKAN. As shown in Fig. 1, we evaluate six model variants: **w/o KAN:** all KAN layers in both the message passing and fusion module are replaced with MLPs; **w/o AF:** removing attention fusion; **w/o KF:** removing KAN-based fusion; **w/o DN:** removing asymmetric DDI network view; **w/o CI:** removing co-interaction view; and **w/o SIM:** removing biochemical similarity view.

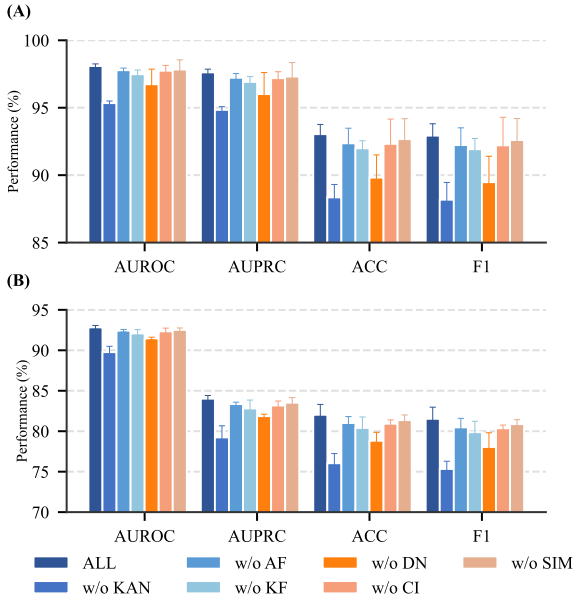

Figure 1: Ablation study results of MGKAN and its variants. **(A)** Performance on **Task 1**. **(B)** Performance on **Task 2**.

The ablation study provides three principal insights consistent with the MGKAN design. First, the removal of KAN results in the most significant performance degradation, confirming its critical role in capturing nonlinear and directional interaction patterns. Second, both fusion modules are essential: removing KAN-based fusion leads to greater degradation, underscoring the importance of nonlinear fusion and the complementarity between the two fusion mechanisms. Third, all three views contribute to performance: the asym-

Method	DS 1				DS 2			
	AUROC	AUPRC	ACC	F1	AUROC	AUPRC	ACC	F1
Gravity VGAE	$88.54_{\pm 0.01}$	$85.26_{\pm 0.03}$	$81.66_{\pm 0.03}$	$82.20_{\pm 0.03}$	$87.90_{\pm 0.01}$	$85.75_{\pm 0.01}$	$79.93_{\pm 0.02}$	$80.57_{\pm 0.03}$
S/T VGAE	$92.14_{\pm 0.08}$	$91.51_{\pm 0.10}$	$84.13_{\pm 0.13}$	$83.72_{\pm 0.22}$	$91.26_{\pm 0.04}$	$90.75_{\pm 0.07}$	$83.23_{\pm 0.12}$	$82.94_{\pm 0.26}$
DiGAE	$94.84_{\pm 0.03}$	$93.73_{\pm 0.06}$	$91.15_{\pm 0.04}$	$91.29_{\pm 0.03}$	$90.57_{\pm 0.02}$	$88.44_{\pm 0.05}$	$85.91_{\pm 0.03}$	$86.14_{\pm 0.02}$
DirGNN	$97.82_{\pm 0.02}$	$97.38_{\pm 0.04}$	$93.24_{\pm 0.03}$	$93.22_{\pm 0.03}$	$95.55_{\pm 0.02}$	$94.99_{\pm 0.03}$	$89.18_{\pm 0.03}$	$89.13_{\pm 0.04}$
MAVGAE	$92.53_{\pm 0.06}$	$91.78_{\pm 0.06}$	$84.39_{\pm 0.07}$	$84.21_{\pm 0.06}$	$91.57_{\pm 0.08}$	$91.57_{\pm 0.08}$	$83.29_{\pm 0.44}$	$82.81_{\pm 0.77}$
DRGATAN	$87.30_{\pm 0.05}$	$86.72_{\pm 0.05}$	$79.82_{\pm 0.08}$	$80.71_{\pm 0.08}$	$83.51_{\pm 0.05}$	$83.88_{\pm 0.05}$	$74.48_{\pm 0.06}$	$76.85_{\pm 0.05}$
DGAT-DDI	$98.55_{\pm 0.03}$	$98.20_{\pm 0.04}$	$94.69_{\pm 0.12}$	$94.67_{\pm 0.14}$	$96.65_{\pm 0.02}$	$95.98_{\pm 0.03}$	$90.91_{\pm 0.04}$	$91.02_{\pm 0.06}$
MGKAN	$\mathbf{99.08_{\pm 0.01}}$	$\mathbf{98.94_{\pm 0.02}}$	$\mathbf{95.05_{\pm 0.14}}$	$\mathbf{94.90_{\pm 0.16}}$	$\mathbf{98.09_{\pm 0.02}}$	$\mathbf{97.62_{\pm 0.03}}$	$\mathbf{93.04_{\pm 0.09}}$	$\mathbf{92.93_{\pm 0.11}}$

Table 1: Performance comparison of MGKAN in **Task 1**. Results are reported as the mean and standard deviation (%) over five-fold cross-validation. The highest value in each column is highlighted in bold.

Method	DS 1				DS 2			
	AUROC	AUPRC	ACC	F1	AUROC	AUPRC	ACC	F1
Gravity VGAE	$69.48_{\pm 0.02}$	$62.33_{\pm 0.06}$	$65.80_{\pm 0.03}$	$71.28_{\pm 0.04}$	$66.33_{\pm 0.02}$	$61.87_{\pm 0.03}$	$61.92_{\pm 0.02}$	$61.92_{\pm 0.02}$
S/T VGAE	$85.05_{\pm 0.21}$	$72.28_{\pm 0.41}$	$69.19_{\pm 0.36}$	$68.33_{\pm 0.21}$	$83.03_{\pm 0.04}$	$83.03_{\pm 0.04}$	$66.49_{\pm 0.11}$	$65.45_{\pm 0.04}$
DiGAE	$90.91_{\pm 0.04}$	$82.12_{\pm 0.09}$	$66.21_{\pm 0.09}$	$66.59_{\pm 0.09}$	$83.48_{\pm 0.02}$	$69.44_{\pm 0.03}$	$58.53_{\pm 0.04}$	$58.53_{\pm 0.04}$
DirGNN	$94.62_{\pm 0.01}$	$88.14_{\pm 0.02}$	$85.31_{\pm 0.05}$	$84.79_{\pm 0.05}$	$89.96_{\pm 0.02}$	$79.37_{\pm 0.05}$	$76.91_{\pm 0.04}$	$76.29_{\pm 0.03}$
MAVGAE	$88.51_{\pm 0.15}$	$77.92_{\pm 0.26}$	$73.50_{\pm 0.26}$	$72.65_{\pm 0.24}$	$85.96_{\pm 0.14}$	$73.52_{\pm 0.22}$	$70.11_{\pm 0.22}$	$69.08_{\pm 0.22}$
DRGATAN	$83.43_{\pm 0.10}$	$70.61_{\pm 0.11}$	$65.42_{\pm 0.07}$	$65.31_{\pm 0.07}$	$78.15_{\pm 0.04}$	$61.62_{\pm 0.05}$	$54.78_{\pm 0.09}$	$54.98_{\pm 0.11}$
DGAT-DDI	$95.50_{\pm 0.06}$	$89.79_{\pm 0.12}$	$87.30_{\pm 0.12}$	$86.86_{\pm 0.10}$	$90.94_{\pm 0.06}$	$80.88_{\pm 0.09}$	$78.18_{\pm 0.11}$	$77.62_{\pm 0.09}$
MGKAN	$\mathbf{96.21_{\pm 0.02}}$	$\mathbf{91.07_{\pm 0.03}}$	$\mathbf{89.32_{\pm 0.10}}$	$\mathbf{88.87_{\pm 0.10}}$	$\mathbf{92.83_{\pm 0.03}}$	$\mathbf{84.02_{\pm 0.05}}$	$\mathbf{82.03_{\pm 0.16}}$	$\mathbf{81.53_{\pm 0.18}}$

Table 2: Performance comparison of MGKAN in **Task 2**. Results are reported as the mean and standard deviation (%) over five-fold cross-validation. The highest value in each column is shown in bold.

metric DDI network view captures the direct drug interaction, while the co-interaction and similarity network views introduce valuable topological and biochemical information. By integrating multiple structural and semantic views, MGKAN learns more expressive embeddings than single-view methods. These results validate the effectiveness of integrating KAN with multi-view complementary fusion for asymmetric DDI prediction.

Visualization of Role-Specific Embeddings

To evaluate the capability of MGKAN for learning directed role-specific embeddings, we randomly selected 500 drugs from the dataset, extracted their corresponding source and target embeddings, and applied t-SNE for visualization. The embedding distributions learned by a standard GCN and MGKAN are visualized in Figure 2 for comparison.

The visualization results show that while the embeddings learned by standard GCN exhibit a degree of clustering, there is substantial overlap between source and target roles, indicating limited capacity for role differentiation. In contrast, MGKAN produces more clearly clustered and well-separated embeddings in both directions, reflecting a stronger awareness of directional semantics. These findings validate the effectiveness of MGKAN's direction-aware design and the GKAN-based encoding architecture in capturing asymmetric role-specific embeddings.

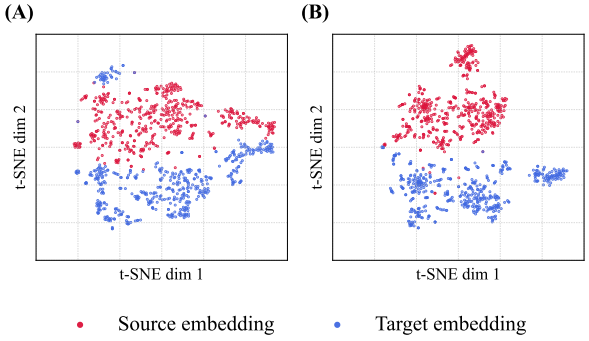

Figure 2: The t-SNE visualization of role-specific embeddings. **(A)** Role-specific embeddings learned by GCN. **(B)** Role-specific embeddings learned by MGKAN.

Case Study

To assess the ability of MGKAN in identifying previously unobserved DDIs, we conducted a transductive prediction on the DS 1 dataset (DrugBank v5.1.2) and selected the top 10 drug pairs ranked by prediction score. These candidate interactions were then validated using the interaction checker in DrugBank v5.1.13.

As shown in Table 3, seven of the predicted pairs were confirmed as asymmetric interactions, two were symmetric, and one had no recorded interaction. These results demon-

Rank	Source drug	Target drug	Direction
1	Calcium carbonate	Dolutegravir	✓
2	Fluvoxamine	Mirtazapine	✓
3	Promethazine	Metyrosine	×
4	Vemurafenib	Miglitol	✓
5	Lidocaine	Arformoterol	✓
6	Labetalol	Clenbuterol	✓
7	Leflunomide	Methotrexate	×
8	Canagliflozin	Pioglitazone	✓
9	Argatroban	Glyburide	N.A.
10	Digoxin	Deslanoside	✓

Table 3: Top 10 asymmetric DDIs predicted by MGKAN. The symbol "✓" denotes a correct directional prediction; "×" denotes an incorrect one; "N.A." indicates that the direction is unknown.

strate the effectiveness of MGKAN in discovering novel asymmetric DDIs and its potential in practical applications.

Theoretical Analysis of Second-Order Similarity Kernels

The spectral properties of second-order similarity kernels are analyzed to understand their propagation behavior. We rigorously justify why the second-order similarity kernels C_{in} and C_{out} yield propagation dynamics that avoid the homogenization typical of Laplacian- or random-walk-based convolutions.

Let $\mathcal{G} = (V, E)$ be a strongly connected digraph with adjacency matrix $A \in \mathbb{R}^{n\times n}$. Define:

$$C_{in}(i,j) = \sum_k \frac{A_{k,i}A_{k,j}}{\sum_v A_{k,v}}, C_{out}(i,j) = \sum_k \frac{A_{i,k}A_{j,k}}{\sum_v A_{v,k}}.$$

Each C_{in}, C_{out} captures degree-normalized two-hop similarity: C_{in} encodes shared predecessors (co-accessibility), while C_{out} encodes shared successors (co-influence). Let their degree matrices be

$$D_{C_{in}} = \mathrm{diag}\Big(\sum_j C_{in}(i,j)\Big), D_{C_{out}} = \mathrm{diag}\Big(\sum_j C_{out}(i,j)\Big).$$

Define the normalized operator:

$$\hat{C}_{in} = D_{C_{in}}^{-1/2} C_{in} D_{C_{in}}^{-1/2}.$$

Proposition 1 (Spectral Structure and Convergence)**.** *For a strongly connected $\mathcal{G}$, the normalized kernel $\hat{C}_{in}$ satisfies:*

1. *Positive Semidefiniteness. C_{in} admits a Gram decomposition:*

$$C_{in} = \Phi^\top \Phi, \qquad \Phi_{k,j} = \frac{A_{k,j}}{\sqrt{\sum_v A_{k,v}}}.$$

Hence C_{in} is symmetric PSD, and $\mathrm{rank}(C_{in}) \le \mathrm{rank}(A)$.

2. *Perron Eigenpair. $\hat{C}_{in}$ is real symmetric, with eigenvalues*

$$1 = \lambda_1 > \lambda_2 \ge \cdots \ge \lambda_n \ge 0.$$

The top eigenvector is explicitly

$$u_1 = \frac{D_{C_{in}}^{1/2}\mathbf{1}}{\|D_{C_{in}}^{1/2}\mathbf{1}\|_2}, \qquad \hat{C}_{in}u_1 = u_1,$$

and all other eigenvalues satisfy $|\lambda_i| < 1$.

3. *Convergence of Propagation. For any $X \in \mathbb{R}^{n\times d}$, iterating*

$$H^{(k+1)} = \hat{C}_{in}H^{(k)}, \qquad H^{(0)} = X,$$

yields

$$H^{(k)} = \sum_{i=1}^n \lambda_i^k u_i u_i^\top X.$$

As $k \to \infty$, all terms with $i \ge 2$ vanish, and

$$H^{(\infty)} = u_1 u_1^\top X.$$

4. *Non-Uniform Steady State. Since*

$$[D_{C_{in}}]_{ii} = \sum_j C_{in}(i,j) = \sum_k \frac{A_{k,i}}{\sum_v A_{k,v}} \sum_j A_{k,j},$$

we have

$$u_1(i) \propto \sqrt{\sum_k \frac{A_{k,i}}{\sum_v A_{k,v}} \sum_j A_{k,j}}.$$

Unless $\mathcal{G}$ is regular, u_1 is non-uniform, and the limit embeddings maintain nonzero variance:

$$\mathcal{V}(H^{(\infty)}) = \frac{1}{n}\sum_{i=1}^n \|H_i^{(\infty)} - \bar{H}^{(\infty)}\|_2^2 > 0.$$

Proof. (1) follows by directly factorizing each term of C_{in} as a rank-one PSD matrix:

$$C_{in} = \sum_k \frac{1}{\sum_v A_{k,v}} (A_{k,:})^\top (A_{k,:}),$$

with Φ defined as above.

(2) Since C_{in} is symmetric, $\hat{C}_{in}$ is real symmetric. Moreover,

$$\hat{C}_{in} D_{C_{in}}^{1/2}\mathbf{1} = D_{C_{in}}^{-1/2} C_{in}\mathbf{1} = D_{C_{in}}^{1/2}\mathbf{1},$$

so $u_1 \propto D_{C_{in}}^{1/2}\mathbf{1}$ is a normalized eigenvector with eigenvalue 1. Strong connectivity implies irreducibility, hence by Perron–Frobenius all other eigenvalues satisfy $|\lambda_i| < 1$.

(3) The eigendecomposition $\hat{C}_{in} = U\Lambda U^\top$ yields the stated expansion. Since $|\lambda_i| < 1$ for $i \ge 2$, only the $i = 1$ term survives as $k \to \infty$.

(4) Substituting the explicit form of $D_{C_{in}}$ gives the stated expression for u_1. Non-regularity implies non-uniform u_1, hence the node-wise embeddings $H^{(\infty)}$ retain structural variance.

□

Remark 1. *This result shows that second-order kernels mitigate over-smoothing: the steady state is not a uniform consensus (as for Laplacians) nor a trivial stationary mixture (as for first-order random walks), but a degree-structured projection that preserves meaningful node-level variation.*

Runtime Efficiency Analysis

To assess runtime efficiency, we measured the average training time per fold (five-fold cross-validation) of each model on Task 1 – DS 2 using a single NVIDIA RTX 4090 GPU.

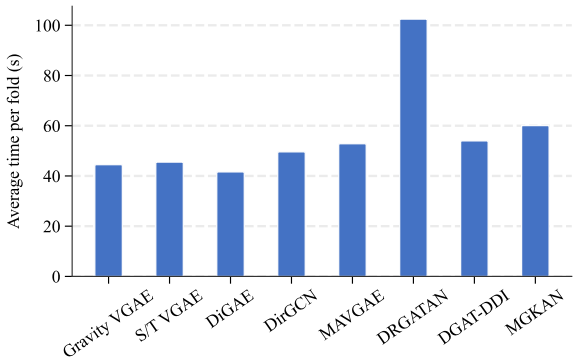

Figure 3: Per-fold runtime of all models (s)

Figure 3 presents the runtime results. DiGAE is the fastest baseline at 41.6 s, followed by Gravity VGAE (44.47s), S/T VGAE (45.51 s), DirGCN (49.58 s), MAVGAE (52.8 1s), and the best-performing baseline DGAT-DDI (53.95 s); DRGATAN is the slowest at 102.44 s. MGKAN requires 60.03 s, incurring an overhead of +18.43 s (+44%) compared to the fastest model and +6.08 s (+11%) over DGAT-DDI, yet it consistently achieves the best performance across all metrics, tasks, and datasets. These results demonstrate that MGKAN delivers state-of-the-art accuracy while maintaining a modest runtime overhead, offering a favorable accuracy–efficiency trade-off for asymmetric DDI prediction.