Supplementary Material for MGKAN

Anonymous submission

Appendix

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

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

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 $\operatorname{rank}(C_{in}) \le \operatorname{rank}(A)$.
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.

$\square$

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.

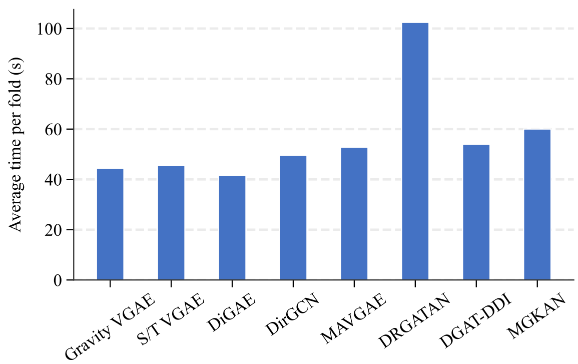

Figure 1: Per-fold runtime of all models (s)

Figure 1 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.