Supplementary Material

A. Fine-Grained Ablation Studies

This section provides detailed ablation studies for the Multi-period Time-Frequency Cooperative Encoder Module (MTFCE) and the Dual-Phase Graph Pooling Module (DPGPM).

1. MTFCE Frequency-Band Analysis

This section presents an ablation study on the Multi-scale Time-Frequency Convolutional Encoder (MTFCE) to dissect the importance of different EEG frequency bands. We trained and evaluated FG-MSTGNN using only a single band as input to isolate its contribution, with all other experimental parameters held constant.

As quantified in Table 1, all bands provided a substantial baseline accuracy, confirming that emotion-related neural correlates are distributed across multiple oscillatory rhythms. The superior performance of the Beta and Gamma bands aligns with existing literature that links high-frequency oscillations to active cognitive processing and emotional arousal. However, the ultimate performance is achieved only when all bands are integrated. The full-spectrum model ("All") outperformed any single-band configuration, surpassing the best individual band (Beta) by 1.98% on SEED and 5.00% on SEED-IV. This clear performance gap demonstrates that the MTFCE's design effectively fuses complementary spectral information, where low-frequency bands (e.g., Delta, Theta) potentially provide contextual or sustained emotional background, and high-frequency bands contribute finer-grained discriminative features.

Table 1. Classification accuracy (%) of FG-MSTGNN using individual frequency bands and full-spectrum input.

Dataset	**Delta**	**Theta**	**Alpha**	**Beta**	**Gamma**	**All**
SEED	91.11	92.26	90.67	92.69	92.41	94.67
SEED-IV	76.58	77.41	77.59	80.28	79.45	85.28

Note: The "All" column represents the performance when using the full multi-band spectrum as input.

2. DPGPM Node-Scoring Analysis

We evaluated the importance of the three scoring metrics—Degree Centrality (DC), Feature Importance (FI), and PageRank (PR), used in the DPGPM module. We ablated each metric individually and in combination, reporting the performance on the SEED and SEED-IV datasets.

Table 2. Classification accuracy (%) of FG-MSTGNN with different combinations of scoring components in the DPGPM module.

Dataset	**DC**	**FI**	**PR**	**DC+FI**	**DC+PR**	**FI+PR**	**All**
SEED	75.83	75.37	76.67	79.91	79.72	81.20	85.28
SEED-IV	93.52	93.52	92.96	94.11	94.26	94.08	94.67

Note:The "All" column represents the performance when integrating all three metrics (DC, FI, and PR).

The results are summarized in Table 2. Removing any of the three metrics leads to a performance drop. The removal of Feature Importance (FI) causes the most significant degradation. The best performance is achieved when all three metrics are integrated, demonstrating that they collectively contribute to stable and optimal graph pruning.

The MTFCE benefits from full-spectrum inputs to leverage information from complementary rhythmic patterns, while the DPGPM achieves optimal stability and performance by integrating all three node-scoring metrics.

B. Additional Dataset Validation: DEAP

Beyond the primary datasets, we assessed the adaptability of FG-MSTGNN on the DEAP dataset to validate its effectiveness under different experimental conditions. The DEAP dataset employs a different emotion model (valence/arousal) and a distinct 32-channel electrode configuration, presenting a stringent test for model generalizability. We followed the standard subject-independent (LOSO) evaluation protocol for a direct comparison with existing works.

As summarized in Table 3, our method consistently outperforms all recent benchmarks on both valence and arousal dimensions. The observed improvements—3.01% for valence and 2.67% for arousal over the previous best methods—are substantial, highlighting the model's ability to capture core emotional features. The fact that FG-MSTGNN, without any paradigm-specific tuning, successfully generalizes to DEAP's structure underscores a key strength: its dynamic graph learning and multi-scale feature extraction mechanisms are inherently adaptable. This suggests that the model learns a generalized representation of emotion-related brain dynamics rather than memorizing dataset-specific artifacts, marking a significant step toward practical, cross-paradigm EEG-based emotion recognition.

Table 3. Classification performance (Accuracy % / Standard Deviation %) on the DEAP dataset.

Method	Year	Valence	Arousal
STFG-CAP	2022	48.22 / 7.83	58.53 / 15.18
TMS-DANN	2023	57.70 / 7.23	61.88 / 5.55
BCJDA	2024	62.98 / 6.38	64.32 / 9.15
DT-EEGNet	2024	63.44 / 4.59	63.87 / 2.38
DCDNet	2025	65.16 / 5.38	66.31 / 8.60
FG-MSTGNN (Ours)	2025	**68.17 / 9.96**	**68.98 / 6.84**