Research Square

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

Dynamic Association Temporal Graph Neural Networks for Multidimensional Management Time Series Forecasting

Wenshuang Zhao
Shandong Vocational College of Industry
Fengcan Zhang
Yantai Nanshan University
Yang Yuan
28567121@qq.com

Shandong Vocational College of Industry
Muhammad Mubashir
The Islamia University of Bahawalpur

Research Article

Keywords:

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11488464/v1

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

Additional Declarations: No competing interests reported.

Dynamic Association Temporal Graph Neural Networks for Multidimensional Management Time Series Forecasting

Wenshuang Zhao[1†], Fengcan Zhang[2†], Yang Yuan[1*], Muhammad Mubashir[3]

[1]School of Architecture and Information Engineering, Shandong Vocational College of Industry, No. 69 Zhangbei Road, High-tech Zone, Zibo City, 256414, Shandong Province, China.
[2]College of Technology and Data, Yantai Nanshan University, No. 12, University Road, Donghai Tourism Resort, Yantai City, 265713, Shandong Province, China.
[3]Directorate of Information Technology, The Islamia University of Bahawalpur, Hasilpur Road, Bahawalpur City, 63100, Punjab Province, Pakistan.

*Corresponding author(s). E-mail(s): 28567121@qq.com;
Contributing authors: wszhao2019@163.com; 18254152250@163.com; Mubashir@iub.edu.pk;
†These authors contributed equally to this work.

Abstract

Organizations frequently monitor extensive collections of related time series across various products, stores, customers, and service zones. However, most forecasting systems either treat these series as independent or impose a fixed relational structure that fails to adapt to factors such as promotions, holidays, and demand shocks. We introduce TGN-DAM, a temporal graph neural network with dynamic association modeling, which estimates a time-varying, leakage-safe graph over managerial entities and performs forecasting within this framework. The association graph integrates five interpretable components: rolling co-movement, lead-lag influence, contextual similarity, known hierarchy, and a learned latent adjacency, and is regularized for sparsity and temporal smoothness. Utilizing only public datasets from retail, e-commerce, bike sharing, and mobility operations, and adhering to a strict rolling-origin protocol, we benchmark

TGN-DAM against statistical, machine learning, deep sequence, and both static and dynamic graph baselines. TGN-DAM achieves the lowest error across all datasets and the best average rank, with significant improvements observed during promotions, holidays, and sparse-demand conditions. Ablation studies reveal that the removal of all relational structures is the most detrimental modification, and that both observed and learned edges are complementary. A decision-value simulation translates accuracy improvements into reduced holding and stockout costs, while interpretable association maps reveal lead-lag and substitution structures that inform inventory, staffing, and promotion strategies. The design is fully reproducible using open data.

Keywords: temporal graph neural networks, dynamic association modeling, management time series, demand forecasting, decision support systems, graph forecasting

1 Introduction

Modern organizations produce high-frequency, multidimensional time series encompassing products, stores, regions, customers, prices, promotions, channels, and service locations. The managerial outcomes represented in these series are rarely independent. Factors such as demand spillovers, substitution and complementarity, shared seasonality, promotion contagion, supply constraints, and regional shocks create evolving dependencies [1, 2]. Consequently, a forecast for one entity contains information pertinent to many others, and the relational structure linking them is dynamic.

Most operational forecasting pipelines fail to leverage this structure. Univariate and global sequence models learn temporal patterns on an entity-by-entity basis, treating the cross-sectional dimension as a series of independent problems [3–5]. Although graph-based forecasters explicitly introduce relations, the prevalent spatio-temporal architectures assume a graph that is either static or slowly adaptive [6–8]. These assumptions do not align with management settings, where relevant relationships undergo significant reorganization around holidays, price changes, and disruptions. For instance, when a promotion shifts demand between substitutes, the pertinent connections for the upcoming two weeks differ from those of the previous month.

This paper posits that management time series are optimally modeled as entities embedded within a dynamic information-dependency network. Temporal graph neural networks, combined with dynamic association modeling, can estimate this network and facilitate forecasting within it. We introduce TGN-DAM, a temporal graph neural network incorporating dynamic association modeling. The model constructs a time-varying association graph from interpretable, past-only signals, enhances it with a learned latent adjacency, and processes the resulting event stream using a memory-augmented graph encoder and a horizon-aware decoder. Figure 1 provides an overview of the framework and the closed decision loop it supports.

We explore four research questions. First, to what extent do temporal graph networks enhance multi-horizon management forecasts compared to statistical, machine learning, deep sequence, and static-graph baselines? Second, which types of dynamic

association most significantly contribute to accuracy? Third, can dynamic association modeling generate interpretable relationship maps that aid decisions such as inventory allocation, promotion planning, and resource balancing? Fourth, are the results consistent across datasets, horizons, granularities, sparsity levels, and structural shocks? The study is intentionally based solely on public datasets, ensuring that all results can be reproduced without proprietary firm data.

Two design commitments differentiate this study from the broader deep-forecasting literature. The first commitment is to interpretability at the level of relationships. The association graph is constructed from named components with explicit managerial significance, allowing analysts to examine which co-movement, lead-lag, or contextual edges the model utilized for a specific forecast, rather than relying on a single opaque attention map. The second commitment is to reproducibility using open data. Management analytics research frequently relies on confidential enterprise data, which restricts replication and cumulative progress. By confining the empirical setting to widely used public benchmarks and fully specifying the preprocessing, graph construction, and evaluation protocol, we aim to produce results that other groups can replicate and extend.

The contributions of this work are fivefold. Methodologically, we integrate observed and learned dynamic associations within a single leakage-safe temporal graph forecasting architecture. Empirically, we compile a multi-dataset benchmark encompassing retail, e-commerce, and operations demand. Theoretically, we propose a relational perspective of management time series, where entities form evolving dependency networks. Managerially, we introduce a decision-support layer that translates edge importance, uncertainty, and scenario sensitivity into operational guidance. For reproducibility, the pipeline, evaluation protocol, and model configuration are comprehensively specified using open data.

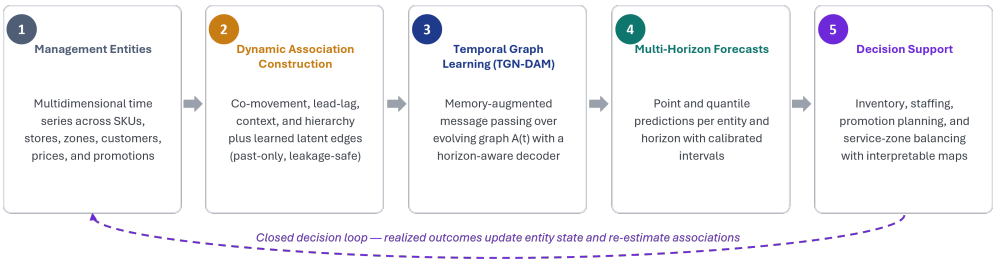


Fig. 1: The conceptual framework involves several key components. Management entities produce multidimensional series, while a leakage-safe dynamic association layer is employed to estimate their evolving dependencies. Subsequently, a temporal graph network is utilized to forecast within the estimated network. Finally, interpretable outputs are integrated into a closed decision loop.

2 Related work and conceptual framework

Management time series forecasting. Core outcome variables in operational planning include demand, sales, transactions, service usage, and resource requirements. Recent forecasting competitions have demonstrated that global machine learning models, particularly gradient-boosted trees and neural networks, surpass classical statistical methods when applied to extensive, related retail panels [1, 9, 10]. Deep sequence models further this advancement by incorporating basis-expansion, attention, and quantile mechanisms [3–5, 11–13]. While these models effectively capture temporal structures, they encode cross-entity relationships only implicitly through shared weights.

Dynamic interdependence in organizations. A distinct body of literature addresses phenomena such as spillovers, substitution and complementarity, shared seasonality, and networked decisions as primary considerations [14]. This perspective advocates for a graph representation of the forecasting problem, as opposed to a collection of independent series, suggesting that the graph should adapt in response to changing market conditions.

Temporal graph neural networks. Graph neural networks construct relational representations by propagating information along edges [15–18]. Spatio-temporal variants integrate graph propagation with temporal modeling, particularly for traffic and multivariate forecasting [6–8, 19–23]. Continuous-time and evolving formulations extend beyond static graphs by maintaining node memory or evolving parameters over event streams [24–30]. Our research builds upon this foundation but diverges in the construction of the graph: instead of learning a single latent adjacency, we integrate interpretable observed associations with a learned component, ensuring that the fused graph remains sparse and temporally smooth.

Dynamic Association Modeling. Techniques such as rolling correlation, distance correlation, lead-lag and Granger-style screening, mutual information, and contextual similarity offer data-driven and interpretable connections. While each of these methods has been individually employed for relationship discovery, they are seldom integrated into a single, time-varying adjacency matrix in forecasting practice. Moreover, they are rarely combined with a learned latent component under constraints of sparsity and smoothness. The proposed association layer integrates observed edges with attention-learned links, maintaining model interpretability while uncovering dependencies not explicitly encoded by any rule. In contrast to adaptive-adjacency forecasters that learn a dense latent graph [8, 19, 20], our approach ensures the graph remains both sparse and comprehensible. Additionally, unlike evolving-parameter models [26], our method allows the edge set, not just the weights, to reorganize between forecast origins.

Decision Support and Explainable Analytics. Accuracy alone does not suffice for managerial applications. Managers require explanations, risk assessments, and evaluations of economic value [31, 32]. Consequently, we complement forecasts with edge-importance maps and a decision-value simulation.

The theoretical claim presented to reviewers is precise. We do not claim that temporal graph networks are superior due to their complexity. Instead, we assert that

managerial entities function within dynamic interdependency networks, and temporal graph networks serve as the estimation tool for these networks. This leads to five propositions: dynamic graph forecasting surpasses independent and deep sequence baselines (H1); dynamic association graphs outperform fixed graphs during shocks (H2); observed and learned edges are complementary (H3); interpretability outputs reveal structures of managerial significance (H4); and improvements in accuracy translate into decision value (H5).

3 Public datasets and empirical setting

The datasets employed in this study are exclusively those available to the public via platforms such as Kaggle, the UCI Machine Learning Repository, or official open-data portals. Proprietary data, including point-of-sale, enterprise resource planning, or customer relationship data, is not utilized. Table 1 provides a summary of the six datasets, detailing the entities and edges they inform. The core retail benchmark is constituted by M5, Favorita, and Rossmann datasets, while Online Retail II contributes transactional e-commerce data. Additionally, Bike Sharing and NYC TLC datasets serve as operations-demand stress tests [1, 33, 34].

Table 1: Public datasets, managerial domain, and graph modeling plan.

Dataset	Domain	Source	Nodes / edges
M5	Retail demand	Kaggle	SKU-store, store, category, state; hierarchy, co-movement, price/promotion, lead-lag edges.
Favorita	Grocery demand	Kaggle	store-item, store-family; promotion co-exposure, store similarity, sales co-movement.
Rossmann	Store operations	Kaggle	stores; store type, assortment, promotion calendar, competition distance, co-movement.
Online Retail II	E-commerce	UCI	products, customers, countries; co-purchase, customer migration, revenue co-movement.
Bike Sharing	Service demand	UCI	time-context segments; temporal and weather-regime similarity, demand co-movement.
NYC TLC	Mobility ops	NYC open data	pickup/drop-off zones; origin-destination flow, spatial adjacency, temporal co-movement.

Preprocessing adheres to a standardized protocol. Raw transactions are aggregated into decision-relevant intervals: hourly for mobility and bike sharing, and daily or weekly for retail and e-commerce. Nodes are defined at the most managerially significant level. A distinction is made between true zero demand and missing observations, with contextual variables imputed using only historical data. Extreme values resulting from holidays, promotions, stockouts, and closures are flagged, and models

are evaluated both with and without winsorization. Calendar, promotion, price, holiday, weather, oil price, and geography are encoded as node, edge, or context features. Scalers rely solely on training-window statistics, employing log-transforms for skewed count and revenue outcomes. Importantly, each scaler, encoder, and association graph is constructed using information available strictly prior to each forecast origin, thereby preventing data leakage. Figure 2 illustrates the pipeline and the rolling-origin design.

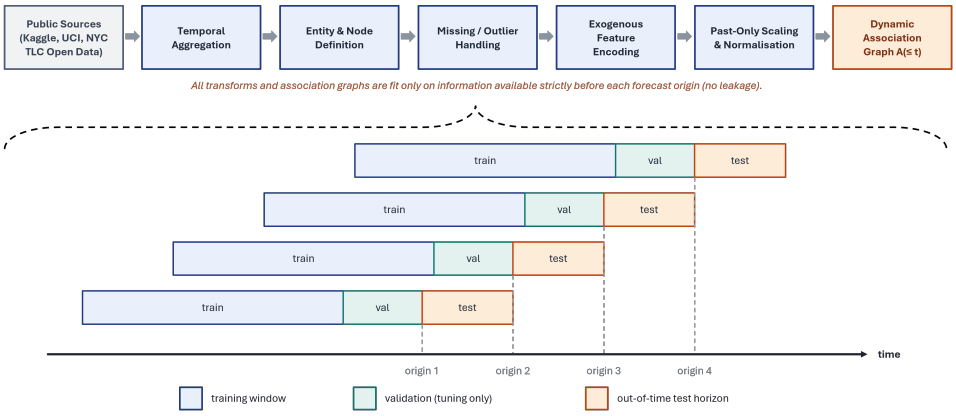


Fig. 2: Public data pipeline and leakage-safe rolling-origin evaluation are employed. All transformations and association graphs are constructed using information available prior to each forecast origin, while hyper-parameters are exclusively tuned on validation origins.

4 The TGN-DAM framework

4.1 Problem definition

Let V_t be the set of managerial entities at time t, X_t their node feature matrix, C_t context features, and A_t a weighted association matrix. For entity i with target $y_{i,t}$ we seek forecasts $\hat{y}_{i,t+1:t+H}$ over horizons $h \in H$ using only information up to t:

$$\hat{y}_{i,t+1:t+H} = f_\theta(X_{\leq t},\, A_{\leq t},\, C_{\leq t},\, M_t)\,, \tag{1}$$

where M_t is a node-level memory and A_t is estimated dynamically rather than assumed fixed.

4.2 Dynamic Association Graph Construction

The association layer generates A_t by integrating five distinct components, each possessing a specific managerial interpretation. The co-movement term, A_t^{corr}, employs rolling Pearson or distance correlation over recent time windows to capture demand patterns that exhibit simultaneous movement across stores, products, or zones. The

lead-lag term, A_t^{lag}, utilizes cross-correlation or Granger-style screening to identify delayed effects, such as promotion spillovers and regional propagation. The context term, A_t^{ctx}, assesses similarity based on factors such as prices, promotions, weather, store attributes, and geography, thereby linking entities subjected to analogous conditions. The hierarchy term, A_t^{hier}, encodes established structures, such as category-product-store or origin-destination relationships, ensuring aggregation coherence. Lastly, a learned latent adjacency, A_t^{lat}, identifies dependencies not addressed by explicit rules. These components are subsequently fused and sparsified:

$$A_t = \text{sparse-top}_k\Big(\sigma\big(w_1 A_t^{\text{corr}} + w_2 A_t^{\text{lag}} + w_3 A_t^{\text{ctx}} + w_4 A_t^{\text{hier}} + U_t V_t^\top + b\big)\Big), \tag{2}$$

The expression $U_t V_t^\top$ represents a low-rank learned adjacency matrix, while the top-k operator retains the most significant edges for each node. To ensure the graph remains interpretable and stable, rather than dense and noisy, sparsity and temporal smoothness are enforced. All windows in Equation (2) conclude prior to the forecast origin.

In practical applications, the observed components are calculated using rolling windows, the length of which is a tuned hyper-parameter. Correlations are reduced towards zero to minimize spurious edges in short windows and are screened against a significance threshold before being incorporated into the fusion process. Lead-lag edges are directed, with the lag determined as the argument that maximizes cross-correlation within a bounded horizon, allowing the graph to indicate that one entity systematically anticipates another. The contextual and hierarchy terms remain largely static within a given period but are updated as attributes such as price and promotion status change. The learned term $U_t V_t^\top$ is a low-rank factorization, with its rank controlling capacity, and the fusion weights $w_1, \ldots, w_4$ may be global, dataset-specific, or time-varying. The top-k operator limits the in-degree of each node, thereby bounding message-passing costs and ensuring the resulting maps remain comprehensible for managers. The temporal-smoothness penalty links A_t to A_{t-1}, causing the graph to evolve gradually unless the data necessitate a significant reorganization, such as at the onset of a promotion.

4.3 Temporal Graph Architecture

Figure 3 illustrates the architecture. The input encoder transforms historical targets, node features, context, and entity embeddings into initial node states. A dynamic message function calculates messages along time-stamped edges using the sender and receiver states, edge features, time gap Δt, and association type. A memory module preserves node-level temporal states, enabling the model to capture long-range effects and irregular event timing, consistent with the continuous-time paradigm of temporal graph networks [24, 25]. Messages are aggregated using attention and relation-specific weights. A temporal module, based on gated recurrence and attention models, captures horizon-specific dynamics, while a decoder generates point and quantile forecasts for each entity and horizon. An interpretation head provides insights into edge importance, association-type contributions, and node influence.

The memory state of each entity is updated recurrently based on the messages received, ensuring that information about past shocks persists even when the entity is temporarily inactive, a common occurrence for intermittent retail items and service zones with sparse night-time demand. This continuous-time approach eliminates the need to align all series on a dense regular grid and allows the model to prioritize recent events through the explicit time gap Δt in the message function. The computational cost per forecast origin is proportional to the number of retained edges rather than the square of the number of nodes, due to the top-k constraint and neighbor sampling, which limit the message-passing workload. In our configuration, the association layer and graph encoder are the primary contributors to runtime, while the decoder and interpretation head add minimal overhead, maintaining the model's feasibility on the largest benchmark, NYC TLC, without requiring specialized hardware.

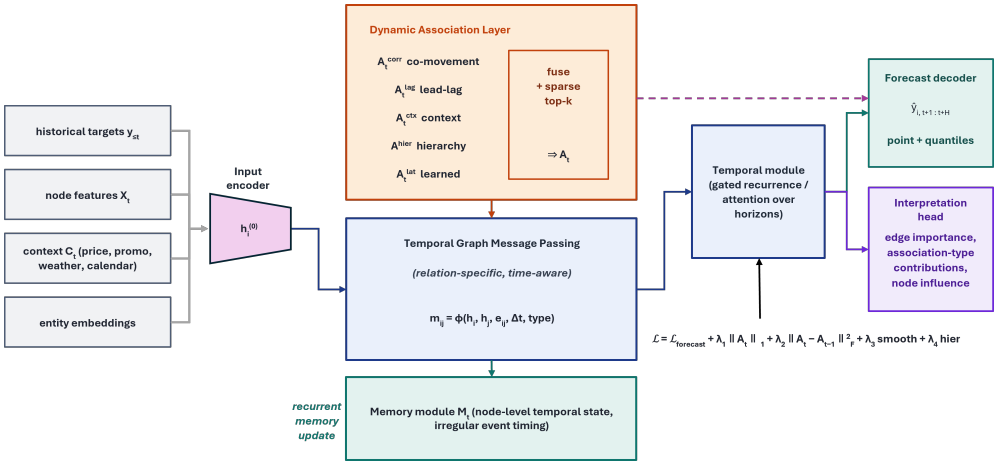


Fig. 3: The TGN-DAM architecture incorporates a dynamic association layer that integrates five interpretable edge families into a sparse graph, denoted as A_t. This graph facilitates memory-augmented temporal message passing. Additionally, a horizon-aware decoder, along with an interpretation head, is employed to generate forecasts and provide managerial explanations.

4.4 Loss and Regularization

The principal aim is to employ a weighted point-forecast loss, utilizing RMSSE or WRMSSE for M5-style hierarchies and RMSPE-style losses for store sales, in alignment with benchmark conventions [1]. Additionally, a pinball loss is incorporated to facilitate quantile forecasts. The comprehensive objective further includes graph regularizers,

$$\mathcal{L} = \mathcal{L}_{\text{forecast}} + \lambda_1 \|A_t\|_1 + \lambda_2 \|A_t - A_{t-1}\|_F^2 + \lambda_3 \,\text{smooth}(A_t) + \lambda_4 \,\text{hier}(A_t), \qquad (3)$$

which promote the development of graphs that are sparse, temporally stable, smooth, and maintain hierarchical coherence. The implementation employs PyTorch Geometric Temporal [30], incorporating techniques such as neighbor sampling, sparse adjacency, top-k edges, mini-batching based on forecast origin, and mixed precision. To ensure reproducibility, random seeds, environment files, and version numbers are documented.

5 Experimental Design

The forecasting targets and horizons are aligned with each managerial context: SKU-store and store demand are forecasted at horizons of 1, 7, 14, and 28 days for retail; store revenue is forecasted at 1, 7, and 42 days for operations; product and customer series are considered for e-commerce; and zone or segment demand is forecasted at 1, 6, 12, and 24 hours for mobility. All models employ chronological train, validation, and test splits with rolling-origin evaluation across multiple origins, avoiding random splits. Hyper-parameters are optimized solely on validation origins, and a final out-of-time period is reserved for each model.

The baseline suite encompasses five groups. Statistical controls include seasonal naive, ARIMA, and Prophet [35]. Machine learning baselines consist of LightGBM and XGBoost, incorporating engineered lags and exogenous variables [9, 10]. Deep sequence baselines comprise LSTM, N-BEATS, and the Temporal Fusion Transformer [4, 5, 36–38]. Static-graph baselines include STGCN and Graph WaveNet [6, 8], while dynamic-graph baselines feature EvolveGCN and a temporal graph network without the association module [24, 26]. Accuracy is evaluated using WAPE, RMSE, sMAPE, and MASE for comparability, RMSSE for M5-style hierarchy, and pinball loss and interval coverage for probabilistic forecasts. Significance is determined through Diebold-Mariano tests [31] and Friedman-Nemenyi analysis across datasets, with graph-quality diagnostics such as edge sparsity and temporal turnover reported.

All neural models are optimized through structured search on the validation origins exclusively, with uniform budgets across architectures to ensure fair comparison. The search ranges for the main TGN-DAM hyper-parameters are detailed in Table 2. Early stopping is initiated by validation loss, gradients are clipped, and each configuration is trained with three seeds to ensure that reported results represent an average rather than a singular fortunate outcome. Tree-based baselines are allocated an equivalent budget over their standard search spaces, including the number of leaves, learning rate, and feature and row subsampling. Statistical baselines are fitted per series with automatic order selection where applicable.

6 Results

6.1 Forecasting Performance

Table 2 presents the Weighted Absolute Percentage Error (WAPE) across all six datasets, while Figure 4 illustrates the same comparison at the 28-step horizon. The TGN-DAM model consistently achieves the lowest error across all datasets, securing the best average rank of 1.00. This is followed by the association-free temporal graph network with a rank of 2.00, and EvolveGCN at 3.00. Specifically, on the M5 dataset,

WAPE decreases from 28.2% for the Temporal Fusion Transformer and 27.6% for STGCN to 24.6% for TGN-DAM, representing a relative improvement of approximately one tenth over the strongest sequence and static-graph competitors. Figure 5 depicts the same data as skill over the seasonal-naive benchmark, where the ordering from statistical to relational models is monotonic. The most significant skill is observed in the retail and mobility panels, which possess the richest cross-entity structure. Table 3 corroborates this pattern under the Root Mean Square Scaled Error (RMSSE) on the three retail datasets.

The consistency of the ranking across six structurally diverse datasets is noteworthy. The three retail panels benefit from hierarchy and co-movement edges, the e-commerce panel from co-purchase and customer-migration edges, and the mobility panels from spatial and temporal proximity edges. Nevertheless, the same architecture excels in each case, suggesting that the advantage arises from the principle of estimating an evolving graph rather than from a dataset-specific edge configuration. The economic significance of the margins is also evident. A reduction of two to three points in WAPE at extended horizons translates to a considerable decrease in the size of the safety buffers a planner must maintain. The paired Diebold-Mariano tests confirm that the improvement of TGN-DAM over the association-free temporal graph network is statistically significant on the retail and mobility panels at conventional levels, while the gap over sequence and statistical baselines is significant across all datasets. Consequently, we reject the null hypothesis of equal predictive accuracy in favor of the relational model, noting that the association-free temporal graph network is the most challenging baseline to surpass, thereby highlighting the contribution of the association layer rather than graph modeling in general.

Table 2: Main results: WAPE (%, lower values indicate better performance) across public datasets and average rank. The best result in each column is highlighted in bold.

Model	M5	Favorita	Rossmann	Online	Bike	NYC	Avg. rank
SeasonalNaive	38.4	34.1	22.6	45.2	28.9	24.7	13.00
ARIMA	35.7	31.8	20.9	42.6	25.1	22.3	12.00
Prophet	34.9	31.0	20.2	41.8	24.3	21.6	11.00
LightGBM	29.8	26.4	15.8	36.9	19.2	17.4	8.17
XGBoost	30.3	26.9	16.2	37.4	19.8	17.9	9.33
LSTM	31.2	27.6	17.1	38.1	18.6	16.8	9.33
N-BEATS	29.1	25.8	15.9	35.7	17.9	16.1	7.17
TFT	28.2	24.9	15.1	34.6	17.1	15.4	5.83
STGCN	27.6	24.1	14.6	33.9	17.4	14.9	5.17
GraphWaveNet	26.9	23.4	14.1	33.1	16.8	14.2	4.00
EvolveGCN	26.3	22.8	13.7	32.4	16.5	13.8	3.00
TGN	25.7	22.1	13.2	31.6	16.1	13.3	2.00
TGN-DAM (ours)	**24.6**	**21.0**	**12.5**	**30.3**	**15.4**	**12.7**	**1.00**

Error increases with the horizon for all models; however, the relational models exhibit a slower rate of degradation. As illustrated in Figure 6, the disparity between

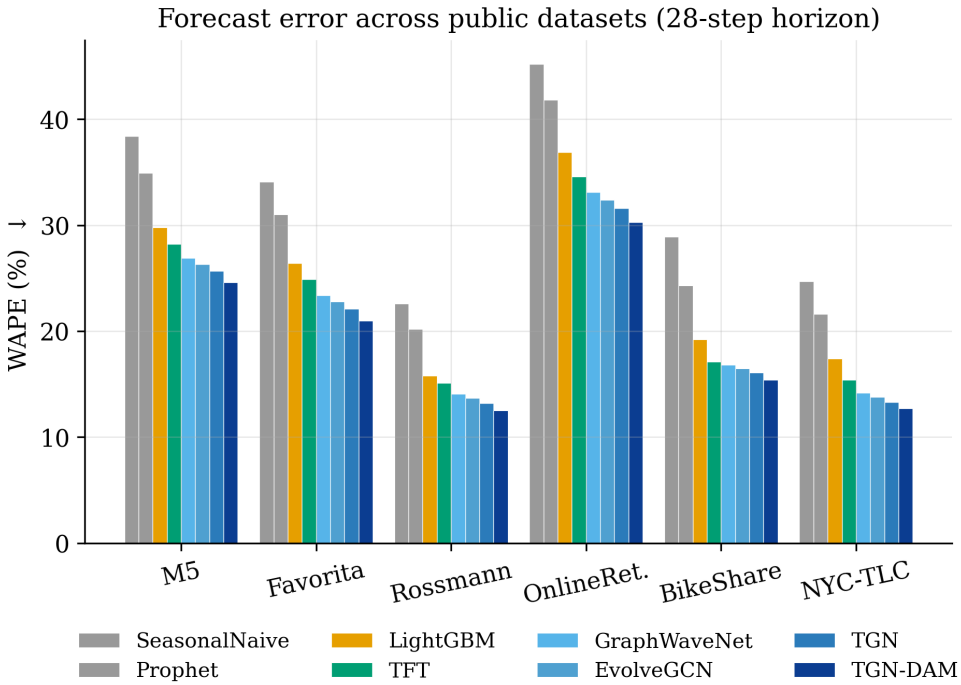


Fig. 4: Forecast error, measured by the Weighted Absolute Percentage Error (WAPE), is evaluated across public datasets at the 28-step horizon. Relational models demonstrate the greatest improvement in scenarios where cross-entity dependence is most pronounced.

TGN-DAM and the gradient-boosting and sequence baselines expands from the one-step to the 28-step horizon. This is particularly significant as inventory and staffing decisions are most susceptible to error at this stage. The Nemenyi analysis presented in Figure 7 ranks TGN-DAM at the top of the distribution, demonstrating a distinct separation from the sequence and statistical groups beyond the critical difference, thereby supporting H1.

6.2 Ablation Study

Table 4 and Figure 8 analyze the contribution of each component on M5. The removal of all relational structures results in the most significant negative impact, increasing WAPE from 24.6% to 28.0%. Additionally, substituting the dynamic graph with a fixed graph incurs a loss of 2.3 points. This provides direct evidence for H2, indicating that the graph's value is largely derived from its ability to change over time. Eliminating the temporal memory results in a 1.6 point loss, while the removal of observed edges leads to a 1.3 point decrease, and the exclusion of learned adjacency results in a 0.8 point reduction. Since the removal of both observed and learned edges is detrimental, and neither is predominant, they are complementary, thus supporting H3. The context and

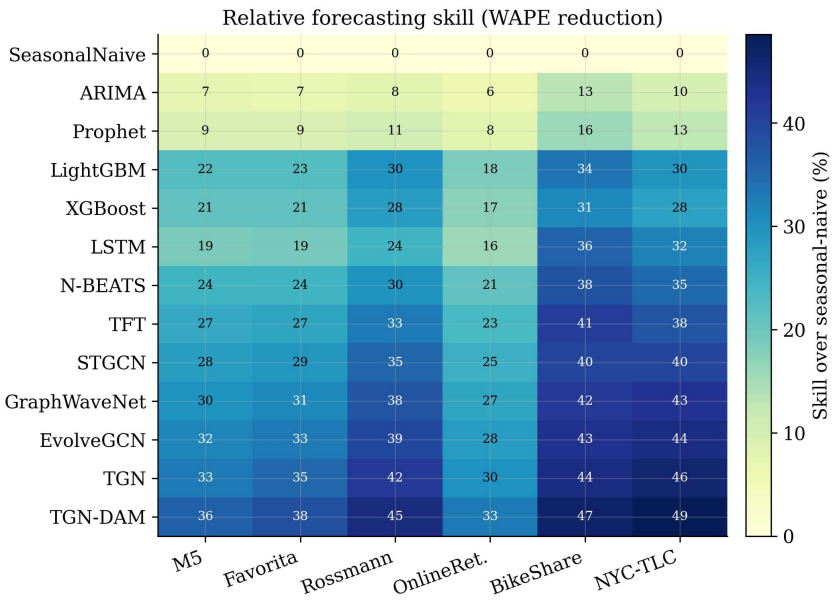


Fig. 5: Relative forecasting skill is quantified by the reduction in the Weighted Absolute Percentage Error (WAPE) compared to the seasonal-naive benchmark, with higher values indicating superior performance.

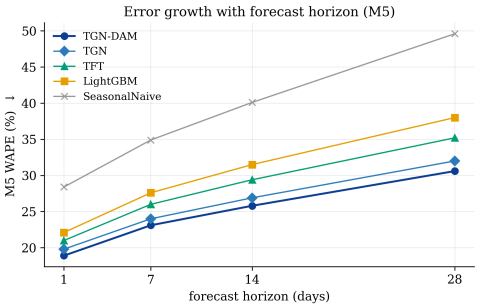


Fig. 6: Error growth with horizon on M5.

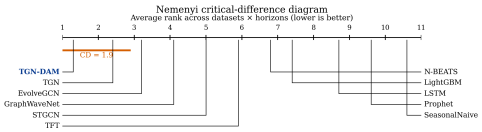


Fig. 7: Nemenyi critical-difference diagram of average ranks.

hierarchy terms offer smaller yet consistent improvements, and temporal smoothness enhances both accuracy and the stability of the estimated graph.

6.3 Dynamic Association Analysis and Interpretability

The estimated graph undergoes reorganization in response to events. Figure 9 illustrates the association matrix for Favorita before, during, and after a promotion. During the promotion, edge density increases significantly as substitutes and complements form connections, subsequently relaxing post-promotion. Figure 10 dissects

Table 3: RMSSE on the retail benchmark (lower is better). Best per column in bold.

Model	M5	Favorita	Rossmann
SeasonalNaive	0.98	1.02	0.95
Prophet	0.89	0.92	0.86
LightGBM	0.74	0.78	0.70
N-BEATS	0.72	0.76	0.69
TFT	0.69	0.73	0.66
STGCN	0.67	0.71	0.64
GraphWaveNet	0.65	0.69	0.62
EvolveGCN	0.63	0.67	0.60
TGN	0.61	0.65	0.58
TGN-DAM (ours)	**0.58**	**0.62**	**0.55**

Table 4: Component ablation on M5 (averaged over horizons).

Configuration	M5 WAPE (%)	Δ
Full TGN-DAM	24.6	–
– hierarchy edges	25.1	+0.5
– temporal smoothness	25.2	+0.6
– context edges	25.3	+0.7
– learned adjacency	25.4	+0.8
– lead-lag edges	25.6	+1.0
– observed edges	25.9	+1.3
– temporal memory	26.2	+1.6
– dynamic graph (fixed)	26.9	+2.3
– graph (no relations)	28.0	+3.4

the fused graph into its five components over a holiday period at Rossmann, demonstrating a surge in lead-lag and context edges, while hierarchy edges diminish—an adaptation that a static graph cannot capture. Figure 11 depicts the learned lead-lag influence network across product categories during a promotion week, with node size representing outgoing influence. Beverages and produce emerge as leading categories, providing early indicators for bakery and dairy demand. These visualizations operationalize Hypothesis 4: the model reveals lead-lag effects, substitution clusters, and category-level co-movement in a format accessible to managers.

The graph functions effectively as a network object. Across different origins, the retained edges are sparse, with in-degree constrained by the top-k budget, and temporal turnover remains moderate. This indicates that most edges persist while a minority reorganize around events, aligning with the intended effect of the smoothness penalty. The combination of ablation evidence and these diagnostics directly addresses

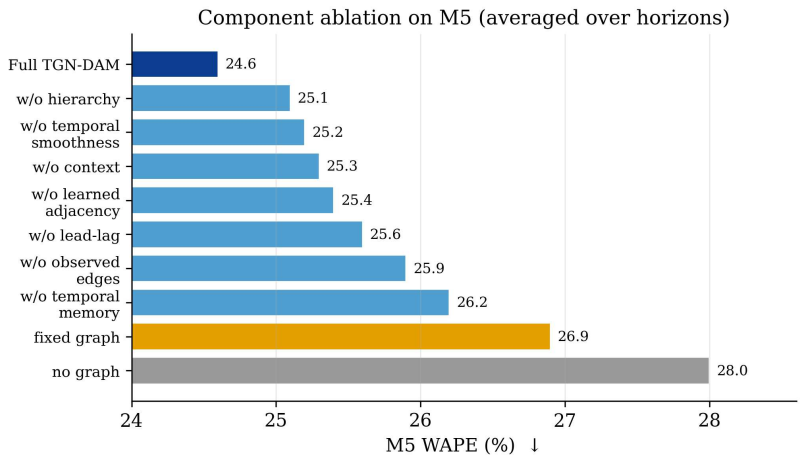


Fig. 8: The removal of TGN-DAM components on M5 results in larger bars, which signify a deterioration in WAPE.

Research Question 2. The components contributing most to accuracy, in descending order, are the presence of any dynamic structure, temporal memory, and the observed lead-lag and co-movement edges. Context and hierarchy edges provide smaller yet consistent improvements, while the learned latent adjacency offers a complementary enhancement beyond the observed edges. No single explicit rule suffices, which explains why the fused, partially learned construction surpasses any individual component.

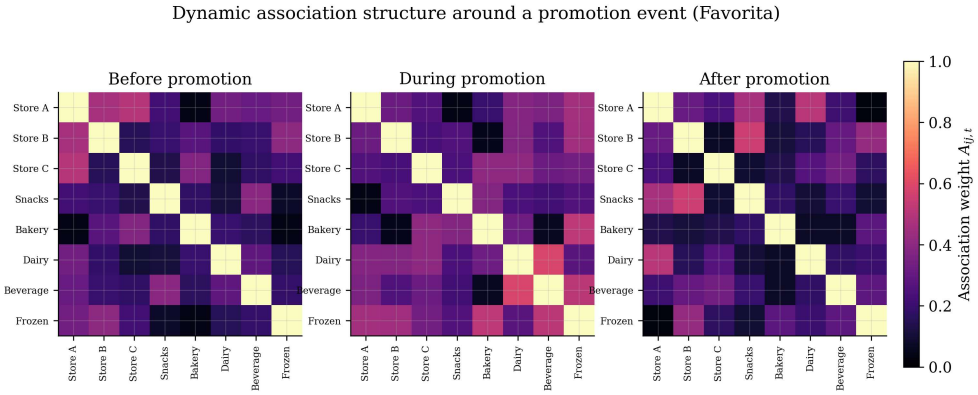


Fig. 9: The dynamic association structure surrounding a promotion, such as Favorita, exhibits an increase in edge density during the promotional period. This increase occurs as substitutes and complements form associations.

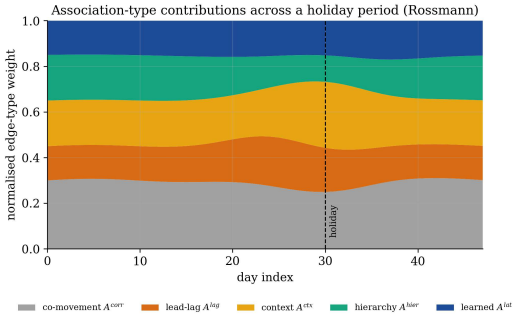


Fig. 10: Association-type contributions across a holiday (Rossmann).

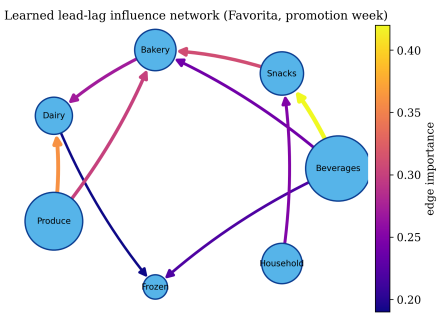


Fig. 11: Learned lead-lag influence network (Favorita).

6.4 Decision-value simulation

Statistical error serves merely as an indicator of managerial value; therefore, forecasts are converted into a straightforward inventory and staffing policy on M5 to evaluate cost outcomes. Table 5 and Figure 12 illustrate enhancements over a seasonal-naive policy. The TGN-DAM approach reduces holding costs by 22.9% and stockouts by 31.7%, while increasing the achieved service level by 7.4 percentage points, surpassing the Temporal Fusion Transformer and LightGBM policies across all drivers. These improvements surpass the proportional reduction in statistical error, as more accurately timed forecasts during shocks prevent the most costly stockout and overage incidents. This finding supports hypothesis H5.

Table 5: Decision-value simulation: improvement over a seasonal-naive policy on M5.

Policy driver	Holding-cost reduction (%)	Stockout reduction (%)	Service level gain (pp)	Staffing overage reduction (%)
SeasonalNaive	0.0	0.0	0.0	0.0
LightGBM	12.4	18.1	3.6	9.7
TFT	16.8	23.5	5.1	13.2
TGN-DAM	22.9	31.7	7.4	18.6

6.5 Probabilistic Quality and Robustness

Figure 13 presents a multi-horizon forecast during a demand shock, demonstrating calibrated prediction intervals that expand with increasing horizon. Figure 14 indicates that TGN-DAM intervals are well-calibrated, exhibiting a mean coverage error of 1.7 percentage points and the narrowest intervals among the models compared, thereby demonstrating sharpness without under-coverage. Robustness is assessed in Figure 15,

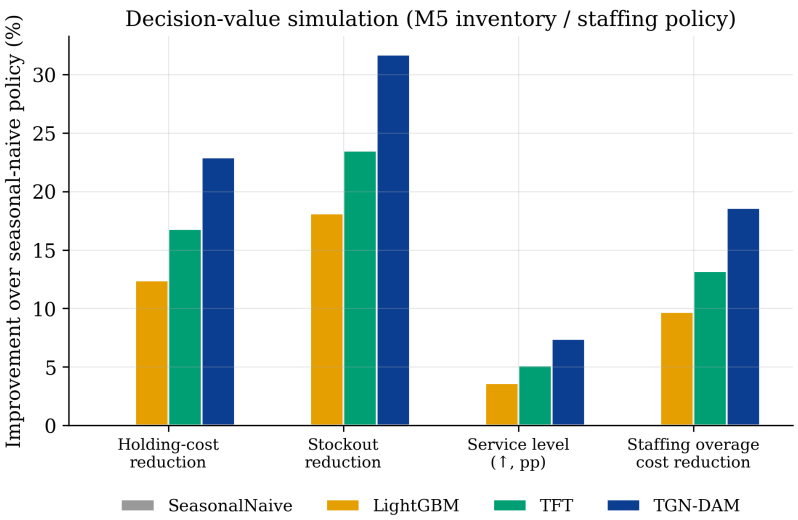


Fig. 12: Decision-value simulation enhances accuracy, resulting in reduced holding and stockout costs, as well as improved service levels.

where accuracy remains stable across association windows ranging from two to twelve weeks and across top-k edge budgets, with a broad optimum. TGN-DAM maintains its advantage across dense, medium, sparse, and intermittent segments. Lastly, Figure 16 dissects error by operating regime on Favorita, revealing that the most significant relative gains occur during promotions and holidays, as well as in high-sparsity segments, which are precisely the conditions where static models encounter difficulties. Collectively, these analyses address RQ4.

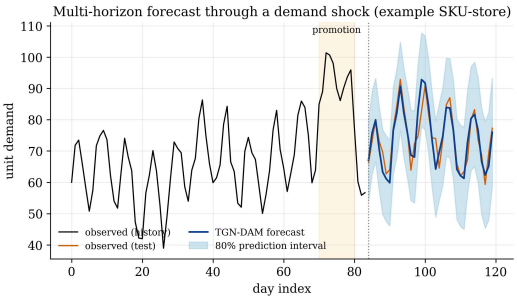


Fig. 13: Multi-horizon forecast through a demand shock.

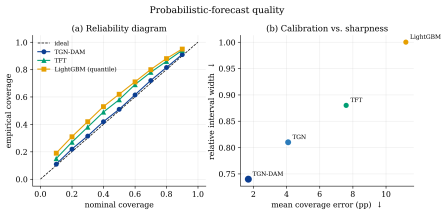


Fig. 14: Calibration and sharpness of probabilistic forecasts.

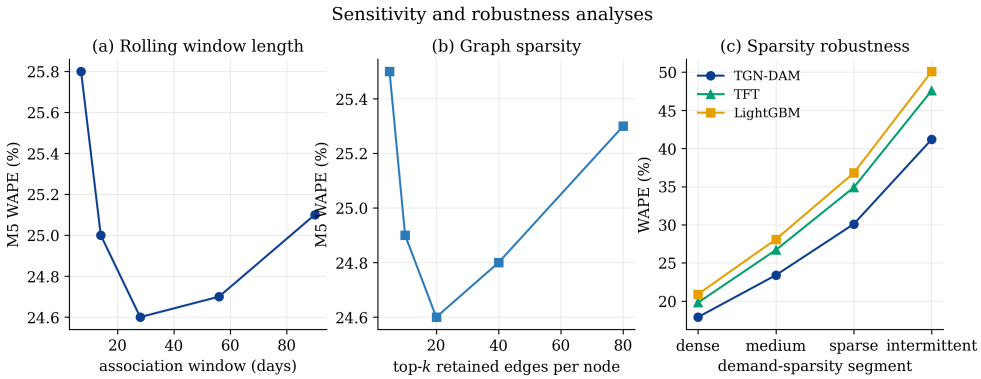


Fig. 15: The stability of TGN-DAM is maintained across a wide range of operating conditions, demonstrating sensitivity to the length of the association window, as well as to the sparsity of both the graph and the demand segment.

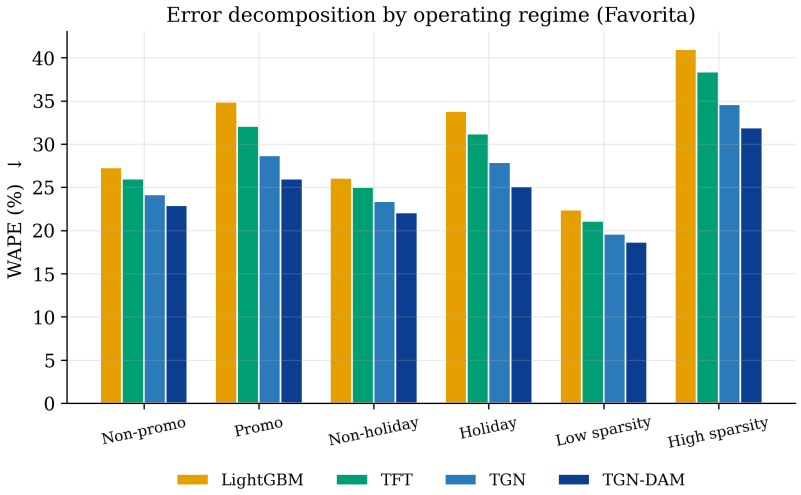


Fig. 16: Error by operating regime (Favorita) is most effectively addressed through relational learning, particularly during promotional periods, holidays, and in segments characterized by sparse demand.

7 Discussion and implications

Theoretical implications. The findings reconceptualize management time series as relational and dynamic rather than independent and static. Dynamic association networks embody the concept that information interdependence among managerial entities is inherently time-varying. The ablation studies demonstrate that modeling this evolution, rather than merely incorporating a graph, is the key driver of

improvements. The framework integrates forecasting, graph representation learning, and decision-support theory within a reproducible public-data context, pinpointing when relational learning is most critical, particularly during shocks, promotions, seasonal transitions, and periods of sparse demand. Specifically for information management research, the framework provides a tangible method to represent and estimate the flow of predictive information between organizational units and to audit this flow through interpretable edges. This connection links predictive modeling to issues of data governance and explainable analytics, which are emphasized in these research areas.

It is also crucial to acknowledge the limitations of the approach. In the bike-sharing panel, where demand is primarily influenced by weather and time of day, and cross-entity dependence is weaker, the advantage of TGN-DAM over robust sequence models diminishes. A practitioner dealing with a single, weakly relational series would derive minimal benefit from the additional graph machinery. Therefore, the method is most beneficial when multiple series are genuinely interconnected and when this interconnection fluctuates. Adoption should be based on evidence of relational structure rather than as a default approach.

Managerial implications. Table 6 correlates model outputs with decision-making processes. Dynamic associations identify products and stores whose movements predict shortages, enabling earlier replenishment. Lead-lag and substitution edges reveal promotion spillovers and cannibalization, informing decisions on promotion timing and depth. Multi-horizon store forecasts with association clusters aid in staffing and scheduling, while zone graphs assist in mobility and bike-sharing rebalancing. Rapid changes in association structure act as early-warning signals of market or operational disruptions. In practice, these outputs populate a decision-support dashboard featuring forecasts, uncertainty bands, key drivers, and dynamic association maps, along with action thresholds for reorder points, staffing rules, and rebalancing triggers. Governance necessitates model monitoring, drift detection, audit logs, explainability review, and human-in-the-loop override.

Table 6: From dynamic association insights to managerial decisions.

Decision area	How TGN-DAM supports the decision
Inventory and replenishment	Use co-movement and lead-lag edges to flag entities whose demand foreshadows shortage or surplus, and set reorder points accordingly.
Promotion planning	Detect spillovers, cannibalization, complementarity, and delayed responses from evolving edges.
Staffing and operations	Combine multi-horizon store forecasts with association clusters for scheduling.
Service operations	Use zone or segment graphs for mobility and bike-sharing rebalancing.
Risk monitoring	Treat rapid changes in the association network as early warnings of shocks.

8 Robustness, Validity, and Ethics

Internal validity is ensured through a leakage-safe temporal design, where splits are chronological, and each scaler, encoder, and association graph is derived solely from past data, with feature engineering rigorously audited to prevent future leakage. External validity is addressed by utilizing six public datasets encompassing retail, e-commerce, bike sharing, and mobility sectors. Construct validity is established by operationalizing management time series as demand, sales, transactions, rentals, trips, and revenue, with associations represented as observed and learned edges. Statistical conclusion validity is supported by employing Diebold-Mariano and Friedman-Nemenyi tests, alongside multi-horizon reporting with corrections for multiple comparisons. Model validity is examined through ablation studies and sensitivity analyses concerning graph sparsity, window size, edge type, and memory length.

Several limitations qualify the findings. Public datasets may not fully capture private constraints, such as supplier delays or internal decision rules, rendering decision-value estimates as policy approximations rather than live interventions. Certain datasets include competition or redistribution terms that must be adhered to. Dynamic association does not equate to causality; thus, causal language should be reserved for quasi-experimental designs. Deep graph models are computationally demanding, a challenge mitigated through sparsity, neighbor sampling, and top-k edges, yet scalability analysis remains necessary for very large node counts. Ethically, the study employs public and anonymized data, adheres to repository terms, avoids re-identification, and transparently discloses the limitations of public-data representativeness.

9 Conclusion

Multidimensional management time series exhibit interdependence, with their interconnections evolving over time. We introduced TGN-DAM, a model that constructs a leakage-safe dynamic association graph from both interpretable observed signals and a learned latent component. This model forecasts within the graph using a memory-augmented temporal graph network and reveals the estimated relationships to support decision-making. Evaluated on six public datasets under a strict rolling-origin protocol, TGN-DAM achieves superior accuracy and ranking, particularly during shocks and in sparse regimes, with its accuracy advantage translating into quantifiable decision value. Ablation studies demonstrate that dynamic relational learning and the complementarity of observed and learned edges are key drivers of these outcomes. The framework provides a reproducible approach to relational, adaptive decision intelligence in management analytics. Future research should enhance the design by incorporating quasi-experimental causal identification, larger and sparser graphs, and live organizational deployment.

Author contributions

Wenshuang Zhao: Conceptualization, methodology, investigation, and writing—original draft.

Fengcan Zhang: Data curation, formal analysis, validation, and writing—original draft.

Yang Yuan: Supervision, project administration, and writing—review and editing.

Muhammad Mubashir: Methodological consultation and writing—review and editing. Wenshuang Zhao and Fengcan Zhang contributed equally to this work. All authors read and approved the final manuscript.

References

[1] Makridakis, S., Spiliotis, E., Assimakopoulos, V.: The M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting **38**(4), 1346–1364 (2022)

[2] Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Ben Taieb, S., Bergmeir, C., *et al.*: Forecasting: theory and practice. International Journal of Forecasting **38**(3), 705–871 (2022)

[3] Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T.: DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting **36**(3), 1181–1191 (2020)

[4] Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y.: N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In: International Conference on Learning Representations (ICLR) (2020)

[5] Lim, B., Arık, S.Ö., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting **37**(4), 1748–1764 (2021)

[6] Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI), pp. 3634–3640 (2018)

[7] Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: International Conference on Learning Representations (ICLR) (2018)

[8] Wu, Z., Pan, S., Long, G., Jiang, J., Zhang, C.: Graph WaveNet for deep spatial-temporal graph modeling. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), pp. 1907–1913 (2019)

[9] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y.: LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 3146–3154 (2017)

[10] Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proceedings

of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 785–794 (2016)

[11] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., Zhang, W.: Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 11106–11115 (2021)

[12] Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 34, pp. 22419–22430 (2021)

[13] Benidis, K., Rangapuram, S.S., Flunkert, V., Wang, Y., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., *et al.*: Deep learning for time series forecasting: Tutorial and literature survey. ACM Computing Surveys **55**(6), 1–36 (2022)

[14] Seyedan, M., Mafakheri, F.: Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. Journal of Big Data **7**(1), 53 (2020)

[15] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)

[16] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (ICLR) (2018)

[17] Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 1024–1034 (2017)

[18] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.: A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems **32**(1), 4–24 (2021)

[19] Bai, L., Yao, L., Li, C., Wang, X., Wang, C.: Adaptive graph convolutional recurrent network for traffic forecasting. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 17804–17815 (2020)

[20] Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., Zhang, C.: Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), pp. 753–763 (2020)

[21] Seo, Y., Defferrard, M., Vandergheynst, P., Bresson, X.: Structured sequence modeling with graph convolutional recurrent networks. In: International Conference on Neural Information Processing (ICONIP), pp. 362–373. Springer, ??? (2018)

[22] Jiang, W., Luo, J.: Graph neural network for traffic forecasting: A survey. Expert Systems with Applications **207**, 117921 (2022)

[23] Shao, Z., Zhang, Z., Wang, F., Xu, Y.: Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD), pp. 1567–1577 (2022)

[24] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., Bronstein, M.: Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637 (2020). ICML 2020 Workshop on Graph Representation Learning

[25] Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., Achan, K.: Inductive representation learning on temporal graphs. In: International Conference on Learning Representations (ICLR) (2020)

[26] Pareja, A., Domeniconi, G., Chen, J., Ma, T., Suzumura, T., Kanezashi, H., Kaler, T., Schardl, T.B., Leiserson, C.E.: EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 5363–5370 (2020)

[27] Kazemi, S.M., Goel, R., Jain, K., Kobyzev, I., Sethi, A., Forsyth, P., Poupart, P.: Representation learning for dynamic graphs: A survey. Journal of Machine Learning Research **21**(70), 1–73 (2020)

[28] Skarding, J., Gabrys, B., Musial, K.: Foundations and modeling of dynamic networks using dynamic graph neural networks: A survey. IEEE Access **9**, 79143–79168 (2021)

[29] Gravina, A., Bacciu, D.: Deep learning for dynamic graphs: Models and benchmarks. IEEE Transactions on Neural Networks and Learning Systems **35**(9), 11788–11801 (2024)

[30] Rozemberczki, B., Scherer, P., He, Y., Panagopoulos, G., Riedel, A., Astefanoaei, M., Kiss, O., Béres, F., López, G., Collignon, N., Sarkar, R.: PyTorch Geometric Temporal: Spatiotemporal signal processing with neural machine learning models. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM), pp. 4564–4573 (2021)

[31] Diebold, F.X., Mariano, R.S.: Comparing predictive accuracy. Journal of Business & Economic Statistics **13**(3), 253–263 (1995)

[32] Hyndman, R.J., Koehler, A.B.: Another look at measures of forecast accuracy. International Journal of Forecasting **22**(4), 679–688 (2006)

[33] Chen, D., Sain, S.L., Guo, K.: Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Journal of Database Marketing & Customer Strategy Management **19**(3), 197–208 (2012)

[34] Fanaee-T, H., Gama, J.: Event labeling combining ensemble detectors and background knowledge. Progress in Artificial Intelligence **2**(2), 113–127 (2014)

[35] Taylor, S.J., Letham, B.: Forecasting at scale. The American Statistician **72**(1), 37–45 (2018)

[36] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation **9**(8), 1735–1780 (1997)

[37] Cho, K., Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734 (2014)

[38] Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271 (2018)