



Dynamic Graph Representation Learning for Spatiotemporal Forecasting under High Data Volatility

Ms. Pratibha Soni¹, Ms. Deepika Agrawal²

¹Assistant Professor, Jagannath University, Jaipur

²Assistant Professor, JECRC University, Jaipur

DOI : <https://doi.org/10.55248/gengpi.07.0926.2901>

ABSTRACT

Spatiotemporal forecasting plays a critical role in managing large-scale physical, economic, and digital infrastructures. Applications range from urban traffic management and financial order book modeling to smart grid load balancing and climate monitoring. Modern spatiotemporal architectures—primarily Spatiotemporal Graph Neural Networks (ST-GNNs)—rely heavily on spatial topology graphs. These graphs are typically constructed using physical proximity, structural connectivity, or long-term historical correlation.

However, in non-stationary real-world environments, systems experience frequent high-volatility events, such as traffic accidents, sudden financial market crashes, localized weather anomalies, or unexpected power grid surges. Under these volatile conditions, static spatial assumptions fail. Spatial dependencies undergo sudden non-linear shifts, causing traditional static and parameter-bound adaptive graph neural networks to experience significant performance degradation.

This paper presents the Volatile Dynamic Graph Network (VDG-Net), an adaptive end-to-end spatiotemporal representation learning framework designed for non-stationary environments under high data volatility. VDG-Net combines a Dynamic Topology Generation Module (DTGM), a Decoupled Spatiotemporal Gating Layer, and a Distributionally Robust Loss (DR-Loss) with adaptive variance regularization. Evaluated across three real-world volatile datasets (METR-LA-VOL, PEMS-BAY-VOL, and NE-GRID), VDG-Net achieves up to a 14.2% reduction in Mean Absolute Error (MAE) and a 16.8% reduction in Root Mean Squared Error (RMSE) over competitive baselines.

1. Introduction

Spatiotemporal forecasting focuses on predicting future feature sequences across a network of interconnected spatial entities over time. Accurate spatiotemporal predictions drive key decision-making pipelines in modern smart cities, automated financial trading systems, distributed logistics networks, and intelligent power infrastructures.

1.1 Background and Motivation

The emergence of Graph Neural Networks (GNNs) led to a major shift in spatiotemporal modeling. By representing spatial relationships as a graph $G = (V, E, A)$, where V represents sensors or observational nodes, E represents edges, and $A \in \mathbb{R}^{(N \times N)}$ represents the adjacency matrix, Spatiotemporal Graph Neural Networks (ST-GNNs) capture spatial correlations via neighborhood message passing. These models combine graph convolutions with sequential architectures such as Recurrent Neural Networks (RNNs) or Temporal Convolutional Networks (TCNs).

Despite these advances, existing spatiotemporal architectures operate under a fundamental assumption: spatial topology remains stationary over time, or evolves smoothly. In practice, real-world systems regularly encounter high volatility, characterized by rapid, non-linear shifts in inter-node relationships:

Urban Transportation: Severe traffic accidents on major freeways cause sudden congestion, altering localized spatial dependencies instantly. Vehicles divert to surface streets, creating strong spatial correlations between previously unrelated road segments.

Financial Order Book Networks: Market volatility spikes or flash crashes cause liquidity correlations across asset classes to mutate within milliseconds, rendering historical covariance metrics uninformative.

Smart Power Grids: High-penetration renewable energy sources cause sudden localized voltage changes during cloud cover or wind drops, shifting inter-substation load dynamics rapidly.

When applied to volatile regimes, traditional models suffer from two primary limitations: (1) Topology Mismatch, where static adjacency matrices force graph convolution layers to aggregate messages along invalid pathways; and (2) Gradient Instability, where standard loss functions like Mean Squared Error (MSE) over-index on transient, extreme outlier spikes.

1.2 Core Contributions

To address the challenge of spatiotemporal forecasting under extreme data volatility, this paper introduces the Volatile Dynamic Graph Network (VDG-Net). The key contributions of this work are summarized as follows:

Dynamic Graph Topology Generation: We design a Dynamic Topology Generation Module (DTGM) that generates continuous-time dynamic adjacency matrices $A^*(t)$ at every time step t using dynamic self-attention constrained by spatial priors.

Decoupled Spatiotemporal Architecture: We introduce a decoupled architecture that separates baseline steady-state trend modeling from volatility adaptation, retaining long-term structural dependencies while adapting to sudden shock events.

Distributionally Robust Loss (DR-Loss): We formulate an adaptive loss function combining Huber Loss with variance regularization to stabilize gradients during extreme volatility spikes.

Comprehensive Empirical Validation: We create three volatile spatiotemporal benchmark datasets (METR-LA-VOL, PEMS-BAY-VOL, and NE-GRID) and run extensive experiments comparing VDG-Net against leading baseline architectures.

2. Literature Review

Spatiotemporal neural networks have evolved from early time-series models to dynamic graph architectures. This section reviews key developments across three primary categories.

Paradigm	Topology Mechanics	Weakness under Volatility
Static ST-GNNs (STGCN, DCRNN)	Fixed physical distance matrix $A_{ij} = \exp(-d^2 / \sigma^2)$	Fails during topological shifts
Adaptive Graph Networks (Graph WaveNet, AGCRN)	Static learned embeddings $A_{\text{adp}} = \text{Softmax}(E_s \cdot E_t^T)$	Parametric freeze post-training; cannot adapt to live shock events
Dynamic Graph Networks (TGN, VDG-Net)	Continuous-time generation $A^*(t) = f(X_t, H_{t-1})$	High computational cost; addressed via linear attention in VDG-Net

2.1 Static Spatiotemporal Graph Neural Networks

Early spatiotemporal models integrated spatial graph convolutions with temporal sequence models. DCRNN modeled spatial dependencies as a bidirectional random walk over a physical distance graph, using Graph Convolutional Recurrent Units for temporal modeling. STGCN replaced recurrent layers with 1D causal temporal convolutions, significantly accelerating training times while maintaining competitive performance. However, both rely on a fixed, predefined adjacency matrix calculated from geographical distances, assuming that physical distance dictates node interactions.

2.2 Adaptive and Learned Graph Neural Networks

To address the limitations of physical distance matrices, recent architectures learn graph structures directly from data. Graph WaveNet introduced parameter-bound adaptive node embeddings, computing an adaptive adjacency matrix via matrix multiplication: $A_{\text{adaptive}} = \text{Softmax}(\text{ReLU}(E_{\text{src}} \cdot E_{\text{tgt}}^T))$. AGCRN learned node-specific patterns alongside traditional recurrent updates. While these models capture hidden spatial correlations without a manual graph, the learned embeddings become fixed after training, rendering them unable to adapt to live inference shifts.

2.3 Dynamic Graph Representation Learning

Dynamic Graph Neural Networks (DGNNs) represent systems where structure evolves over time, divided into Discrete-time Dynamic Graphs (DTDGs) and Continuous-time Dynamic Graphs (CTDGs). While CTDGs like Temporal Graph Networks (TGN) excel at event-driven interactions, applying them directly to dense spatiotemporal sensor grids presents computational challenges ($O(N^2)$ at every step) and lacks explicit mechanisms to separate baseline signal trends from temporary high-frequency volatility spikes.

3. Mathematical Framework & Problem Formulation

3.1 Notation and Definitions

Let $G_t = (V, E_t, A^t(t))$ represent a continuous dynamic spatiotemporal graph snapshot at time t , where V is the set of $N = |V|$ observational nodes, $E_t \subseteq V \times V$ represents dynamic functional edges at time t , $A^t(t) \in \mathbb{R}^{(N \times N)}$ is the non-negative weighted dynamic adjacency matrix, and $X_t \in \mathbb{R}^{(N \times F)}$ represents the feature tensor across N nodes.

3.2 Formal Spatiotemporal Forecasting Problem

Given an observed historical signal tensor $X_{-(t-P+1:t)} \in \mathbb{R}^{(P \times N \times F)}$ and a static physical spatial prior matrix $A_{\text{static}} \in \mathbb{R}^{(N \times N)}$, the spatiotemporal forecasting objective is to learn a parameterized function $f_{\Theta}(\cdot)$ that maps historical sequences to target values $Y_{-(t+1:t+H)} \in \mathbb{R}^{(H \times N \times C)}$ over a future forecast horizon H :

Spatiotemporal Forecast Equation:

$$\hat{Y}_{-(t+1:t+H)} = f_{\Theta}(X_{-(t-P+1:t)}; A_{\text{static}})$$

3.3 Volatility Definition and Metrics

To quantify spatial and temporal volatility mathematically, we establish two volatility indices used for evaluation:

Temporal Volatility Index (V_{temp}):

Measures localized feature shifts across time steps relative to node-level standard deviation: $V_{\text{temp}}(i, t) = |X_{-}(t, i) - X_{-}(t-1, i)| / (\sigma_i + \epsilon)$

Spatial Volatility Index (V_{spat}):

Quantifies sudden spatial divergence between a node v_i and its structural neighborhood $N(i)$: $V_{\text{spat}}(i, t) = (1 / |N(i)|) \sum |X_{-}(t, i) - X_{-}(t, j)|$

4. Proposed Methodology: VDG-Net

This section details the internal architecture of the Volatile Dynamic Graph Network (VDG-Net).

4.1 Dynamic Topology Generation Module (DTGM)

The Dynamic Topology Generation Module (DTGM) continuously infers instantaneous spatial adjacency matrices $A^t(t)$ from raw node observations.

Step 1: Feature Projection and Scaled Attention: At time step t , input features X_t are projected into query and key spaces $Q_t, K_t = X_t W_Q, X_t W_K$. Instantaneous similarity $S^t(t)$ is computed via scaled dot-product attention: $S^t(t) = \text{Softmax}(Q_t K_t^T / \sqrt{d_k})$.

Step 2: Adaptive Prior Gating: To prevent overfitting to localized noise, DTGM incorporates static physical priors A_{static} and parameter-learned background graphs A_{base} using a learnable spatial gating parameter $\gamma_t = \sigma(X_t W_{\gamma} + b_{\gamma})$. The final matrix is: $A^t(t) = (1 - \gamma_t) \odot (\alpha A_{\text{static}} + (1 - \alpha) A_{\text{base}}) + \gamma_t \odot S^t(t)$.

4.2 Decoupled Spatiotemporal Aggregation Layer

VDG-Net stacks multiple Decoupled Spatiotemporal Blocks (DST-Blocks), combining Spatial Graph Convolutional Networks (S-GCN) with Dynamic Gated Temporal Convolutional Networks (T-TCN).

4.3 Distributionally Robust Loss (DR-Loss)

VDG-Net employs a Distributionally Robust Loss (DR-Loss) with an adaptive variance regularization penalty:

DR-Loss Formulation:

$$L_{\text{DR}}(\Theta) = (1 / (N \cdot H)) \sum \sum \ell_{\delta}(Y_{-(t+h)}^i, \hat{Y}_{-(t+h)}^i) + \lambda_{\text{var}} \cdot \text{Var}(\{ \ell_{\delta}(Y_{-(t+h)}^i, \hat{Y}_{-(t+h)}^i) \})$$

where $\ell_{\delta}(\cdot)$ is the robust Huber Loss with threshold δ set dynamically to the 90th percentile of running epoch residual errors.

5. Experimental Setup & Protocol

To benchmark model performance under volatile conditions, we constructed three volatile spatiotemporal datasets:

Dataset	Nodes (N)	Time Steps	Sampling	Target Variable
METR-LA-VOL	207	34,272	5 minutes	Speed (mph)
PEMS-BAY-VOL	325	52,116	5 minutes	Speed (mph)
NE-GRID	118	26,280	15 minutes	Load (MW)

6. Results & Performance Analysis

Performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Dataset	Metric	HA	STGCN	DCRNN	WaveNet	TGN	VDG-Net
METR-LA-VOL	MAE	9.12	7.25	6.81	6.12	5.95	5.10 (14.2%↓)
METR-LA-VOL	RMSE	14.30	12.10	11.45	10.82	10.40	8.92 (14.2%↓)
METR-LA-VOL	MAPE	18.50%	14.12%	13.20%	11.85%	11.20%	9.65% (13.8%↓)
PEMS-BAY-VOL	MAE	6.80	4.95	4.22	3.85	3.62	3.01 (16.8%↓)
PEMS-BAY-VOL	RMSE	11.50	8.90	7.95	7.15	6.85	5.70 (16.7%↓)
PEMS-BAY-VOL	MAPE	14.20%	10.10%	8.80%	7.90%	7.30%	6.10% (16.4%↓)
NE-GRID	MAE	12.45	9.30	8.85	7.92	7.50	6.18 (17.6%↓)
NE-GRID	RMSE	21.10	16.40	15.20	13.80	12.90	10.25 (20.5%↓)
NE-GRID	MAPE	16.80%	12.30%	11.50%	10.10%	9.40%	7.80% (17.0%↓)

7. Ablation Study

To evaluate the contribution of each architectural component within VDG-Net, we conducted an ablation study on METR-LA-VOL (H=12):

Architecture Variant	MAE (60 min)	RMSE (60 min)	Performance Impact
Variant A (Full VDG-Net)	5.10	8.92	Base Baseline
Variant B (w/o Dynamic Topology)	6.45	11.20	+26.4% Error
Variant C (w/o Spatial Prior Gate)	5.88	10.45	+15.3% Error
Variant D (w/o DR-Loss)	5.68	9.95	+11.3% Error
Variant E (w/o T-TCN Layers)	5.82	10.15	+14.1% Error

8. Conclusion & Future Work

This paper presented the Volatile Dynamic Graph Network (VDG-Net) for spatiotemporal forecasting in non-stationary environments. By combining dynamic topology generation, decoupled spatiotemporal gating, and distributionally robust loss functions, VDG-Net achieves state-of-the-art accuracy under extreme data volatility across transportation and energy grid benchmarks.