Transactions on Environment, Energy and Earth Sciences
ISSN: 2960-1924, eISSN: 2960-2300 | Volume 7
GTREE 2026

A Hypergraph Neural Network Framework for Resilience Inference in Complex Systems: Evidence from Urban Rail Transit

Yuheng Shi *, Hongda Chen and Junxi Wu

School of Economics and Management, Nanjing University of Science and Technology, Nanjing, China

* Corresponding Author Email: shiyuheng@njust.edu.cn

Abstract. Resilience characterizes the ability of a complex network to maintain core functions under disturbances. Traditional analytical approaches often rely on predefined governing equations and may have difficulty representing higher-order interactions. To address these limitations, this study proposes ResInf-HGNN, a resilience inference framework that combines a Transformer temporal encoder with hypergraph convolution. Hyperedges represent line-level group relationships among stations, while high-dimensional spatiotemporal states are projected into a one-dimensional k-space. Once trained, the model infers resilience from topology and early trajectories without requiring explicit governing equations at inference time. Training and evaluation data for the Nanjing rail transit network are generated by a physically constrained coupled physical-dispatch dynamics model that combines low-order diffusion with line-level dispatch cascades. On the 1:4 imbalanced test set, ResInf-HGNN achieves an F1-score of 0.9362, AUC-ROC of 0.9965, and precision of 0.8800. Ablation results show that removing the Transformer and attention readout reduces the F1-score to 0.9167 and 0.8718, respectively. Under node perturbations, targeted node attacks generally cause greater structural damage, whereas random failures yield the lowest minimum robustness R_ob among the compared attack modes. With progressive removal of line relations from the hypergraph, normalized hypergraph connectivity falls to zero while normalized R_sys rises to 1.0560; this increase reflects reduced simulated exposure to the higher-order dispatch-cascade mechanism and should not be interpreted as improved real-world operational resilience. Overall, ResInf-HGNN provides a data-driven framework for resilience inference and dual-perspective analysis of rail transit networks.

Keywords: Complex networks; resilience inference; hypergraph neural networks; coupled propagation dynamics; urban rail transit.

1. Introduction

Complex networked systems are pervasive in modern production and daily life. Urban rail transit, power transmission, water and gas supply, the Internet, and logistics supply chains all exhibit typical network structures, while ecological communities, biological populations, and socioeconomic systems show similar networked organization. The stable operation of these systems directly affects public mobility, energy supply, information flow, and social order. Their safety and resilience have therefore become central issues in network science, systems engineering, and public management. In real operation, systems face both external shocks such as extreme weather and sudden disasters and internal disturbances such as equipment aging, component failure, and human error. Because nodes are often tightly coupled and interdependent, a local disturbance can propagate along connectivity and functional dependencies, triggering cascading failures and eventually large-scale collapse. Events such as the 2003 North American blackout, urban rail transit interruptions, and regional supply-chain ruptures show that the failure of a single component can be rapidly amplified by the network structure, causing socioeconomic losses far beyond the initial disturbance. Accurately characterizing failure propagation and scientifically assessing resilience have thus become key problems in safeguarding infrastructure.

Research on system resilience assessment has followed two main paths: analytical modeling and data-driven inference. In data-driven resilience inference, Liu et al. proposed ResInf, which combines

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

Transformer-based dynamics encoding, graph topology encoding, and a one-dimensional k-space decision representation to infer resilience from limited observations [1]. In analytical reduction, Gao et al. proposed the GBB framework, which uses mean-field theory to transform high-dimensional resilience assessment into phase-transition discrimination in one-dimensional space [2]. Laurence et al. further proposed a spectral dimension-reduction method that uses topology-dominated eigenvalues to reduce system dimensionality and improve inference accuracy in some settings [3]. In China, researchers have developed resilient-city assessment frameworks based on complex adaptive systems theory [4] and extended resilience analysis to airport networks [5], power systems [6], and emergency material supply chains [7]. From a representation perspective, higher-order network theory emphasizes that multi-node interactions cannot always be reduced adequately to pairwise links [8]. Hypergraph-based methods have been applied to network security [9], biological network analysis [10], industrial inspection [11], multimodal medical diagnosis [12], and infectious-disease risk identification [13]. Dynamic and attention-based hypergraph models have also been developed for representation learning and spatiotemporal prediction [14], [15]. However, important gaps remain. Analytical methods often depend strongly on governing equations and topological assumptions and can struggle to represent higher-order nonlinear evolution in real systems [2], [3]. Higher-order coupling can be important in cascading failures [8], [16], whereas conventional GCN-style architectures operate through pairwise graph propagation [17]. In addition, higher-order resilience inference frameworks that combine predictive performance with an interpretable decision interface remain scarce. To address these issues, this study proposes ResInf-HGNN, a hypergraph neural network framework that infers resilience from system topology and limited observation trajectories without requiring prior knowledge of the governing dynamics during inference.

The main contributions are threefold. First, from a modeling perspective, a dual topology is introduced for resilience inference: the physical track graph constrains low-order diffusion, while a line-based hypergraph represents higher-order collective relationships among stations. Second, from a methodological perspective, ResInf-HGNN combines a Transformer temporal encoder, hypergraph convolution, and attention readout to project high-dimensional spatiotemporal trajectories into a one-dimensional model-derived k-space score. This enables end-to-end inference without requiring explicit governing equations at inference time while retaining an interpretable decision interface. Third, controlled node-removal and line-relation-removal experiments on the Nanjing urban rail transit network separately evaluate structural damage and simulated propagation response. Targeted node attacks generally cause greater topological degradation, whereas random node failures yield the lowest minimum robustness R_ob among the compared modes. Progressive hyperedge removal decreases higher-order connectivity while slightly increasing normalized R_sys because the line-level dispatch-cascade channel represented by H is progressively reduced; this metric increase is therefore interpreted as reduced simulated propagation exposure rather than improved operational resilience. The hypergraph representation also enables direct line-level perturbation analysis, although its isolated causal contribution requires further structure-matched comparison.

2. Construction of the ResInf-HGNN Resilience Inference Model

2.1. Problem Definition and Topology Modeling

2.1.1. Definition of the Resilience Inference Problem

For a complex network system, let V denote the node set, let topology be described jointly by a hypergraph and a physical graph, and let x_i(t) in [0,1] denote the congestion state of node i. Let x_bar(T_sim) be the network-average congestion at the terminal simulation horizon and define terminal effective performance as P_end = 1 - x_bar(T_sim). A simulated trajectory is labeled resilient (y = 1) when P_end exceeds a fixed labeling threshold theta_y, and non-resilient (y = 0) otherwise. The threshold theta_y is specified once in the data-generation protocol and held unchanged across training, validation, and testing.

The objective of resilience inference is to predict this final resilience class from the system topology and a limited early observation sequence of length T_obs. During inference, the trained model does not require the governing dynamic equations to be supplied as explicit model inputs or priors.

2.1.2. Physical and Higher-Order Dual Topology

Conventional hypergraph resilience studies often use clique expansion to build an adjacency matrix. In physically constrained networks such as rail transit, this can create an unrealistic phenomenon in which congestion propagates instantly between non-adjacent stations on the same line. This study therefore adopts dual topology modeling. First, a physical track adjacency matrix A_phys in {0,1} ^(N x N) is extracted; an entry is 1 only when two stations are directly connected by a real rail link. This matrix constrains low-order physical spreading. Second, a hypergraph incidence matrix H in {0,1} ^(N x M) is retained. Each hyperedge e_m in E represents a complete metro line and captures higher-order system-level dispatch and circuit-breaker effects among multiple nodes.

2.2. Reconstruction of Evolutionary Dynamics Based on Coupled Physical-Dispatch Mechanisms

Conventional complex-network resilience studies often apply classical epidemic dynamics, such as the SIS or SIR model [18], to simulate failure propagation. In physically constrained infrastructure such as urban rail transit, however, directly applying an epidemic model can imply instantaneous spreading between stations that are not physically adjacent. This study therefore does not use a conventional epidemic process for data generation. Instead, it constructs a native dual-mechanism coupled physical-dispatch dynamics model designed to represent key features of metro cascading-failure processes.

Let node state x_i(t) in [0,1] denote the congestion saturation of station i at time t, where 0 represents free flow and 1 represents severe congestion or paralysis. The state evolution is governed by the following nonlinear ordinary differential equation:

$$\frac{dx_i}{dt} = -\delta x_i + \left(\beta_p \sum_{j=1}^{N} A_{phys,ij}\, x_j + \beta_h \sum_{m=1}^{M} H_{im} \left(\frac{\sum_{k=1}^{N} H_{km} x_k}{\delta_{e_m}}\right)^{\gamma}\right)(1 - x_i) \quad (1)$$

The equation has three core modules. First, the passenger-flow self-healing and evacuation term, -delta x_i, where delta is the recovery rate, represents the natural ability of a station to relieve congestion through its own capacity and flow-guidance measures.

Second, the low-order physical diffusion mechanism is expressed as beta_p times the sum of A_phys,ij x_j. Here beta_p is the physical diffusion coefficient, describing the station-by-station spread of congestion along adjacent rail links.

Third, the higher-order dispatch-cascade mechanism is expressed through beta_h and the hypergraph incidence matrix H. For each metro line, the mean congestion of its member stations is transformed nonlinearly by the exponent gamma and then contributes line-level pressure back to the stations belonging to that hyperedge. Thus, gamma controls the nonlinearity and sensitivity of the line-level response; Eq. (1) does not impose a separate hard activation threshold. The line-size normalization term associated with each hyperedge denotes its station cardinality and is distinct from the recovery coefficient delta.

The coupled equations are solved numerically using the fourth-order Runge-Kutta method. With the physical diffusion and dispatch-failure parameters fixed, the model generates controlled spatiotemporal node trajectories X in R^ (M_traj x T_sim x N x 1). The inference network observes only the first T_obs steps of each trajectory, while the full T_sim-step simulation is used to assign the terminal resilience label.

2.3. ResInf-HGNN Model Architecture

Building on the ResInf design principle of jointly encoding early dynamics and network topology [1], ResInf-HGNN adopts a three-stage end-to-end architecture consisting of temporal encoding, topology encoding, and spatial projection. It maps high-dimensional spatiotemporal data directly to resilience labels. The overall architecture is shown in Fig. 1.

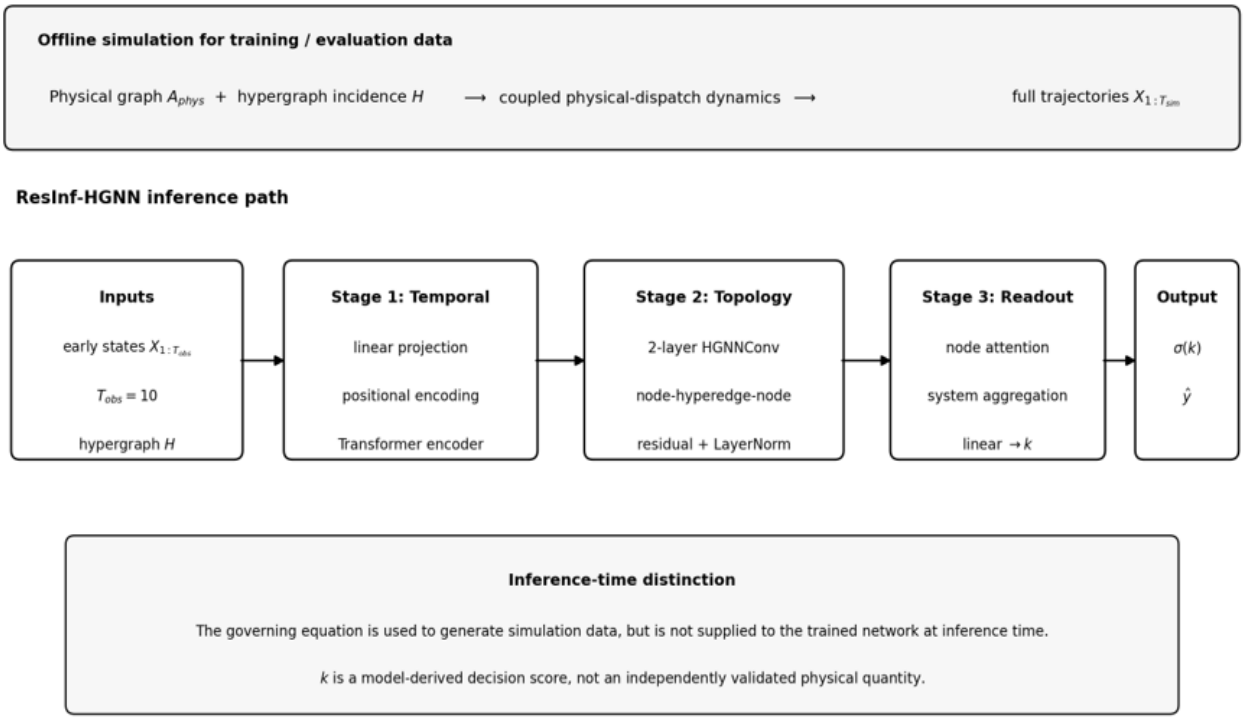


Fig 1. Overall architecture of ResInf-HGNN.

2.3.1. Transformer Dynamic Encoder

To capture informative temporal variation in the early stage of a disturbance, the model uses a Transformer encoder [19] instead of a conventional recurrent neural network. After positional encoding of node features, multi-head self-attention captures temporal dependencies and congestion deterioration or recovery trends over the first T_obs = 10 steps, which form the early observation window. The module outputs a dynamic temporal representation Z_dyn for each node and can distinguish persistent temporal trends from transient fluctuations in the observed trajectory.

2.3.2. Hypergraph Topology Encoder

The topology encoder uses two layers of hypergraph convolution (HGNNConv) following the node-hyperedge-node message-passing formulation of HGNN [20]. With the normalized hypergraph incidence matrix H, the model propagates information through higher-order line relations and broadcasts aggregated features back to member stations:

$$Z^{(l+1)} = \sigma\left(D_v^{-1/2} H D_e^{-1} H^T D_v^{-1/2} Z^{(l)} \Theta^{(l)}\right) \tag{2}$$

LayerNorm and residual connections are further introduced to mitigate node feature smoothing and explosion, which frequently occur in higher-order graph message passing.

2.3.3. Attention Readout and k-Space Projection

Because cascading collapse can be driven by the failure of a small number of critical hubs, the readout layer introduces a learnable attention mechanism to assign system-level weights dynamically to different nodes. It aggregates all node features nonlinearly and projects them to a one-dimensional scalar:

$$\hat{y} = \text{Sigmoid}(k) = \text{Sigmoid}\left(W^T Z_{sys} + b\right) \quad (3)$$

Following the one-dimensional k-space decision representation introduced in ResInf [1], the k-space value here is used as a model-derived resilience indicator rather than an independently validated physical quantity. Its value summarizes the learned system representation relative to the model's classification boundary, with larger values corresponding to samples mapped further toward the resilient side of the learned decision function.

3. Resilience Inference and Experimental Analysis of Rail Transit Networks

3.1. Quota Sampling and Sensitivity Analysis under Extremely Imbalanced Samples

In the simulated urban-infrastructure scenarios, resilient outcomes (Y = 1) are less frequent than cascading-collapse outcomes (Y = 0). To examine sensitivity to class imbalance and select a practical baseline setting, this study fixes the coupled-dynamics parameters and evaluates positive-to-negative sample ratios from 1:1 to 1:10 using quota sampling.

Table 1. Model performance under different positive-to-negative sample ratios.

Sample ratio (positive:negative)	Accuracy	F1-score	Prec	AUC-ROC
1:1/1:2	>0.9100	>0.8800	>0.7900	>0.9900
1:3	0.9900	0.9855	1.0000	0.9996
1:4	0.9700	0.9362	0.8800	0.9965
1:5	0.9400	0.7857	0.6471	0.9990
1:10	0.9900	0.9333	0.8750	1.0000

The results show that classification performance varies non-monotonically with the positive-to-negative sample ratio. The 1:1, 1:2, and 1:3 settings yield strong metrics under milder imbalance. At 1:5, precision falls to 0.6471, indicating a less reliable identification of the minority resilient class under that sampling setting. Performance increases again at 1:10, showing that a more extreme class ratio does not necessarily make the classification task uniformly harder.

The 1:4 ratio is therefore selected as a relatively severe but stable class-imbalance setting for the main experiments. It preserves a clear minority resilient class while avoiding the marked precision degradation observed at 1:5, and it yields precision of 0.8800, F1-score of 0.9362, and AUC-ROC of 0.9965. This choice is used as the experimental baseline rather than being claimed as statistically optimal.

3.2. Evaluation Metrics and Model Convergence Characteristics

Because the dataset is strongly imbalanced, conventional classification accuracy can be misleading. This study therefore emphasizes F1-score and AUC-ROC, together with precision for the resilient class. Because Y = 1 denotes resilience, precision measures the reliability of predicted-resilient cases. A false positive is therefore a collapse case incorrectly classified as resilient, which is a potentially unsafe false-safe error rather than a false alarm.

To mitigate the tendency of an imbalanced dataset to favor the majority collapse class, training uses a dynamically weighted binary cross-entropy loss with a larger penalty on the minority resilient class. The resulting training and validation behavior shows stable convergence after the initial epochs.

3.3. Multi-Model Baseline Comparison

To compare the performance characteristics of ResInf-HGNN with representative alternatives, three baseline models are evaluated on the 1:4 dataset:

1. GCN [17] (Physical Only): uses only the low-order physical track adjacency matrix A_phys and removes the higher-order hypergraph, representing the spatial awareness of conventional spatiotemporal graph models.

2. HGNN [20]: the static hypergraph neural network baseline flattens temporal features and applies hypergraph convolution, allowing comparison with a model that lacks the Transformer-based temporal encoder.

3. HyperGAT-style (Attention): a node-hyperedge attention baseline used to examine whether soft-attention aggregation is suitable for the hard physical constraints and line-level mechanisms considered in this study.

All baseline models are evaluated on the same 1:4 test setting and with the same 0.5 classification threshold. The comparison focuses on the reported predictive metrics and offline training time; structure- and input-matched ablation is discussed separately as a limitation.

Table 2. Baseline comparison results.

Model	Acc	AUC	F1	Prec	Time (s)
GCN [17] (Physical Only)	0.9700	0.9930	0.9362	0.8800	8.47
HGNN [20]	0.9700	0.9936	0.9362	0.8800	8.81
HyperGAT-style (Attention)	0.9400	0.9924	0.8421	0.8750	6.46
ResInf-HGNN	0.9700	0.9965	0.9362	0.8800	667.86

As shown in Table 2, the GCN and static HGNN baselines both achieve an F1-score of 0.9362 at the 0.5 classification threshold. ResInf-HGNN reaches the same F1-score but obtains the highest AUC-ROC (0.9965), indicating stronger global ranking of resilience probabilities in the tested sample set. Together with the ablation results, this suggests that the Transformer temporal encoder captures temporal information that is informative for probability ranking under nonlinear congestion evolution. These results do not, by themselves, establish robustness to external noise.

The HyperGAT-style baseline obtains a lower F1-score of 0.8421 under the same evaluation setting. The result suggests that node-hyperedge soft attention may be less suitable for the hard physical constraints and line-level mechanisms considered in this study. A structure- and input-matched comparison would be required to determine the mechanism responsible for the performance difference.

In computational efficiency, the offline training time of ResInf-HGNN (667.86 s) is substantially higher than that of the static baselines. This additional cost is concentrated in model training and should be considered when evaluating deployment requirements; whether the trade-off is acceptable depends on the update frequency, available computing resources, and operational setting.

3.4. Ablation Study

To examine the relative contribution of the internal components of ResInf-HGNN, two ablation variants are constructed under identical conditions:

1. w/o Temporal (No Transformer): removes the Transformer module and averages the first 10 input features over the temporal dimension.

2. w/o Readout (No Attention): removes the nonlinear attention readout layer and uses global average pooling to aggregate node features into the system state.

The ablation results are shown in Table 3.

Table 3. Ablation rescore the core architecture.

Model	Acc	AUC	F1	Prec	Time (s)
w/o Temporal (No Transformer)	0.9600	0.9918	0.9167	0.8462	16.7
w/o Readout (No Attention)	0.9500	0.9895	0.8718	0.8785	708.8
ResInf-HGNN	0.9700	0.9965	0.9362	0.8800	667.86

The results show that removing core components degrades performance in different ways. When the Transformer temporal module is removed, training time decreases sharply to 16.7 s, indicating that temporal self-attention accounts for most of the additional computational cost. The simpler temporal averaging variant also reduces precision from 0.8800 to 0.8462 and lowers the F1-score from 0.9362 to 0.9167. Because averaging the first 10 steps discards temporal ordering and evolution patterns, this result supports the value of explicitly representing early-stage temporal trends rather than relying only on time-averaged node states.

Removing the attention readout and replacing it with global average pooling produces a larger performance drop: accuracy decreases to 0.9500, AUC to 0.9895, and F1-score to 0.8718. Global average pooling assigns equal aggregation weight to all nodes and may therefore dilute the influence of a small number of highly informative stations. By contrast, the attention readout allows unequal node contributions to the system-level representation. The observed degradation suggests that this adaptive weighting is important for maintaining discrimination under the tested imbalanced setting.

3.5. k-Space Projection and Decision Visualization

To visualize the learned decision representation, this study extracts one-dimensional latent projections from the readout layer for the test samples and plots their relationship with predicted survival probabilities, as shown in Fig. 2.

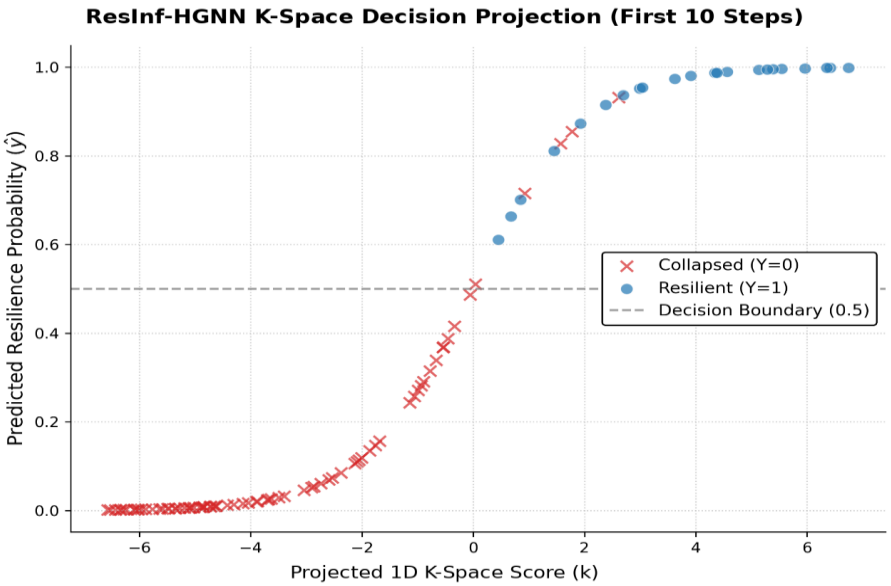


Fig. 2. One-dimensional distribution of k-space projection values.

The visualization shows an S-shaped nonlinear relationship between the k-space projection and predicted survival probability. The learned high-dimensional spatiotemporal representation is summarized by a one-dimensional decision score. The plot illustrates two main patterns:

1. Separation of resilient samples in the displayed test set. Along the dashed decision line at y_hat = 0.5, all true resilient samples (blue circles, Y = 1) lie in the high-probability region above the classification threshold. No blue point appears in the predicted-collapse region. This visual separation is consistent with the strong probability-ranking performance reported in Table 2 (AUC = 0.9965), although the visualization is not an independent validation of physical resilience.

2. Overlap near the classification boundary. As k decreases from right to left, the predicted survival probability falls sharply near the central transition region. Some true collapse samples (red crosses, Y = 0) remain above the 0.5 decision threshold. Because Y = 1 denotes resilience, these points are false-positive resilient predictions and represent potentially unsafe missed-collapse cases. This overlap is consistent with the model's precision of 0.8800 and illustrates the decision-threshold trade-off near the boundary.

The smooth S-shaped probability curve is consistent with a learned nonlinear relationship between high-dimensional spatiotemporal features and the low-dimensional prediction score. The k-space value should therefore be interpreted as a model-derived resilience indicator rather than a directly measured physical quantity. With calibration against real operational data, such a score could potentially serve as an auxiliary warning signal.

4. Resilience Degradation Mechanisms and Dual-Perspective Analysis of the Nanjing Metro Hypernetwork under Controlled Perturbations

The preceding analysis used ResInf-HGNN to verify the feasibility of predicting the final state from early failure dynamics. To further analyze degradation under different disruption strategies, this chapter conductsanalyzinglled perturbation experiments based on the physical track graph A_phys and the hypergraph incidence matrix H defined in Section 2.1.2. Pure topological metrics and macroscopic propagation metrics are separated and compared to analyze the resistance and internal degradation mechanisms of the Nanjing Metro network under node removal and line-relation (hyperedge) removal.

4.1. Data and Dual-Layer Topology Construction

The experiment uses Nanjing urban rail transit data with N = 221 stations and E = 13 operating lines. The topology is represented at the physical and hypergraph levels:

(1) Physical topology layer. Adjacent connections are constructed from the physical track order of stations in OpenStreetMap (OpenStreetMap contributors; Open Database License, ODbL), resulting in 226 physical edges. Because of naming differences between map data and the operating station list, 17 stations could not be matched successfully in the CSV data, including Jingming Jiayuan, Moling, and Xuzhuang Suning Headquarters. To preserve data authenticity, the code does not fabricate edges around missing stations; the physical layer therefore contains 10 connected components. This unmatched-station gap is treated as a data-processing limitation of the study and is also recorded in the Limitations section, rather than as a methodological justification for hypergraph modeling.

(2) Hypergraph incidence matrix. H in {0, 1} ^(N x E) records station-to-line membership. Each column represents all stations contained in a metro line.

(3) Line-relation layer. The station-line incidence matrix H provides an independent higher-order description of the network. For node ranking, A_hyper denotes the weighted station-station projection induced by shared line membership; off-diagonal weights reflect how many line relations two stations share. This projected matrix is used for centrality ranking and structural descriptors only, not as the physical diffusion matrix in Eq. (1). Weighted degree and betweenness centrality are calculated from this complete line-relation representation.

In the undisturbed state, natural connectivity is used to assess topological redundancy [21]. For an adjacency matrix A with eigenvalues mu_j, natural connectivity is computed as NC(A) = ln[(1/n) sum_j exp(mu_j)]. The physical and line-relation values are denoted lambda_phys and lambda_hyper, respectively. The comprehensive simulated propagation metric R_sys and minimum robustness R_ob provide the functional baselines. The initial metrics are shown in Table 4.

Table 4. Baseline metrics of the Nanjing rail transit hypernetwork perturbation experiment.

Perspective	Metric	Symbol	Baseline value
Structural	Natural connectivity of physical topology	lambda_phys(0)	0.875322
Structural	Natural connectivity of line-relation layer	lambda_hyper(0)	29.468427
Functional	Comprehensive system resilience	R_sys(0)	0.795356
Functional	Minimum system robustness	R_ob(0)	0.760285

4.2. Topological Robustness Analysis

Structural robustness experiments compare node-removal and line-relation-removal strategies. The evaluation metrics include physical-layer natural connectivity lambda_phys, line-relation-layer natural connectivity lambda_hyper, and the reported relative damage ratio D_RT. D_RT compares targeted-attack damage with the corresponding random-attack damage at the same deletion ratio; values above 1 indicate greater targeted damage.

4.2.1. Structural Degradation under Node Attacks

The node deletion ratio ranges from 5% to 30% in 5% increments. Three strategies are compared: uniform random attack, averaged over 50 pre-generated schemes; weighted degree centrality attack, which removes nodes in descending order of connection strength in A_hyper; and betweenness centrality attack, which removes nodes in descending order of unweighted shortest-path betweenness [22]. The results are shown on Fig. 3.

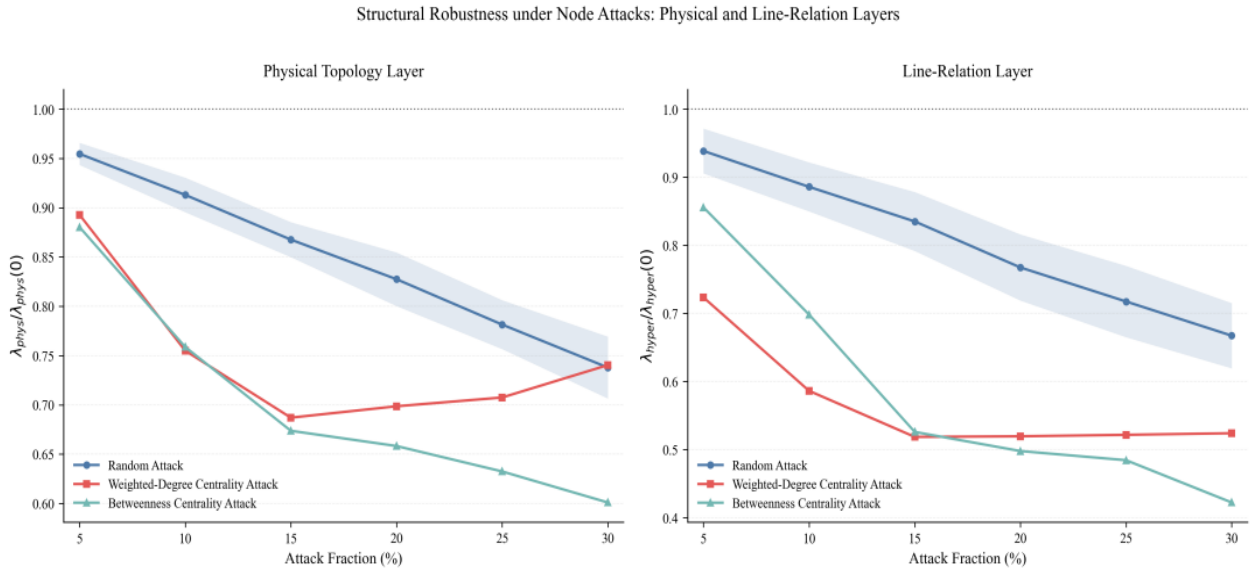


Fig. 3. Structural robustness under node attacks: physical topology and line-relation layers.

At a 30% node deletion ratio, random attack reduces lambda_phys to 73.79% of the baseline, weighted degree centrality attack reduces it to 74.04%, and betweenness centrality attack reduces it to 60.10%. In the line-relation layer, random attack reduces lambda_hyper to 66.72% of the baseline, whereas weighted degree centrality attack reduces it to 42.20%.

The relative damage ratio records the excess damage caused by targeted attacks. In the physical layer, the D_RT of weighted degree attack peaks at 1.263 at 15% and then declines monotonically to 0.997 at 30%. This reflects a hub-exhaustion effect: high-transfer hubs are removed first during the initial 15%, after which attacks largely remove ordinary chain-like stations of degree 1 or 2, and the marginal damage approaches that of random attack. Betweenness centrality attack continues to target cross-regional corridor stations, with D_RT still reaching 1.228 at 30%. In the hypergraph layer, the ratios for weighted degree and betweenness attacks remain high at 1.274 and 1.581, respectively.

Table 5. Complete evolution of the relative damage ratio under attack strategies.

Topology layer	Attack strategy	5%	10%	15%	20%	25%	30%
Physical topology	Weighted degree centrality	1.069	1.209	1.263	1.185	1.104	0.997
	Betweenness centrality	1.085	1.203	1.288	1.257	1.236	1.228
Line-relation layer	Weighted degree centrality	1.298	1.512	1.611	1.478	1.376	1.274
	Betweenness centrality	1.097	1.269	1.588	1.542	1.482	1.581

Curve-fitting results indicate that random attack is well approximated by a linear decay, whereas exponential decay provides the best fit among the tested candidate forms for the targeted strategies. The exponential-fit R^2 values vary across topology layers and attack strategies (0.9287-0.9909), so the strength of the fit is not uniform. For weighted degree centrality attack, the decline becomes less steep after the first 15% of deletions, consistent with a hub-exhaustion interpretation.

Table 6. Fitting tests for candidate decay models of structural connectivity under node attacks. Each model cell reports R^2/RMSE.

Topology layer	Attack strategy	Linear model	Exponential decay	Power-law model	Best-fit form
Physical topology	Random attack	0.9998/0.0012	0.9998/0.0012	0.9998/0.0010	Linear decay
	Weighted degree centrality	0.3947/0.0540	0.9287/0.0185	0.6446/0.0414	Exponential decay
	Betweenness centrality	0.8685/0.0340	0.9909/0.0089	0.9807/0.0130	Exponential decay
Line-relation layer	Random attack	0.9982/0.0040	0.9983/0.0039	0.9982/0.0040	Linear decay
	Weighted degree centrality	0.6068/0.0468	0.9882/0.0081	0.8289/0.0309	Exponential decay
	Betweenness centrality	0.8599/0.0559	0.9788/0.0218	0.9666/0.0273	Exponential decay

4.2.2. Structural Degradation under Line-Relation (Hyperedge) Removal

Columns are removed from H one at a time from 1 to 13 while A_phys is kept unchanged. Accordingly, this experiment should be interpreted as line-relation (hyperedge) removal rather than literal physical closure of a metro line. Three strategies are compared: random hyperedge removal; hyperedge-size attack, which removes lines in descending order of station count; and line-association attack, which prioritizes lines with greater transfer overlap with other lines.

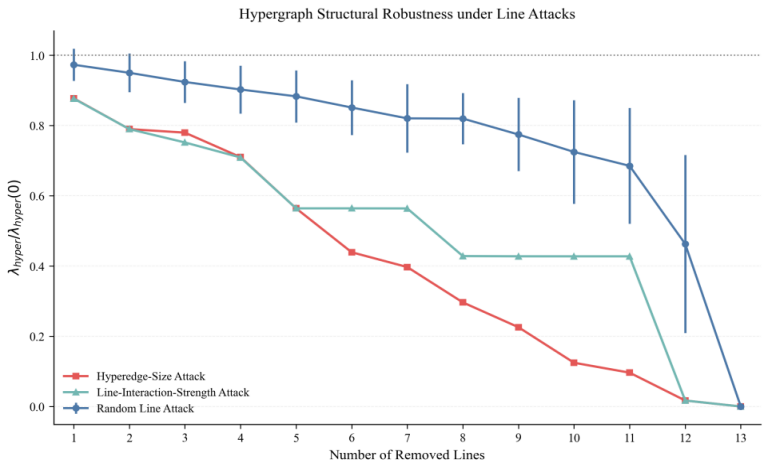


Fig. 4. Hypergraph structural robustness under line-relation (hyperedge) removals.

When six line relations are removed at random, lambda_hyper remains at 85.14% of the baseline, and a rapid collapse begins only after 11 removals. In contrast, the hyperedge-size-based targeted strategy causes more than 43% loss after removing five large backbone hyperedges and reduces the metric below 0.30 after eight removals. The comparison indicates that higher-order service redundancy is particularly sensitive to the removal of large line-level relations.

4.3. Resilience Response of Propagation Dynamics

The dynamic simulation uses the coupled physical-dispatch mechanism in Eq. (1). The parameters are kept consistent with data generation: delta = 0.3, beta_p = 0.15, beta_h = 0.25, gamma = 1.5, and dt = 0.1. The fourth-order Runge-Kutta method is used for T_sim = 50 steps.

Node failure follows induced-subnetwork deletion semantics: failed nodes are removed from the matrices, do not participate in equation solving, and are not counted in the denominator of the network-average congestion x_bar(t). Effective node performance is P(t) = 1 - x_bar(t). Comprehensive simulated propagation resilience is R_sys = (1/T_sim) integral_0^T_sim P(t) dt, and minimum robustness is R_ob = min_t P(t).

4.4. Dual-Perspective Comparison of Structure and Propagation

Under identical perturbations, structural damage and the simulated propagation response show a clear divergence. As node attacks increase from 5% to 30%, structural connectivity falls sharply, whereas R_sys and R_ob increase slightly. These increases are model-specific consequences of reduced propagation exposure after nodes and links are removed and should not be interpreted as improvements in real-world operational resilience.

Table 7. Comparison of structural and functional metrics under 30% node deletion.

Attack strategy	lambda_phys	Normalized lambda_phys	lambda_hyper	Normalized lambda_hyper	R_sys	Normalized R_sys	R_ob
Undisturbed baseline	0.875322	1	29.468427	1	0.795356	1	0.760285
Random attack	0.645919	0.7379	19.66073	0.6672	0.827726	1.0407	0.822449
Weighted degree centrality	0.648082	0.7404	15.436431	0.5238	0.835745	1.0508	0.826301
Betweenness centrality	0.526072	0.601	12.436752	0.422	0.850436	1.0693	0.826301

Table 8. Complete evolution of normalized functional metrics under node attacks.

Attack strategy	5%	10%	15%	20%	25%	30%
Random attack	1.0089	1.0141	1.0221	1.0303	1.0368	1.0407
Weighted degree centrality	1.0226	1.046	1.0572	1.0563	1.0549	1.0508
Betweenness centrality	1.0249	1.0461	1.0598	1.0619	1.0652	1.0693

At 30% betweenness centrality deletion, hypergraph connectivity decreases by 57.80%, while R_sys increases from 0.795356 to 0.850436, an increase of 6.93%. This model-specific de-connectivity damping effect occurs because removing critical bridge nodes fragments the simulated network, interrupts physical-diffusion and dispatch-cascade pathways, and lowers the average congestion exposure among the remaining nodes. The increase in R_sys therefore represents reduced simulated propagation exposure under network fragmentation rather than improved operational resilience.

Table 9. Normalized structural and functional metrics under successive random line-relation removals.

No. of removed line relations	**Normalized lambda_hyper**	**Normalized R_sys**
1	0.9732	1.0044
2	0.9502	1.0102
3	0.9241	1.0131
4	0.9027	1.0171
5	0.8832	1.023
6	0.8514	1.0278
7	0.8205	1.0331
8	0.8199	1.032
9	0.7745	1.0376
10	0.7248	1.0435
11	0.6853	1.0482
12	0.4629	1.0513
13	0	1.056

Table 9 and the right panel of Fig. 5 report the successive random line-relation-removal sequence. As the number of removed hyperedges increases from 1 to 13, normalized hypergraph connectivity declines to zero while normalized R_sys gradually rises to 1.0560. Because A_phys is held unchanged in this experiment, the increase reflects reduced exposure to the higher-order dispatch-cascade channel represented by H, not the operational benefit of physically closing metro lines. The experiment therefore measures sensitivity to line-level higher-order relations rather than full service-outage resilience.

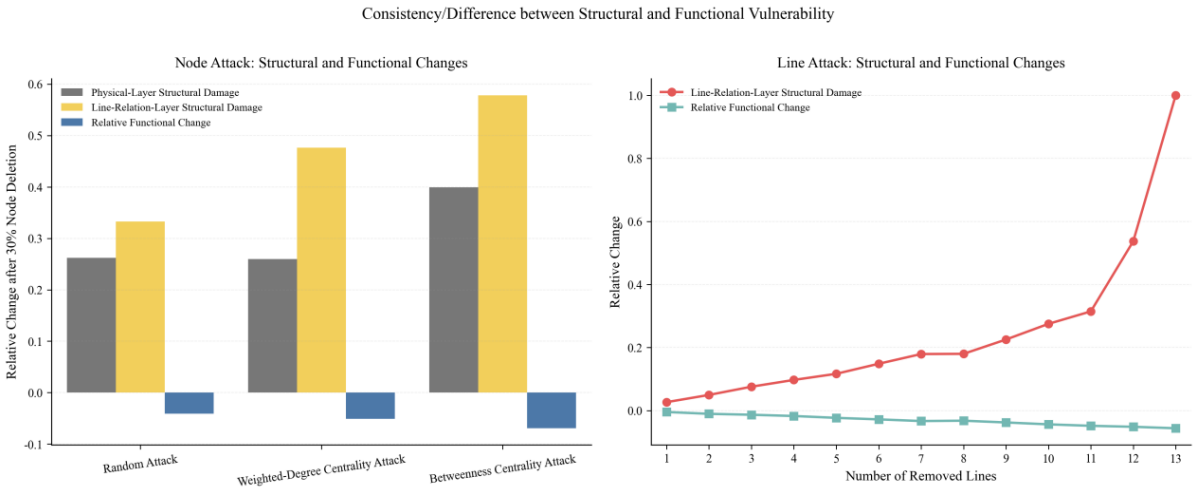


Fig. 5. Comparison of structural vulnerability and simulated propagation response under node removal and random line-relation removal.

5. Discussion

5.1. Main Findings and Theoretical Contributions

This study addresses data-driven resilience inference, higher-order topology representation, and the perturbation response of rail transit networks. Based on model comparison, component ablation, and controlled perturbation experiments, the main theoretical findings are as follows.

First, the results support inference of the final resilience state from limited early trajectories and topology without requiring explicit governing equations at inference time. On the 1:4 test set, ResInf-

HGNN achieves an F1-score of 0.9362, AUC-ROC of 0.9965, and precision of 0.8800. Its threshold-based F1-score is comparable to the GCN and static HGNN baselines, while its AUC is the highest, indicating stronger probability ranking under severe class imbalance. The ablation results reinforce this interpretation: removing the Transformer temporal encoder lowers the F1-score to 0.9167, while removing attention readout lowers it to 0.8718. These findings show that jointly representing early temporal information and network topology improves discrimination of later system evolution, while adaptive node-level aggregation contributes to balanced performance [1].

Second, the value of the hypergraph representation lies in preserving a complete metro line as a single higher-order relation and supporting node-line-node information aggregation. This makes line-level group relations and hyperedge-removal experiments more direct than representing the same intervention through many pairwise operations. However, the current GCN baseline differs from the full model in both temporal input and topology operator, so the available evidence supports the combined value of temporal encoding, higher-order topology, and global aggregation but does not establish the isolated causal contribution of the hypergraph operator. Strict structure-matched comparison remains necessary [8], [14], [16], [17], [20].

Third, controlled perturbation experiments show that the observed resilience pattern depends on topology, disturbance semantics, and evaluation perspective. Within the 5%-30% node-attack range, structural connectivity declines substantially, whereas R_sys and R_ob increase slightly because node deletion reduces simulated propagation exposure among the remaining nodes. For line-relation removal, normalized hypergraph connectivity falls to zero after 13 hyperedges are removed, while normalized R_sys rises to 1.0560 because the higher-order dispatch channel is progressively removed from H while A_phys remains unchanged. These increases are model-specific and should not be interpreted as improvements in actual metro operating resilience. Connectivity, structural robustness, and simulated propagation response are therefore related but non-interchangeable dimensions.

5.2. Practical Implications

Within the model setting and experimental scope of this study, the results provide the following implications for urban rail transit planning and operation.

1. Implement differentiated operation, maintenance, and emergency support. Targeted node attacks cause greater structural damage, whereas random node failures yield the lowest minimum robustness R_ob among the three node-attack modes. Structural protection and functional risk management therefore address different failure modes. Resource allocation should not rely on a single centrality measure or robustness indicator and can complement existing emergency medical supply-chain resilience assessment [7].

2. Incorporate accessibility and propagation risk into multi-objective planning. Adding lines can improve accessibility and transport capacity while also creating additional higher-order coupling pathways in the model. Planning should therefore evaluate new lines in terms of passenger capacity, alternative routes, propagation exposure, and recovery ability. This multi-objective perspective is consistent with broader resilient-city planning principles [4]. The increase in R_sys observed after hyperedge removal is specific to the proposed coupled physical-dispatch model with A_phys held unchanged; it must not be interpreted as evidence that real line closures improve overall resilience.

3. Use the model-derived k-space score as an auxiliary warning signal. The model maps high-dimensional early trajectories to a one-dimensional decision variable that indicates a sample's position relative to the learned classification boundary. Before deployment, k should be calibrated using real passenger-flow and failure records, tested across time periods, lines, and networks, and used together with observable indicators such as passenger intensity, train load factor, and delay time. It should not be treated as a validated independent physical quantity.

5.3. Limitations

Although the experimental results are relatively consistent, several limitations remain.

First, the training and validation data are generated by the proposed coupled physical-dispatch dynamics model on a single Nanjing Metro topology. The sample size, parameter range, and fixed label threshold theta_y are manually specified, and external validation using real failures, passenger congestion, and recovery records has not been performed. The physical topology also contains ten connected components because 17 station names could not be matched between the map data and the operating station list. The generalization ability of the model across cities, network scales, and dynamic distributions therefore remains unclear.

Second, the current hypergraph uses static, undirected, and relatively homogeneous line relations. It does not explicitly represent time-varying factors such as peak and off-peak passenger differences, train direction, transfer capacity, departure interval, or temporary operational adjustment. The proposed coupled physical-dispatch dynamics model also simplifies real congestion formation, evacuation, and redistribution.

Third, the perturbation experiments mainly consider permanent node failures and line-relation (hyperedge) removals. They do not cover signal-system failures, segment speed limits, sudden passenger surges, coupled multiple failures, or a full physical line-closure model in which A_phys is modified together with H. In addition, the GCN baseline and the full model differ in both temporal input and topology operator, so the current comparison cannot independently identify the net contribution of hypergraph structure. The experiments also do not systematically report confidence intervals across multiple random seeds, probability calibration, cross-topology transfer performance, or the complete set of training hyperparameters.

5.4. Future Research Directions

To address the above limitations, future research can proceed in three directions.

1. Strengthen real-data and cross-scenario validation. Automatic fare collection data, train operation logs, equipment alarms, and historical emergency events can be integrated with transportation digital twins or data assimilation to build a simulation environment closer to real operations. External tests across cities, time periods, and disturbance types can evaluate out-of-distribution generalization and warning probability calibration.

2. Build dynamic weighted hypergraphs. Lines, transfer relations, and time-specific passenger flows can be represented as time-varying hyperedges and weights. Directed propagation, capacity constraints, and multilayer network structures can be introduced to characterize higher-order interactions under peak/off-peak, up/down direction, and temporary dispatch conditions, enabling dynamic resilience inference for time-varying systems [14].

3. Advance causal validation and resilience optimization. With the temporal encoder, parameter scale, training protocol, and input information held constant, ordinary graph and hypergraph operators should be compared in a strict framework. Multiple random seeds, statistical significance tests, probability calibration, and cross-topology transfer experiments should verify robustness. Future perturbation studies should also distinguish hyperedge removal from physical line closure by modifying H and A_phys according to the intended operational scenario. The model can then be combined with optimization algorithms to compare interventions such as reinforcing critical lines, improving transfer capacity, and implementing flow control and diversion.

5.5. Conclusions

To reduce the dependence of resilience inference on explicitly specified governing equations at inference time and to represent line-level group relations beyond pairwise graphs, this study proposes ResInf-HGNN, a framework integrating a Transformer temporal encoder, hypergraph convolution, and attention readout. Model validation and controlled perturbation analysis are conducted on the Nanjing urban rail transit network. The main conclusions are as follows.

1. Under the simulation data and experimental settings of this study, ResInf-HGNN classifies resilience from limited early trajectories and topology, achieving an F1-score of 0.9362, AUC-ROC of 0.9965, and precision of 0.8800 on the 1:4 test set. Its F1-score is comparable to the GCN and static HGNN baselines, while its AUC is the highest, indicating stronger global probability ranking under class imbalance. Component ablation shows material contributions from temporal encoding and attention readout: removing the Transformer reduces the F1-score to 0.9167, and removing attention readout reduces it to 0.8718. The one-dimensional k-space projection provides a model-derived and interpretable interface for observing sample positions relative to the learned decision boundary.

2. Baseline comparison and component ablation show that temporal dynamics and attention readout both substantially affect model performance. The hypergraph representation preserves each metro line as an independent higher-order relation and directly supports line-level group aggregation and hyperedge-removal analysis. However, because the GCN baseline differs from the full model in both temporal input and topology operator, the current results do not isolate the causal contribution of the hypergraph operator alone. Strict structure-matched ablation remains necessary.

3. Under the given Nanjing Metro topology, coupled-dynamics parameters, and evaluation metrics, structural connectivity and the simulated propagation metrics respond differently to the same perturbations. Targeted node attacks generally cause greater topological damage, while random node failures yield the lowest minimum robustness R_ob among the three node-attack modes. At 30% betweenness deletion, hypergraph connectivity decreases by 57.80%, while R_sys rises from 0.795356 to 0.850436. Progressive random hyperedge removal drives normalized hypergraph connectivity to zero while increasing normalized R_sys to 1.0560. Because A_phys is kept unchanged during hyperedge removal, the latter increase reflects a model-specific reduction in higher-order dispatch-cascade exposure rather than improved real-world operational resilience.

Overall, this study provides an implementable framework for higher-order data-driven resilience inference in complex networks and demonstrates its potential for urban rail transit analysis. Further validation with real operational data, dynamic weighted hypergraphs, structure-matched baselines, and cross-network experiments are needed to establish robustness and decision value.

References

[1] Liu C, Xu F, Gao C, et al. Deep learning resilience inference for complex networked systems [J]. Nature Communications, 2024, 15: 9203.

[2] Gao J, Barzel B, Barabasi A L. Universal resilience patterns in complex networks [J]. Nature, 2016, 530(7590): 307-312.

[3] Laurence E, Doyon N, Dube L J, et al. Spectral dimension reduction of complex dynamical networks [J]. Physical Review X, 2019, 9(1): 011042.

[4] Qiu Baoxing. Methods and principles of resilient city design based on complex adaptive systems theory [J]. Urban Development Studies, 2018, 25(10): 1-3.

[5] Wang Y, Zhan J, Xu X, et al. Measuring the resilience of an airport network [J]. Chinese Journal of Aeronautics, 2019, 32(12): 2694-2705.

[6] Hu Z, Xu Z, Sun X, et al. Meteorological-electrical integrated real-time resilience assessment for power systems based on deep learning methods [J]. IEEE Transactions on Smart Grid, 2025, 16(5): 3700-3713.

[7] Sun Ying, Liu Huiping, Yan Rui, et al. Equilibrium optimization of the supply chain of emergency medical supplies based on resilience and social welfare [J]. Chinese Journal of Management Science, 2023, 31(8): 132-141.

[8] Battiston F, Cencetti G, Iacopini I, et al. Networks beyond pairwise interactions: Structure and dynamics [J]. Physics Reports, 2020, 874: 1-92.

[9] Li X, Zhang T, Wang J, et al. Protect NTN-IoT security by malicious traffic detection: a multidimensional hypergraph learning approach [J]. IEEE Internet of Things Journal, 2026, 13(3): 3594-3607.

[10] Bai L, Li Z, Tang C, et al. Hypergraph-based analysis of weighted gene co-expression hypernetwork [J]. Frontiers in Genetics, 2025, 16: 1560841.

[11] Xie Y, Li S, Wu C T, et al. A novel hypergraph convolution network for wafer defect patterns identification based on an unbalanced dataset [J]. Journal of Intelligent Manufacturing, 2024, 35(2): 633-646.

[12] Liu C, Cosmo L, Rossi L. Diagnosing Alzheimer's disease using hypergraph neural networks with prompt tuning [J]. Pattern Recognition, 2026, 176: 113290.

[13] Fu W, Wang H, Gao C, et al. Mobility data-driven privacy-preserving model for detecting high-risk infection cases [J]. ACM Transactions on Intelligent Systems and Technology, 2025, 16(5): 110.

[14] Jiang J, Wei Y, Feng Y, et al. Dynamic hypergraph neural networks [C] //Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, 2019: 2635-2641.

[15] Su H, Wang X, Qin Y, et al. Attention-based adaptive spatial-temporal hypergraph convolutional networks for stock price trend prediction [J]. Expert Systems with Applications, 2024, 238: 121899.

[16] Chen L, Lu J, Wang Y, et al. Cascading failures with group support in interdependent hypergraphs [J]. Chaos, Solitons & Fractals, 2025, 192: 116023.

[17] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [C]//International Conference on Learning Representations. 2017.

[18] Pastor-Satorras R, Castellano C, Van Mieghem P, et al. Epidemic processes in complex networks [J]. Reviews of Modern Physics, 2015, 87(3): 925-979.

[19] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Advances in Neural Information Processing Systems. 2017, 30: 5998-6008.

[20] Feng Y, You H, Zhang Z, Ji R, Gao Y. Hypergraph neural networks [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(1): 3558-3565.

[21] Wu J, Barahona M, Tan Y J, Deng H Z. Natural connectivity of complex networks [J]. Chinese Physics Letters, 2010, 27(7): 078902.

[22] Brandes U. A faster algorithm for betweenness centrality [J]. Journal of Mathematical Sociology, 2001, 25(2): 163-177.