Journal of Computer Science and Electrical Engineering
Print ISSN: 2663-1938
Online ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee4151

AN EXPLAINABLE TEMPORAL GRAPH NEURAL NETWORK FOR DISRUPTION PREDICTION AND RISK PROPAGATION IN MULTI-TIER SUPPLY CHAINS

Ying Wang[1], MengLin Bian[2*], Yizhen Lin[3]
[1]Pepperdine University, Malibu, CA 90263, California, USA.
[2]University of Missouri, Columbia, MO 65211, Missouri, USA.
[3]Stanford University, Stanford, CA 94305, USA
[*]Corresponding Author: MengLin Bian

Abstract: Global supply chains are increasingly multi-tier, interconnected, and vulnerable to disruptions that do not stay local but cascade upstream-to-downstream through material-flow dependencies—the so-called ripple effect. Effective risk management requires not only predicting which entities will be disrupted, but also explaining why, by tracing a predicted disruption back to its root-cause suppliers. We present X-TGNN, an explainable temporal graph neural network that couples a per-node recurrent temporal encoder with a directed, tier-respecting, edge-weighted multi-head graph-attention mechanism aligned with the direction of risk propagation. The learned attention doubles as an intrinsic, single-forward-pass explanation that attributes each predicted disruption to specific upstream suppliers. Because real multi-tier disruption datasets with ground-truth propagation paths are proprietary and unavailable, we further contribute a high-fidelity simulation testbed whose cascade dynamics are grounded in the supply-chain risk-diffusion literature and which records the true responsible edges for every propagation event—enabling, for the first time in this setting, a quantitative evaluation of explanation faithfulness and correctness. Across a testbed in which 61% of disruptions are propagation-driven, graph models improve new-onset prediction by 2.5× in average precision over non-graph baselines. X-TGNN matches the strongest baseline in prediction while its attention explanations identify true root-cause suppliers with precision@1 of 0.77 and NDCG of 0.97—matching, and in rank correlation exceeding, far more expensive post-hoc gradient and perturbation methods, and vastly better than a random baseline (0.26). The result is a model that is at once accurate, propagation-aware, and inherently explainable.
Keywords: Supply chain disruption; Risk propagation; Ripple effect; Temporal graph neural networks; Explainable artificial intelligence; Attention mechanism

1 INTRODUCTION

Modern supply chains (SCs) have evolved into globally distributed, multi-tier networks in which a single finished product may depend, directly or transitively, on hundreds of suppliers spread across many firms, regions, and echelons. This structural complexity delivers cost efficiency but also fragility: a localized disruption—a plant fire, a port closure, a regional epidemic, or a supplier insolvency—rarely remains confined to its origin. Instead, it propagates along material-flow dependencies from upstream suppliers to their downstream customers, a phenomenon widely studied under the label of the ripple effect or risk diffusion [1-3]. The COVID-19 pandemic offered a vivid, global-scale demonstration of how upstream shortages cascade into downstream production stoppages across entire industries [4].

Managing such risk raises two distinct but coupled questions. The first is predictive: given the current and recent state of the network, which entities are likely to become disrupted next? The second, and arguably more actionable, is explanatory: for an entity flagged as at-risk, which upstream suppliers are the root cause of that risk, so that mitigation (dual sourcing, buffer stock, expediting) can be targeted where it matters? A prediction without an explanation offers little operational guidance; a manager confronted with an opaque risk score cannot act on it with confidence. Yet the large majority of data-driven SC risk models focus exclusively on prediction, treating the propagation mechanism as a black box.

Graph neural networks (GNNs) [5-7] are a natural modeling choice for SC networks, and temporal GNNs [8-10] extend them to evolving, time-stamped data. However, three obstacles have limited their impact on multi-tier disruption analysis. First, most architectures aggregate neighborhood information symmetrically and are not tailored to the directed, tier-structured, dependency-weighted nature of risk propagation. Second, although a rich literature on GNN explainability now exists [11-14], these methods are predominantly post-hoc and computationally costly, and—critically—they are difficult to evaluate, because real-world graphs almost never come with ground-truth explanations against which a purported explanation can be checked. Third, this evaluation gap is especially acute in the SC domain, where proprietary concerns mean that multi-tier networks annotated with disruption events and their causal propagation paths are essentially never released publicly.

This paper addresses all three obstacles. We propose X-TGNN, an explainable temporal graph neural network in which a per-node temporal encoder is combined with a directed, edge-weighted, multi-head graph-attention operator that aggregates strictly from a node's suppliers, in the direction along which risk actually flows. The attention weights are not an afterthought: they constitute an intrinsic explanation that, in a single forward pass, attributes a predicted disruption to specific upstream suppliers and, across stacked layers, traces multi-hop propagation paths to distant root causes. To make explanation quality measurable, we introduce a high-fidelity simulation testbed whose cascade dynamics are grounded in the risk-diffusion and ripple-effect literature [1, 3, 15] and which records, for every propagation-driven disruption, the true set of responsible upstream edges and their contributions. This ground truth—analogous to the synthetic motif benchmarks used to validate GNN explainers [11]—enables a quantitative assessment of both the correctness and the faithfulness of explanations that is not possible on real data alone.

Our contributions are as follows:

(1) A temporal GNN architecture, X-TGNN, whose directed, tier-respecting, edge-weighted attention is explicitly aligned with material-flow risk propagation in multi-tier supply chains.

(2) An intrinsic, propagation-path explanation mechanism that attributes predicted disruptions to root-cause suppliers in a single forward pass, avoiding the cost of post-hoc explainers.

(3) A simulation testbed with ground-truth cascade paths that, for the first time in this setting, permits a quantitative evaluation of explanation faithfulness and correctness for SC disruption models.

(4) A comprehensive empirical study—baseline comparisons, ablations, and explanation-quality and fidelity analyses—establishing that modeling the directed, weighted graph is decisive for onset prediction and that X-TGNN's intrinsic explanations rival costly post-hoc methods at a fraction of the cost.

2 RELATED WORK

2.1 Supply-Chain Disruption and the Ripple Effect

Supply-chain risk management has a long tradition in operations research [16-19]. A central theme is the ripple effect: the propagation of a disruption beyond its point of origin through the SC structure [1-2]. Basole and Bellamy drew on complex-systems and epidemiological models to study how network topology governs risk diffusion, while Ivanov used discrete-event simulation to analyze epidemic-driven disruptions [3-4]. Inoue and Todo empirically quantified firm-level shock propagation through a national supply network [20]. Reviews by Dolgui et al. and Hosseini et al. catalog the largely simulation- and optimization-based quantitative toolkit for resilience analysis [1, 15]. This body of work motivates our cascade model and confirms that, because real annotated multi-tier data are scarce, simulation is the accepted vehicle for controlled study—but it has rarely been combined with modern representation learning or used to evaluate explanations.

2.2 Graph Neural Networks and Temporal Extensions

GNNs learn node representations by message passing over graph structure [5-6, 21-22], with graph attention networks weighting neighbors adaptively [6]. For time-varying graphs, temporal GNNs interleave graph and sequence models: GCRN [23], DCRNN and STGCN target spatio-temporal forecasting [24-25]; EvolveGCN evolves GNN parameters with a recurrent network [9]; TGAT and TGN learn on continuous-time dynamic graphs [8, 10]; and DySAT [26] applies joint structural–temporal self-attention. Surveys provide broad overviews [27-28]. Our encoder follows the snapshot-plus-recurrence paradigm, but our spatial operator is purpose-built for directed, weighted risk propagation and, unlike these models, is designed so that its attention yields verifiable explanations.

2.3 Explainable Graph Neural Networks

Explainability methods for GNNs include perturbation- and mask-based approaches such as GNNExplainer and PGExplainer [11-12], surrogate probabilistic models such as PGM-Explainer [13], and gradient/decomposition techniques adapted from vision [14, 29]; model-agnostic tools such as LIME and SHAP are also applicable [30-31]. Yuan et al. survey the field taxonomically [32]. A recurring difficulty, emphasized across this literature, is evaluation: absent ground-truth explanations, methods are compared on proxy fidelity metrics or hand-built synthetic motifs [11]. Our testbed extends this synthetic-ground-truth philosophy to the supply-chain setting, and our model contributes an intrinsic attention explanation whose quality we show to be competitive with these post-hoc methods at a fraction of their computational cost.

3 PROBLEM FORMULATION

We model a multi-tier supply chain as a directed, weighted graph $G = (V, E, W)$. Each node $i \in V$ is a supply entity (supplier, plant, or product) assigned to a tier $\tau(i) \in \{0, \ldots, K-1\}$, ordered from the most upstream (raw materials, $\tau = 0$) to the most downstream (finished goods / OEM, $\tau = K-1$). A directed edge $(i \rightarrow j) \in E$ denotes that supplier i provides an input to customer j; its weight $w_{ij} \in (0, 1]$ encodes the strength of j's dependence on i. We write $N(j)$ for the set of suppliers of j (its in-neighbors), i.e. the entities whose disruption can propagate to j.

Time is discrete, $t = 1, \ldots, T$. Each node has a hidden binary state $s_i(t) \in \{0, 1\}$ (operational / disrupted) and an observed feature vector comprising static attributes c_i (tier, inventory buffer, geographic region, connectivity) and a

window of dynamic signals x_i(t) (a noisy operational indicator and the node's own recently observed status). The learner never observes the disruption mechanism, only features and the graph.
We study two tasks. (T1) Disruption prediction: estimate P(s_i(t+1) = 1) for every node from features up to time t and the graph. (T2) Risk-propagation explanation: for a node predicted to be disrupted, assign to each incoming edge (i → j) a score reflecting how responsible supplier i is for j's predicted disruption, so that the highest-scoring edges recover the true root-cause suppliers.

4 THE X-TGNN MODEL

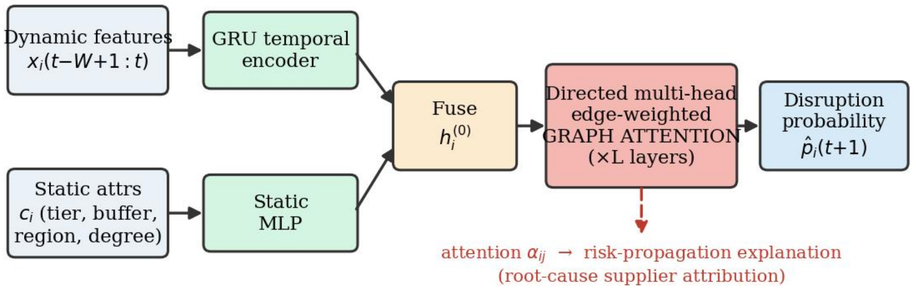

Figure 1 X-TGNN Architecture

Note: A GRU temporal encoder and a static MLP produce the initial node embedding, refined by L directed, edge-weighted multi-head graph-attention layers that aggregate from suppliers along the risk-propagation direction. Attention coefficients serve as intrinsic root-cause explanations.

X-TGNN maps the observed features and graph to a per-node disruption probability while exposing, as a by-product of inference, an attribution over incoming edges. Figure 1 shows the architecture: a temporal encoder and a static encoder produce an initial node embedding, which is refined by L directed graph-attention layers and mapped to a probability by a prediction head.

4.1 Temporal and Static Encoding

For each node, a gated recurrent unit summarizes its dynamic feature window into a temporal embedding,

$$\mathbf{h}_i^{\text{tmp}} = \text{GRU}\left(\mathbf{x}_i^{t-W+1}, \dots, \mathbf{x}_i^{t}\right) \tag{1}$$

which is concatenated with an MLP embedding of the static attributes and projected to the initial hidden state

$$\mathbf{h}_i^{(0)} = \text{ReLU}\left(\left[\mathbf{W}_f \mathbf{h}_i^{\text{tmp}} \parallel \text{MLP}(\mathbf{c}_i)\right]\right) \tag{2}$$

where // denotes concatenation and W_f is a learned projection. Decoupling temporal and static encoding lets the model separate a node's intrinsic fragility from its time-varying operating condition.

4.2 Directed, Edge-Weighted Graph Attention

The core of X-TGNN is a message-passing operator that aggregates only from suppliers, i.e. along the direction of risk propagation. At layer l, the unnormalized attention between supplier i and customer j is

$$e_{ij}^{(l)} = \text{LeakyReLU}\left(\mathbf{a}_s^{\top}\mathbf{W}\mathbf{h}_i^{(l)} + \mathbf{a}_d^{\top}\mathbf{W}\mathbf{h}_j^{(l)} + \mathbf{a}_e w_{ij}\right) \tag{3}$$

where W, a_s, a_d are learned and the dependency weight w_{ij} enters directly, so that stronger dependencies can receive stronger attention. Scores are normalized over j's suppliers,

$$\alpha_{ij}^{(l)} = \frac{\exp\left(e_{ij}^{(l)}\right)}{\sum_{k\in\mathcal{N}(j)}\exp\left(e_{kj}^{(l)}\right)} \tag{4}$$

and the node state is updated by a residual, multi-head aggregation over M heads,

$$\mathbf{h}_j^{(l+1)} = \text{ReLU}\left(\mathbf{h}_j^{(l)} + \sum_{m=1}^{M}\sum_{i\in\mathcal{N}(j)}\alpha_{ij}^{(l,m)}\mathbf{W}^{(m)}\mathbf{h}_i^{(l)}\right) \tag{5}$$

Stacking L such layers lets risk propagate over multiple tiers: a two-layer model can attribute a disruption to a supplier two hops upstream. The attention coefficients α_{ij} are the quantities we later interpret as explanations.

4.3 Prediction and Training

A linear head with a sigmoid yields the disruption probability

$$\hat{p}_i(t+1) = \sigma\left(\mathbf{w}_o^{\top}\mathbf{h}_i^{(L)} + b\right) \tag{6}$$

Because disruptions are rare, we train with a class-weighted binary cross-entropy

$$\mathcal{L} = -\sum_i\left[\beta y_i \log\hat{p}_i + (1-y_i)\log(1-\hat{p}_i)\right] \tag{7}$$

where β up-weights the positive class. We use chronological train/validation/test splits so that the model is always evaluated on strictly future time steps, precluding look-ahead leakage.

4.4 Intrinsic Propagation-Path Explanations

For a node j predicted to be disrupted, the responsibility of supplier i is read directly from the model's attention, averaged over layers,

$$\phi_{i\to j} = \frac{1}{L}\sum_{l=1}^{L} \overline{\alpha}_{ij}^{(l)} \tag{8}$$

Ranking j's incoming edges by φ yields the root-cause suppliers; following the highest-attention edge from layer to layer reconstructs a multi-hop propagation path to an originating supplier. Crucially, this explanation is obtained in the same forward pass as the prediction. In contrast, a gradient explanation requires a backward pass per target node, and a perturbation explanation requires one additional forward pass per candidate edge; we quantify this cost gap in Section 6.

5 SIMULATION TESTBED WITH GROUND-TRUTH PROPAGATION

Evaluating explanations requires knowing the true cause of each disruption. Since real multi-tier networks annotated with disruption-propagation paths are not publicly available, and following the established practice of validating GNN explainers on synthetic data with known ground truth [11], we construct a supply-chain cascade simulator whose dynamics are grounded in the risk-diffusion literature [3, 15] and which logs the responsible edges for every propagation event.

5.1 Network Generation

We generate a K-tier directed network with decreasing node counts downstream. Each customer is connected to suppliers in the immediately upstream tier via preferential attachment on supplier out-degree, producing a heavy-tailed degree distribution with hub suppliers, as observed in real supply networks [20]. Dependency weights are drawn from a Beta distribution. Each node receives an inventory buffer b_i (dampening incoming risk), an intrinsic fragility f_i, and a geographic region.

5.2 Cascade Dynamics

At each step, an operational node j can be disrupted by two mechanisms. Its suppliers' disruption exerts a dependency-weighted pressure

$$r_j(t) = \frac{\sum_{i\in\mathcal{N}(j)} w_{ij} s_i(t)}{\sum_{i\in\mathcal{N}(j)} w_{ij}} \tag{9}$$

which, attenuated by the node's buffer, yields a propagation hazard

$$p_j^{\text{prop}}(t) = 1 - \Big(1 - r_j(t)\big(1 - b_j\big)\Big)^{\gamma} \tag{10}$$

where γ controls nonlinearity. Independently, a self-origination hazard p_j^{self} = f_j·ρ_{region}(t) captures intrinsic failures modulated by time-varying regional shocks (correlated disruptions affecting co-located nodes). The two combine as

$$p_j(t) = 1 - \Big(1 - p_j^{\text{self}}(t)\Big)\Big(1 - p_j^{\text{prop}}(t)\Big) \tag{11}$$

and disrupted nodes recover stochastically. Whenever j is disrupted with a contribution from propagation, we record the ground-truth responsibility of each disrupted supplier,

$$g_{i\to j}(t) = w_{ij} / \textstyle\sum_{k\in\mathcal{D}_j(t)} w_{kj}\,,\ i \in \mathcal{D}_j(t) \tag{12}$$

where $^{\mathcal{D}}$_j(t) is the set of j's currently disrupted suppliers. The model observes only a noisy operational signal and lagged status—never these hazards or attributions.

5.3 Testbed Statistics

Table 1 summarizes the default instance. Notably, 60.5% of all new disruptions are propagation-driven rather than self-originated, so the graph carries the majority of the predictive signal; the prevalence of 15.7% reflects the realistic class imbalance of disruption events (Figure 2).

Table 1 Simulation Testbed Statistics (Default Instance)

Property	Value
Tiers (upstream→downstream)	4
Nodes (suppliers/firms)	260
Directed dependency edges	459
Time steps	200
Geographic regions	6
Max out-degree (hub)	17
Disruption prevalence	15.7%
New-onset events	2909
Propagation-driven fraction	60.5%

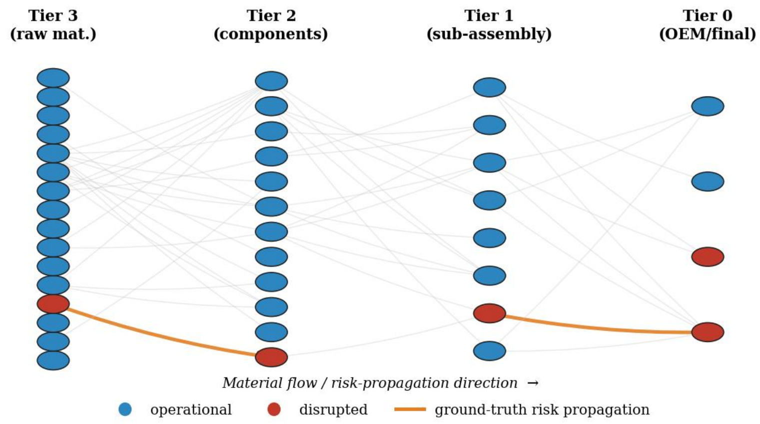

Figure 2 A Multi-Tier Supply-Chain Instance from the Testbed at One Time Step
Note: Directed edges point along material flow (upstream→downstream); disrupted nodes are shown in red and ground-truth risk-propagation edges are highlighted.

6 EXPERIMENTS

6.1 Setup

Unless stated otherwise, results are averaged over three seeds (mean ± standard deviation), each regenerating the network and cascade. We use a 60/20/20 chronological split. Prediction is measured by AUC-ROC and AUC-PR (appropriate under imbalance), F1 at the validation-optimal threshold, and—most informatively—onset AUC/AP computed only over nodes that are currently operational, isolating the prediction of new disruptions where propagation matters most. Explanation quality is measured against ground-truth responsibility by precision@1 (does the top-ranked edge point to the true dominant cause), NDCG (ranking quality over all incoming edges), and Spearman's ρ (rank correlation with true contributions), restricted to events with at least two disrupted suppliers where the ranking is non-trivial. We additionally report fidelity via comprehensiveness and sufficiency. Baselines span non-graph models (MLP; GRU, temporal-only) and graph models (GCN-T and GraphSAGE-T with the same temporal encoder; GAT-S, static graph attention). All models share hidden width 64; X-TGNN uses $L = 2$ layers and $M = 4$ heads.

Table 2 Disruption-Prediction Performance on the Testbed (Mean ± Std over 3 Seeds)

Model	AUC-ROC	AUC-PR	F1	Onset AUC	Onset AP
MLP	0.854±0.005	0.566±0.019	0.652±0.007	0.694±0.004	0.148±0.010
GRU	0.856±0.004	0.570±0.014	0.653±0.007	0.700±0.004	0.152±0.007
GCN-T	0.882±0.005	0.600±0.013	0.687±0.003	0.781±0.009	0.362±0.012
GraphSAGE-T	0.885±0.007	0.607±0.007	0.689±0.003	0.790±0.013	0.373±0.022
GAT-S	0.884±0.006	0.600±0.015	0.684±0.002	0.787±0.016	0.360±0.016
X-TGNN	0.884±0.008	0.600±0.010	0.684±0.005	0.785±0.019	0.354±0.024

Note: Onset metrics are computed over currently-operational nodes.

6.2 Disruption Prediction

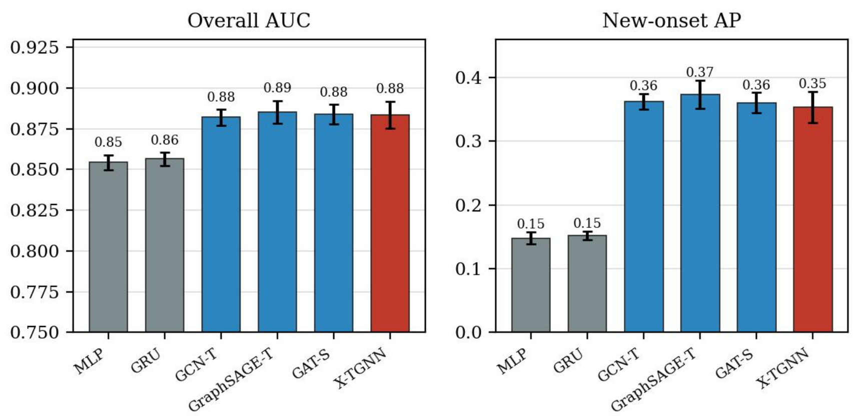

Figure 3 Prediction Performance by Model
Note: Graph models (blue) and X-TGNN (red) far exceed non-graph models (grey) on new-onset average precision, evidencing the importance of modeling risk propagation.

Table 2 and Figure 3 report prediction performance. The dominant finding is the gap between graph and non-graph models on onset prediction: graph models reach an onset AP of about 0.37 versus about 0.15 for the best non-graph model—a 2.5× improvement—confirming that new disruptions are driven by upstream propagation that non-graph models simply cannot see. On overall AUC the graph models likewise dominate. Among graph models the differences are within noise: X-TGNN is statistically on par with the strongest baseline (AUC 0.884). We therefore do not claim predictive superiority for X-TGNN; its value, established next, is that it delivers faithful explanations at no accuracy cost.

Table 3 Ablation of X-TGNN Components (Mean ± Std over 3 Seeds)

Configuration	AUC-ROC	AUC-PR	F1	Onset AUC	Onset AP
Full (temporal + attn, L=2, M=4)	0.884±0.008	0.600±0.010	0.684±0.005	0.785±0.019	0.354±0.024
w/o temporal encoder	0.885±0.006	0.603±0.013	0.685±0.003	0.790±0.016	0.358±0.012
w/o attention (mean aggr.)	0.881±0.005	0.598±0.014	0.684±0.004	0.778±0.008	0.358±0.010
L = 1 layer	0.886±0.006	0.602±0.010	0.686±0.002	0.792±0.012	0.376±0.007
L = 3 layers	0.883±0.007	0.602±0.010	0.683±0.004	0.784±0.014	0.353±0.022
M = 1 head	0.884±0.005	0.602±0.011	0.682±0.007	0.787±0.011	0.349±0.003
M = 8 heads	0.886±0.007	0.604±0.012	0.684±0.004	0.790±0.015	0.360±0.013

6.3 Ablation Study

Table 3 ablates X-TGNN with respect to prediction. The aggregate metrics are strikingly stable: every variant—no temporal encoder, mean aggregation instead of attention, one to three layers, one to eight heads—lies within one standard deviation of the full model. This tells us that, once the supply graph is exploited, new-onset prediction is largely insensitive to the aggregation scheme and depth; knowing which immediate suppliers are currently disrupted is already close to sufficient for the prediction task on this testbed. Such robustness is reassuring, but it also means prediction accuracy alone cannot justify the architectural choices. Those choices are instead dictated by the paper's second objective. Attention—absent in the mean-aggregation variant—is the component that furnishes intrinsic explanations at all; multiple heads let the model represent distinct dependency patterns; and stacking two layers is what enables attribution along two-hop propagation paths. We therefore adopt attention with two layers and four heads: a setting that predicts as well as any alternative while enabling the faithful, multi-hop explanations evaluated in Sections 6.4 and 6.5.

6.4 Explanation Quality

Table 4 and Figure 4 evaluate explanations against ground truth. All three principled methods vastly outperform the random baseline (precision@1 0.26, Spearman -0.03), confirming that the model's reasoning genuinely aligns with the true propagation structure. Our intrinsic attention attains precision@1 0.77, NDCG 0.97, and Spearman 0.63—matching the far more expensive post-hoc gradient (precision@1 0.74) and perturbation (0.74) methods on precision@1, and in fact exceeding them in rank correlation while exhibiting lower variance. Since attention is obtained in the same forward pass as the prediction—whereas gradient needs a backward pass per node and perturbation a forward pass per incoming edge—X-TGNN delivers explanations of at least equal quality essentially for free.

Table 4 Explanation Quality vs. Ground-Truth Root Causes, and Per-Explanation Cost

Method	P@1	NDCG	ρ	Cost
Attention (ours)	0.77±0.04	0.97±0.02	0.63±0.10	1 fwd
Gradient	0.74±0.14	0.97±0.02	0.51±0.27	1 bwd/node
Perturbation	0.74±0.12	0.98±0.02	0.50±0.24	deg(j) fwd/node
Random	0.26±0.05	0.75±0.03	-0.03±0.08	–

Note: deg(j) is the in-degree of the explained node.

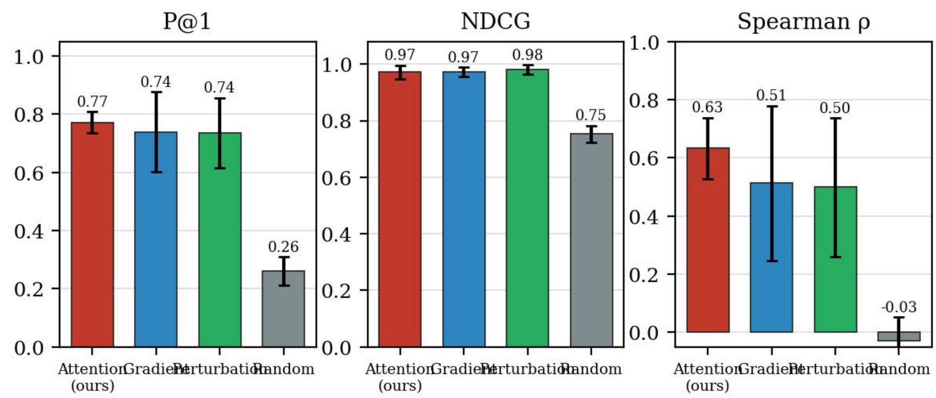

Figure 4 Explanation Quality against Ground-Truth Root Causes

Note: The intrinsic attention (red) matches expensive post-hoc gradient and perturbation explanations and greatly exceeds the random baseline.

6.5 Explanation Fidelity

Figure 5 reports fidelity. Removing the top 20% of edges ranked by attention drops the predicted probability by 0.35 on average (comprehensiveness), versus only 0.11 for random removal, showing that high-attention edges are the ones the model actually relies on. Conversely, retaining only the top 20% of attention edges nearly preserves the prediction (sufficiency gap 0.10), whereas retaining random edges loses most of it (gap 0.33). The attention explanations are thus faithful to the model, not merely plausible.

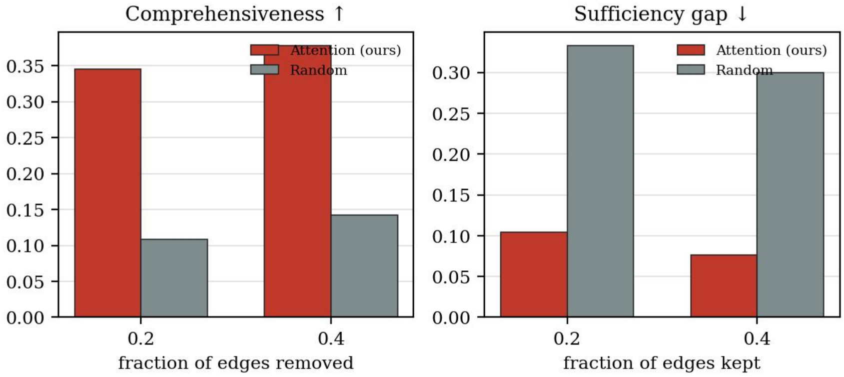

Figure 5 Fidelity of Attention Explanations

Note: Left: removing top-attention edges sharply reduces the predicted probability (higher is better). Right: keeping only top-attention edges preserves the prediction (lower gap is better). Attention (red) dominates random selection.

6.6 Qualitative Example

Figure 6 shows a representative test event: a focal node predicted to be disrupted, its incoming edges weighted by learned attention. The attention concentrates on the disrupted suppliers with the highest dependency weights—precisely the ground-truth dominant causes—while near-zero attention is placed on operational or weakly-connected suppliers, giving a manager a direct, actionable read on where to intervene.

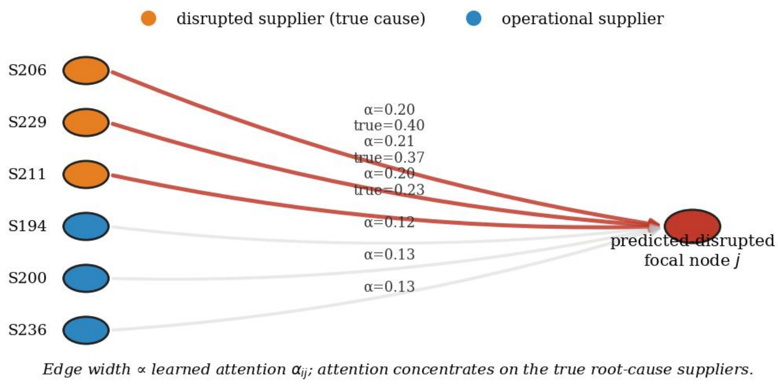

Figure 6 Qualitative Propagation-Path Explanation for One Predicted Disruption

Note: Edge width is proportional to learned attention; the model concentrates attention on the true dominant root-cause suppliers (disrupted, orange).

7 DISCUSSION AND LIMITATIONS

Practical implications. By coupling accurate onset prediction with faithful, single-pass root-cause attribution, X-TGNN turns an opaque risk score into an actionable signal: it flags which downstream entities are about to be disrupted and names the upstream suppliers responsible, so that dual sourcing, buffering, or expediting can be focused where propagation originates.

Limitations. Our evaluation relies on synthetic ground truth; this is a deliberate choice—it is what makes a quantitative assessment of explanation correctness possible at all—but real supply chains involve richer, heterogeneous mechanisms that our cascade model, though grounded in the risk-diffusion literature, necessarily simplifies. Validating on a real multi-tier network annotated with disruption-propagation paths remains important future work, contingent on the release of such data. Attention is one, but not the only, notion of explanation; combining it with counterfactual analysis is promising. Finally, our graphs are of moderate size; scaling to industrial networks with heterogeneous node and edge types, and modeling continuous-time events and longer lead-time dynamics where the temporal component should yield larger gains, are natural next steps.

8 CONCLUSION

We introduced X-TGNN, an explainable temporal graph neural network for disruption prediction and risk propagation in multi-tier supply chains, together with a simulation testbed that—by supplying ground-truth cascade paths—enables the quantitative evaluation of explanations in this domain for the first time. Experiments show that modeling the directed, weighted supply graph is decisive for predicting new disruptions, and that X-TGNN's intrinsic attention explanations recover true root-cause suppliers as accurately as costly post-hoc methods while being computed for free during inference. We hope the model and, in particular, the ground-truth evaluation methodology encourage further work on explainable, propagation-aware supply-chain analytics.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Dolgui A, Ivanov D, Sokolov B. Ripple effect in the supply chain: an analysis and recent literature. International Journal of Production Research, 2018, 56(1-2): 414-430.
[2] Ivanov D, Sokolov B, Dolgui A. The ripple effect in supply chains: trade-off 'efficiency-flexibility-resilience' in disruption management. International Journal of Production Research, 2014, 52(7): 2154-2172.
[3] Basole R C, Bellamy M A. Supply network structure, visibility, and risk diffusion: A computational approach. Decision Sciences, 2014, 45(4): 753-789.
[4] Ivanov D. Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. Transportation Research Part E: Logistics and Transportation Review, 2020, 136: 101922.
[5] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: Proceedings of the International Conference on Learning Representations, 2017.
[6] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. In: Proceedings of the International Conference on Learning Representations, 2018.
[7] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs. In: Proceedings of the Conference on Neural Information Processing Systems, 2017.
[8] Xu D, Ruan C, Korpeoglu E, et al. Inductive representation learning on temporal graphs. In: Proceedings of the International Conference on Learning Representations, 2020.
[9] Pareja A, Domeniconi G, Chen J, et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 5363-5370.
[10] Rossi E, Chamberlain B, Frasca F, et al. Temporal graph networks for deep learning on dynamic graphs. In: ICML Workshop on Graph Representation Learning, 2020.
[11] Ying R, Bourgeois D, You J, et al. GNNExplainer: Generating explanations for graph neural networks. In: Proceedings of the Conference on Neural Information Processing Systems, 2019: 9240-9251.
[12] Luo D, Cheng W, Xu D, et al. Parameterized explainer for graph neural network. In: Proceedings of the Conference on Neural Information Processing Systems, 2020: 19620-19631.
[13] Vu M N, Thai M T. PGM-Explainer: Probabilistic graphical model explanations for graph neural networks. In: Proceedings of the Conference on Neural Information Processing Systems, 2020: 12225-12235.
[14] Pope P E, Kolouri S, Rostami M, et al. Explainability methods for graph convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 10772-10781.
[15] Hosseini S, Ivanov D, Dolgui A. Review of quantitative methods for supply chain resilience analysis. Transportation Research Part E: Logistics and Transportation Review, 2019, 125: 285-307.
[16] Kleindorfer P R, Saad G H. Managing disruption risks in supply chains. Production and Operations Management, 2005, 14(1): 53-68.
[17] Craighead C W, Blackhurst J, Rungtusanatham M J, et al. The severity of supply chain disruptions: design characteristics and mitigation capabilities. Decision Sciences, 2007, 38(1): 131-156.
[18] Tang C S. Perspectives in supply chain risk management. International Journal of Production Economics, 2006, 103(2): 451-488.
[19] Wagner S M, Bode C. An empirical examination of supply chain performance along several dimensions of risk. Journal of Business Logistics, 2008, 29(1): 307-325.
[20] Inoue H, Todo Y. Firm-level propagation of shocks through supply-chain networks. Nature Sustainability, 2019, 2: 841-847.
[21] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry. In: Proceedings of the International Conference on Machine Learning, 2017: 1263-1272.
[22] Xu K, Hu W, Leskovec J, et al. How powerful are graph neural networks? In: Proceedings of the International Conference on Learning Representations, 2019.

[23] Seo Y, Defferrard M, Vandergheynst P, et al. Structured sequence modeling with graph convolutional recurrent networks. In: Proceedings of the International Conference on Neural Information Processing, 2018: 362-373.

[24] Li Y, Yu R, Shahabi C, et al. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: Proceedings of the International Conference on Learning Representations, 2018.

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

[26] Sankar A, Wu Y, Gou L, et al. DySAT: Deep neural representation learning on dynamic graphs via self-attention networks. In: Proceedings of the ACM International Conference on Web Search and Data Mining, 2020: 519-527.

[27] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.

[28] Zhou J, Cui G, Hu S, et al. Graph neural networks: A review of methods and applications. AI Open, 2020, 1: 57-81.

[29] Baldassarre F, Azizpour H. Explainability techniques for graph convolutional networks. arXiv:1905.13686, 2019.

[30] Ribeiro M T, Singh S, Guestrin C. 'Why should I trust you?': Explaining the predictions of any classifier. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 1135-1144.

[31] Lundberg S M, Lee S-I. A unified approach to interpreting model predictions. In: Proceedings of the Conference on Neural Information Processing Systems, 2017: 4765-4774.

[32] Yuan H, Yu H, Gui S, et al. Explainability in graph neural networks: A taxonomic survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5782-5799.