Research Square

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

Physics-informed graph learning of collapse distance in complex networks

Xin Lu

xin.lu.lab@outlook.com

National University of Defense Technology https://orcid.org/0000-0002-3547-6493

Jie Zhang

National University of Defense Technology

Tao Wang

Yatai Ji

Hua He

National University of Defense Technology

Zhengqiu Zhu

National University of Defense Technology

Bo-quan Zhang

https://orcid.org/0009-0003-8742-3777

Xin Zhou

National University of Defense Technology

Changjun Fan

National University of Defense Technology

Bin Chen

University of Electronic Science and Technology of China

Manlio De Domenico

University of Padua

Article

Keywords:

Posted Date: July 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10222350/v1

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

Additional Declarations: There is **NO** Competing Interest.

Physics-informed graph learning of collapse distance in complex networks

Jie Zhang[1,#], Tao Wang[1,#], Yatai Ji[1,2], Hua He[1], Zhengqiu Zhu[1,2], Boquan Zhang[1], Xin Zhou[1], Changjun Fan[1], Bin Chen[3], Manlio De Domenico[4,*], Xin Lu[1,*]

[1]*College of Systems Engineering, National University of Defense Technology, Changsha, 410073, China*

[2]*State Key Laboratory of Digital Intelligent Modeling and Simulation, National University of Defense Technology, Changsha, 410073, China*

[3]*The Institute of Intelligent Computing, University of Electronic Science and Technology of China, Chengdu 611731, China*

[4]*Department of Physics and Astronomy "Galileo Galilei", University of Padua, via F. Marzolo 8, 35131 Padova, Italy*

[#]These authors contributed equally to this work.

[*]Correspondence: Manlio De Domenico (`manlio.dedomenico@unipd.it`), Xin Lu (`xin.lu.lab@outlook.com`).

Abstract

Complex networks can undergo abrupt transitions from functional operation to systemic collapse under progressive failures or attacks, yet existing robustness metrics provide no universal measure of how close a network is to failure. Here, we introduce collapse distance, a target-configurable safety-margin metric defined as the minimum fraction of additional node removals required to trigger a prescribed collapse response. To estimate it efficiently, we develop TCR-GIN, a physics-informed graph learning framework trained with dismantling-derived upper-bound supervision and degradation-law regularization. We theoretically establish a sufficient condition under which TCR-GIN can outperform any individual heuristic in exact collapse distance and derive an error bound. Across synthetic benchmarks and diverse real-world networks, TCR-GIN occupies the accuracy–efficiency Pareto front among 28 baselines and generalizes to unseen real systems. Under sustained attacks, its trajectories are more stable and interpretable than those of competing baselines and widely used robustness metrics, enabling a configurable early-warning signal. With collapse distance as a measure of network failure proximity, our framework shifts robustness from retrospective evaluation to real-time safety management.

Introduction

From biological systems that sustain life to engineering infrastructures that support modern society [1–4], the functional stability of complex systems is fundamentally governed by their

topological structure [5,6] and its interplay with intervening dynamical processes [7–9]. Although these systems have acquired a certain degree of robustness through long-term evolution or engineering design, they can remain highly vulnerable to specific perturbations [10]. For example, the loss of key species in an ecosystem can trigger cascading extinctions [11], while the failure of a few critical lines in a power grid, such as in response to environmental stressors [12], can lead to widespread blackouts [13]. These risks motivate real-time monitoring of network structural robustness, where the key quantity is the remaining safety margin before a prescribed breakdown threshold (or collapse target) is reached [14–18].

Existing studies have largely quantified network structural robustness through connectivity-based indicators, particularly those derived from the giant component, or largest connected component (LCC) [19]. Yet quantities that are useful for robustness assessment do not necessarily serve as effective warning indicators during ongoing degradation. As illustrated in Fig. 1, conventional metrics can occupy markedly different numerical ranges near collapse across networks, attack scenarios, and collapse targets, making consistent interpretation difficult and hindering the construction of configurable early-warning signals. One class of metrics is based on the snapshot, such as LCC size and natural connectivity [20–22]. Because they characterize only the instantaneous topological state, these metrics do not directly reflect the remaining structural safety margin and may respond sharply only when collapse is already imminent [23,24] (Fig. 1c and g). Moreover, they are not naturally defined with respect to prescribed collapse targets, which limits their ability to measure the distance to a desired level of retained functionality [25–27] (Fig. 1d and h).

A second class relies on attack simulations, exemplified by the R-value proposed by Schneider et al. [28], which quantifies robustness by computing the area under the LCC degradation curve during a sustained attack. The validity of such metrics depends on the chosen simulation strategy [10,16,29], thereby coupling robustness metrics to a hypothesized attacker rather than to the network itself. However, existing strategies, such as degree- or betweenness-based attacks, exhibit pronounced inconsistency because their performance varies substantially across network topologies, and their sensitivity to collapse targets differs across settings [17] (Fig. 1c and g). Moreover, attack-induced R-value trajectories can exhibit pronounced non-monotonic fluctuations and abrupt changes (Fig. 1e and i), reducing their interpretability as warning signals. Such simulations are also computationally expensive, limiting their practicality for real-time monitoring [17,22].

To address those challenges, we define **collapse distance** (CD) as the minimum fraction of node removals required to drive a network to a prescribed, application-specific collapse target. Beyond its semantic interpretation, we prove that CD obeys monotonicity and bounded variation under single-node removal and scale-transformed additivity under fragmentation. Our objective is not to construct an intervention sequence, but to estimate a graph-level distance to failure for monitoring and early warning (Fig. 1o and p).

Computing CD exactly is NP-hard, precluding its direct use in online settings. To overcome this limitation, we develop Topology-Consistency Regularized Graph Isomorphism Network (TCR-GIN) (Fig. 2), a physics-informed graph learning framework that maps network topology

to collapse distance in a single forward pass. Here, physics-informed refers to incorporating theoretically derived degradation laws for the discrete network collapse process, specifically the monotonicity and bounded-variation behavior of CD under node removal. TCR-GIN is trained with dismantling-derived upper-bound supervision and regularized by degradation laws of CD to encourage physically consistent predictions under progressive node removal. Because every dismantling heuristic furnishes only an upper bound on CD, this supervision scheme introduces systematic one-sided bias. The theoretical analysis addresses three questions that arise from this one-sided supervision design. First, can a learner trained on biased labels outperform its own supervision? We prove a sufficient condition under which TCR-GIN achieves a smaller true-label mean absolute error (MAE) than any individual heuristic, where true labels (TL) are exact collapse distances. This explains why a smooth predictor that aggregates multiple upper-bound supervisions can outperform each heuristic, even though none provides true labels for training. Second, does the guarantee improve as additional dismantling heuristics are incorporated? We derive an explicit upper bound on the true-label MAE and show that its supervision-bias components contract monotonically with heuristic-pool expansion, so the estimator can co-evolve with advances in dismantling algorithms upon relabeling and retraining. Third, what role does degradation-law regularization play beyond label quality? We show that topological-consistency regularization reduces a supervision-bias component that persists regardless of training-set size, providing a benefit that neither more data nor a larger heuristic pool alone can substitute. Further, we bound the additivity error of TCR-GIN predictions under network fragmentation. The overall framework runs in $O(N + M)$ time per forward pass for a network with N nodes and M edges.

We evaluate TCR-GIN on synthetic benchmarks, real-world network datasets, and sustained-attack trajectories. Across multiple benchmarks, the model provides favorable accuracy–efficiency trade-offs relative to advanced algorithms, such as GDMR [14], FINDER [30], and DomiRank [31]. On small networks with true labels, TCR-GIN achieves lower true-label MAE than all 28 baselines in the heuristic pool. TCR-GIN also shows zero-shot transfer from synthetic training networks to REDDIT and benefits from domain-specific data augmentation on infrastructure networks. The predicted CD trajectories remain stable and interpretable across diverse networks, attack scenarios, and collapse targets. Further, on two real-world infrastructure networks, they provide configurable early-warning signals that conventional metrics such as LCC size, natural connectivity [22], and the R-value [28] are failing. For larger systems beyond the direct training scale, we use late-stage component-wise reconstruction after fragmentation, exploiting the additivity structure of CD once component sizes enter the trained regime. Beyond the fixed-target scalar setting considered here, we further explore a lightweight extension from learning a single collapse-distance value to predicting a discrete collapse profile. Overall, these results position collapse distance as a practical and interpretable quantity for monitoring structural collapse and show how one-sided heuristic supervision can be combined with discrete degradation laws to build a theoretically grounded graph-learning estimator.

Results

physics-informed graph learning enables collapse distance for early warning in complex networks

Theoretical formulation of collapse distance

We define collapse distance for a given network state, whether undisturbed or already degraded by a sequence of attacks, as the minimum fraction of additional node removals required to reach a prescribed collapse target. This metric is designed to provide a unified and configurable measure of structural robustness for complex systems across different scenarios. Formally, consider a network $G = (V, E)$ with $N = |V|$ nodes. Let $\Phi(G) \in [0, 1]$ denote a normalized structural measure (in this work, we adopt the LCC size). Given a prescribed collapse target $\tau \in [0, 1]$, the collapse distance $D_\tau(G)$ is defined as:

$$D_\tau(G) := \min_{\mathcal{R} \subseteq V} \left\{ \frac{|\mathcal{R}|}{N} \,\middle|\, \Phi(G \backslash \mathcal{R}) \leq \tau \right\}, \tag{1}$$

where $\mathcal{R}$ denotes the set of removed nodes. A smaller CD means that the system is closer to collapse, whereas a larger CD indicates a greater remaining structural safety margin. Collapse distance has two key properties. First, it is target-configurable because its value explicitly depends on the scenario-specified threshold τ. Second, it is strategy-independent because the minimization ranges over all possible removal sets, so collapse distance captures the worst-case safety margin determined solely by network topology and is independent of any particular attack strategy. This strategy independence fundamentally distinguishes collapse distance from the critical removal fraction [10].

CD further obeys two structural properties that are central to both estimation and deployment. First, under single-node removal, the scale-corrected decrease in collapse distance is bounded, with $\Delta \in \{0, 1/N\}$. This bounded variation yields a well-behaved degradation trajectory and

rules out abrupt jumps. Second, when the network fragments into disconnected components, the global collapse distance can be written as a size-weighted aggregation of component-wise collapse distances evaluated under appropriately rescaled collapse targets. This scale-transformed additivity provides the basis for component-wise estimation and for extending the framework to fragmented large networks. These results are formalized as Theorem S1 and Theorem S2 in the Supplementary Note I, and they jointly motivate both the regularization scheme and the component-wise reconstruction strategy of the estimator developed next.

The TCR-GIN framework

Exact computation of the collapse distance is NP-hard and becomes impractical at scale. We therefore formulate collapse-distance estimation as a graph-level regression problem. The exact collapse distance is unavailable for large networks. For training, we run a pool of dismantling algorithms $\mathcal{H}$ and define the observed label (OL) as $\tilde{D}_\tau(G;\mathcal{H}) = \min_{h\in\mathcal{H}} \hat{D}^h_\tau(G)$, where $\hat{D}^h_\tau(G)$ is the collapse distance obtained by heuristic h. We learn a predictor f_θ from a network snapshot G to the predicted collapse distance $\hat{D}_\tau(G)$. The input comprises the network topology and node structural features.

We develop TCR-GIN, a topological-consistency-regularized graph isomorphism network (Fig. 2a). The framework integrates a graph neural encoder with constraints derived from the degradation laws of CD. Specifically, we adopt the graph isomorphism network (GIN) [32,33] as the backbone to embed network structure into high-dimensional representations. Because the collapse

distance is a global network property, we introduce a virtual node to extend the receptive field of local message passing. The virtual node connects to all real nodes and enables layer-wise global information exchange via permutation-invariant aggregation. TCR-GIN uses a Long Short-Term Memory (LSTM)-based multi-scale aggregation module to route layer-wise node embeddings into an attention network for fusion. We additionally incorporate residual connections to mitigate over-smoothing. Additional implementation details are provided in Data and Methods. The full theoretical analysis is presented in Supplementary Note II.

Accuracy and efficiency analysis on synthetic and real networks

We compare TCR-GIN against 28 representative baselines on both synthetic and real-world networks (Supplementary Note IV). The synthetic dataset contains 84,000 networks generated by four standard models, including Barabási–Albert (BA) [34,35], Erdős–Rényi (ER) [36], Watts–Strogatz (WS) [37,38], and Lancichinetti–Fortunato–Radicchi (LFR) [39]. We group networks by size into four regimes, with small, medium, large, and extra-large corresponding to $N < 300$, $300 \leq N < 700$, $700 \leq N < 1100$, and $1100 \leq N \leq 2000$, respectively. Unless otherwise stated, we use the collapse target $\tau = 5/N$. Additional dataset construction and split details are provided in the Methods and Supplementary Note V. We measure prediction accuracy using MAE, and the definition is provided in Supplementary Note VI. Unless otherwise stated, MAE is calculated relative to the observed labels.

Figure 3 summarizes the accuracy–efficiency trade-off across all methods, with detailed

numerical results reported in Table S1. TCR-GIN consistently achieves the lowest prediction error across all four synthetic size regimes, with MAEs of 0.011 ± 0.0002, 0.008 ± 0.0005, 0.005 ± 0.0007, and 0.005 ± 0.0010 for small, medium, large, and extra-large networks, respectively (mean ± s.d. across networks and five independent training seeds). The same convention is used below. Under the CPU-only setting, TCR-GIN consistently lies in the low-error, low-latency region of the scatter plots, showing a consistent advantage in the accuracy–efficiency trade-off. Because TCR-GIN is implemented using sparse operations, it is well-positioned to benefit further from GPU acceleration, whereas many heuristic baselines are less amenable to comparable acceleration.

At comparable inference-time scales, TCR-GIN is substantially more accurate. For example, in the extra-large size range, DC, BC, and TCR-GIN all operate around 10^{-1} s per network, but their MAEs are 0.276, 0.266, and 0.005, respectively. This corresponds to an approximately $50\times$ reduction in error at a similar computational cost.

At a given MAE level, TCR-GIN delivers orders-of-magnitude faster inference. This advantage is most pronounced for large networks. For example, in the extra-large size range, the Min-Sum reinsertion baseline (MSR) attains an MAE of 0.003 with an inference time of 29.89 s per network, whereas TCR-GIN reaches an MAE of 0.005 in only 0.19 s per network, corresponding to an $\sim 160\times$ speedup. The same advantage is observed in the other size ranges.

On the real-world REDDIT dataset, TCR-GIN achieves an MAE of 0.008 ± 0.0008 with an end-to-end runtime of 0.027 s per network, showing a clear advantage over baseline methods. To further assess generalization, we directly evaluate on REDDIT using a model trained only on

synthetic networks, obtaining an MAE of 0.012 ± 0.0034 (the red open star in Fig. 3e). These results indicate that the model retains useful transferability from synthetic training networks to real-world networks.

We further assess TCR-GIN on small networks where exact collapse distances can be computed by exhaustive search and used as true labels for evaluation. By construction, OL satisfies $\mathrm{OL} \geq \mathrm{TL}$ for each network, so the training supervision contains a one-sided upward bias rather than symmetric label noise. Experimentally, we find that despite being trained on OL, TCR-GIN achieves lower TL MAE than every individual heuristic baseline (Fig. 3f–g), consistent with our theoretical analysis under one-sided supervision bias (Supplementary Note II.A). Specifically, on small networks with $N \in [20, 40]$, TCR-GIN attains a TL-based MAE of 0.019 ± 0.0012, compared with an OL-based MAE of 0.023 ± 0.0009. TCR-GIN also outperforms every individual baseline under TL. For example, the best heuristic, betweenness centrality removal (BCR), has a TL-based MAE of 0.034, which TCR-GIN reduces by 47.8%. As shown in Fig. 3f, the observed labels systematically overestimate the true labels, whereas TCR-GIN can predict below OL on part of the test set and thereby move closer to TL. Consistently, many heuristics in Fig. 3g lie below the $y = x$ dashed line, meaning that their MAE with respect to OL is smaller than their MAE with respect to TL. Thus, for dismantling methods, low error with respect to observed labels does not necessarily indicate high accuracy with respect to the true labels.

Online estimation of collapse distance trajectories under sustained attack

Unless otherwise stated, we use the same experimental setting as above and assess both the online accuracy of TCR-GIN and the quality of its predicted trajectories under sustained attacks. Our experiments cover both synthetic networks generated by multiple models and real-world infrastructure networks with diverse structural properties. Compared with static evaluation, online monitoring places greater emphasis on stability and inference efficiency, and it benefits from prediction trajectories that are monotone, smooth, and consistent with component-wise additivity after fragmentation. We compare different methods in terms of estimation accuracy and trajectory consistency across networks and attack scenarios (Fig. 4). Definitions of the evaluation metrics are provided in Supplementary Note VI.

Fig. 4 summarizes aggregated results over seven networks spanning synthetic and real-world networks. To emphasize overall cross-network performance, we aggregate the metric values by summing them across datasets before comparison. Overall, TCR-GIN consistently ranks among the top methods on all three metrics. The stronger baselines are mainly advanced dismantling heuristics and learning-based methods, including GDMR, GDM, CoreGDM, and CoreHD, which can slightly outperform TCR-GIN on a single metric, but they tend to underperform on the others. GDMR achieves the best overall performance across the three metrics. However, its runtime is 2–3 orders of magnitude slower than that of TCR-GIN (Fig. 3).

To illustrate trajectory behavior in real-world systems, we further present the collapse-

distance time series for the Transport and Power networks under three representative attack types (degree-based attacks, betweenness-based attacks, and random attacks) (see Fig. 5; more experimental results are provided in Fig. S2–S7). The experimental results across networks and attack scenarios exhibit similar trends. Within the same attack trajectory, there is a significant discrepancy among collapse-distance estimates yielded by different heuristic baselines. In contrast, the trajectories predicted by TCR-GIN remain nearly monotonic throughout the attack process and stay close to the lower envelope formed by the baseline set. More importantly, this trajectory-level stability makes the estimated collapse distance more interpretable and reliable for online monitoring. Operational and maintenance teams can use the continuous decrease in collapse distance to trigger measures such as node protection and network reconfiguration, rather than passively responding to sudden changes near single-point critical thresholds [40,41].

Under sustained attacks, real-world systems can enter a fragmentation state in which the remaining network is split into multiple disjoint connected components $\{\mathcal{C}_{t,j}\}_{j=1}^{k_t}$. This setting allows us to directly test whether the predictor remains consistent with the component-wise additivity described in Theorem S2. We therefore evaluate snapshots with $k_t > 1$ and compare the model's direct prediction $\hat{D}(G_t)$ with the component-wise aggregated prediction $\hat{D}^{\mathrm{agg}}(G_t)$ obtained by combining predictions over connected components according to Theorem S2. We quantify their agreement using the additivity mean absolute error (A-MAE). As shown in Fig. S1, TCR-GIN exhibits small additivity errors on both the synthetic and real-world networks spanning diverse generative models and scales.

We further examine a Route network whose scale substantially exceeds the regime covered during training and test whether TCR-GIN can still be applied through late-stage monitoring and component-wise reconstruction. Supported by the additivity of CD itself (Theorem S2) and the additivity-error bound of TCR-GIN in Supplementary Note II.D, this strategy enables scalable monitoring of larger networks without modifying the model architecture. The window-selection rule for Route and a detailed case study are provided in Supplementary Note IX.

Configurable early warning signal

Collapse distance enables configurable early warning under sustained attacks. To make this operational, we introduce a user-defined warning target n/N_0, where n denotes the minimum desired lead time before collapse, measured in attack steps, and N_0 is the initial network size. Under the one-node-per-step attack setting considered here, this is equivalent to n node removals. Across the four representative trajectories in Fig. 6, using $n = 3$ and $\tau = 0.3$, the warning consistently precedes collapse, with observed lead times ranging from 11 to 66 attack steps. This behavior is observed across most attack scenarios, with the full results reported in Fig. S12–S15.

The warning target has a conservative temporal interpretation. When CD reaches n/N_0, the current remaining network is already close enough to collapse that a collapse-inducing set of size at most n exists under a worst-case continuation. The realized attack sequence, however, may not coincide with such a set. This explains why the system can remain above the collapse target for more than n additional attack steps, as illustrated by the mismatch between the red collapse-

inducing sets and the orange future attack nodes in Fig. 6c,d,g,h. In this sense, the warning step marks a conservative safety threshold. At this point, the remaining network can be driven to collapse by at most n further removals under a worst-case continuation, so intervention is warranted even if the realized attack sequence is less damaging and collapse occurs later.

This early-warning property reflects a fundamental difference in how LCC size and collapse distance evolve under sustained attacks. LCC size can remain nearly unchanged for several steps before dropping abruptly, whereas collapse distance decreases in bounded increments under one-node removal. Accordingly, LCC mainly reflects realized functional loss after the critical structure fails, while the collapse distance provides a more interpretable estimate of the remaining margin to collapse. Plateau regions also differ in meaning. A flat LCC trajectory can mask ongoing redundancy erosion, whereas a flat collapse-distance trajectory indicates that recent removals have not further shortened the minimum path to collapse. Learning-based estimates may introduce mild local jitter, but this does not change the overall decreasing trend.

Comparison with conventional robustness metrics further clarifies this advantage. We evaluated LCC size, natural connectivity $\bar{\lambda}$, two R-based measures, and collapse distance estimated online by TCR-GIN on the same attack sequences under $\tau = 0.3$ and $\tau = 0.5$ (Fig. 7, and additional cases are provided in Fig. S16-S19). Natural connectivity captures redundancy loss, but its near-collapse magnitudes are not calibrated to a fixed remaining safety margin. Across the Transport trajectories, $\bar{\lambda}$ remains below 1, whereas in the Power network it reaches several tens (Fig. 7). Moreover, even within the same network, its value at the last pre-collapse snapshot varies

with attack type and collapse target. For example, in the Power network under DC attack, the value at the last pre-collapse snapshot shifts from 29.9 to 39.1 when τ increases from 0.3 to 0.5 (Fig. 7v, af). Thus, although $\bar{\lambda}$ is informative about redundancy loss, a fixed threshold in $\bar{\lambda}$ cannot be assigned a transferable remaining-collapse margin across settings. The R-based measures show a similar limitation. Their values at the last pre-collapse snapshot vary substantially across attack scenarios. For example, in the Power network, $R(\mathrm{DCR})$ is 0.064 under DC attack and 0.126 under BC attack for $\tau = 0.3$, and remains separated at 0.056 and 0.099, respectively, for $\tau = 0.5$ (Fig. 7x, ac, ah, am). In addition, the online trajectories of both $R(\mathrm{rand})$ and $R(\mathrm{DCR})$ commonly exhibit visible oscillations and local rebounds, weakening the correspondence between signal level and risk stage during monitoring.

By contrast, collapse distance is calibrated by the number of additional removals needed to reach collapse. As attacks progress, the trajectories approach the one-removal level on the order of $1/N_0$, providing a direct reference for how close the system is to collapse. This calibration allows the same warning target n/N_0 to retain the same interpretation as a worst-case margin of n further removals across networks, attack scenarios, and collapse targets. Together with its configurable lead time, collapse distance therefore offers a practical basis for online warning and cross-scenario comparison.

Discussion

Complex networks underpin the connectivity of critical infrastructures such as power grids, transportation systems, and communications. In these settings, risk management and operational scheduling require a quantifiable metric that directly links to operational thresholds and answers a simple operational question: *how far from failure?* Collapse distance addresses this need by measuring the minimum fraction of node removals required to reach a prescribed collapse target, providing a safety margin that is anchored to the operational threshold while remaining independent of any particular attack strategy. To make CD practically computable, we develop TCR-GIN, a graph-learning estimator that combines dismantling-derived upper-bound supervision with degradation-law regularization, turning a theoretically appealing yet NP-hard quantity into a single-forward-pass predictor.

We provide a theoretical analysis of both the metric itself and the estimation framework. For collapse distance, we prove two structural laws. First, one of the most remarkable features of our framework is that CD obeys bounded variation under progressive node removal (Theorem S1), providing the theoretical basis for a stable, non-lagging monitoring signal. This guarantee also ensures that CD values near collapse share the same magnitudes of $1/N_0$ across networks, attack scenarios, and collapse targets, representing a significant conceptual advantage over existing snapshot-based or simulation-driven robustness metrics (Fig. 7). Second, CD satisfies scale-transformed linear additivity under fragmentation (Theorem S2), so that the system-level value can be reconstructed from component-level values via a weighted sum, justifying scalable, component-

wise evaluation on larger networks whose fragments fall within the trained size regime. For TCR-GIN, three theoretical results explain why the supervision design works despite label bias. The sufficient condition for true-label MAE improvement (Theorem S3) shows that the learner needs only fit the observed labels comparably while inheriting less one-sided upward bias, enabling a smooth predictor to move below the lower envelope of its own supervision, although the bias threshold can currently only be verified against exact solutions on small networks. The true-label MAE bound contracts monotonically with heuristic-pool expansion (Theorem S5), so TCR-GIN can absorb future stronger dismantling algorithms by relabeling and retraining alone. The bounded-variation regularization analysis (Theorem S6) translates the structural law of Theorem S1 into a soft constraint on the learned predictor, identifying and reducing a bias that persists in the infinite-data limit under fixed biased labels and that cannot be removed by pool expansion alone. Beyond the supervision mechanism, the bounded additivity error of the learned predictor (Theorem S7) underpins component-wise reconstruction for networks beyond the directly trained scale, connecting the estimator to the exact additivity law of CD. In addition, TCR-GIN scales linearly with network size owing to the message-passing architecture, making it applicable to large-scale infrastructure networks.

The empirical results support and extend the theoretical picture. Across four synthetic size regimes and one real-world network dataset, TCR-GIN consistently occupies the accuracy–efficiency Pareto front relative to 28 representative baselines (Fig. 3a–e). On small networks where exact collapse distances are computable, TCR-GIN attains smaller true-label MAE than every individual heuristic used as its own supervision, even though its observed-label MAE exceeds that

of some heuristics (Fig. 3f–g), confirming the phenomenon predicted by Theorem S3. The model transfers zero-shot from synthetic training networks to the real-world REDDIT dataset (Fig. 3e). With domain-specific data augmentation, online CD trajectories on infrastructure networks are nearly monotone and hug the lower envelope of competing estimators throughout the attack process (Figs. 4–5). The additivity MAE remains low upon fragmentation (Fig. S1), enabling component-wise reconstruction to handle networks well beyond the trained size regime. For instance, the ROUTE network with approximately $6{,}000$ nodes is evaluated using a model trained on networks of up to $2{,}000$ nodes (Fig. S11). Translated into a configurable early-warning signal, CD provides stable early warnings on two infrastructure networks under two collapse targets across 30 attack scenarios (Fig. 6, full results in Supplementary Figs. S12–S15). Figure 7 explains why this capability is difficult to obtain from conventional metrics. Beyond the fixed-target scalar setting, a lightweight collapse-profile extension (Supplementary Note X) replaces the scalar regression head with a vector-valued head over a discrete target grid, indicating that collapse distance can serve not only as a configurable scalar warning signal but also as a profile-based characterization of structural vulnerability.

We emphasize that our objective differs from prior network dismantling studies, which seek to identify an attack sequence or a minimum dismantling set [14,17,30]. Instead, we focus on the minimum fraction required to reach a prescribed collapse target, leveraging it as an interpretable indicator for monitoring and early warning. Our setting is also different from resilience inference and critical-transition early-warning paradigms that focus on node activity trajectories [11,42–44]. Those approaches typically target changes in recovery capacity or dynamical

tipping points, whereas here we characterize a topology-defined safety boundary. Beyond this conceptual separation, the warning-step formalization developed here casts structural-robustness assessment into an actionable early-warning signal, linking topological analysis with real-time monitoring and decision-making.

Several directions could extend the practical scope of structural robustness research. First, the discrete collapse-profile predictor explored here could be extended to a continuous target-conditioned estimator $f(G, \tau)$, allowing collapse distance to be queried at arbitrary operational thresholds and further combined with task-specific functional failure criteria. Second, the failure setting could be generalized from node removals to edge failures, mixed node–edge failures, and coupled degradation in multilayer or interdependent networks [45,46]. Third, CD could be used not only for passive monitoring but also as an objective or feedback signal for proactive intervention, including protection prioritization, network rewiring, repair scheduling, and robustness optimization [47–50]. Together, these directions would close the loop between structural risk assessment, early warning, and active system restoration. To conclude, TCR-GIN makes collapse distance practically usable by providing an efficient estimator for a theoretically appealing yet NP-hard quantity. More broadly, this work illustrates how degradation-law constraints derived from a problem's combinatorial structure can be embedded as inductive biases into graph learning and helps move structural robustness analysis from qualitative risk awareness toward quantitative, proactive safety management.

Data and Methods

TCR-GIN

Let $G = (V, E)$ be a networked system with $N = |V|$ nodes, represented by an adjacency matrix $\mathbf{A} \in \{0,1\}^{N\times N}$ and a structural feature matrix $\mathbf{X} \in \mathbb{R}^{N\times d_{\text{in}}}$. The feature matrix $\mathbf{X}$ comprises three basic network descriptors, including node degree, local clustering coefficient, and k-core value. Our goal is to train a graph neural network $f_\theta : (\mathbf{A}, \mathbf{X}) \mapsto [0, 1]$ to estimate the collapse distance.

Encoder. We adopt a multi-layer message-passing architecture enhanced with virtual nodes to capture global contextual information. Let $\mathbf{h}_i^{(l)} \in \mathbb{R}^h$ denote the embedding of node i at layer l, and let $\mathbf{g}^{(l)} \in \mathbb{R}^h$ represent the embedding of the virtual node. The initial embedding vector of the virtual node, $\mathbf{g}^{(0)}$, is initialized from a standard normal distribution and then learned jointly with the model parameters. To facilitate bidirectional information flow, the global virtual node embedding from the previous layer is incorporated into the current local node features, ensuring that the local aggregation process captures global context. The node-enhanced feature $\widetilde{\mathbf{h}}_i^{(l-1)}$, which incorporates global information, is expressed as,

$$\widetilde{\mathbf{h}}_i^{(l-1)} = \mathbf{h}_i^{(l-1)} + \mathbf{g}^{(l-1)}. \tag{2}$$

Based on the enhanced features, we use GIN convolution layers to aggregate local neighborhood information. To facilitate gradient propagation in deep networks and prevent over-smoothing,

we introduce residual connections for the enhanced features after the convolution operation. The state update formula for node i is,

$$\mathbf{h}_i^{(l)} = \sigma\left(\mathrm{BN}\left(\widetilde{\mathbf{h}}_i^{(l-1)} + \sum_{j\in\mathcal{N}(i)} \widetilde{\mathbf{h}}_j^{(l-1)}\right)\right) + \widetilde{\mathbf{h}}_i^{(l-1)}, \tag{3}$$

where $\mathcal{N}(i)$ denotes the neighbor set of node i, $\mathrm{BN}(\cdot)$ refers to Batch Normalization, and $\sigma(\cdot)$ is the GeLU activation function.

After all real nodes complete their local updates, the virtual node updates its global state by aggregating the latest representations of all nodes at the current layer,

$$\mathbf{g}^{(l)} = \mathbf{g}^{(l-1)} + \mathrm{MLP}\left(\sum_{i\in V} \mathbf{h}_i^{(l)}\right). \tag{4}$$

Multi-scale aggregation. The collapse distance is jointly determined by topological signals spanning multiple scales. To adaptively fuse these hierarchical representations, we employ a Jumping Knowledge (JK) architecture with an LSTM-attention mechanism. Specifically, we treat the layer-wise representations of node i as a temporal sequence $S_i = (\mathbf{h}_i^{(1)}, \ldots, \mathbf{h}_i^{(L)})$ and feed it into an LSTM to capture inter-layer dependencies,

$$\mathbf{o}_i^{(l)},\, \mathbf{c}_i^{(l)} = \mathrm{LSTM}\Big(\mathbf{h}_i^{(l)}, \Big(\mathbf{o}_i^{(l-1)}, \mathbf{c}_i^{(l-1)}\Big)\Big), \tag{5}$$

where $\mathbf{o}_i^{(l)}$ and $\mathbf{c}_i^{(l)}$ denote the LSTM output state and memory cell at step l, respectively. The memory cell $\mathbf{c}_i^{(l)}$ propagates long-range structural dependencies across layers, while the output state $\mathbf{o}_i^{(l)}$ encodes contextual features at the current hierarchy level. To mitigate over-smoothing and automatically select the most informative hierarchical signals, we further compute an attention

weight $\alpha_i^{(l)}$ for each layer,

$$\alpha_i^{(l)} = \frac{\exp\left(\mathbf{w}^T \mathbf{o}_i^{(l)}\right)}{\sum_{k=1}^{L} \exp\left(\mathbf{w}^T \mathbf{o}_i^{(k)}\right)}, \tag{6}$$

where $\mathbf{w}$ is a learnable attention parameter vector. The final node embedding $\mathbf{h}_i^{\text{final}}$ is defined as the weighted sum of the LSTM outputs across all layers,

$$\mathbf{h}_i^{\text{final}} = \sum_{l=1}^{L} \alpha_i^{(l)} \, \mathbf{o}_i^{(l)}. \tag{7}$$

This mechanism dynamically balances the preservation of critical local details with the capture of global semantics by adaptively assigning weights.

Regression. Finally, we obtain a graph-level representation via global mean pooling and map it to a scalar prediction through an MLP,

$$\hat{D}_\tau(G) = \text{MLP}\left(\frac{1}{N}\sum_{i \in V} \mathbf{h}_i^{\text{final}}\right). \tag{8}$$

At inference time, after converting the regression output back to the original collapse-distance scale, we projected the prediction onto the physically valid range, $\hat{D}_\tau(G) = \min\{1, \max\{0, \text{MLP}\left(\frac{1}{N}\sum_{i \in V} \mathbf{h}_i^{\text{final}}\right)\}\}$. Thus, predictions below 0 were set to 0, predictions above 1 were set to 1, and predictions within $[0, 1]$ were left unchanged.

Training loss function. The training process is driven by a composite loss function $\mathcal{L} = (1 - \lambda)\,\mathcal{L}_{\text{sup}} + \lambda\,\mathcal{L}_{\text{consist}}$, which aims to balance data fitting accuracy and topological consistency, where $\lambda \in [0, 1]$. Here, $\mathcal{L}_{\text{sup}}$ is the mean squared error (MSE) with respect to the observed labels, while

$\mathcal{L}_{\text{consist}}$ encourages the monotonicity and smoothness of the predictions. In this discrete network setting, the physics-informed component is not a partial differential equation (PDE) residual as in classical physics-informed neural networks (PINNs) [51,52], but a theoretically derived one-step degradation law that constrains how CD can change after node removal.

Supervised loss ($\mathcal{L}_{\text{sup}}$). We adopt MSE to minimize the discrepancy between the prediction $\hat{D}_\tau(G)$ and the observed label $\tilde{D}_\tau(G)$,

$$\mathcal{L}_{\text{sup}} = \mathbb{E}_{G\in\mathcal{B}}\left[\left(\hat{D}_\tau(G) - \tilde{D}_\tau(G)\right)^2\right], \tag{9}$$

where $\mathcal{B}$ denotes a mini-batch of graph samples.

Topological consistency loss ($\mathcal{L}_{\text{consist}}$). Consider an anchor graph G with N nodes. Given its node set V, we randomly sample K nodes $\{v_1, \ldots, v_K\}$ and construct K node-deleted perturbed graphs by removing a single node, denoted as $\{G'_k\}_{k=1}^K$, where $G'_k = G \setminus \{v_k\}$. Since the collapse distance is a normalized metric, directly comparing predictions on subgraphs (with denominator $N-1$) and the original graph (with denominator N) may introduce scale mismatch. Following Theorem S1, we introduce a scale correction factor $\gamma = \frac{N-1}{N}$ to map the subgraph prediction back to the scale of the original graph. Guided by Theorem S1, the corrected difference $\Delta_k = \hat{D}(G) - \gamma \cdot \hat{D}(G'_k)$ should ideally belong to the discrete set $\left\{0, \frac{1}{N}\right\}$. Accordingly, we define the consistency error function $E(G, G'_k)$ as the minimum distance between the prediction gap and the theoretically feasible values,

$$E(G, G'_k) = \min\left(\left|\Delta_k - 0\right|^2, \left|\Delta_k - \frac{1}{N}\right|^2\right). \tag{10}$$

To capture the key structures that most significantly affect system robustness, we adopt a worst-case consistency error strategy, i.e., selecting the maximum error among the K samples for backpropagation,

$$\mathcal{L}_{\text{consist}} = \mathbb{E}_{G \in \mathcal{B}} \left[\max_{k \in \{1,\ldots,K\}} E(G, G'_k) \right]. \tag{11}$$

This regularization incorporates discrete degradation laws of network collapse into the learning objective.

Datasets

We compare TCR-GIN with 28 state-of-the-art baseline methods (provided in Supplementary Note IV) on a broad range of synthetic networks and real-world networks. All datasets are stratified and split into training, validation, and test sets with a ratio of $8:1:1$. More details on dataset construction are provided in Supplementary Note V.

Data Availability

The datasets, TCR-GIN source code, trained model checkpoints, generated result files, and data underlying the figures and tables are available through Zenodo at `https://doi.org/10.5281/zenodo.20287619`.

Code Availability

The source code for TCR-GIN is publicly available at `https://github.com/DistinZhang/TCR-GIN`. The GitHub repository contains the implementation, configuration files, analysis scripts, and web-application source code. It does not include the large datasets, trained models, or generated result files, which are deposited in the Zenodo archive listed above. A lightweight web-based demonstration system for online collapse-warning analysis is available at `https://tcr-gin-early-warning.streamlit.app/`. Supplementary Movie 1 illustrates the use of this prototype for interactive collapse-warning analysis.

References

1. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).

2. Albert, R., Jeong, H. & Barabási, A.-L. Diameter of the world-wide web. *nature* **401**, 130–131 (1999).

3. Suweis, S., Simini, F., Banavar, J. R. & Maritan, A. Emergence of structural and dynamical properties of ecological mutualistic networks. *Nature* **500**, 449–452 (2013).

4. Barthelemy, M. The statistical physics of cities. *Nature Reviews Physics* **1**, 406–415 (2019).

5. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).

6. Newman, M. E. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).

7. Ghavasieh, A., Stella, M., Biamonte, J. & De Domenico, M. Unraveling the effects of multiscale network entanglement on empirical systems. *Communications Physics* **4**, 129 (2021).

8. Ghavasieh, A. & De Domenico, M. Diversity of information pathways drives sparsity in real-world networks. *Nature Physics* **20**, 512–519 (2024).

9. Zhang, Z., Ghavasieh, A., Zhang, J. & De Domenico, M. Coarse-graining network flow through statistical physics and machine learning. *Nature Communications* **16**, 1605 (2025).

10. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *nature* **406**, 378–382 (2000).

11. Liu, C. *et al.* Deep learning resilience inference for complex networked systems. *Nature Communications* **15** (2024).

12. Scagliarini, T., Artime, O. & De Domenico, M. Assessing the vulnerability of empirical infrastructure networks to natural catastrophes. *Chaos, Solitons & Fractals* **191**, 115813 (2025).

13. Bashan, A., Berezin, Y., Buldyrev, S. V. & Havlin, S. The extreme vulnerability of interdependent spatially embedded networks. *Nature Physics* **9**, 667–672 (2013).

14. Grassia, M., De Domenico, M. & Mangioni, G. Machine learning dismantling and early-warning signals of disintegration in complex systems. *Nature Communications* **12** (2021).

15. Lou, Y., Wang, L. & Chen, G. Structural robustness of complex networks: A survey of a posteriori measures [feature]. *IEEE Circuits and Systems Magazine* **23**, 12–35 (2023).

16. Freitas, S., Yang, D., Kumar, S., Tong, H. & Chau, D. H. Graph vulnerability and robustness: A survey. *IEEE Transactions on Knowledge and Data Engineering* **35**, 5915–5934 (2022).

17. Artime, O. *et al.* Robustness and resilience of complex networks. *Nature Reviews Physics* **6**, 114–131 (2024).

18. Liu, Z. *et al.* Early predictor for the onset of critical transitions in networked dynamical systems. *Physical Review X* **14** (2024).

19. Zhu, T. *et al.* A small set of critical hyper-motifs governs heterogeneous flow-weighted network resilience. *Nature Communications* **16** (2025).

20. Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Physical review letters* **85**, 4626 (2000).

21. Callaway, D. S., Newman, M. E., Strogatz, S. H. & Watts, D. J. Network robustness and fragility: Percolation on random graphs. *Physical review letters* **85**, 5468 (2000).

513 22. Jun, W., Barahona, M., Yue-Jin, T. & Hong-Zhong, D. Natural connectivity of complex
514 networks. *Chinese Physics Letters* **27**, 078902 (2010).

515 23. Achlioptas, D., D'souza, R. M. & Spencer, J. Explosive percolation in random networks.
516 *science* **323**, 1453–1455 (2009).

517 24. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of
518 failures in interdependent networks. *Nature* **464**, 1025–1028 (2010).

519 25. Yamashita, K., Yasuda, Y., Nakamura, R. & Ohsaki, H. Predictability of network robustness
520 from spectral measures. *Journal of Information Processing* **28**, 551–561 (2020).

521 26. Wu, C. *et al.* A multitask network robustness analysis system based on the graph isomorphism
522 network. *IEEE Trans. Cybern.* **54**, 6630–6642 (2024).

523 27. Lou, Y., Wang, L. & Chen, G. Toward robust network controllability: Insights and future
524 directions. *Europhysics Letters* **149**, 41003 (2025).

525 28. Schneider, C. M., Moreira, A. A., Andrade Jr, J. S., Havlin, S. & Herrmann, H. J. Mitigation
526 of malicious attacks on networks. *Proceedings of the National Academy of Sciences* **108**,
527 3838–3841 (2011).

528 29. Wu, C. *et al.* Spp-cnn: An efficient framework for network robustness prediction. *IEEE*
529 *Transactions on Circuits and Systems I: Regular Papers* **70**, 4067–4079 (2023).

530 30. Fan, C., Zeng, L., Sun, Y. & Liu, Y.-Y. Finding key players in complex networks through deep
531 reinforcement learning. *Nature Machine Intelligence* **2**, 317–324 (2020).

31. Engsig, M., Tejedor, A., Moreno, Y., Foufoula-Georgiou, E. & Kasmi, C. Domirank centrality reveals structural fragility of complex networks via node dominance. *Nature Communications* **15** (2024).

32. Xu, K., Hu, W., Leskovec, J. & Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).

33. Xie, X., Sun, Y., Liu, Y., Zhang, M. & Tan, K. C. Architecture augmentation for performance predictor via graph isomorphism. *IEEE Transactions on Cybernetics* **54**, 1828–1840 (2024).

34. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *science* **286**, 509–512 (1999).

35. Barabási, A.-L. Scale-free networks: A decade and beyond. *Science* **325**, 412–413 (2009).

36. Erdős, P. & Rényi, A. On the strength of connectedness of a random graph. *Acta Mathematica Hungarica* **12**, 261–267 (1961).

37. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *nature* **393**, 440–442 (1998).

38. Newman, M. & Watts, D. Renormalization group analysis of the small-world network model. *Physics Letters A* **263**, 341–346 (1999).

39. Lancichinetti, A., Fortunato, S. & Radicchi, F. Benchmark graphs for testing community detection algorithms. *Physical Review E* **78** (2008).

40. Zhao, J., Li, D., Sanhedrai, H., Cohen, R. & Havlin, S. Spatio-temporal propagation of cascading overload failures in spatially embedded networks. *Nature Communications* **7** (2016).

41. Yu, F., Guo, Q., Wu, J., Qiao, Z. & Sun, H. Early warning and proactive control strategies for power blackouts caused by gas network malfunctions. *Nature Communications* **15** (2024).

42. Clements, C. F., McCarthy, M. A. & Blanchard, J. L. Early warning signals of recovery in complex systems. *Nature Communications* **10** (2019).

43. Panahi, S., Do, Y., Hastings, A. & Lai, Y.-C. Rate-induced tipping in complex high-dimensional ecological networks. *Proceedings of the National Academy of Sciences* **120** (2023).

44. Liu, C., Ding, J., Song, Y. & Li, Y. Tdnetgen: Empowering complex network resilience prediction with generative augmentation of topology and dynamics. In Baeza-Yates, R. & Bonchi, F. (eds.) *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, 1875–1886 (ACM, 2024).

45. Lou, Y. *et al.* A learning convolutional neural network approach for network robustness prediction. *IEEE Transactions on Cybernetics* **53**, 4531–4544 (2023).

46. Lou, Y., Wu, C. & Chen, B.-Y. Exploring graph representations in machine learning for network robustness evaluation. *2024 International Joint Conference on Neural Networks (IJCNN)* 1–8 (2024).

47. Shargel, B., Sayama, H., Epstein, I. R. & Bar-Yam, Y. Optimization of robustness and connectivity in complex networks. *Physical Review Letters* **90** (2003).

48. Lou, Y., Wu, C., Li, J., Wang, L. & Chen, G. Network robustness prediction: Influence of training data distributions. *IEEE Transactions on Neural Networks and Learning Systems* **35**, 13496–13507 (2023).

49. Milisav, F., Bazinet, V., Betzel, R. F. & Misic, B. A simulated annealing algorithm for randomizing weighted networks. *Nature Computational Science* **5**, 48–64 (2025).

50. Zhu, B., Zhu, T., Gao, J., Havlin, S. & Li, D. Design of robust networks via reinforcement learning prompts the emergence of multi-backbones. *Nature Communications* (2026).

51. Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nature Reviews Physics* **3**, 422–440 (2021).

52. Wang, F., Zhai, Z., Zhao, Z., Di, Y. & Chen, X. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. *Nature Communications* **15**, 4332 (2024).

Acknowledgements This work is supported by the National Natural Science Foundation of China (72421002, 92467302, 72474223, 72301285), the National Science and Technology Major Project for Brain Science and Brain-like Intelligence Technology (2025ZD0215700), the Hunan Science and Technology Plan Project (2023JJ40685, 2024RC3133), and the Major Program of Xiangjiang Laboratory (24XJJCYJ01001).

Author Contributions Statement T.W. conceived and designed the project. B.C. and Z.Z. managed the project. J.Z., X.Z., and Y.J. performed the experiments. H.H. and B.Z. conducted the theoretical analysis. J.Z., T.W., M.D.D., and X.L. contributed to discussing the results and writing the manuscript.

Competing Interests Statement The authors declare no competing interests.

Correspondence Correspondence and requests for materials should be addressed to M.D.D. (manlio.dedomenico@unipd.it) and X.L. (xin.lu.lab@outlook.com).

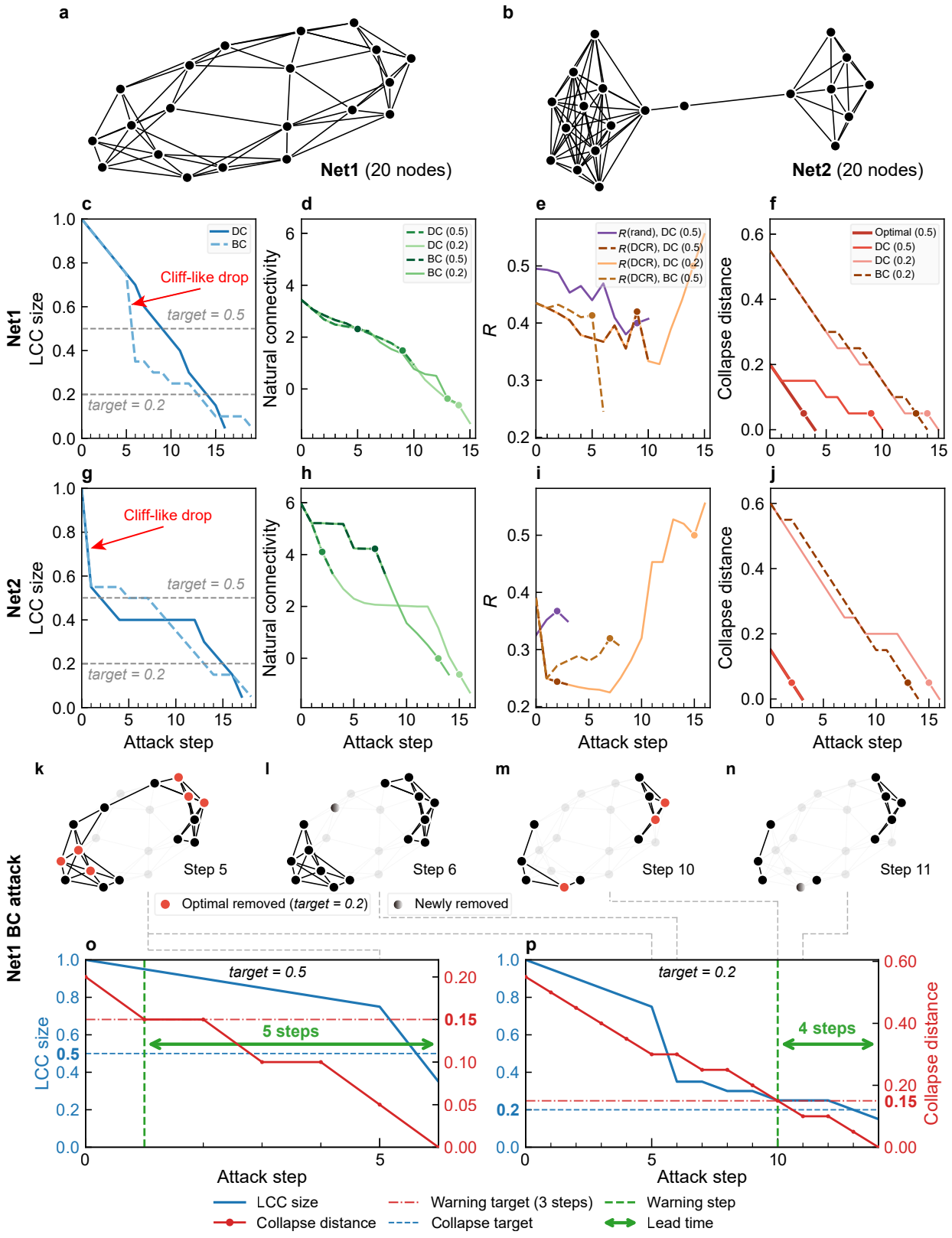

a
Net1 (20 nodes)
b
Net2 (20 nodes)
c
d
e
f
Net1
LCC size
DC
BC
Cliff-like drop
target = 0.5
target = 0.2
Natural connectivity
DC (0.5)
DC (0.2)
BC (0.5)
BC (0.2)
R
R(rand), DC (0.5)
R(DCR), DC (0.5)
R(DCR), DC (0.2)
R(DCR), BC (0.5)
Collapse distance
Optimal (0.5)
DC (0.5)
DC (0.2)
BC (0.2)
g
h
i
j
Net2
Attack step
k
l
m
n
Step 5
Step 6
Step 10
Step 11
Net1 BC attack
Optimal removed (target = 0.2)
Newly removed
o
p
5 steps
4 steps
LCC size
Collapse distance
Warning target (3 steps)
Warning step
Collapse target
Lead time

Figure 1 Toy examples comparing robustness metrics and illustrating how collapse distance supports configurable early warning. **(a, b)** Two toy networks with 20 nodes, Net1 and Net2. **(c, g)** The LCC size may exhibit hard-to-anticipate cliff-like drops under attack. **(d, h)** Natural connectivity decreases monotonically but remains snapshot-based and independent of the collapse target τ. Consequently, within the same attack scenario, curves for different τ overlap and thus do not reflect a target-oriented remaining safety margin. **(e, i)** The online evolution of the R value depends on the assumed simulation strategy (e.g., random dismantling $R(\mathrm{rand})$ versus iterative degree-centrality dismantling $R(\mathrm{DCR})$), producing strategy-dependent trajectories with level drift and local rebounds. **(f, j)** Under sustained attacks, CD exhibits a monotonically non-increasing, step-like descent. The per-step change is restricted to $\{0,\ 1/N_0\}$ (where N_0 is the initial network size), rendering the variation predictable and precluding cliff-like jumps. Open circles in **(d–f, h–j)** mark the last pre-collapse snapshot, that is, the final step before the LCC size reaches or falls below τ. At this state, natural connectivity and the R-based measures occupy different numerical ranges across networks, attack scenarios, and collapse targets, whereas CD reaches $1/N_0$, corresponding to one additional node removal before collapse. **(k–n)** Remaining-network snapshots of Net1 under betweenness centrality (BC) attack. Red circles denote the optimal collapse-inducing node set at the given step (for $\tau = 0.2$; steps 5, 6, 10, and 11). These snapshots elucidate two asynchronous phenomena between LCC size and CD. At step 6, the LCC size drops while CD remains constant. The removed node does not belong to the previous optimal set, thus not compressing the remaining safety margin along the shortest collapse path. At step 11, the LCC size remains unchanged while

CD drops because the removed node belongs to the previous optimal set and shortens the minimum path to collapse. **(o, p)** Translation of CD into a configurable early-warning signal for the same Net1 BC attack under two collapse targets, with **(o)** $\tau = 0.5$ and **(p)** $\tau = 0.2$. The red horizontal dash-dotted line marks the warning target n/N_0 with $n = 3$. Under the one-node-per-step setting, this corresponds to 3 attack steps. The warning step is the first step at which CD reaches or falls below the warning target, and the collapse step is the first step at which LCC size reaches or falls below τ. The interval between them defines the observed lead time, yielding 5 attack steps for $\tau = 0.5$ and 4 attack steps for $\tau = 0.2$.

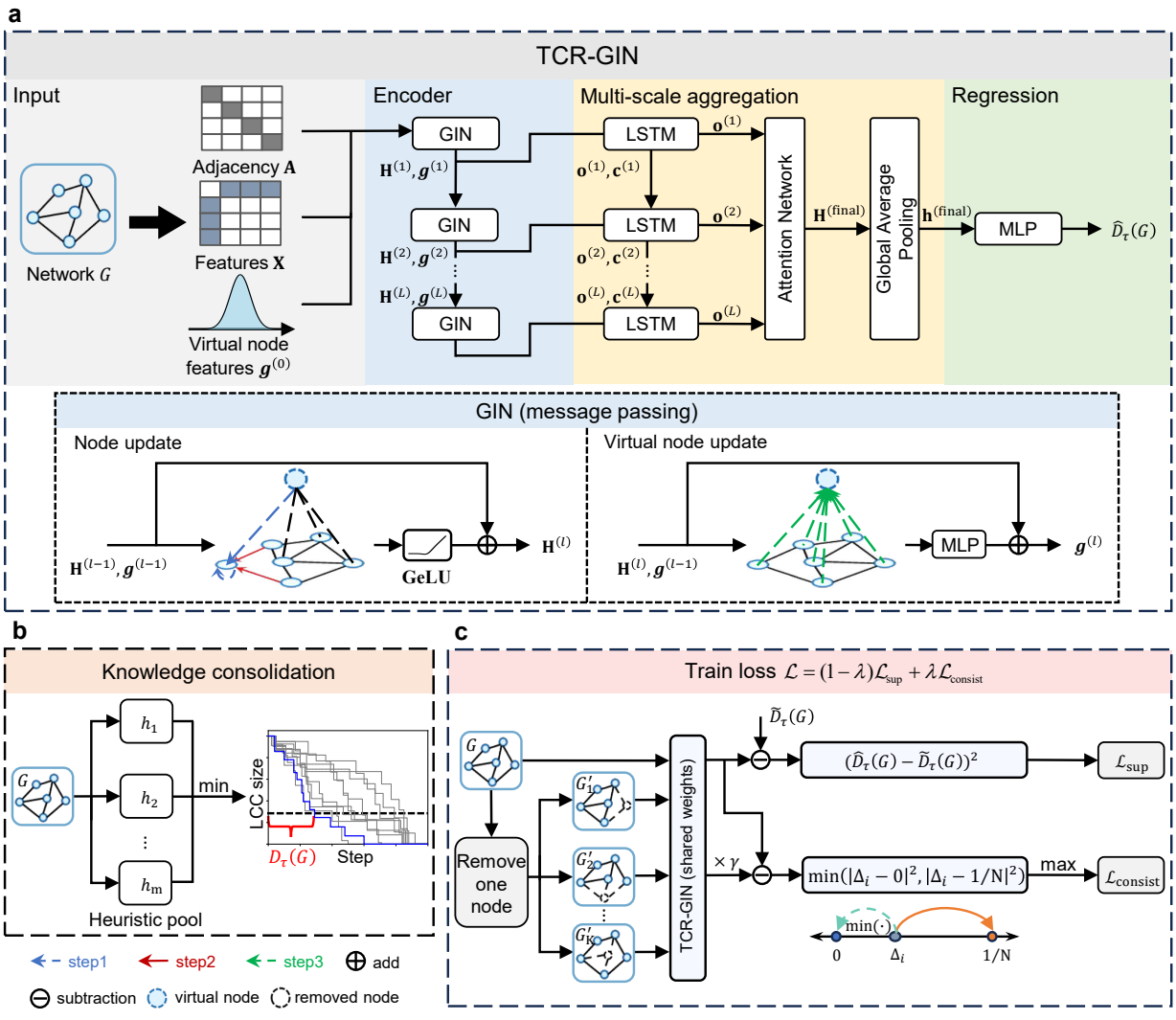

a
TCR-GIN
Input
Network G
Adjacency $\mathbf{A}$
Features $\mathbf{X}$
Virtual node features $\boldsymbol{g}^{(0)}$
Encoder
GIN
GIN
GIN
$\mathbf{H}^{(1)}, \boldsymbol{g}^{(1)}$
$\mathbf{H}^{(2)}, \boldsymbol{g}^{(2)}$
$\mathbf{H}^{(L)}, \boldsymbol{g}^{(L)}$
Multi-scale aggregation
LSTM
LSTM
LSTM
$\mathbf{o}^{(1)}, \mathbf{c}^{(1)}$
$\mathbf{o}^{(2)}, \mathbf{c}^{(2)}$
$\mathbf{o}^{(L)}, \mathbf{c}^{(L)}$
$\mathbf{o}^{(1)}$
$\mathbf{o}^{(2)}$
$\mathbf{o}^{(L)}$
Attention Network
$\mathbf{H}^{(\text{final})}$
Global Average Pooling
$\mathbf{h}^{(\text{final})}$
Regression
MLP
$\widehat{D}_\tau(G)$
GIN (message passing)
Node update
$\mathbf{H}^{(l-1)}, \boldsymbol{g}^{(l-1)}$
GeLU
$\mathbf{H}^{(l)}$
Virtual node update
$\mathbf{H}^{(l)}, \boldsymbol{g}^{(l-1)}$
MLP
$\boldsymbol{g}^{(l)}$
b
Knowledge consolidation
h_1
h_2
h_m
min
LCC size
$D_\tau(G)$
Step
Heuristic pool
step1
step2
step3
add
subtraction
virtual node
removed node
c
Train loss $\mathcal{L} = (1-\lambda)\mathcal{L}_{\text{sup}} + \lambda\mathcal{L}_{\text{consist}}$
$\widetilde{D}_\tau(G)$
$(\widehat{D}_\tau(G) - \widetilde{D}_\tau(G))^2$
$\mathcal{L}_{\text{sup}}$
Remove one node
TCR-GIN (shared weights)
$\times\gamma$
$\min(|\Delta_i - 0|^2, |\Delta_i - 1/\mathrm{N}|^2)$
max
$\mathcal{L}_{\text{consist}}$
min(·)
0
Δ_i
1/N

Figure 2 TCR-GIN framework and training process. (a) Network architecture. Inputs include the graph topology, node structural features (degree, clustering coefficient, and k-core number), and a learnable virtual node initialized from a standard normal distribution. A GIN encoder with residual connections produces layer-wise node and virtual-node embeddings. An LSTM-based multi-scale aggregation module, together with an attention network, fuses representations across depths, and an MLP regressor maps the globally pooled graph embedding vector to the predicted collapse distance $\hat{D}_\tau(G)$. **(b) Knowledge consolidation.** When exact labels are infeasible to compute at scale, we run a set of dismantling algorithms $(h_1, \ldots, h_m)$ and use the minimum estimated collapse distance as the observed label. **(c) Train loss.** The loss combines a supervised term $\mathcal{L}_{\text{sup}}$ with a physics-inspired consistency regularizer $\mathcal{L}_{\text{con}}$. Based on Theorem S1, $\mathcal{L}_{\text{con}}$ constrains the scale-corrected prediction change Δ_i under single-node perturbation to the discrete set $\{0, 1/N\}$, where N is the number of nodes.

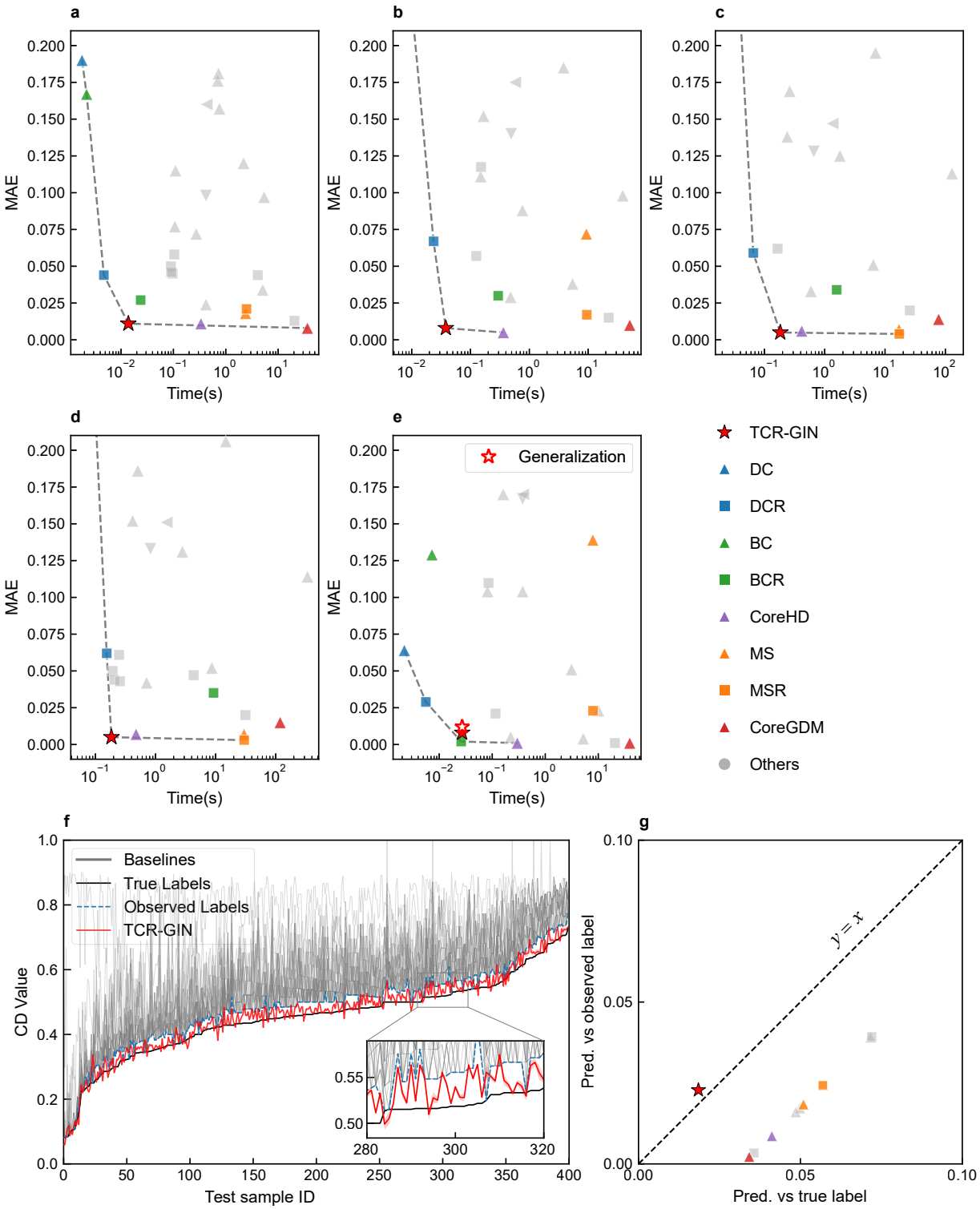

a
b
c
d
e
f
g
MAE
Time(s)
Generalization
TCR-GIN
DC
DCR
BC
BCR
CoreHD
MS
MSR
CoreGDM
Others
Baselines
True Labels
Observed Labels
TCR-GIN
CD Value
Test sample ID
Pred. vs observed label
Pred. vs true label
y = x

Figure 3 Accuracy–efficiency trade-off for collapse distance prediction. (a–d) Accuracy–efficiency scatter plots for four size ranges (Small/Medium/Large/Extra-large) on synthetic networks. The x-axis is the mean end-to-end inference time per graph (seconds, log scale), including data/model loading, feature construction, and model inference. The y-axis is the MAE with respect to the observed labels. All runtimes in **(a–e)** were measured on CPU (Intel Xeon Platinum 8470Q). For clarity of presentation, we show only mean MAE values in the figure. Standard deviations for results obtained over repeated runs are omitted because they are too small to visualize clearly at the plotting scale. Full numerical results are provided in Table S1. For clarity, some baseline results exceeding the y-axis limit are omitted. TCR-GIN (red solid star) consistently lies in the low-error, low-latency region across all scales. **(e)** Accuracy–efficiency scatter plot on the real-world REDDIT dataset. The red solid star reports results when training and testing on the same dataset, while the red open star denotes cross-distribution generalization, i.e., training on synthetic networks and testing on REDDIT. **(f)** Ground-truth analysis on small graphs (true labels obtained by exhaustive search). Grey curves correspond to baselines; the blue dashed curve shows observed labels; the black curve shows true labels; and the red curve shows TCR-GIN predictions trained on observed labels. The inset highlights that TCR-GIN predictions can fall below the observed labels in a local range, thereby being closer to the true labels. **(g)** Scatter plot comparing TL-MAE and OL-MAE. The x-axis is MAE with respect to true labels and the y-axis is MAE with respect to observed labels; the dashed line indicates $y = x$. Baselines may achieve a small observed-label MAE but a much larger true-label MAE, whereas TCR-GIN attains

better performance under the TL-MAE criterion.

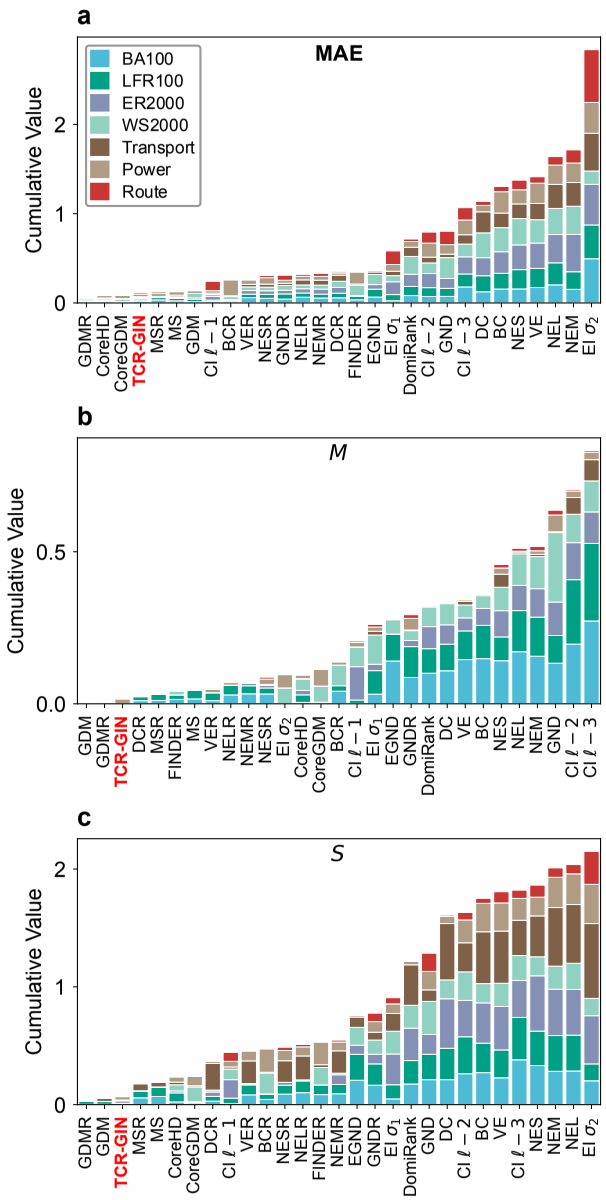

a
MAE
BA100
LFR100
ER2000
WS2000
Transport
Power
Route
Cumulative Value
0
1
2
GDMR
CoreHD
CoreGDM
TCR-GIN
MSR
MS
GDM
CI ℓ − 1
BCR
VER
NESR
GNDR
NELR
NEMR
DCR
FINDER
EGND
EI σ1
DomiRank
CI ℓ − 2
GND
CI ℓ − 3
DC
BC
NES
VE
NEL
NEM
EI σ2
b
M
Cumulative Value
0.0
0.5
GDM
GDMR
TCR-GIN
DCR
MSR
FINDER
MS
VER
NELR
NEMR
NESR
EI σ2
CoreHD
CoreGDM
BCR
CI ℓ − 1
EI σ1
EGND
GNDR
DomiRank
DC
VE
BC
NES
NEL
NEM
GND
CI ℓ − 2
CI ℓ − 3
c
S
Cumulative Value
0
1
2
GDMR
GDM
TCR-GIN
MSR
MS
CoreHD
CoreGDM
DCR
CI ℓ − 1
VER
BCR
NESR
NELR
FINDER
NEMR
EGND
GNDR
EI σ1
DomiRank
GND
DC
CI ℓ − 2
BC
VE
CI ℓ − 3
NES
NEM
NEL
EI σ2

Figure 4 Cross-network and cross-attack comparison of algorithmic performance under sustained attacks. **(a)** MAE; **(b)** mean frequency of monotonicity violations (M); **(c)** frequency of smoothness violations (S). Results cover two 100-node synthetic networks (BA100 and LFR100), two 2000-node synthetic networks (ER2000 and WS2000), and three real-world networks (Transport, Power, and Route), averaged over 30 attack scenarios. For each algorithm, the bar height equals the cumulative sum of the corresponding metric over all datasets, and stacked segments indicate each dataset's contribution to the total. In each panel, the algorithms on the x-axis are ordered by the cumulative value of that metric, from left to right (better to worse). For Route, the three metrics are computed only after the largest connected component size enters a range close to the covered training distribution (see Supplementary Note IX for details). For the other networks, metrics are computed over the full trajectory. Error bars are omitted. The related repeated-run variability of TCR-GIN is reported in Table S5.

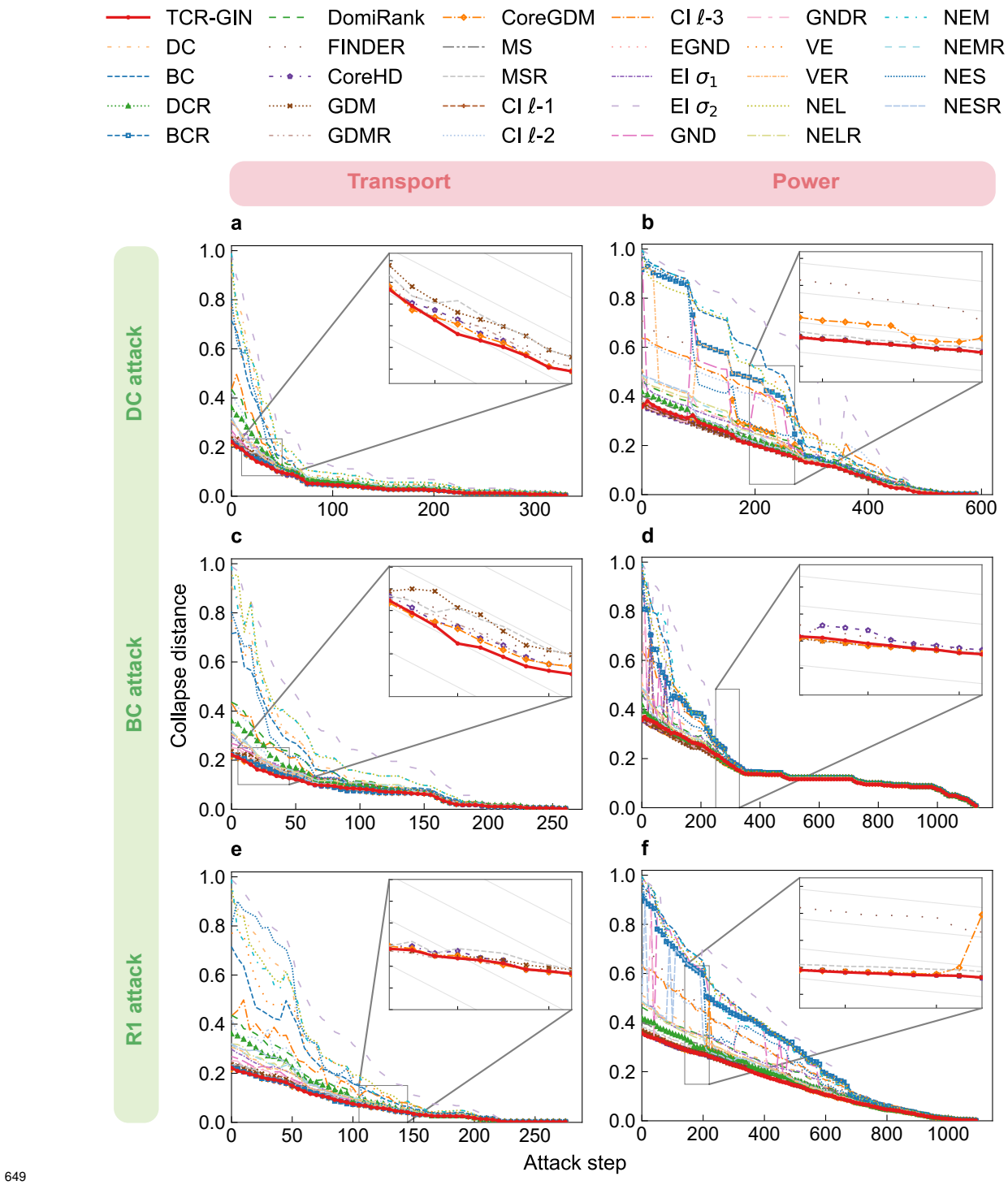

TCR-GIN
DC
BC
DCR
BCR
DomiRank
FINDER
CoreHD
GDM
GDMR
CoreGDM
MS
MSR
CI ℓ-1
CI ℓ-2
CI ℓ-3
EGND
EI σ1
EI σ2
GND
GNDR
VE
VER
NEL
NELR
NEM
NEMR
NES
NESR
Transport
Power
DC attack
BC attack
R1 attack
a
b
c
d
e
f
Collapse distance
Attack step

Figure 5 Online collapse-distance trajectories of real-world networks under representative sustained attacks. (a, c and e) Transport network and **(b, d and f)** Power network. For each network, results are reported for degree-based, betweenness-based, and random attacks. Grey curves represent the collapse-distance trajectories given by baseline algorithms, and red curves represent the trajectories predicted by TCR-GIN. Insets show a magnified view of the low-error region. The insets include only TCR-GIN and six advanced baselines, including GDMR, GDM, CoreGDM, CoreHD, MSR, and FINDER. The reference grid in each inset has a slope of $-1/N_0$, where N_0 is the initial number of nodes used for sequence normalization, visualizing the theoretical upper bound of the single-step change in collapse distance under one-node removal. Baseline algorithms show substantial disagreement on real networks, whereas TCR-GIN remains continuous, exhibits an almost monotonic decrease with controlled jitter, and broadly follows the lower envelope of the baseline set, yielding a more stable estimate of collapse distance.

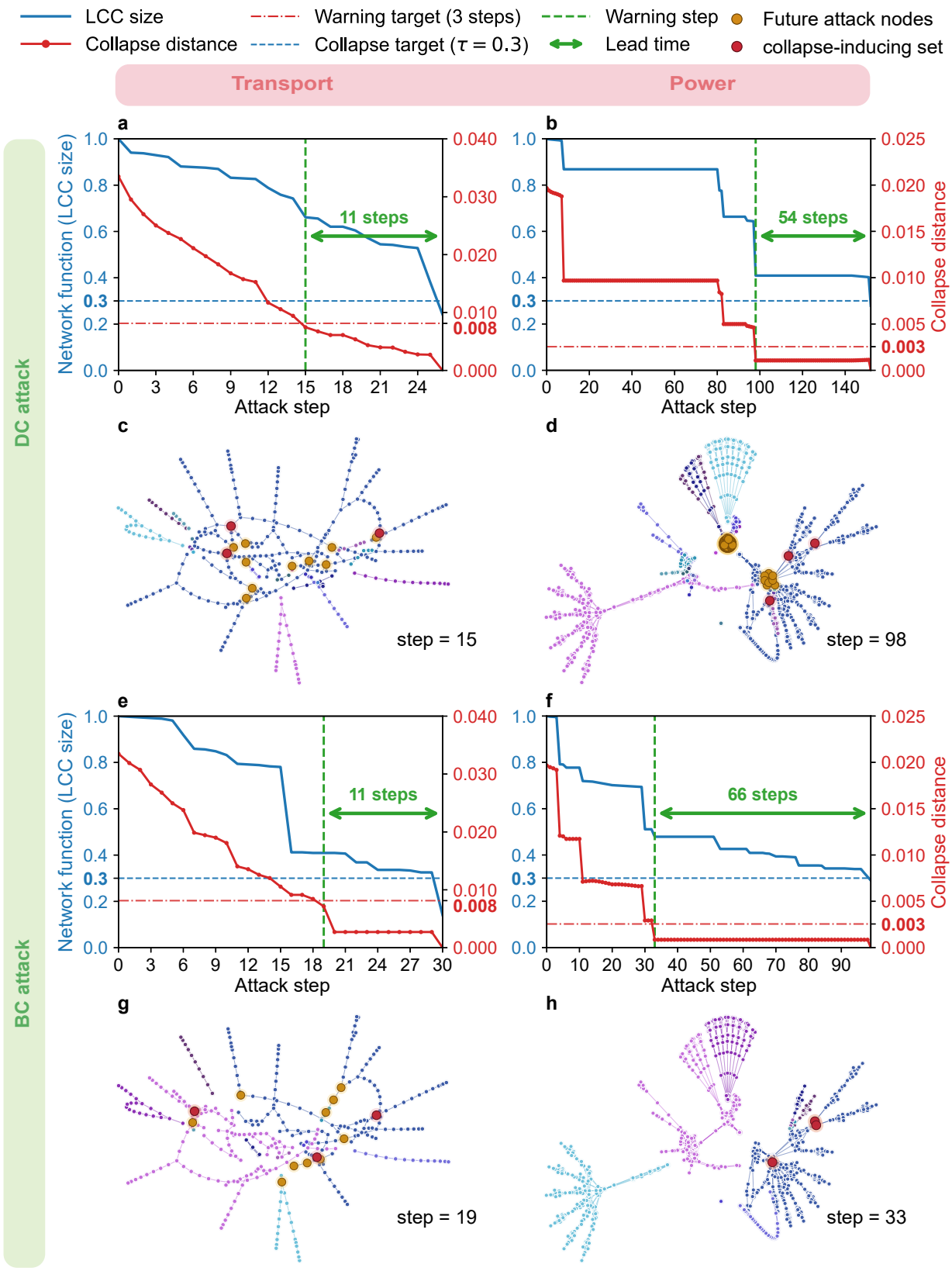

LCC size
Collapse distance
Warning target (3 steps)
Collapse target (τ = 0.3)
Warning step
Lead time
Future attack nodes
collapse-inducing set
Transport
Power
DC attack
BC attack
a
b
c
d
e
f
g
h
Network function (LCC size)
Collapse distance
Attack step
11 steps
54 steps
66 steps
step = 15
step = 98
step = 19
step = 33

Figure 6 Early-warning signal of real-world networks under representative sustained attacks. (a, b) Degree-based attacks and **(e, f)** betweenness-based attacks on the Transport (**a, e**) and Power (**b, f**) networks. Blue curves show network functionality measured by LCC size, and red curves show collapse distance estimated online by TCR-GIN. The blue horizontal dashed line marks the collapse target ($\tau = 0.3$), and the red horizontal dash-dotted line marks the warning target n/N_0 with $n = 3$ under the one-node-per-step setting. Green vertical dashed lines indicate the warning step, and green double-headed arrows indicate the observed lead time. **(c, d, g, and h)** Remaining network snapshots at the corresponding warning steps. Red nodes denote a collapse-inducing set of size n obtained by exhaustive search at the warning snapshot, whereas orange nodes denote nodes subsequently removed by the realized attack sequence before collapse. The mismatch between the red and orange nodes illustrates why the realized lead time can exceed the requested warning horizon.

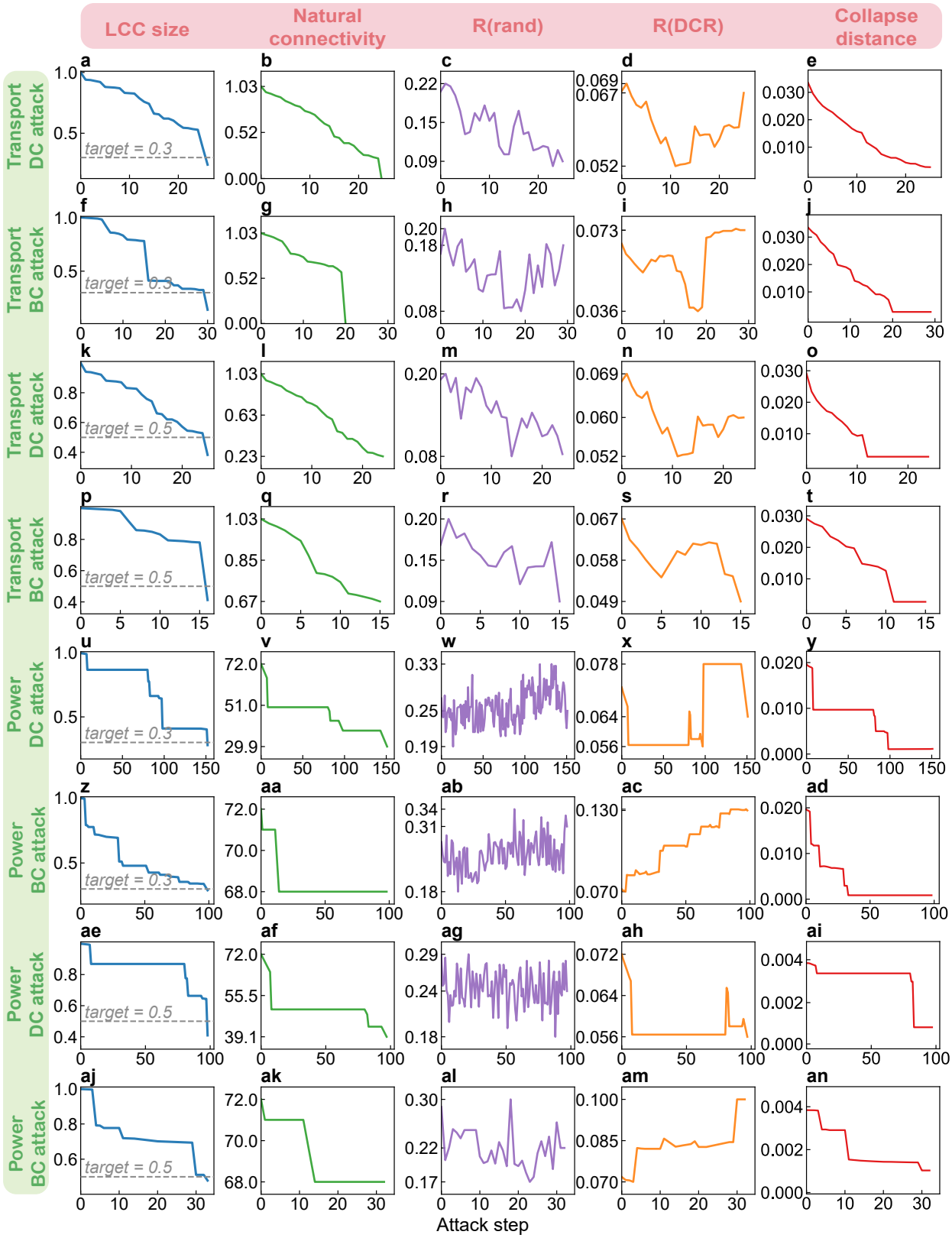

LCC size
Natural connectivity
R(rand)
R(DCR)
Collapse distance
Transport DC attack
Transport BC attack
Transport DC attack
Transport BC attack
Power DC attack
Power BC attack
Power DC attack
Power BC attack
target = 0.3
target = 0.5
Attack step

Figure 7 Online degradation trajectories of different robustness metrics under representative sustained attacks. From top to bottom, we report the online trajectories of the Transport and Power networks, under two collapse targets ($\tau = 0.3$ and $\tau = 0.5$), subjected to two targeted attacks (DC and BC). From left to right, each panel sequentially presents the LCC size (the dashed line indicates τ), natural connectivity $\bar{\lambda}$, $R(\mathrm{rand})$, $R(\mathrm{DCR})$, and the collapse distance estimated by TCR-GIN, all as functions of the attack step. The LCC trajectory is shown through the collapse step, whereas the other metrics are plotted only up to the last pre-collapse snapshot. Robustness R is defined as the area under the robustness curve $S(p)$. For the online trajectory, we update R at each step by regarding the remaining nodes as a new network and simulating its dismantling process using a specified heuristic (random or DCR). Natural connectivity decreases monotonically, but its absolute scale varies markedly across networks and shifts with τ and the attack strategy. The R-based trajectories exhibit setting-dependent level shifts and visible fluctuations. In contrast, the TCR-GIN-estimated collapse distance remains within a more consistent range near the collapse across different scenarios. For visualization on a common reference scale, whenever a metric definition involves network size (e.g., normalization by N), we use the initial network size N_0 as a fixed reference (instead of the remaining size N_t).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMovieStepbystepearlywarningofnetworkbreakdownwithcollapsedistance.mp4
- SupplementaryInformation.pdf