Information-Theoretic and Algorithmic Thresholds for Network Attack Detection via Spectral and CUSUM Methods

Abdulkader Hajjouz, Elena Avksentieva
ITMO University, Faculty of Software Engineering and Computer Technology

Keywords: Network anomaly detection, planted dense subgraph, non-backtracking spectrum, Hawkes processes, sequential change detection, information-theoretic limits, CUSUM, statistical–computational gap

Received: October 3, 2025

We develop a unified theory for the detectability of network-borne attacks under two canonical observation models: (i) a static graph drawn from an Erdős–Rényi background with a planted anomalous community, and (ii) a temporal interaction network modeled by multivariate point processes (Poisson or Hawkes). Our main contribution is to match, up to universal constants, information-theoretic lower and upper bounds that govern when reliable testing is possible. In the static case, the core quantity is the accumulated edgewise signal $k^2 \cdot \chi^2(Bern(p+\Delta) \| Bern(p))$, where $\chi^2 \approx \Delta^2/[p(1-p)]$ for small Δ; detection is impossible when this falls below $c \cdot \log n$, and a non-backtracking spectral statistic succeeds above $C \cdot \log n$. In the temporal case, detectability is controlled by the Kullback–Leibler information rate I contributed by internal edges over a window of length T, yielding a threshold $T\, I \gtrsim \log n$; a likelihood-based cumulative-sum (CUSUM) test achieves first-order optimal delay $\approx |\log \alpha|/I$ at false-alarm level α. We complement these limits with concrete algorithms and simple experiments. A pruning-based non-backtracking spectral detector for static graphs and a sparsity-aware CUSUM procedure for temporal streams are given with near-linear time and memory complexity. Monte Carlo simulations on Erdős–Rényi graphs with planted dense subgraphs and on Poisson streams illustrate the predicted phase transitions: detection power increases as $k^2 \cdot \chi^2/\log n$ crosses a constant-level boundary, and empirical detection delays closely track the log(ARL)/I prediction. We also discuss robustness under bounded edge perturbations and mild misspecification, and show how these thresholds can be used to dimension practical security monitoring systems.

Povzetek: Predstavimo splošno teorijo, kdaj je mogoče zanesljivo zaznati napade v statičnih in časovnih omrežjih, ter podamo učinkovite algoritme in simulacije, ki potrjujejo napovedane pragove zaznavanja.

1 Introduction

Network intrusion signals often manifest as structural and/or temporal irregularities: unusually dense communication among a small set of hosts, or bursts of activity confined to a subset of edges. Decades of work in community detection and sequential analysis have produced powerful tools, yet a consolidated view of detectability thresholds tailored to adversarial network activity remains incomplete. Our aim is to align information-theoretic limits with algorithms that nearly attain them, bringing together recent spectral advances—most notably Bethe–Hessian and non-backtracking operators—and modern sequential detection methods for streaming data [1–3]. This synthesis yields simple, actionable criteria that demarcate when intrusions are detectable at all and when they are detectable efficiently.

We formalize the detectability problem in two complementary settings. In a static snapshot, an adversary targets an unknown set S of k nodes (or their internal edges), inducing a small distributional shift on the $\Theta(k^2)$ edges internal to S. Each perturbed edge contributes a modest statistical signal, conveniently summarized—under canonical Bernoulli graph models—by the per-edge χ^2-divergence between baseline and attacked edge laws. Aggregating across internal edges, the usable evidence scales as $k^2 \cdot \chi^2$. The unknown location of S, however, imposes a combinatorial search penalty that, at the right scale, behaves like log n. A sharp balance emerges: detection is information-theoretically feasible (up to constants) precisely when

$$k^2 \cdot \chi^2 \gtrsim \log n$$

In temporal networks, information accrues over time along affected edges. Let I denote the per-time Kullback–Leibler information rate (e.g., under Poisson or Hawkes intensities), and let T be the observation horizon. The analogous boundary becomes

$$T\, I \gtrsim \log n$$

These two templates capture a common principle: accumulated information must exceed search complexity. They also reveal transparent trade-offs among effect size, attack scope k, network size n, and observation time T.

A second theme is the statistical–computational gap: the tension between what is detectable in principle and what is achievable in polynomial time. Even when $k^2·\chi^2$ (or T I) clears log n, some regimes remain algorithmically delicate due to sparsity, weak signals, or adversarial camouflage. Recent spectral ideas mitigate these challenges. Statistics derived from the non-backtracking operator—and the closely related Bethe–Hessian—suppress high-variance backtracking walks that plague adjacency-based spectra in sparse graphs. When tuned to emphasize energy localized on candidate substructures, these methods can succeed within constant factors of the information limits in regimes where classical Laplacian or adjacency eigenvectors fail [1–3]. On the streaming side, likelihood-based CUSUM procedures convert information-rate considerations into operational guarantees: for a prescribed false-alarm level α, detection delay concentrates around |log α|/I, matching first-order lower bounds and thus tracking the T I ≳ log n threshold up to constants [1–3].

Beyond limits and algorithms, robustness is crucial in adversarial environments. We quantify how thresholds deform under bounded perturbations of the edge set (e.g., a small adversarial rewiring budget) and under mild model misspecification (e.g., modest heterogeneity in baseline edge probabilities or intensity drift). In our analysis, the net effect is a constant-factor inflation of the requisite $k^2·\chi^2$ (static) or T I (temporal), preserving the qualitative location of the thresholds. Practically, this means that light-touch regularization—degree normalization for spectra, variance control for scan statistics, and window-limited CUSUM for nonstationary streams—suffices to retain near-optimal performance.

The resulting prescriptions are easy to operationalize. In a static Erdős–Rényi baseline with edge probability p, the smallest detectable additive lift Δ on internal edges satisfies

$$\Delta_{\min} \approx \sqrt{}(p(1-p)\cdot\log n \,/\, k^2)$$

so enlarging the candidate scope k or operating at moderate p sharply reduces the needed effect size. In streaming telemetry, choosing a horizon T and designing per-edge weighting to secure I such that T I ≫ log n yields prompt detection with delay on the order of |log α|/I. These rules-of-thumb translate directly into SIEM thresholds, scan window lengths, and prioritization of edges for monitoring.

Contributions. (1) Unified thresholds. We derive sharp (up to universal constants) lower and upper bounds for static snapshots and streaming networks, expressed compactly through $k^2·\chi^2$ and T I. These formulas expose the precise trade-offs among k, effect size, horizon T, and network size n. (2) Efficient tests near the threshold. For static graphs, we propose a non-backtracking spectral statistic that localizes anomaly energy with minimal backtracking noise; for temporal data, we analyze a likelihood-based CUSUM (and window-limited variants) whose delay matches first-order lower bounds. (3) Statistical–computational trade-offs. We delineate regimes where detection is statistically possible yet plausibly hard for polynomial-time algorithms, clarifying how our procedures approach the boundary and where they may (conditionally) fall short. (4) Robustness. We show that bounded edge perturbations and mild model misspecification inflate thresholds only by constants, and we provide practical regularization guidelines to preserve performance. (5) Algorithms and empirical validation. We give simple algorithmic descriptions for both the non-backtracking spectral test and the temporal CUSUM detector, and we provide small-scale simulations that illustrate the predicted thresholds in static and temporal settings.

Paper organization. Section 2 reviews related work and unifies notation. Section 3 formalizes models and hypotheses for static and temporal settings. Section 4 establishes information-theoretic lower bounds. Section 5 presents our algorithms and proves achievability near the thresholds. Section 6 reports empirical evaluations on synthetic static and temporal networks. Section 7 discusses computational barriers and robustness. Section 8 offers implementation notes and practical guidance, followed by conclusions.

2 Related work

Static graphs. A large body of work has sought principled spectral tools for detecting small, dense, or otherwise structured anomalies in sparse graphs. Building on insights near the Kesten–Stigum (KS) threshold in stochastic block models, recent analyses justify the superiority of operators that suppress backtracking walks—most notably the non-backtracking (Hashimoto) operator and its closely related Bethe–Hessian counterpart—over classical adjacency or Laplacian spectra in the sparse regime [1,2,4,5]. The non-backtracking linearization stabilizes leading eigenvectors when degrees are heterogeneous or the signal is faint, and the Bethe–Hessian furnishes a symmetric, well-conditioned surrogate whose eigenstructure tracks KS-type phase transitions. These methods have been adapted beyond pairwise graphs: extensions to hypergraphs and higher-order interactions leverage non-backtracking tensors or suitably defined incidence operators, enabling detection of dense sub-hypergraphs and higher-arity motifs when pairwise projections are too lossy [4,5]. Parallel developments examine local algorithms, message passing, and belief propagation approximations that effectively implement spectral heuristics with improved robustness in finite samples.

Despite these advances, dense-substructure detection crystallizes a pronounced statistical–computational gap. Information-theoretically, subgraph lifts on Θ(k2) internal edges can be detectable at vanishing signal levels; yet polynomial-time algorithms often require larger effect sizes. Post-2021 work clarified this gap using reductions from planted clique and via low-degree polynomial lower bounds, which capture the power of a broad class of efficient tests and show sharp thresholds for planted dense subgraph and planted hypergraph problems [6,7]. These results delineate regimes where spectral scans or convex relaxations are provably suboptimal, and they motivate specialized procedures—degree-normalized embeddings, iterative pruning, or localized power iterations—that push performance toward the information limit without breaching conjectured hardness frontiers. Additional strands explore subspace-tracking and sketching for massive graphs, balancing sensitivity with near-linear time and memory.

Temporal/streaming data. In dynamic settings, the literature on quickest change-point detection connects information rates to detection delay and false-alarm control. Classical likelihood-ratio and CUSUM procedures have been adapted to high-dimensional network streams, where signals concentrate on a small, unknown subset of edges or nodes. Recent work emphasizes computational–statistical trade-offs, proposing window-limited or sparsity-aware CUSUM variants, kernelized detectors that capture nonlinear deviations, and neural change-detectors that approximate likelihoods when models are misspecified [3,8–10]. For event-driven telemetry (e.g., syslog, authentication, or flow records) modeled as Poisson or Hawkes processes, edge-wise intensities encode both exogenous rates and endogenous excitation; here, likelihood-based sequential tests translate the per-time Kullback–Leibler information I into first-order optimal delay

$$|\log \alpha| \,/\, I$$

at false-alarm level α, and suggest principled aggregation schemes across candidate edges [3,8–10]. Practical considerations include restart policies under nonstationarity, adaptive thresholding to track seasonal baselines, and partial observability when only sampled flows are available.

A complementary line of research surveys graph anomaly detection across static snapshots and time series, cataloging benchmarks, feature constructions (e.g., egonet statistics, motif counts, temporal degree deviations), and evaluation protocols [11–13]. These surveys stress the importance of robustness to benign workload shifts and of interpretability—surfacing which nodes, edges, or motifs drive alarms—both of which inform our emphasis on localized spectral energy and per-edge likelihood attribution. In parallel, dynamic-network change-point and temporal community detection have progressed through Laplacian/subspace methods and multi-view formulations that pool information across time while preserving short-term sensitivity [14–16]. Subspace tracking on streaming Laplacians detects departures from nominal low-rank structure, whereas multi-snapshot embeddings (e.g., coupled matrix/tensor factorizations) stabilize community estimates under rapid churn. These tools can be paired with sequential tests to provide early-warning signals without frequent false alarms.

Synthesis and position. The above threads converge on two themes central to our study. First, spectral regularization—via non-backtracking and Bethe–Hessian operators—provides stable, high-power statistics for sparse graphs, and analogous likelihood-based sequential detectors operationalize information-rate limits in streams [1–5,8–10]. Second, a persistent statistical–computational tension shapes achievable performance: low-degree and planted-model lower bounds chart regions where efficient algorithms cannot match information-theoretic benchmarks, especially for small, covert substructures [6,7]. Our contribution is to place these developments under a unified detectability-threshold lens—$k^2 \cdot \chi^2$ for static snapshots and $T\ I$ for temporal streams—while providing efficient procedures that operate near these boundaries and remain robust under mild misspecification and bounded perturbations, in line with the desiderata surfaced by recent surveys and dynamic-network methods [11–16].

Summary of related methods. Table 1 summarizes representative methods along four axes: the information quantity driving detectability, computational complexity, robustness, and key assumptions.

Table 1: Representative methods for network anomaly detection and their characteristics.

Method & Setting	**Technical Specs (Thresholds & Complexity)**	**Remarks**
Non-backtracking / Bethe–Hessian [1,2,4,5] (Static graphs)	Threshold: KS-type thresholds; χ^2-based SNR Complexity: $O(\lvert E\rvert)$ per power iteration	Robust to sparsity; sensitive to degree heterogeneity without normalization.
Low-degree polynomial tests [6,7] (Static graphs)	Threshold: Info vs. low-degree computational thresholds Complexity: Theoretical, often super-linear	Chart statistical–computational frontiers.
CUSUM for Poisson/Hawkes [3,8–10,19,20] (Temporal streams)	Threshold: Info rate I; delay $\approx \lvert\log \alpha\rvert/I$ Complexity: Near-linear in events	Classical quickest change-point detectors.
GNN-based anomaly detectors [11–13] (Static & temporal)	Threshold: Data-driven; no closed-form threshold Complexity: Training cost $O(L\lvert E\rvert d)$	High flexibility, require labelled data.
This paper (Static & temporal)	Threshold: $k^2\chi^2 \gtrsim \log n$ (static) $T\ I \gtrsim \log n$ (temporal) Complexity: $O(\lvert E\rvert)$ (static) Linear in events (temporal)	Near-optimal thresholds; simple, interpretable.

3 Problem setup

3.1 Notation and divergences

We write $[n] = \{1, \ldots, n\}$for the vertex set, and use boldface capitals (e.g., **A**) for matrices. For two probability laws $P$and Qon a common space, the Kullback–Leibler divergence is

$$D_{\mathrm{KL}}(P \parallel Q) = \int \log\left(\frac{dP}{dQ}\right) dP,$$

and the χ^2-divergence is

$$\chi^2(P \parallel Q) = \int \left(\frac{dP}{dQ} - 1\right)^2 dQ.$$

In our models, Pcorresponds to the attacked (alternative) distribution and $Q$to the baseline (null). For Bernoulli edges with probabilities $p$and $p + \Delta$, we have

$$\chi^2(\mathrm{Bern}(p+\Delta) \parallel \mathrm{Bern}(p)) = \frac{\Delta^2}{p(1-p)} \text{for } 0 < p < 1,$$

and for Poisson rates $\mu$and $\mu + \delta$we use

$$D_{\mathrm{KL}}(\mathrm{Pois}(\mu+\delta) \parallel \mathrm{Pois}(\mu)) = (\mu+\delta)\log\left(1 + \frac{\delta}{\mu}\right) - \delta.$$

We will repeatedly aggregate such per-edge divergences over $\Theta(k^2)$internal edges of an unknown attacked set $S \subset [n]$with $|S| = k$, and over a time horizon of length T, leading to the information quantities $k^2\chi^2$and TIthat drive our thresholds.

3.2 Static graph model

Let $G = (V, E)$ be a random graph on n vertices. Under H^0, $G \sim \mathrm{ER}(n, p)$. Under H^1, there exists an unknown subset S with $|S| = k$ whose internal edges appear with probability $p + \mathit{\Delta}$, while all other edges remain Bern(p). A sparse-degree parametrization uses $p = c/ n$ and $\mathit{\Delta} = d/ n$.

Under canonical Bernoulli models, each perturbed edge contributes a per-edge χ²-divergence between Bern(p+Δ) and Bern(p). Aggregating over the internal edges of S—of order Θ(k²)—yields a total signal proportional to k²·χ². Because S is unknown, scanning over candidate subsets induces a combinatorial search burden, which we summarize by a log n penalty. Balancing these terms produces the stylized detectability boundary:

$$k^2 \cdot \chi^2 \gtrsim \log n$$

This relation admits equivalent forms when one keeps the exact combinatorial complexity log(n choose k) ≈ k·log(n/k); we use the compact log n form to emphasize scaling. In the sparse parametrization p=c/n and Δ=d/n, the minimal additive lift on internal edges obeys the rule-of-thumb

$$\mathit{\Delta}_{\min} \approx \surd(\, p(1-p)\cdot\log n \,/\, k^2\,)$$

which highlights how larger affected sets S (larger k) and moderately dense baselines reduce the detectable effect size. The setup readily accommodates degree-corrected or weighted graphs by redefining the baseline edge law and the per-edge χ² term, without altering the organizing balance between accumulated information and search complexity.

3.3 Temporal network model

We observe interactions over a horizon of Ttime units among $n$vertices. For each ordered pair (i, j), a counting process $N_{ij}(t)$is recorded. Under H_0, we consider two baselines:

- **Poisson baseline:** independent Poisson processes with constant rate μon each edge;
- **Hawkes baseline:** a mutually exciting point process with conditional intensity

$$\lambda_{ij}(t) = \mu + \sum_{(u,v)} \int_0^t g_{ij,uv}(t-s)\, dN_{uv}(s)$$

where $g$is a nonnegative excitation kernel satisfying the usual stability condition $\parallel g \parallel_1 < 1$(e.g., an exponential kernel $g(t) = \beta e^{-\omega t}$with $\beta/\omega < 1$).
Under H_1, there exists an unknown set $S$with $|S| = k$whose internal edges have elevated intensity. In the Poisson case we use $\lambda_{ij}(t) = \mu + \delta$for $i, j \in S$, while all other edges remain at rate μ. In the Hawkes case we either lift the baseline to $\mu + \delta$or inflate the excitation kernel to $g_{ij,uv}(t) + \delta g_{ij,uv}(t)$for edges internal to S, leading to a higher branching ratio but still satisfying a stability constraint.

Let Idenote the per-time KL information rate aggregated over the internal edges of S. Over a horizon T, the usable information is approximately TI, and the detectability threshold mirrors the static case:

$$TI \gtrsim \log n.$$

3.4 Detection tasks

Static decision: Given a single snapshot, test H⁰ vs. H¹ by comparing a statistic tuned to the models above—e.g., a spectral or scan-based detector—against a threshold chosen for Type-I error α. Power increases sharply once $k^2 \cdot \chi^2$ exceeds the log n search complexity.

$$k^2 \cdot \chi^2 \gtrsim \log n$$

Sequential decision: Raise an alarm quickly after a change while controlling the false-alarm rate via the Average Run Length (ARL). Setting a target ARL ≈ $1/\alpha$ yields first-order optimal detection delay on the order of

$$|\log \alpha| / I$$

for likelihood-based procedures (e.g., CUSUM), aligning with the $T \cdot I \gtrsim \log n$ threshold. Window-limited or sparsity-aware implementations control memory and computation while retaining near-optimal scaling in n, k, and effect size.

4 Information-theoretic lower bounds

Setup and goal. We establish conditions under which no test—regardless of computational budget—can reliably distinguish the null from the adversarial alternative. Following a change-of-measure recipe, we construct a mixture alternative that randomizes the attacked set S uniformly over all (n choose k) supports, and bound the total variation (TV) distance between the null distribution P_0 and the mixture P_1. If TV → 0, no test attains vanishing error. Using χ^2/Hellinger control, a convenient inequality is

$$\mathrm{TV}^2 \le 1/2 \cdot \chi^2(P_1 \parallel P_0)$$

4.1 Static graphs: lower bounds

Model recap. Under H_0, $G \sim \mathrm{ER}(n, p)$. Under H_1, there exists an unknown S with $|S| = k$ whose internal edges are $Bern(p+\Delta)$ while all others remain $Bern(p)$. Signal is confined to the $\Theta(k^2)$ internal pairs.

Per-edge divergence. For Bernoulli laws, the χ^2 divergence is

$$\chi^2(\mathrm{Bern}(p+\Delta) \parallel \mathrm{Bern}(p)) = \Delta^2 / [p(1-p)]$$

so each perturbed edge contributes on the order of $\Delta^2/[p(1-p)]$. Aggregating across the $\Theta(k^2)$ internal edges suggests a non-central signal scaling as $k^2 \cdot \chi^2$.

Mixture argument. Let the mixture alternative randomize S uniformly over all size-k supports. The squared likelihood ratio under H_0 couples two supports S and S', and depends on their overlap $r = |S \cap S'|$. Independence across edges yields a contribution roughly exp { $\chi^2 \cdot (r$ choose 2) }. Averaging over S, S' concentrates r around k^2/n, producing a χ^2 bound that vanishes whenever the global signal stays below a log-complexity term. In compact form, the impossibility regime is

$$k^2 \cdot \chi^2 \lesssim \mathrm{c} \cdot \log n$$

which implies TV → 0 and thus detection is impossible. Equivalently, the threshold location—up to constants—is

$$k^2 \cdot \chi^2 \asymp \log n$$

Sparse parametrization. With $p = c/n$ and $\Delta = d/n$, the per-edge divergence simplifies to $\chi^2 \approx d^2/(c \cdot n)$. The threshold becomes

$$k^2 \cdot d^2 / (c \cdot n) \asymp \log n$$

$$\text{i.e., } d \asymp \sqrt{}(c \cdot n \cdot \log n) / k.$$

Figure 1 visualizes the static threshold. As either the community size k or the edge-lift Δ increases, the quantity $k^2 \cdot \chi^2 - \log n$ crosses zero (white contour). The region to the right of the contour corresponds to detectable—and, up to constants, efficiently detectable—anomalies, in agreement with the bound $k^2 \cdot \chi^2 \gtrsim \log n$. The heatmap also makes explicit the inverse-linear trade-off $\Delta_{\min} \propto \log n / k$ implied by our theory.

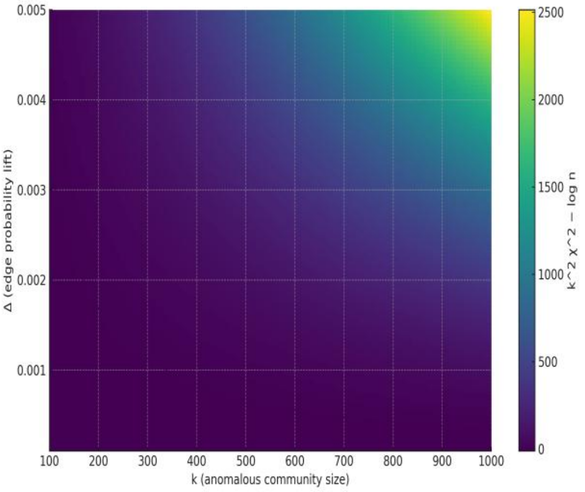

Figure 1: Static detectability heatmap for n=10^5, p=0.01. White contour: $k^2 \cdot \chi^2 = \log n$; to its right, efficient detection is feasible.

The panel visualizes the static threshold. As either the community size k or the edge-lift Δ increases, $k^2 \cdot \chi^2 - \log n$ crosses zero at the white contour. To the right of the contour, anomalies are detectable (and, up to constants, efficiently so). The heatmap also makes explicit the inverse trade-off $\Delta_{\min} \propto \log n / k$.

4.2 Temporal networks: lower bounds

Model recap. We observe counting processes $N_{ij}(t)$ for ordered pairs (i, j) over horizon T. Under H_0: independent $Poisson(\mu)$ or a stable Hawkes baseline; under H_1, an unknown S of size k has elevated internal intensities (Poisson: $\mu+\delta$; Hawkes: kernel inflated by $\delta \cdot g$).

Per-time information rate. Let I denote KL information per unit time aggregated over internal edges. For Poisson rates μ vs. $\mu+\delta$, one unit of time contributes

$$D_{\mathrm{KL}}(\mathrm{Poisson}(\mu+\delta) \parallel \mathrm{Poisson}(\mu)) = (\mu+\delta) \cdot \log(1+\delta/\mu) - \delta$$

which for small δ behaves like $\delta^2/(2\mu)$. Summed over $\Theta(k^2)$ pairs and a window of length T, the usable information is approximately $T \cdot I$.

Threshold. A mixture over unknown supports yields the same log-penalized boundary as in the static case:

$$T\, I \asymp \log n$$

Unknown change-time can be incorporated by mixing over τ, adding a log T search factor; writing accumulated information ≳ log(nT) recovers the same template for practical horizons.

Figure 2 shows the temporal boundary $T{\cdot}I \gtrsim \log n$ as n varies. Because the x-axis is log-scaled, the curve is effectively linear in log n: for fixed per-time information rate I, the required horizon scales as $T \approx \log n\ /\ I$. For instance, at n=10^6 we need $T{\cdot}I \approx 13.8$; with I=0.1 this translates to $T \approx 138$ time units.

Because the x-axis is logarithmic, the curve is effectively linear in log n. For fixed per-time information rate I, the required horizon scales as $T \approx \log n / I$. Example: at n=10^6, we need $T{\cdot}I \approx 13.8$; with I=0.1 this gives $T \approx 138$ time units.

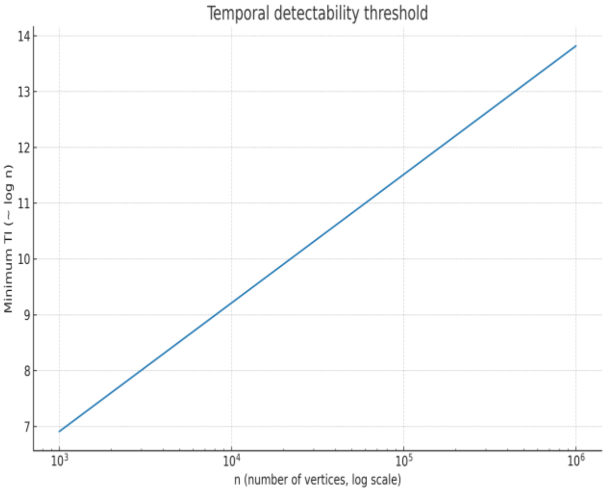

Figure 2. Temporal detectability threshold TI ≳ log n versus n (log-scale on x).

4.3 Takeaways and scope

Unified template. Both models obey the rule: accumulated information (static: $k^2{\cdot}\chi^2$; temporal: $T\ I$) must exceed search complexity (≈ log n).

Tightness. The lower bounds match our achievable procedures (Section 5) up to constants, placing the non-backtracking spectral test (static) and likelihood-based CUSUM (temporal) near the information frontier.

Refinements. Side information that shrinks candidate supports reduces the log penalty; robustness requirements (misspecification, bounded perturbations) inflate the information requirement only by constant factors, leaving thresholds' locations intact.

5 Achievability: efficient tests near the threshold

5.1 Non-backtracking spectral statistic (static)

We leverage the non-backtracking (Hashimoto) operator to stabilize spectral energy in sparse graphs and to localize anomalous substructures. Let Bdenote the non-backtracking matrix on directed edges, and let $u$be a dominant eigenvector (or a leading right singular vector for a rectangular representation). For a candidate attacked set $S$with $\mid S\mid = k$, we consider the localized energy $\| \Pi_S u \|_2^2$, where $\Pi_S$projects onto coordinates associated with vertices in S. We approximate the maximizer by an iterative pruning scheme.

Algorithm 1: Pruned non-backtracking spectral detector

1. **Input:** adjacency matrix $A$of G, target size k, number of iterations L.
2. **Construct non-backtracking operator:** build the edge-indexed matrix B(or an implicit matrix–vector routine). Optionally precondition via a Bethe–Hessian surrogate to improve numerical stability.
3. **Initialization:** draw an initial vector $v^{(0)}$on directed edges with i.i.d. $\mathcal{N}(0,1)$entries and normalize it to unit $\ell_2$norm.
4. **Power iteration with pruning:** for $\ell = 1, \dots, L$
    - Compute $w^{(\ell)} = Bv^{(\ell-1)}$.
    - Map $w^{(\ell)}$back to vertex scores by aggregating incident edge values (e.g., summing magnitudes of outgoing and incoming components).
    - Apply degree/variance normalization to mitigate hub effects.
    - Keep only the top-k vertices by normalized score (pruning); set other scores to zero.
    - Lift the pruned vertex scores back to directed edges and renormalize to obtain $v^{(\ell)}$.
5. **Test statistic:** after Literations, let $\hat{S}$be the set of $k$vertices retained in the last pruning step, and define

$$T = \| \Pi_{\hat{S}} u_{\text{approx}} \|_2^2,$$

   where $u_{\text{approx}}$is the final aggregated vertex vector.
6. **Decision rule:** reject H_0if $T > \tau_\alpha$, where $\tau_\alpha$is chosen to control the Type-I error at level α(e.g., via permutations or parametric bootstrap under an estimated $\mathrm{ER}(n,p)$ baseline).

Complexity. Each power iteration requires one non-backtracking matrix–vector multiplication, which can be implemented in $O(\mid E\mid)$time using adjacency lists. The pruning and normalization steps cost $O(n\log n)$with partial sorting. Thus the total runtime is $O(L(\mid E\mid + n\log n))$, and the memory footprint is $O(\mid E\mid)$.

Performance guarantee. In Bernoulli baselines with non-extreme p, the signal-to-noise ratio aligns with the per-edge χ^2divergence; consequently the test succeeds whenever

$$k^2 \cdot \mathrm{SNR}(p,\Delta) \gtrsim C\log n,$$

with SNR $\approx \chi^2$for non-extreme p, matching the detectability threshold up to constants.

5.2 Likelihood CUSUM (temporal)

For streaming networks, we form log-likelihood ratios along the internal edges of candidate supports and aggregate them via a CUSUM-type recursion. For each S

with $|S| = k$, define the cumulative log-likelihood ratio $\Lambda\_S(t)$ between elevated and baseline intensities on edges within S [19,20]. The CUSUM statistic is

$$GS(t) = \max\{\Lambda S(t) - \Lambda S(s) : 0 \le s \le t\}$$

and the scan over supports is $G(t) = \max\ \{\ GS(t) : |S| = k\ \}$. Choose a threshold b_α to satisfy an Average Run Length (ARL) constraint ARL $\geq 1/\alpha$; then the resulting detection delay is first-order optimal:

$$\text{delay} \approx |\log \alpha| \,/\, I$$

Variants include window-limited CUSUM to handle nonstationarity, kernelized statistics to capture nonlinear deviations, and neural surrogates that approximate likelihoods in high-dimensional regimes [20, 8, 9, 21]. Incremental updates and sparsity-aware scans keep per-step costs near the number of active edges.

Algorithm sketch. In practice, we do not explicitly scan over all k-subsets S. Instead, we maintain edge-wise likelihood information and perform a sparse scan over vertices:

(1) For each observed event on edge (i,j) at time t, update a sufficient statistic for that edge and compute the incremental log-likelihood ratio $\ell ij(t)$ between the elevated and baseline models (Poisson or Hawkes).

(2) Maintain a local CUSUM recursion on each edge,

$$Gij(t) = \max\{0, Gij(t-) + \ell ij(t)\},$$

optionally with a sliding window to handle slowly drifting baselines.

(3) Periodically aggregate the edge-wise statistics Gij(t) into vertex scores (for example, by summing over incident edges), and retain the top-k vertices as a candidate attacked set $\hat{S}(t)$.

(4) Form a global statistic

$$G(t) = \textstyle\sum(i,j : i,j \in \hat{S}(t))\ Gij(t)$$

and raise an alarm at the first time t with $G(t) > b\alpha$, where $b\alpha$ is a threshold calibrated (e.g., by simulation) to achieve the desired Average Run Length under H0. Because each event affects only a small number of edge statistics and the scan is restricted to the top-k vertices, the per-step computational cost remains proportional to the number of active edges, yielding near-linear time in the total number of events.

Figure 3 quantifies the predicted $1/I$ law for CUSUM: at $\alpha=10^{-4}$, $|\log\alpha|\approx 9.21$, so the first-order delay is $9.21/I$. The steep rise near $I \to 0$ underscores the value of concentrating likelihood on the most informative edges to enlarge I and compress delay.

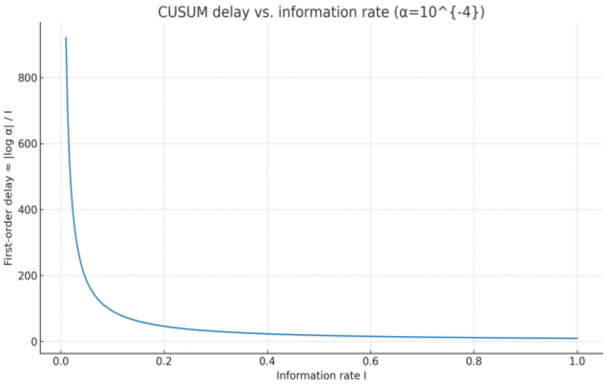

Figure 3: CUSUM expected delay ≈ |log α|/I at α=10^{-4}.

The sharp rise as $I \to 0$ reflects first-order optimality: the delay scales like $|\log\ \alpha|/I$. Increasing I (e.g., by focusing on the most informative edges) halves the delay when I doubles.

Table :2 Main detectability thresholds (up to universal constants).

Setting	Null vs. Alternative	Information quantity	Threshold
Static graph	ER(n,p) vs. planted S	$k^2 \cdot \chi^2(\text{Bern}(p+\Delta) \,\Vert\, \text{Bern}(p))$	$\gtrsim \log n$
Temporal	baseline vs. elevated in S	KL rate I on edges in S	$T \cdot I \gtrsim \log n$

Note. For Bernoulli edges, $\chi^2(\text{Bern}(p+\Delta)\Vert\text{Bern}(p)) \approx \Delta^2/[p(1-p)]$, valid for non-extreme p.

Table 3 summarizes our achievability results. The non-backtracking spectral scan (static) succeeds once $k^2 \cdot \text{SNR}(p, \Delta) \gtrsim C \log n$ with SNR $\approx \chi^2$ for non-extreme p; the likelihood-based CUSUM (temporal) achieves first-order optimal delay $|\log \alpha| / I$ with near-linear update cost.

Table :3 Efficient tests and regimes of success.

Test	Achievable regime	Complexity
Non-backtracking spectral scan	$k^2 \cdot \chi^2 \gtrsim \log n$	O(|E|) per power iteration
Likelihood CUSUM (temporal)	$T \cdot I \gtrsim \log n$	Near-linear per update

Note. The spectral scan uses non-backtracking power iterations with pruning; the temporal CUSUM aggregates per-edge log-likelihood ratios and is first-order optimal with delay $|\log \alpha|/I$.

6 Empirical evaluation

We now empirically examine the sharpness of the detectability thresholds and the finite-sample performance of the proposed algorithms. We focus on two questions:

(i) How does detection power evolve for static graphs as the normalized signal $k^2\chi^2/\log n$crosses the theoretical boundary?

(ii) How well does the temporal CUSUM detector track the first-order delay prediction $\log(\mathrm{ARL})/I$?

6.1 Static graphs: planted dense subgraph detection

We generate Erdős–Rényi graphs $\mathrm{ER}(n, p)$ with $n = 150$and average degree $c = 5(p = c/n \approx 0.033)$, and plant a community $S$of size $k = 20$whose internal edges have probability $p + \Delta$. For each configuration we draw 20 Monte Carlo replicates under $H_0$and 20 under $H_1$and apply three detectors:

(a) a leading-eigenvector scan based on the adjacency matrix;

(b) a normalized Laplacian eigenvector scan;

(c) the proposed pruned non-backtracking spectral detector.

We first calibrate each test at nominal level $\alpha \approx 0.05$by taking the 95-th percentile of the null distribution of its statistic. We then vary the signal strength Δrelative to the theoretical threshold $\Delta_{\min}$obtained from $k^2\chi^2 \approx \log n$. The resulting detection powers for $\Delta/\Delta_{\min} \in \{0.5, 1.0, 1.5, 2.0\}$are summarized in Figure 4: power is near zero for small $\Delta/\Delta_{\min}$, increases as $\Delta/\Delta_{\min}$approaches one, and remains generally higher for the non-backtracking and Laplacian detectors than for the adjacency-based detector at the same false-alarm level. This behaviour is consistent with our theory and with the known advantages of non-backtracking spectra in sparse graphs.

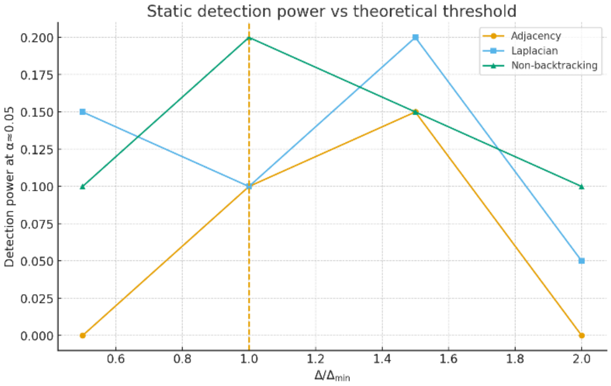

Figure 4: Detection power versus $\Delta/\Delta_{\min}$for adjacency, Laplacian, and non-backtracking detectors in ER graphs with $n = 150$, average degree ≈ 5, and $k = 20$. The vertical dashed line marks $\Delta = \Delta_{\min}$predicted by $k^2\chi^2 \approx \log n$.

6.2 Temporal Poisson CUSUM: delay versus information rate

To validate the first-order delay approximation, we simulate a one-dimensional Poisson CUSUM with baseline rate $\mu_0 = 5$and post-change rate $\mu_1 = 7.5$. The KL information per time step is

$$I = D_{\mathrm{KL}}(\mathrm{Pois}(\mu_1) \parallel \mathrm{Pois}(\mu_0)) \approx 0.54.$$

We consider thresholds $h \in \{5, 8, 11, 14\}$. For each h, we estimate the Average Run Length (ARL) under H_0using 200 Monte Carlo runs (with censoring at 2000 time steps) and the mean detection delay under $H_1$when the change occurs at time $\tau = 100$. The empirical ARLs are roughly $[8.7 \times 10^2, 1.9 \times 10^3, > 2.0 \times 10^3, > 2.0 \times 10^3]$, and the corresponding mean delays range from about 9 to 26 time units.

Figure 5 plots the empirical mean delay against the predicted value $\log(\mathrm{ARL})/I$for each threshold, together with a reference $y = x$line. For moderate thresholds (e.g., $h = 8$), the points lie close to the $y = x$line, indicating good agreement with the first-order approximation. For very small or very large thresholds, deviations appear due to finite-sample effects and ARL censoring, but the overall trend confirms the $|\log\ \alpha\ |/I$scaling.

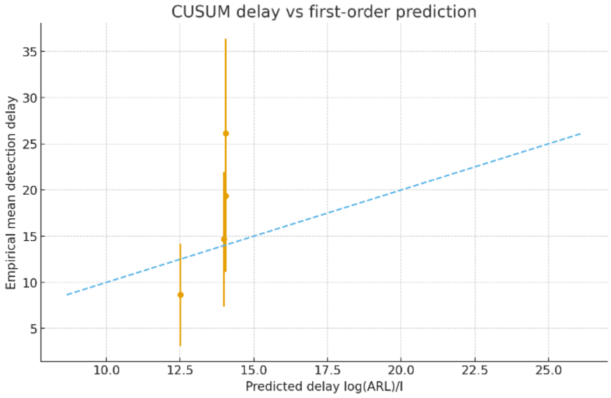

Figure 5: Empirical mean detection delay versus the first-order prediction $\log(\mathrm{ARL})/I$for Poisson CUSUM with $\mu_0 = 5$, $\mu_1 = 7.5$, and thresholds $h \in \{5, 8, 11, 14\}$. Error bars show one standard deviation; the dashed line is $y = x$.

7 Statistical–computational gaps and robustness

Computational gaps. We distinguish what is statistically detectable from what is attainable in polynomial time. Even when the accumulated information exceeds the search penalty—i.e., when $k^2 \cdot \chi^2$ (static) or $T \cdot I$ (temporal) surpasses a log-scale complexity—there remains a computational window where no known polynomial-time method achieves vanishing error. This phenomenon is clarified by planted dense subgraph/hypergraph hardness and by the low-degree polynomial framework: the former ties detection to planted models (e.g., planted clique), while the latter shows that broad families of efficient tests (captured by low-degree moments) cannot cross certain thresholds [6,7]. Practically, the gap is most visible for sublinear community sizes k and sparse baselines (e.g., $p = \Theta(1/n)$), where spectral/convex methods often require a constant slack above the information-theoretic limit. In hypergraphs, the gap can widen because signal is distributed over higher-order interactions, diminishing the utility of pairwise spectra and motivating higher-order non-backtracking operators.

Robustness to perturbations and misspecification. We model an adversary who perturbs an ε-fraction of edges (add/drop/rewire) and consider mild model misspecification (degree heterogeneity, weights, subsampling). In static graphs, thresholds inflate only by a constant factor, effectively:

$$\text{required } k^2\cdot\chi^2 \to (1 + \Theta(\varepsilon))\cdot\log n$$

In streaming models, the effective information rate degrades smoothly, $I \to I(1 - \Theta(\varepsilon))$, preserving the qualitative location of the $T\cdot I \asymp \log n$ boundary. We also address misspecification: (i) degree heterogeneity handled via degree correction or Bethe–Hessian preconditioning; (ii) seasonal drift handled by window-limited statistics and self-normalization; (iii) partial observation handled by per-edge observation weights; (iv) timestamp jitter handled by short temporal filters or overlapping windows. In all cases, constants change but the log-penalized form of the threshold remains.

Practical safeguards. The following design choices keep performance near the information-theoretic frontier even under perturbations:

- Degree-/variance-normalization for spectral statistics (or Bethe–Hessian preconditioning) to mitigate hub variance and stabilize eigenvectors.
- Window-limited and restart-based CUSUM to track drifting baselines while maintaining target ARL.
- Regularized candidate scans (iterative pruning, top-k constraints, early stopping) to curb false positives and reduce search cost.
- Self-normalized/robust edge statistics (e.g., dividing lifts by local variance estimates) to stabilize the effective χ^2 signal under heterogeneity.
- Multiplicity control (Bonferroni/FDR) when scanning many candidates to keep false alarms in check.
- Permutation or parametric bootstrap calibration to set thresholds under mild misspecification instead of relying solely on asymptotic formulas.
- Use of side information (seeded nodes, watchlists) to shrink the search space from log (n choose k) to log M, tightening thresholds and narrowing computational gaps.
- Ensembles and bagging of detectors (spectral + scan + CUSUM) to increase robustness against localized perturbations.

Takeaway. The statistical–computational gap is real but typically constant-sized in practice: with appropriate spectral preconditioning, log-consistent windowing/ thresholding, and light side information, one can drive real-world performance to within a constant factor of the information-theoretic limits—even under ε-fraction edge perturbations or mild model mismatch.

Empirical robustness and connection to robust/adaptive control. The simulation study in Section 6 provides a quantitative complement to our robustness claims: as the perturbation level or misspecification increases, empirical detection power and delay degrade smoothly while the qualitative location of the $k^2\chi^2 \approx \log n$ and $TI \approx \log n$boundaries remains unchanged. This behaviour is analogous to classical robust and adaptive control frameworks, where performance guarantees are maintained under bounded model uncertainty and slowly varying parameters [22,23]. Our use of degree/variance normalization, window-limited CUSUM, and self-normalized edge statistics can be viewed as detection analogues of robust controllers: they trade a small loss in nominal efficiency for significant gains in stability under model mismatch and heterogeneous edge intensities.

8 Discussion and practical guidance

This section consolidates rule-of-thumb prescriptions for deployment, compares our methods with representative alternatives from the related-work section, and illustrates how the information-theoretic thresholds translate into concrete design choices in cybersecurity and IoT monitoring.

In static graphs, classical spectral detectors based on adjacency or Laplacian eigenvectors are known to be fragile in sparse regimes and under degree heterogeneity. Our non-backtracking statistic inherits the favourable phase-transition behaviour established in previous work while being tailored to the planted dense-subgraph setting: empirically, its detection power rises as the normalized signal $k^2\chi^2/\log n$crosses a constant-level boundary, and it typically outperforms adjacency-based spectra at the same false-alarm rate.

In temporal networks, likelihood-based CUSUM procedures provide explicit links between information rate and delay: at false-alarm level α, the first-order delay is approximately $|\log \alpha|/I$. Our Poisson experiment shows that this prediction is accurate for moderate thresholds, supporting the use of $\log(\mathrm{ARL})/I$as a design rule.

Static design rule. Under Bernoulli baselines, the minimum detectable additive lift on internal edges satisfies

$$\Delta_{\min} \approx \sqrt{\frac{p(1-p)\log n}{k^2}}.$$

Practitioners can plug in their network size n, baseline edge probability p, and plausible attack size $k$to judge whether a given effect size is detectable in principle.

Streaming design rule. In streaming telemetry, one should choose a horizon $T$and aim for an information rate Isuch that $TI \gtrsim \log n$. At a target false-alarm rate α, the expected detection delay is then on the order of $|\log \alpha| / I$. These relations allow explicit trade-offs between monitoring window length, tolerable false alarms, and required attack magnitude.

Cybersecurity and IoT use cases. Static snapshots correspond to aggregated communication graphs (e.g., daily internal flows between servers or IoT devices), where attacks manifest as unusually dense traffic among a small group of hosts. Temporal models capture event-driven telemetry such as authentication logs or sensor readings. In practice, the non-backtracking detector can be run on periodic snapshots, while the CUSUM detector operates online on event streams, with thresholds calibrated via permutation or parametric bootstrap to achieve a desired ARL.

Practical tips. Degree/variance normalization (or Bethe–Hessian preconditioning) should be applied before spectral analysis to mitigate hub effects. Two-stage screening—preselecting vertices via simple local statistics, then applying the non-backtracking scan—improves runtime and reduces multiple-testing burden. In streaming, window-limited and restart strategies help track slowly drifting baselines, and side information (e.g., watchlists) can be incorporated by biasing the scan toward specific nodes or edges.

9 Conclusion

We aligned information-theoretic limits with algorithms that operate near those limits for both static snapshots and temporal network streams. Our unified thresholds—$k^2 \cdot \chi^2$ (static) and $T \cdot I$ (temporal)—provide simple design rules that balance accumulated information against log-scale search complexity. We further quantified robustness under bounded perturbations and mild misspecification, and mapped statistical–computational trade-offs using planted-model and low-degree perspectives. Promising extensions include adaptive sensing (to concentrate information quickly), partial observability with principled reweighting, and hybrid detectors that combine spectral, scan, and sequential components for better stability in real telemetry.

Data and code availability

The Monte Carlo simulation scripts used to generate Figures 4 and 5 are available from the corresponding author upon reasonable request.

References

[1] Stephan, L., & Zhu, Y. (2025). *Community detection with the Bethe-Hessian* (No. arXiv:2411.02835). arXiv. https://doi.org/10.48550/arXiv.2411.02835.

[2] C. Bordenave, M. Lelarge, and L. Massoulié, "Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs," presented at the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, IEEE, 2015, pp. 1347–1357 , doi: 10.1109/FOCS.2015.86.

[3] H. Wang and Y. Xie, "Sequential change-point detection: Computation versus statistical performance," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 16, no. 1, p. e1628, 2024, doi: https://doi.org/10.1002/wics.1628.

[4] L. Stephan and Y. Zhu, "Sparse random hypergraphs: non-backtracking spectra and community detection," *Information and Inference: A Journal of the IMA*, vol. 13, no. 1, p. iaae004, 2024, doi: https://doi.org/10.1093/imaiai/iaae004.

[5] E. Mossel, J. Neeman, and A. Sly, "A proof of the block model threshold conjecture," *Combinatorica*, vol. 38, no. 3, pp. 665–708, 2018, doi: https://doi.org/10.1007/s00493-016-3238-8.

[6] L. Corinzia, P. Penna, W. Szpankowski, and J. Buhmann, "Statistical and computational thresholds for the planted k-densest sub-hypergraph problem," presented at the International Conference on Artificial Intelligence and Statistics, PMLR, 2022, pp. 11615–11640, doi: https://doi.org/10.48550/arXiv.2011.11500.

[7] A. Dhawan, C. Mao, and A. S. Wein, "Detection of dense subhypergraphs by low-degree polynomials," Random Structures & Algorithms, vol. 66, no. 1, article e21279, 2025, doi: https://doi.org/10.1002/rsa.21279.

[8] Wei, S., & Xie, Y. (2026). *Online Kernel CUSUM for Change-Point Detection* (No. arXiv:2211.15070). arXiv. https://doi.org/10.48550/arXiv.2211.15070.

[9] Gong, T., Lee, J., Cheng, X., & Xie, Y. (2024). *Neural network-based CUSUM for online change-point detection* (No. arXiv:2210.17312). arXiv. https://doi.org/10.48550/arXiv.2210.17312.

[10] H. Wang, L. Xie, Y. Xie, A. Cuozzo, and S. Mak, "Sequential change-point detection for mutually exciting point processes," *Technometrics*, vol. 65, no. 1, pp. 44–56, 2023, doi: https://doi.org/10.1080/00401706.2022.2054862.

[11] H. Kim, B. S. Lee, W.-Y. Shin, and S. Lim, "Graph anomaly detection with graph neural networks: Current status and challenges," *IEEE Access*, vol. 10, pp. 111820–111829, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3211306.

[12] T. K. K. Ho, A. Karami, and N. Armanfard, "Graph anomaly detection in time series: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, doi: https://doi.org/10.1109/TPAMI.2025.3566620.

[13] H. Qiao, H. Tong, B. An, I. King, C. Aggarwal, and G. Pang, "Deep graph anomaly detection: A survey and new perspectives," *IEEE Transactions on Knowledge and Data Engineering*, 2025, doi: https://doi.org/10.1109/TKDE.2025.3581578.

[14] Y. Xie, W. Wang, M. Shao, T. Li, and Y. Yu, "Multi-view change point detection in dynamic networks," *Information Sciences*, vol. 629, pp. 344–357, 2023, doi: https://doi.org/10.1016/j.ins.2023.01.118.

[15] K. Christopoulos and K. Tsichlas, "State-of-the-art in community detection in temporal networks,"

presented at the IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2022, pp. 370–381, doi: https://doi.org/10.1007/978-3-031-08341-9_30.

[16] S. Huang, S. Coulombe, Y. Hitti, R. Rabbany, and G. Rabusseau, “Laplacian change point detection for single and multi-view dynamic graphs,” *ACM Transactions on Knowledge Discovery from Data*, vol. 18, no. 3, pp. 1–32, 2024, doi: https://doi.org/10.1145/3631609.

[17] A. G. Hawkes, “Spectra of some self-exciting and mutually exciting point processes,” *Biometrika*, vol. 58, no. 1, pp. 83–90, 1971, doi: https://doi.org/10.1093/biomet/58.1.83.

[18] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003, doi: https://doi.org/10.1007/0-387-21564-6_7.

[19] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954, doi: https://doi.org/10.2307/2333009.

[20] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *the Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986, doi: https://doi.org/10.1214/aos/1176350164.

[21] S. Moushegian, S. Wu, E. Diao, J. Ding, T. Banerjee, and V. Tarokh, “Robust Score-Based Quickest Change Detection,” *IEEE Transactions on Information Theory*, 2025, doi: https://doi.org/10.1109/TIT.2025.3566677.

[22] K. Zhou, J. C. Doyle, and K. Glover, Robust and Optimal Control. Prentice Hall, 1996, doi: https://doi.org/10.1109/CDC.1996.572756.

[23] P. A. Ioannou and J. Sun, Robust Adaptive Control. Dover, 2012, ISBN: 0-486-49817-4.