Early Detection and Attribution of Structural Changes in Dynamic Networks

Izhar Ali
Department of Computer Science
Rowan University
Glassboro, USA
aliizh94@rowan.edu

Shen-Shyang Ho
Department of Computer Science
Rowan University
Glassboro, USA
hos@rowan.edu

***Abstract*—Dynamic networks evolve through structural changes that significantly impact their functionality. Early detection of these changes remains a fundamental challenge due to the inherent trade-offs between detection delay and false alarm control. Current methods accumulate evidence of these changes only after they occur, creating unavoidable detection delays that limit practical utility in time-sensitive applications.**

We address this limitation through horizon martingales that leverage statistical forecasting to accumulate evidence from predicted future network states, enabling detection before changes fully manifest. We establish two fundamental results: (1) horizon martingales preserve the martingale property under proper forecasting calibration, and (2) they maintain rigorous false alarm control with probability bounds of $1/\lambda$ for detection threshold λ. Comprehensive evaluation across diverse synthetic network types and real-world social interaction data demonstrates consistent improvements in detection speed with 13–25% reductions in detection delay while maintaining strict statistical guarantees.

***Index Terms*—Dynamic Networks, Early Change Detection, Conformal Prediction, Martingales, Feature Attribution**

I. Introduction

Complex systems are naturally modeled as networks where nodes represent system components and edges capture their interactions [1]. A dynamic network is formalized as a temporal sequence of graph snapshots $\{G_t\}_{t=1}^{T}$ where $G_t = (V_t, E_t)$ denotes the graph at time t with vertex set V_t and edge set E_t. The temporal evolution of these snapshots reveals structural changes that signal fundamental system transitions [2].

Structural changes occur at time τ when the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$ (where d is the dimensionality of the parameter space) governing the network generative process undergoes a shift:

$$\boldsymbol{\theta}_t = \begin{cases} \boldsymbol{\theta}_0 & \text{for } t < \tau \\ \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0 & \text{for } t \geq \tau \end{cases} \tag{1}$$

where $\boldsymbol{\theta}_t$ controls network generation through parameters such as connection probabilities, community structures, and growth dynamics. These parametric shifts induce distributional changes in observable graph-theoretic features, including degree distributions, clustering coefficients, and spectral characteristics, enabling detection through statistical monitoring of feature sequences.

This research is supported by NSF Award No. 2213658.

The detection of these transitions is essential across multiple domains. In social networks, it reveals community formation and dissolution [3]; in financial systems, it identifies market disruptions and regime shifts [4]; in computer networks, it detects intrusion patterns and attacks [5]; and in infrastructure networks, it signals failures or reconfigurations [6].

Formally, the change detection problem seeks to minimize expected detection delay while controlling false alarms:

$$\min_{\tau^*} \mathbb{E}[(\tau^* - \tau)^+ \mid \tau^* \geq \tau] \quad \text{subject to} \quad \mathbb{P}(\tau^* < \tau) \leq \alpha$$

where τ^* is the stopping time for the detection rule, $(\tau^* - \tau)^+ = \max\{0, \tau^* - \tau\}$ represents detection delay, and $\alpha \in (0, 1)$ is the acceptable false alarm probability.

Three fundamental challenges arise in solving this optimization problem:

1) **Non-parametric detection**: Network feature sequences follow complex, often unknown distributions. Parametric assumptions about pre- and post-change distributions are frequently unrealistic, necessitating distribution-free approaches that maintain statistical rigor.
2) **Feature attribution**: Determining which structural properties drive detected changes requires maintaining explicit mathematical connections between detection statistics and underlying graph features to ensure interpretable results.
3) **Speed-accuracy trade-off**: Achieving early detection while maintaining false alarm control requires minimizing $\mathbb{E}[(\tau^* - \tau)^+]$ while ensuring $\mathbb{P}(\tau^* < \tau) \leq \alpha$. This often necessitates predictive approaches that accumulate evidence before changes fully manifest.

We address these challenges through horizon martingales that accumulate predictive evidence from future network states, enabling detection before changes fully manifest while maintaining rigorous statistical guarantees.

II. Related Work

The literature on change detection in dynamic networks spans classical statistical methods, network-specific approaches, and recent advances in martingale-based detection. We organize this review around the three fundamental challenges identified above.

Classical Statistical Methods. Traditional statistical process control methods, including CUSUM [7] and EWMA [8], require explicit specification of pre- and post-change distributions. While these methods provide asymptotic optimality under Lorden's criterion [9], their parametric assumptions limit applicability to complex networks with unknown or time-varying distributions. The distributional complexity of network features makes these classical approaches impractical for real-world dynamic networks.

Network-Specific Detection Methods. Several approaches have been developed specifically for network change detection: (i) *Community-based methods* [2] monitor posterior distributions of community structure but require specifying generative models a priori, limiting their applicability when the underlying network generation process is unknown; (ii) *Spectral approaches* [6] track eigenvalues $\lambda_i(L_t)$ (where $i \in \{1, 2, \ldots, |V_t|\}$) of the graph Laplacian L_t, effectively capturing global connectivity patterns but potentially missing localized changes that preserve spectral properties; (iii) *Latent space methods* [10] project networks into lower-dimensional manifolds, improving computational efficiency but often sacrificing interpretability of detected changes; (iv) *Similarity-based approaches* like DeltaCon [11] compute distance functions between consecutive graph snapshots but lack formal control over false alarm rates, a critical limitation for practical deployment.

Martingale-Based Detection Framework. Recent work by Ho et al. [12] established martingale-based change detection in dynamic networks with distribution-free false alarm guarantees and exact feature attribution via the Martingale-Shapley equivalence. However, this framework accumulates evidence only after changes occur, creating inherent detection delays.

Early Detection and Forecasting Approaches. The fundamental trade-off between detection delay and false alarm rate remains a central challenge in sequential detection theory. Traditional methods accumulate evidence only post-change [13], creating delays dependent on the magnitude of distributional shifts and signal-to-noise ratios. Statistical forecasting approaches [14] have shown promise in univariate time series but face significant challenges when extended to graph-structured data while maintaining theoretical guarantees on false alarm control.

Our contributions directly address these challenges:

1) **Horizon Martingales**: We extend the existing martingale framework of Ho et al. [12] to integrate statistical forecasting, enabling evidence accumulation from predicted future states. Theorem 3 proves that horizon martingales maintain the martingale property under proper forecasting calibration, while Theorem 4 provides rigorous false alarm control, ensuring early detection does not compromise theoretical guarantees.
2) **Interpretable Feature Attribution**: We establish the applicability of the Martingale-Shapley equivalence [12] within the horizon martingale framework.
3) **Empirical Validation**: We demonstrate consistent improvements in detection speed (13–25% delay reduction) across diverse network types while maintaining rigorous false alarm control.

III. Martingale Change Detection Framework

A martingale-based change detection framework [12] detects structural changes in dynamic networks through monitoring distributional shifts in graph-theoretic features.

Definition 1 (Null and Alternative Hypotheses)**.** *Let $\{G_t\}_{t=1}^T$ be a sequence of graph snapshots. We define:*

- ***Null Hypothesis*** *$\mathcal{H}_0$: No structural change occurs, i.e., $\{G_t\}_{t=1}^T$ is generated from a stationary process with parameter vector $\boldsymbol{\theta}_0$ for all $t \in \{1, 2, \ldots, T\}$.*
- ***Alternative Hypothesis*** *$\mathcal{H}_1$: A structural change occurs at time $\tau \in \{1, 2, \ldots, T\}$, where the generative process shifts from parameter $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$ as described in Equation 1.*

Under $\mathcal{H}_0$, the graph sequence is exchangeable with stationary distribution $\mathbb{P}_{\boldsymbol{\theta}_0}$. Under $\mathcal{H}_1$, the distribution changes to $\mathbb{P}_{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \tau}$ at the change-point τ.

A. Graph Representation and Feature Extraction

Let $\mathcal{G}$ denote the space of all possible graphs. For a graph $G = (V, E) \in \mathcal{G}$ with vertex set V and edge set $E \subseteq V \times V$, the standard representation is via an adjacency matrix $A \in \{0, 1\}^{|V| \times |V|}$ where $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise. We extract the following feature statistics from a graph snapshot using a feature function $f : \mathcal{G} \to \mathbb{R}^d$.

- **Local Connectivity Features** quantify immediate neighborhood structure. The *degree* $\deg(v) = \sum_{u \in V} A_{vu}$ counts edges incident on node v, representing how well-connected the node is locally. The *clustering coefficient* measures how densely interconnected v's neighbors are:

$$C(v) = \frac{2|\{(u, w) \in E : u, w \in N(v)\}|}{\deg(v)(\deg(v) - 1)}$$

  for $\deg(v) \geq 2$ (else 0), where $N(v) = \{u \in V : (v, u) \in E\}$ denotes the neighborhood of v.
- **Information Flow Features** quantify communication efficiency and node importance in network paths. *Betweenness centrality* identifies nodes that act as bridges or bottlenecks in information flow:

$$BC(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

  where σ_{st} is the number of shortest paths from s to t, and $\sigma_{st}(v)$ counts those passing through v. *Closeness centrality* measures how quickly information can spread from a node to the entire network:

$$CC(v) = \frac{|V| - 1}{\sum_{u \neq v} d(v, u)}$$

  where $d(v, u)$ is the shortest-path distance (setting $CC(v) = 0$ if v is isolated). *Eigenvector centrality* $EC(v)$ is the component corresponding to vertex v in the dominant eigenvector of adjacency matrix A, capturing influence through high-quality connections—being

connected to important nodes matters more than having many connections.

- **Spectral Features** are global metrics derived from the graph Laplacian $L = D - A$, where A is the adjacency matrix and $D = \text{diag}(\deg(v_1), \ldots, \deg(v_{|V|}))$ is the degree matrix. The Laplacian eigenvalues satisfy $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|}$. The *algebraic connectivity* λ_2 measures how well-connected the network is overall (higher values indicate harder to disconnect), while the *spectral gap* $\lambda_3 - \lambda_2$ reveals how clearly separated different communities or clusters are.

B. Single-Feature Martingale Detection

We briefly overview key concepts from the martingale-based change detection framework of Ho et al. [12] in Sections III-B, III-C, and III-D that are essential for understanding our horizon martingale extension.

For a feature function $f : \mathcal{G} \to \mathbb{R}^d$, the corresponding feature sequence $\{X_t\}_{t=1}^T$ is defined as $X_t = f(G_t)$ for $t = 1, 2, \ldots, T$. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$ such that the feature sequence $\{X_t\}_{t\geq 1}$ is adapted to this filtration, with X_t being $\mathcal{F}_t$-measurable. Under the null hypothesis of no structural change, the feature sequence is assumed to be exchangeable [15], a weaker condition than i.i.d. that allows for certain types of temporal dependence while maintaining distributional invariance under permutation. This assumption is critical for network data, where complete independence is rarely satisfied due to inherent temporal dependencies. Ho et al. [12] transforms the feature sequence into a non-negative martingale $\{M_t\}_{t\geq 0}$ with $\mathbb{E}[M_t] = 1$ under the null hypothesis through the following sequential steps:

1) Compute a non-conformity score S_t measuring the deviation of X_t from the empirical distribution of $\{X_s\}_{s=1}^{t-1}$ using $\mathcal{K}$-means clustering with $\mathcal{K} = 1$ as a robust location estimator
2) Transform S_t into a conformal p-value statistic p_t that quantifies the statistical significance of the deviation
3) Update martingale M_t through a betting function that determines evidence accumulation

The non-conformity score measures how "strange" or unusual the current observation is compared to historical patterns. When structural changes occur, feature values should deviate significantly from past behavior, resulting in large non-conformity scores that signal potential change points.

Definition 2 (Non-conformity Score)**.** *Given $\{X_t\}_{t=1}^T$ where $X_t \in \mathbb{R}^d$, we construct a non-conformity function that quantifies deviation from historical patterns. We employ $\mathcal{K}$-means clustering with $\mathcal{K} = 1$ as a robust location estimator. Let $C_t = \arg\min_{c\in\mathbb{R}^d} \sum_{s=1}^{t-1} \|X_s - c\|^2$ denote the cluster center computed from observations up to time $t-1$. The non-conformity score at time t is defined as:*

$$S_t = \|X_t - C_t\| \tag{2}$$

where $\|\cdot\|$ denotes an appropriate distance metric in $\mathbb{R}^d$.

The choice of distance metric $\|\cdot\|$ is critical for the non-conformity score computation. Distance metrics including the Euclidean, Mahalanobis, Cosine, and Chebyshev distances were explored [12], each with different sensitivity properties for various types of network structural changes.

The non-conformity score quantifies deviation from historical patterns. To establish statistical significance, we transform it into a p-value through a distribution-free approach. The conformal p-value represents the empirical probability that past observations were at least as "strange" as the current one, with small p-values indicating potential structural changes.

Definition 3 (Conformal P-value)**.** *Let $\{S_s\}_{s=1}^{t-1}$ be the sequence of non-conformity scores computed up to time $t-1$. The conformal p-value at time t is:*

$$p_t = \frac{\#\{s : S_s > S_t\} \; + \; \theta_t \, \#\{s : S_s = S_t\}}{t} \tag{3}$$

where $\theta_t \sim \text{Unif}(0, 1)$ is drawn independently at time t to break ties. p_t represents the fraction of past observations with non-conformity scores at least as extreme as the current one.

Finally, a test martingale is constructed through a betting function that accumulates evidence against the null hypothesis. A test martingale is a stochastic process that remains balanced (expected value of 1) under the null hypothesis but grows when changes occur, providing a principled way to accumulate statistical evidence.

Definition 4 (Test Martingale)**.** *A non-negative stochastic process $\{M_t\}_{t\geq 0}$ adapted to filtration $\{\mathcal{F}_t\}_{t\geq 0}$ is a test martingale under the null hypothesis if*

1) $M_0 = 1$ *(initial condition),*
2) $\mathbb{E}[|M_t|] < \infty$, $\forall t \geq 0$ *(integrability),*
3) $\mathbb{E}[M_t \mid \mathcal{F}_{t-1}] = M_{t-1}$, $\forall t \geq 1$ *(martingale property)*

Definition 5 (Valid Betting Function)**.** *A function $g : [0, 1] \to [0, \infty)$ is a valid betting function [16] if and only if:*

$$\int_0^1 g(p)dp = 1 \tag{4}$$

The value $g(p)$ represents the multiplicative update to the martingale when observing a p-value p.

Ho et al. [12] explored the following parametric families of betting functions:

- **Power betting**: $g(p; \epsilon) = \epsilon p^{\epsilon-1}$ for $\epsilon \in (0, 1)$, where smaller ϵ values emphasize extreme p-values (near 0) while larger values provide more uniform sensitivity across the p-value range.
- **Mixture betting** offers adaptivity across different ranges of the p-value spectrum by combining $k \geq 1$ power betting functions:

$$g(p; \{w_i\}, \{\epsilon_i\}) = \sum_{i=1}^{k} w_i \epsilon_i p^{\epsilon_i - 1}$$

where $\sum_{i=1}^k w_i = 1$ and $\epsilon_i \in (0, 1)$ for all $i \in \{1, 2, \ldots, k\}$.

- **Beta betting** allows targeting specific intervals of p-values through the shape parameters α and β of the beta function $B(\alpha, \beta)$:

$$g(p; \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$

for $\alpha, \beta > 0$, where $B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$.

Given the definition of a test martingale and a valid betting function, a martingale sequence can be constructed as:

Definition 6 (Single-Feature Martingale)**.** *For a single feature sequence $\{X_t\}_{t=1}^T$ with corresponding p-values $\{p_t\}_{t=1}^T$ and a valid betting function g, the single-feature martingale sequence $\{M_t\}_{t\geq 0}$ is defined as:*

$$M_t = \prod_{i=1}^{t} g(p_i) \tag{5}$$

with $M_0 = 1$.

Theorem 1 (Martingale Property)**.** *Let $\{X_t\}_{t=1}^T$ be exchangeable under the null hypothesis, and let $\{p_t\}_{t=1}^T$ be the corresponding p-values. Then for any valid betting function g, the sequence $\{M_t\}_{t\geq 0}$ is a test martingale with respect to filtration $\{\mathcal{F}_t\}_{t\geq 0}$.*

Proof. We verify the three martingale properties:

1) **Adapted**: M_t is $\mathcal{F}_t$-measurable as it depends only on observations up to time t
2) **Integrable**: $\mathbb{E}[|M_t|] = \mathbb{E}[M_t] = 1 < \infty$ by construction
3) **Martingale Property**: Under the null hypothesis with exchangeable $\{X_t\}$, the p-values satisfy $p_t \mid \mathcal{F}_{t-1} \overset{d}{=} \text{Unif}(0, 1)$, so:

$$\begin{aligned}\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] &= M_t \cdot \mathbb{E}[g(p_{t+1}) \mid \mathcal{F}_t] \\ &= M_t \cdot \int_0^1 g(p)dp = M_t \cdot 1 = M_t\end{aligned}$$

□

Test martingales provide rigorous control of false alarms through Ville's inequality [17], which provides an upper bound for the probability of the martingale ever exceeding any threshold $\lambda > 0$ under the null hypothesis.

Theorem 2 (Ville's Inequality)**.** *Let $\{M_t\}_{t\geq 0}$ be a test martingale. Then for any $\lambda > 0$:*

$$\mathbb{P}\left(\sup_{t\geq 0} M_t \geq \lambda\right) \leq \frac{1}{\lambda} \tag{6}$$

C. Multi-Feature Martingale Detection

The framework extends to multiple features by constructing individual martingales for each feature and combining them through summation. The sum $M_t^A = \sum_{i=1}^K M_t^{(i)}$ is also a martingale [12] (Corollary 1), offering several benefits: (1) it aggregates weak signals from multiple features that might individually be insufficient for detection, (2) it maintains theoretical control over false alarms through Ville's inequality, and (3) it provides a natural framework for attributing detected changes to specific features.

D. Shapley Values for Feature Attribution

After detecting a structural change, we need to understand which network properties contributed most to the detection. Shapley values provide fair attribution by satisfying four fundamental axioms: (1) **Efficiency**: individual attributions sum to the total martingale value; (2) **Symmetry**: features with identical marginal contributions receive equal attribution; (3) **Dummy**: features contributing zero marginal evidence receive zero attribution; (4) **Additivity**: attribution is linear when combining multiple detection games.

Computing exact Shapley values traditionally requires evaluating 2^K possible coalitions, where K is the number of features, which becomes computationally intractable. The SHAP (SHapley Additive exPlanations) method [18] provides efficient Shapley value approximations. In particular, [12] proved that each feature's martingale value at detection time is mathematically equivalent to its Shapley value, providing exact feature attributions with no additional computational overhead.

This equivalence provides immediate computational advantages and fairness guarantees. Shapley value computation reduces from $O(K \cdot 2^K)$ complexity to $O(K)$ by directly reading martingale values. The martingale values provide attribution that automatically satisfies the four fundamental Shapley axioms: (1) **Efficiency**: $\sum_{k=1}^K M_t^{(k)} = M_t^A$; (2) **Symmetry**: features with identical betting functions and p-value sequences receive equal attribution; (3) **Dummy**: features with $g(p) = 1$ for all p contribute $M_t^{(k)} = 1$; (4) **Additivity**: attribution is preserved under martingale combination.

The relative contribution of feature k to the detection event at time t is given by:

$$\psi_k(t) = \frac{M_t^{(k)}}{M_t^A} \times 100\% \tag{7}$$

This normalized measure provides an immediate ranking of feature importance for interpretation, with the property that $\sum_{k=1}^K \psi_k(t) = 100\%$.

IV. Horizon Martingale Change Detection

Existing martingale detection [12] accumulates evidence only after changes occur, creating detection delays. If a structural change begins at time τ, the martingale process $\{M_t\}_{t=\tau}^{\tau+d}$ must accumulate sufficient evidence before crossing the detection threshold λ at time $\tau + d$, creating a fundamental lower bound on achievable detection delay of at least d timesteps.

We propose horizon martingales that leverage statistical forecasting to accumulate evidence from predicted future states.

A. Horizon Martingale Construction

For a feature sequence $\{X_t^{(k)}\}_{t=1}^T$, we denote the information available at time t as $\Gamma_t = \{X_1^{(k)}, X_2^{(k)}, \ldots, X_t^{(k)}\}$. A forecasting function f_h maps this historical information to a prediction h steps ahead:

$$\hat{X}_{t+h}^{(k)} = f_h(\Gamma_t) \tag{8}$$

where $h \geq 1$ is the prediction horizon parameter. This prediction is used to construct a predictive non-conformity score:

$$S_{t,h}^{(k)} = \|\hat{X}_{t+h}^{(k)} - C_t\| \tag{9}$$

where C_t is the cluster center computed from observations $\{X_s^{(k)}\}_{s=1}^t$. This score quantifies the expected deviation of the future feature vector from historical patterns. We transform this predictive score into a p-value by comparing it against the empirical distribution of historical non-conformity scores:

$$p_{t,h}^{(k)} = \frac{|\{s \in \{1, \ldots, t\} : S_s^{(k)} \geq S_{t,h}^{(k)}\}| + 1}{t+1} \tag{10}$$

which represents the fraction of historical observations with non-conformity scores at least as large as the predicted score. Unlike classical p-values [12] that quantify present deviations, $p_{t,h}^{(k)}$ measures the anticipated statistical significance h steps ahead, enabling preemptive evidence accumulation.

Definition 7 (Horizon Martingale)**.** *For a feature sequence $\{X_t^{(k)}\}_{t=1}^T$ with corresponding martingale values $\{M_t^{(k)}\}_{t\geq 0}$ [12], prediction horizon $h \geq 1$, and predictive p-values $\{p_{t,h}^{(k)}\}_{t\geq h}$, the horizon martingale is defined as:*

$$M_{t,h}^{(k)} = M_{t-1}^{(k)} \cdot g(p_{t,h}^{(k)}) \tag{11}$$

where $M_{t-1}^{(k)}$ is the martingale value from the previous timestep and g is a valid betting function.

The key insight is that the horizon martingale builds upon the accumulated evidence from the existing martingale [12], rather than starting from zero. This approach leverages the already-built statistical evidence from the corresponding martingale and extends it forward in time through forecasting. For the horizon martingale to maintain statistical guarantees, the forecasting function must be properly calibrated. Calibration ensures that under the null hypothesis of no structural change, the predicted feature values behave statistically similar to observed historical data, preserving the uniform distribution property of p-values that is essential for the martingale construction. Poor calibration would bias the p-values away from uniformity, violating the theoretical assumptions and potentially leading to false alarms or reduced detection power.

Definition 8 (Properly Calibrated Forecasting Function)**.** *A forecasting function f_h is properly calibrated under the null hypothesis $\mathcal{H}_0$ if the predictive p-values $\{p_{t,h}^{(k)}\}$ are asymptotically uniform:*

$$\lim_{t\to\infty} \sup_{x\in[0,1]} \left| \mathbb{P}(p_{t,h}^{(k)} \leq x \mid \mathcal{F}_{t-1}) - x \right| = 0 \tag{12}$$

In practice, we require approximate calibration: $\sup_{x\in[0,1]} |\mathbb{P}(p_{t,h}^{(k)} \leq x \mid \mathcal{F}_{t-1}) - x| \leq \epsilon_{cal}$ for small calibration tolerance $\epsilon_{cal} > 0$.

Calibration quality directly impacts detection performance. With perfect calibration, predictive p-values remain uniform under $\mathcal{H}_0$ while becoming stochastically smaller under changes, leading to optimal detection power with minimal delay. Poor calibration leads to two failure modes: (1) if predictive p-values cluster near 1 under $\mathcal{H}_0$, then $\mathbb{E}[g(p_{t,h}^{(k)}) \mid \mathcal{H}_0] < 1$, causing $M_{t,h}^{(k)}$ to decay over time; (2) if they cluster near 0, then $\mathbb{E}[g(p_{t,h}^{(k)}) \mid \mathcal{H}_0] > 1$, leading to false alarms. We address calibration challenges through periodic recalibration using recent stable periods.

Theorem 3 (Horizon Martingale Property)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_t\}_{t\geq 0}$ be the natural filtration generated by $\{X_s^{(k)}\}_{s=1}^t$. Under the null hypothesis $\mathcal{H}_0$ with a properly calibrated forecasting function f_h, if $\{M_t^{(k)}\}_{t\geq 0}$ is a test martingale, then the horizon martingale sequence $\{M_{t,h}^{(k)}\}_{t\geq h}$ defined by Definition 7 is also a test martingale with respect to $\{\mathcal{F}_t\}_{t\geq 0}$.*

Proof. We establish the three fundamental properties required for a test martingale:

1) **Adapted**: Since $M_{t-1}^{(k)}$ is $\mathcal{F}_{t-1} \subset \mathcal{F}_t$-measurable and $p_{t,h}^{(k)}$ is $\mathcal{F}_t$-measurable because the forecasting function f_h depends only on $\{X_s^{(k)}\}_{s=1}^t \in \mathcal{F}_t$, the horizon martingale $M_{t,h}^{(k)}$ is $\mathcal{F}_t$-measurable.
2) **Integrable**: Since $M_{t,h}^{(k)} = M_{t-1}^{(k)} \cdot g(p_{t,h}^{(k)}) \geq 0$ with $M_{t-1}^{(k)}$ integrable and $g \geq 0$, we have $\mathbb{E}[|M_{t,h}^{(k)}|] < \infty$.
3) **Martingale Property**: For $t \geq h$, we compute:

$$\begin{aligned} \mathbb{E}[M_{t,h}^{(k)} \mid \mathcal{F}_{t-1}] &= \mathbb{E}[M_{t-1}^{(k)} \cdot g(p_{t,h}^{(k)}) \mid \mathcal{F}_{t-1}] \\ &= M_{t-1}^{(k)} \cdot \mathbb{E}[g(p_{t,h}^{(k)}) \mid \mathcal{F}_{t-1}] \end{aligned} \tag{13}$$

where the factorization in (13) holds because $M_{t-1}^{(k)}$ is $\mathcal{F}_{t-1}$-measurable. Under proper calibration (Definition 8), the conditional distribution of $p_{t,h}^{(k)}$ given $\mathcal{F}_{t-1}$ converges to uniform on $[0, 1]$. For a valid betting function g with $\int_0^1 g(p)dp = 1$:

$$\lim_{t\to\infty} \mathbb{E}[g(p_{t,h}^{(k)}) \mid \mathcal{F}_{t-1}] = \int_0^1 g(p)dp = 1 \tag{14}$$

For finite samples with approximate calibration tolerance ϵ_{cal}:

$$\left| \mathbb{E}[g(p_{t,h}^{(k)}) \mid \mathcal{F}_{t-1}] - 1 \right| \leq \|g\|_\infty \cdot \epsilon_{\text{cal}} \tag{15}$$

where $\|g\|_\infty = \sup_{p\in[0,1]} |g(p)|$. Thus, $\mathbb{E}[M_{t,h}^{(k)} \mid \mathcal{F}_{t-1}] \approx M_{t-1}^{(k)}$, establishing the martingale property up to calibration error.

□

We now establish that the horizon martingale maintains the same rigorous false alarm guarantees as classical martingales. This is crucial to ensure that improved detection speed does not come at the cost of increased false alarms.

Theorem 4 (False Alarm Control for Horizon Martingales)**.** *Let $\tau_h^{(1)} = \inf\{t \geq h : M_{t,h}^{(k)} \geq \lambda\}$ be the first crossing time for the horizon-h martingale with detection threshold $\lambda > 0$.*

Under the null hypothesis $\mathcal{H}_0$, the probability of false alarm for the first detection satisfies:

$$\mathbb{P}(\tau_h^{(1)} < \infty \mid \mathcal{H}_0) \leq \frac{1}{\lambda} \tag{16}$$

Proof. We establish the false alarm bound through Ville's inequality applied to the horizon martingale sequence. By Theorem 3, $\{M_{t,h}^{(k)}\}_{t \geq h}$ is a test martingale under $\mathcal{H}_0$ with $\mathbb{E}[M_{h-1,h}^{(k)}] < \infty$. For the first crossing time $\tau_h^{(1)}$, the event $\{\tau_h^{(1)} < \infty\}$ is equivalent to $\{\sup_{t \geq h} M_{t,h}^{(k)} \geq \lambda\}$ since both represent the martingale ever exceeding the threshold. Applying Ville's inequality:

$$\mathbb{P}(\tau_h^{(1)} < \infty \mid \mathcal{H}_0) = \mathbb{P}\left(\sup_{t \geq h} M_{t,h}^{(k)} \geq \lambda \mid \mathcal{H}_0\right) \tag{17}$$

$$\leq \frac{\mathbb{E}[M_{h-1,h}^{(k)} \mid \mathcal{H}_0]}{\lambda} \tag{18}$$

Since $M_{h-1}^{(k)}$ represents the accumulated evidence up to time $h-1$ and $\mathbb{E}[M_{h-1}^{(k)} \mid \mathcal{H}_0] = 1$ for the baseline martingale, we have $M_{h-1}^{(k)} \approx 1$ under $\mathcal{H}_0$, yielding the desired bound. □

Corollary 1 (Multi-Feature Horizon Martingale)**.** *For multiple features $k \in \{1, 2, \ldots, K\}$, the additive combination:*

$$M_{t,h}^{A} = \sum_{k=1}^{K} M_{t,h}^{(k)} \tag{19}$$

preserves both the martingale property (by linearity of conditional expectation) and false alarm guarantees.

Feature Attribution. The Martingale-Shapley equivalence from Section III-D applies directly to horizon martingales since Theorem 3 establishes that they satisfy the fundamental martingale property, and the equivalence depends only on the additive structure $M_t^{A,h} = \sum_{k=1}^{K} M_t^{(k),h}$ and martingale properties—both preserved in the horizon setting. When horizon detection occurs at time t, each feature's contribution is given by $\psi_k^h(t) = M_t^{(k),h} / M_t^{A,h} \times 100\%$, providing attribution for anticipated structural changes with $O(K)$ computational complexity.

B. Statistical Forecasting of Graph Features

The efficacy of horizon martingales depends critically on forecasting quality. Accurate forecasting enables detection of changes before they fully manifest, reducing detection delay. We develop two complementary approaches: feature-space forecasting operates on extracted network metrics, while graph-space forecasting works directly with adjacency matrices to preserve structural relationships.

a) ***Feature-Space Forecasting****:* For each feature sequence $\{X_t^{(k)}\}_{t=1}^{T}$, we construct a prediction function $f_\phi^{(k)}$ (with parameters ϕ) that maps $p_{\text{hist}} \geq 1$ historical observations to a prediction h steps ahead:

$$\hat{X}_{t+h}^{(k)} = f_\phi^{(k)}(X_t^{(k)}, X_{t-1}^{(k)}, \ldots, X_{t-p_{\text{hist}}+1}^{(k)}) \tag{20}$$

We implement this through exponentially weighted moving averages (EWMA), which balance robustness and adaptivity:

$$\hat{X}_{t+h}^{(k)} = \sum_{j=1}^{p_{\text{hist}}} w_j X_{t-j+1}^{(k)}, \quad \text{where } w_j = \frac{\alpha^j}{\sum_{l=1}^{p_{\text{hist}}} \alpha^l} \tag{21}$$

The decay parameter $\alpha \in (0, 1)$ controls the trade-off between prediction inertia and responsiveness. Smaller values prioritize recent observations, while larger values weight the overall trend more heavily. EWMA is particularly suitable for network features as it smooths out stochastic fluctuations while remaining responsive to emerging shifts.

For calibration maintenance, we employ a three-step validation process: (1) identify stable periods using a sliding window variance test with window size $w_{\text{cal}} = 0.1T$ (where smaller windows capture local stability but may miss gradual trends, while larger windows provide robustness at the cost of delayed adaptation), (2) compute empirical p-value distributions over these periods, and (3) adjust forecasting parameters to maintain approximate uniformity. In practice, we monitor calibration through Kolmogorov-Smirnov tests against uniform distribution and trigger recalibration when the test statistic exceeds 0.15, ensuring sustained detection performance across varying network dynamics.

b) ***Graph-Space Forecasting with Structural Constraints****:* While feature-space forecasting is computationally efficient, it may miss complex structural dependencies. For direct adjacency matrix forecasting, we employ a constrained optimization approach that enforces topological invariants. Let $\mathcal{A}(n, \kappa)$ denote the space of adjacency matrices with n vertices and sparsity level κ. The constrained graph prediction at horizon h combines temporal and structural components:

$$\hat{A}_{t+h} = \beta_t \hat{A}_{t+h}^{(T)} + (1 - \beta_t) \hat{A}_{t+h}^{(S)} \tag{22}$$

where:

- $\hat{A}_{t+h}^{(T)} = \sum_{j=1}^{p_{\text{hist}}} w_j A_{t-j+1}$ captures recent evolution trends through exponentially decaying weights w_j defined in EWMA
- $\hat{A}_{t+h}^{(S)} = \arg\min_{A \in \mathcal{A}(n,\kappa_t)} \|\hat{A}_{t+h}^{(T)} - A\|_F + \gamma R(A)$ balances temporal fidelity with structural plausibility, where $\|\cdot\|_F$ is the Frobenius norm and $\gamma > 0$ is the regularization weight
- $\beta_t = (1 + e^{\delta_t})^{-1} \in (0, 1)$ is the adaptive blending weight based on recent prediction errors $\delta_t = \|\hat{A}_{t-1}^{(T)} - A_t\|_F - \|\hat{A}_{t-1}^{(S)} - A_t\|_F \in \mathbb{R}$

The regularization functional $R : \mathcal{A}(n, \kappa_t) \to \mathbb{R}_+$ enforces structural coherence through three components for the preservation of network degree $R_1(A)$, community structure $R_2(A)$,

and sparsity constraint $R_3(A)$:

$$R_1(A) = \frac{1}{|V|} \sum_{v \in V} \left(\frac{d_A(v) - \mu_t(d(v))}{\sigma_t(d(v))} \right)^2 \tag{23}$$

$$R_2(A) = \sum_{C \in \mathcal{C}_t} \left| \frac{e_C(A)}{|C|^2} - \frac{e_C(A_{\text{hist}})}{|C|^2} \right| \tag{24}$$

$$R_3(A) = \left| \frac{\|A\|_0}{n(n-1)} - \kappa_t \right| \tag{25}$$

where $\mu_t(d(v))$ and $\sigma_t(d(v))$ are the empirical mean and standard deviation of degree for node v over the historical window $\{G_s\}_{s=t-w+1}^{t}$ of size w, $\mathcal{C}_t$ is the set of communities detected at time t using modularity optimization, $e_C(A)$ counts edges within community C, $|C|$ denotes the size of community C, $\kappa_t = \frac{1}{w} \sum_{s=t-w+1}^{t} \frac{\|A_s\|_0}{n(n-1)}$ is the average sparsity level over the window, $\|A_s\|_0$ denotes the number of non-zero entries (edges) in adjacency matrix A_s, and A_{hist} represents the historical community structure. The total regularization is $R(A) = \omega_1 R_1(A) + \omega_2 R_2(A) + \omega_3 R_3(A)$, where the weights $\{\omega_j\}_{j=1}^{3}$ with $\sum_{j=1}^{3} \omega_j = 1$ balance the relative importance of degree preservation, community structure, and sparsity constraints.

Algorithm 1 Horizon Martingale Detection

Require: $\{G_t\}_{t=1}^{T}$, features $\{f_k\}_{k=1}^{K}$, threshold λ, betting g, horizon h, history size w
Ensure: Change-points $\mathcal{C}$

1: Initialize: $\mathcal{C} \leftarrow []$
2: Initialize: $M_0^{(k)} \leftarrow 1$, $M_0^{(k),h} \leftarrow 1 \ \forall k$
3: **for** $t = 1$ to T **do**
4: **for** $k = 1$ to K **do**
5: $X_t^{(k)} \leftarrow f_k(G_t)$
6: $S_t^{(k)} = \|X_t^{(k)} - C_t^{(k)}\|$ ▷ Def. 2
7: $p_t^{(k)}$ from empirical CDF ▷ Def. 3
8: $M_t^{(k)} \leftarrow M_{t-1}^{(k)} \cdot g(p_t^{(k)})$ ▷ Def. 6
9: **if** $t \geq w$ **then**
10: $\hat{X}_{t+h}^{(k)} \leftarrow f_h(\{X_s^{(k)}\}_{s=t-w+1}^{t})$ ▷ Sec. IV-B
11: $S_{t,h}^{(k)} \leftarrow \|\hat{X}_{t+h}^{(k)} - C_t^{(k)}\|$
12: $p_{t,h}^{(k)}$ from empirical CDF
13: $M_t^{(k),h} \leftarrow M_{t-1}^{(k)} \cdot g(p_{t,h}^{(k)})$ ▷ Def. 7
14: **end if**
15: **end for**
16: $M_t^A \leftarrow \sum_{k=1}^{K} M_t^{(k)}$
17: $M_t^{A,h} \leftarrow \sum_{k=1}^{K} M_t^{(k),h}$
18: **if** $M_t^A \geq \lambda$ **or** ($t \geq w$ **and** $M_t^{A,h} \geq \lambda$) **then**
19: $\mathcal{C} \leftarrow \mathcal{C} \cup \{t\}$
20: Reset: $M_t^{(k)} \leftarrow 1 \ \forall k$
21: Reset: $M_t^{(k),h} \leftarrow 1 \ \forall k$
22: **end if**
23: **end for**
24: **return** $\mathcal{C}$

Algorithm 1 maintains both baseline and horizon detection streams in parallel. Steps 4–6 construct classical martingales $M_t^{(k)} = M_{t-1}^{(k)} \cdot g(p_t^{(k)})$ following Definition 6, ensuring baseline statistical guarantees. Steps 8–10 implement horizon martingales $M_t^{(k),h} = M_{t-1}^{(k)} \cdot g(p_{t,h}^{(k)})$ as specified in Definition 7, where the predictive p-values $p_{t,h}^{(k)}$ are computed from forecasted features $\hat{X}_{t+h}^{(k)}$. Steps 11–12 form additive martingales $M_t^A = \sum_{k=1}^{K} M_t^{(k)}$ and $M_t^{A,h} = \sum_{k=1}^{K} M_t^{(k),h}$ preserving the martingale property by linearity (Corollary 1). Step 13 applies the detection rule with threshold λ, creating stopping times $\tau^{(1)} = \inf\{t : M_t^A \geq \lambda\}$ and $\tau_h^{(1)} = \inf\{t : M_t^{A,h} \geq \lambda\}$ that satisfy the false alarm bounds of Theorem 4.

V. Experimental Setup

We evaluate this approach on four synthetic networks: Stochastic Block Model (SBM) [19], Barabási-Albert (BA) [20], Erdős-Rényi (ER) [21], and Newman-Watts-Strogatz (NWS) [22], and one real-world MIT Reality dataset [3] to assess the detection performance across diverse structural change scenarios. Each synthetic sequence consists of $T_{\text{synth}} = 200$ graph snapshots with 1–2 randomly placed structural change points to avoid temporal bias where model parameters shift according to predefined ranges (Table I). We enforce a minimum separation of $\Delta_{\text{min}} = 40$ timesteps (equivalent to a time window of $0.2T$ for sequences of length $T = 200$) between consecutive changes to ensure sufficient observations for establishing baseline behavior in each stable regime. All synthetic networks use $N_{\text{nodes}} = 50$ vertices. For scalability beyond moderate-sized networks, the bottleneck lies in feature extraction ($O(n^3)$ for centrality metrics) rather than martingale computation ($O(K)$ per timestep), suggesting that dimensionality reduction or sampling-based feature approximation would be necessary for networks with $n > 10^3$ nodes. The synthetic configurations use extreme parameter ranges for rigorous stress testing. The model parameters control distinct structural properties: SBM parameters govern community structure through intra-block (p_{intra}) and inter-block (p_{inter}) connection probabilities with n_{blocks} communities; ER parameters regulate overall edge probability (p); BA parameters control preferential attachment strength via the number of edges added per node (m); and NWS parameters manage small-world topology through neighborhood size (k) and rewiring probability (p_{rewire}).[1]

TABLE I
Synthetic network generation parameters & change scenarios

Model	Baseline Params	Change Params	Length	Nodes
SBM	$n_{\text{blocks}} = 2$ $p_{\text{intra}} = 0.95$ $p_{\text{inter}} = 0.01$	$p_{\text{intra}} \in [0.3, 0.95]$ $p_{\text{inter}} \in [0.01, 0.3]$	200	50
ER	$p = 0.05$	$p \in [0.05, 0.4]$	200	50
BA	$m = 1$	$m \in [1, 6]$	200	50
NWS	$k = 6$ $p_{\text{rewire}} = 0.1$	$k \in [4, 8]$ $p_{\text{rewire}} \in [0.05, 0.15]$	200	50

For real-world validation, we use the MIT Reality Dataset [3], which captures Bluetooth proximity interactions among

[1]Code and data are available at GitHub.

$N_{\text{students}} = 94$ dormitory students over $T_{\text{MIT}} = 289$ days (September 2008 to June 2009). This dataset provides ground truth annotations for significant academic events (holiday periods, semester transitions) that manifest as observable network structural changes due to student mobility patterns and social interaction shifts (Table II). We construct daily network snapshots where edges represent proximity interactions within 24-hour windows.

TABLE II
MIT REALITY DATASET ACADEMIC EVENTS

Day Index	Date	Academic Event
23	2008-10-12	Columbus Day / Fall Break
68	2008-11-26	Thanksgiving Holiday
94-100	2008-12-25 / 2009-01-01	Christmas / New Year
173	2009-03-15	Spring Break
234	2009-05-15	End of Spring Semester

All experimental results in this paper are computed over $N_{\text{trials}} = 10$ independent trials for statistical robustness. We systematically evaluate prediction horizon effects by varying $h \in \{1, 3, 5, 7, 10\}$ and find optimal performance at $h = 5$, where forecast accuracy balances with detection gain (shorter horizons provide limited predictive advantage while longer horizons accumulate forecast errors). We evaluate detection performance using three standard metrics:

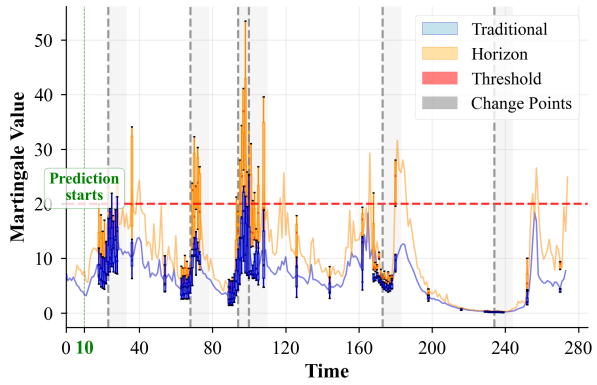

Fig. 1. MIT Reality Dataset detection performance comparison. Martingale (blue) vs. Horizon martingale (orange) over 289 days. Gray dashed vertical lines mark ground truth academic events (holidays, semester transitions), red horizontal line shows detection threshold $\lambda = 20$. Horizon approach achieves 22% average delay reduction (14.4 to 11.2 days) while improving TPR from 88% to 96%.

- **True Positive Rate (TPR)**: Proportion of actual change points successfully identified within a detection window:

$$\text{TPR} = \frac{|\{i : \exists j \text{ s.t. } \tau_j \in [t_i, t_i + \Delta]\}|}{|\{t_i\}|} \tag{26}$$

where $\{t_i\}$ are actual change times, $\{\tau_j\}$ are detection times, and $\Delta = 20$ timesteps is the maximum acceptable detection delay parameter. Sensitivity analysis reveals that smaller windows ($\Delta = 10$) increase precision but reduce recall, while larger windows ($\Delta = 30$) show diminishing returns in detection gain, confirming $\Delta = 20$ as a balanced choice.

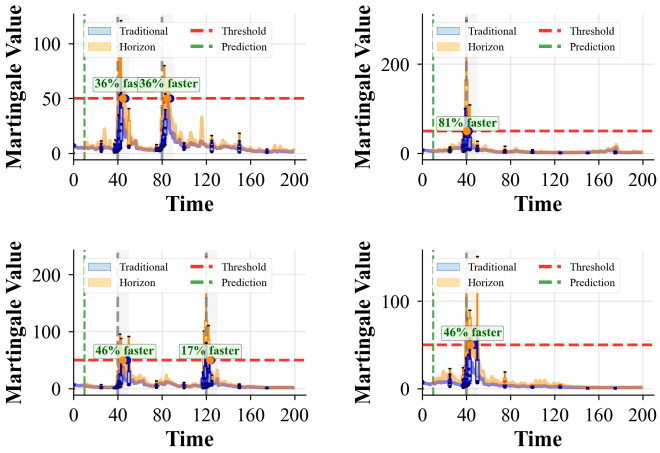

Fig. 2. Network-specific martingale detection performance across synthetic models at $\lambda = 50$. (Top-left) SBM with change points at $t \in \{40, 80\}$ achieving 36% delay reduction; (Top-right) ER with change point at $t = 40$ achieving 81% reduction; (Bottom-left) BA with change points at $t \in \{40, 120\}$ achieving 46% and 17% reduction respectively; (Bottom-right) NWS with change point at $t = 40$ achieving 46% reduction. Horizon approach consistently outperforms martingale detection across all network types.

- **False Positive Rate (FPR)**: Proportion of non-change periods incorrectly flagged as containing changes:

$$\text{FPR} = \frac{|\{\tau_j : \tau_j \notin \bigcup_i [t_i, t_i + \Delta]\}|}{T - |\{t_i\}| \cdot (\Delta + 1)} \tag{27}$$

where T is the total sequence length and the denominator represents the number of non-change timesteps.
- **Average Detection Delay (ADD)**: Mean time elapsed between change-points and successful detections:

$$\text{ADD} = \frac{1}{|\mathcal{D}|} \sum_{(i,j) \in \mathcal{D}} (\tau_j - t_i) \tag{28}$$

where $\mathcal{D} = \{(i, j) : \tau_j \in [t_i, t_i + \Delta]\}$ is the set of correctly matched change-detection pairs, with t_i representing the i-th true change time and τ_j the j-th detection time.

A. Baseline Methods

We compare this approach against the following methods:

- **CUSUM** [7]: A parametric approach that accumulates evidence of distributional shifts by tracking the cumulative sum of log-likelihood ratios. We implement CUSUM with control limit $h = 5$ (detection threshold) and reference value $k = 0.5$ (slack parameter), optimized for each network type.
- **EWMA** [8]: Exponentially Weighted Moving Average (EWMA) applies temporal smoothing with exponentially decaying weights ($\lambda_{\text{EWMA}} = 0.3$ smoothing parameter and control limit $L = 3$ for detection threshold).
- **Martingale** [12]: The basic martingale approach without horizon enhancement, serving as an ablation study to isolate the contribution of horizon prediction. This baseline uses the same multi-feature framework and betting

functions but accumulates evidence only from current observations, not predicted future states.

B. Parameter Sensitivity Analysis

We identify optimal configurations for different network types through a comprehensive grid search. Table III summarizes the findings on parameter effectiveness across network structures, with the best overall parameters highlighted in gray.

Table III reveals that power mixture betting achieves perfect TPR (1.000) for SBM while maintaining zero FPR, and achieves strong performance across all network types (TPR: 0.750–0.950). Beta betting shows consistently poor performance with TPR values below 0.900 and often near zero. For power betting, optimal ϵ values vary: 0.7 works well for SBM and ER (TPR: 0.950–1.000), while 0.9 is effective for NWS (TPR: 0.900). The globally optimal configuration is mixture power betting with Mahalanobis distance and threshold $\lambda = 50$, which provides the best balance of high TPR and low FPR across network types. For the MIT Reality dataset, we use $\lambda = 20$ as it achieves superior performance (TPR=0.792) for real-world academic event detection.

In Table IV, we design specific change scenarios for each network model and evaluate detection performance using the best combination of parameters found in Table III (i.e., Mixture power betting, Mahalanobis distance, and $\lambda = 50$).

SBM scenarios target community dynamics: *Community Merge* transitions from strong isolated communities ($p_{\text{intra}} = 0.95$, $p_{\text{inter}} = 0.01$) to weaker, more interconnected structures (increasing p_{inter} up to 0.3), simulating organizational restructuring; *Density Change* modifies overall connectivity while preserving community structure; *Mixed Changes* combines multiple simultaneous structural shifts with two change-points.

ER scenarios focus on global connectivity: *Density Increase* represents systematic network growth from sparse ($p = 0.05$) to dense ($p = 0.4$) configurations; *Density Decrease* captures network pruning in the reverse direction.

BA scenarios examine hub formation: *Parameter Shift* changes preferential attachment strength (m from baseline 1 to range 1–6), altering hub prominence; *Hub Addition* dramatically increases hub formation from tree-like structures ($m = 1$) to highly centralized networks ($m = 6$).

NWS scenarios manipulate small-world properties: *Rewiring Increase* transitions from regular lattice structure to random connectivity (rewiring probability from 0.05 to 0.15), destroying small-world characteristics; *K Parameter Shift* changes neighborhood size (from 4 to 8 nearest neighbors), affecting local clustering while maintaining rewiring probability.

VI. Results and Discussion

Table IV demonstrates several important patterns:

1) **Network-Specific Performance:** Complex networks (SBM, NWS) show greater benefits from prediction-enhanced detection. Table IV shows delay reductions of 13.1–23.8% for SBM and 21.2–24.8% for NWS, compared to 16.9% for ER and 14.5% for BA networks.
2) **Distance Metric Selection:** Network structural properties dictate optimal distance metrics, though performance varies significantly. Mahalanobis distance shows moderate performance across networks (TPR: 0.275–0.535), while Euclidean (TPR: 0.358–0.552) and Cosine (TPR: 0.335–0.547) performs similarly. Chebyshev achieves competitive performance for some networks (TPR up to 0.475) with consistently low ADD values.
3) **Threshold Configuration:** Lower thresholds ($\lambda = 20$) generally achieve higher TPR (0.750–0.950) but may increase FPR, while higher thresholds ($\lambda = 100$) provide more conservative detection (TPR: 0.542–0.775) with better FPR control. For the comparative analysis in Table IV, we employed the optimal configuration identified from Table III (mixture betting with Mahalanobis distance and $\lambda = 50$) across all synthetic networks. However, for the MIT Reality dataset, we used $\lambda = 20$ as it achieved superior performance (TPR=0.792) for real-world academic event detection.
4) **Feature Attribution:** The Martingale-Shapley equivalence provides precise attribution of detected changes. Table IV shows that spectral features dominate detection for most synthetic networks (contributing 20.2–35.4%), while betweenness centrality drives detection in real-world academic events (21.6%).
5) **Detection Improvement:** Horizon martingales consistently achieve earlier detection while maintaining statistical guarantees. Table IV shows delay reductions ranging from 13.1% to 24.8% across network types, with TPR improvements up to 7.9% for real-world data.

VII. Conclusions and Future Work

We address early detection in dynamic networks through horizon martingales that accumulate evidence from predicted future states. Our key results establish that horizon martingales preserve the martingale property (Theorem 3) and maintain strict false alarm control (Theorem 4). The framework maintains interpretability through the Martingale-Shapley equivalence and achieves 13–25% delay reduction with TPR > 0.88 and FPR bounded by $1/\lambda$.

The computational complexity is $O(K \cdot T)$ for martingale updates plus $O(n^3 \cdot T)$ for feature extraction, where the latter dominates for large networks. Future work should explore: (1) integrating advanced forecasting models (graph neural networks, temporal transformers) to improve prediction accuracy beyond current EWMA-based results; (2) developing adaptive parameter selection algorithms that dynamically adjust ϵ values based on network state; (3) extending to continuous-time networks and attributed graphs; (4) investigating hybrid approaches combining martingales with e-processes for enhanced detection power; (5) addressing scalability for large networks through distributed implementations; (6) developing privacy-preserving variants for sensitive applications; and (7) optimizing reset strategies to balance detection sensitivity with false alarm control, leveraging the multiple detection framework established in Theorem 4.

TABLE III

PARAMETER SENSITIVITY ANALYSIS SHOWING RELATIVE EFFECTIVENESS OF DIFFERENT CONFIGURATIONS ACROSS NETWORK TYPES

Network	**Metric**	**Power Betting**					**Beta Betting** (a, b)			**Distance Metric**				**Threshold** (λ)		
		ϵ=0.2	ϵ=0.5	ϵ=0.7	ϵ=0.9	**Mixture**	(0.2,2.5)	(0.4,1.8)	(0.6,1.2)	**Euc.**	**Mah.**	**Cos.**	**Cheb.**	**20**	**50**	**100**
SBM	TPR	0.000	0.925	0.950	0.875	1.000	0.000	0.225	0.750	0.552	0.535	0.547	0.475	0.950	0.938	0.775
	FPR	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.005	0.005	0.008	0.003	0.000	0.000	0.000
	ADD	–	12.8	9.4	13.2	7.7	–	11.6	3.8	8.2	14.1	9.0	7.0	14.2	6.7	12.8
ER	TPR	0.000	0.825	1.000	1.000	0.925	0.000	0.125	0.900	0.470	0.487	0.465	0.425	0.750	0.863	0.600
	FPR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.040	0.013	0.050	0.000	0.000	0.000	0.000
	ADD	–	15.0	5.9	11.9	9.2	–	5.6	13.0	13.2	10.8	12.5	12.9	16.4	16.6	13.2
BA	TPR	0.000	0.575	1.000	0.625	0.750	0.000	0.062	0.850	0.507	0.485	0.335	0.460	0.750	0.862	0.725
	FPR	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.000	0.013	0.017	0.057	0.000	0.037	0.050	0.000
	ADD	–	8.0	9.2	14.3	12.2	–	18.5	17.9	11.4	13.2	9.0	5.1	18.3	18.8	11.0
NWS	TPR	0.000	0.725	0.550	0.900	0.950	0.000	0.025	0.838	0.362	0.455	0.443	0.463	0.775	0.825	0.738
	FPR	0.000	0.000	0.050	0.000	0.000	0.000	0.000	0.013	0.037	0.025	0.003	0.010	0.050	0.000	0.013
	ADD	–	13.0	8.2	14.2	11.7	–	20.0	15.4	14.0	10.9	4.9	10.6	12.4	19.1	10.6
MIT	TPR	0.000	0.625	0.875	0.333	0.875	0.000	0.042	0.583	0.358	0.275	0.450	0.275	**0.792**	0.625	0.542
	FPR	0.000	0.113	0.201	0.181	0.201	0.000	0.000	0.088	0.068	0.056	0.082	0.056	**0.157**	0.123	0.069
	ADD	–	21.5	14.7	14.5	17.1	–	17.0	21.5	18.4	18.2	13.4	18.2	**18.3**	20.6	21.6

TABLE IV

DETECTION DELAY REDUCTION (↓) AND TPR IMPROVEMENT (↑) COMPARE HORIZON MARTINGALE TO MARTINGALE

Network Type	**Martingale**			**Horizon Martingale**			**CUSUM**			**EWMA**			**Improvements**		**Attribution**	
	TPR	**FPR**	**ADD**	**TPR**	**FPR**	**ADD**	**TPR**	**FPR**	**ADD**	**TPR**	**FPR**	**ADD**	**Delay ↓**	**TPR ↑**	ψ_k	**Driver**
Stochastic Block Model (SBM)																
Community Merge	1.000	0.000	9.9	1.000	0.000	8.6	1.000	0.237	9.1	1.000	0.087	7.5	13.1%	0.0%	20.7%	Spectral
Density Change	1.000	0.000	6.3	1.000	0.001	4.8	1.000	0.231	9.5	1.000	0.016	5.4	23.8%	0.0%	27.8%	Spectral
Mixed Changes	0.98	0.001	10.9	1.000	0.001	8.4	1.000	0.125	9.5	1.000	0.019	5.8	22.9%	2.0%	34.8%	Clustering
Erdős-Rényi (ER)																
Density Increase	0.97	0.001	13.0	1.000	0.001	10.8	1.000	0.200	9.2	1.000	0.006	4.8	16.9%	2.9%	20.2%	Spectral
Density Decrease	0.96	0.001	13.0	1.000	0.001	10.8	1.000	0.200	9.2	1.000	0.006	4.8	16.9%	4.0%	20.2%	Spectral
Barabási-Albert (BA)																
Parameter Shift	0.94	0.003	5.5	1.000	0.002	4.7	1.000	0.128	9.3	1.000	0.013	6.5	14.5%	5.9%	35.4%	Spectral
Hub Addition	0.97	0.002	5.5	1.000	0.000	4.7	1.000	0.331	9.5	1.000	0.028	4.9	14.5%	2.9%	35.4%	Spectral
Newman-Watts-Strogatz (NWS)																
Rewiring Increase	1.000	0.000	11.8	1.000	0.001	9.3	1.000	0.275	10.2	1.000	0.009	4.5	21.2%	0.0%	37.6%	Closeness
k Parameter Shift	1.000	0.000	12.1	1.000	0.000	9.1	1.000	0.194	10.6	1.000	0.012	4.5	24.8%	0.0%	55.7%	Laplacian
MIT Reality Dataset (using $\lambda = 20$)																
Academic Events	0.88	0.040	14.4	0.96	0.070	11.2	0.833	0.607	7.6	1.000	0.549	6.5	22.2%	7.9%	21.6%	Betweenness

REFERENCES

[1] C. D. Barros, M. R. Mendonça, A. B. Vieira, and A. Ziviani, "A survey on embedding dynamic graphs," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–37, 2021.

[2] L. Peel and A. Clauset, "Detecting change points in the large-scale structure of evolving networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.

[3] A. Madan, M. Cebrian, S. Moturu, K. Farrahi *et al.*, "Sensing the" health state" of a community," *IEEE Pervasive Computing*, vol. 11, no. 4, pp. 36–45, 2011.

[4] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: a survey," *Data mining and knowledge discovery*, vol. 29, pp. 626–688, 2015.

[5] D. Eswaran, C. Faloutsos, S. Guha, and N. Mishra, "Spotlight: Detecting anomalies in streaming graphs," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1378–1386.

[6] S. Huang, Y. Hitti, G. Rabusseau, and R. Rabbany, "Laplacian change point detection for dynamic graphs," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 349–358.

[7] E. Page, "Cumulative sum charts," *Technometrics*, vol. 3, no. 1, pp. 1–9, 1961.

[8] S. Roberts, "Control chart tests based on geometric moving averages," *Technometrics*, vol. 1, no. 3, pp. 239–250, 1959.

[9] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[10] Y. Wang, A. Chakrabarti, D. Sivakoff, and S. Parthasarathy, "Fast change point detection on dynamic social networks," in *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017, pp. 2992–2998.

[11] D. Koutra, N. Shah, J. T. Vogelstein, B. Gallagher, and C. Faloutsos, "Deltacon: Principled massive-graph similarity function with attribution," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 3, pp. 1–43, 2016.

[12] S.-S. Ho, T. T. Kairamdonda, and I. Ali, "Detecting and explaining structural changes in an evolving graph using a martingale," *Pattern Recognition*, p. 111855, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320325005151

[13] I. McCulloh and K. M. Carley, "Detecting change in longitudinal social networks," *Journal of social structure*, vol. 12, no. 3, pp. 1–37, 2011.

[14] G. E. Box and G. M. Jenkins, *Time series analysis: forecasting and control*. Holden-day, 1970.

[15] S.-S. Ho and H. Wechsler, "A martingale framework for detecting changes in data streams by testing exchangeability," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, 2010.

[16] D. Volkhonskiy, E. Burnaev, I. Nouretdinov, A. Gammerman, and V. Vovk, "Inductive conformal martingales for change-point detection," in *Conformal and Probabilistic Prediction and Applications*. PMLR, 2017, pp. 132–153.

[17] J. Ville, *Étude critique de la notion de collectif*. Paris: Gauthier-Villars, 1939.

[18] S. M. Lundberg and S.-I. Lee, "An approach to interpreting model predictions," *Advances in neural information processing systems*, 2017.

[19] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983.

[20] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[21] P. Erdős, A. Rényi *et al.*, "On the evolution of random graphs," *Publ. math. inst. hung. acad. sci*, vol. 5, no. 1, pp. 17–60, 1960.

[22] M. E. Newman and D. J. Watts, "Renormalization group analysis of the small-world network model," *Physics Letters A*, vol. 263, no. 4-6, pp. 341–346, 1999.