Learning Changes in Graphon Attachment Network Models

Xinyuan Fan [1] **Bufan Li** [1] **Chenlei Leng** [2] **Weichi Wu** [1]

Abstract

This paper introduces Graphon Attachment Network Models (GAN-M), a novel framework for modeling evolving networks with rich structural dependencies, grounded in graphon theory. GAN-M provides a flexible and interpretable foundation for studying network formation by leveraging graphon functions to define attachment probabilities, thereby combining the strengths of graphons with a temporal perspective. A key contribution of this work is a methodology for learning structural changes in these networks over time. Our approach uses graph counts—frequencies of substructures such as triangles and stars—to capture shifts in network topology. We propose a new statistic designed to learn changes in the resulting piecewise polynomial signals and develop an efficient method for change detection, supported by theoretical guarantees. Numerical experiments demonstrate the effectiveness of our approach across various network settings, highlighting its potential for dynamic network analysis.

1. Introduction

The learning, understanding, and application of networks have become increasingly crucial with the exponential growth of datasets represented as networks or graphs. The rise of large-scale networks has driven the development of mathematical frameworks for analyzing and interpreting these structures, with graph theory playing a central role (Lovász, 2012). A significant breakthrough in this area is the graphon framework—short for "graph functions"—which provides a unified and powerful tool for understanding large networks. This framework is particularly appealing due to its functional characterization of network limits, represented as a bivariate function that captures the asymptotic properties of graphs. However, traditional graphon models are inherently static, limiting their applicability to dynamic systems where the network structure evolves over time.

A defining characteristic of many real-world networks is their dynamic nature, where new nodes are introduced sequentially over time, leading to a final network with an inherent node registration order. Such networks naturally arise in various contexts, including the World Wide Web, email communication, paper citation networks, and more, where new participants continuously join the system. For a comprehensive survey on dynamic graph models, we refer to Zaki et al. (2016). A widely studied model for describing this attachment process is the preferential attachment model (Barabási & Albert, 1999). In this model, new nodes are more likely to connect to existing nodes with higher degrees, encapsulating the "rich get richer" phenomenon. This mechanism produces networks with degree distributions that follow power laws, making it particularly suitable for modeling certain types of real-world networks. Nonetheless, preferential attachment focuses primarily on degree heterogeneity and overlooks higher-order structural dependencies or temporal changes in the underlying network properties.

Motivated by the complementary strengths of graphons for characterizing large networks and attachment models for describing growth processes, we propose a novel framework called the Graphon Attachment Network Model (GAN-M). GAN-M models the attachment process using time-evolving graphons, effectively bridging the gap between static graphon theory and dynamic network growth. A defining advantage of this framework lies in flexibility: when the underlying graphon remains constant, the resulting network conforms to a standard graphon model. However, when the graphon evolves over time, the framework introduces structural changes, enabling the modeling of more complex and realistic network dynamics. This capability provides a unified approach to analyzing networks, addressing both static and dynamic regimes, and opens the door to modeling a broader range of real-world applications, such as evolving citation networks or online social platforms.

Building on this foundation, we address the challenge of learning changes in the dynamic process of graphons. Change detection in such dynamic models has far-reaching implications for network analysis and has garnered increas-

[1]Department of Statistics and Data Science, Tsinghua University, Beijing, China [2]Department of Statistics, University of Warwick, Coventry, UK. Correspondence to: Weichi Wu <wuweichi@mail.tsinghua.edu.cn>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

ing attention in recent years. For example, within the preferential attachment framework, several studies have investigated change-point detection (Bhamidi et al., 2018; Bet et al., 2023; Kaddouri et al., 2024). Similarly, learning change points in time series of networks has been explored extensively (Lévy-Leduc & Roueff, 2009; Peel & Clauset, 2015; Wang et al., 2021; Zhang et al., 2024; Fan & Wu, 2024). We refer to Aminikhanghahi & Cook (2017); Chen & Chu (2023); Zhou et al. (2024) for reviews on change-point detection methods. However, these approaches are largely confined to static networks or rely on assumptions of independence, making them unsuitable for dynamic networks characterized by cumulative growth and evolving graphon structures. Our work extends these efforts by developing methodologies tailored to the GAN-M framework, providing a novel perspective on dynamic networks and their structural evolution.

From a methodological perspective, classical change-point detection methods such as CUSUM and MOSUM are not directly applicable to our setting due to three fundamental challenges: (i) the adjacency matrices of the networks vary in size at different time points, reflecting the dynamic growth of the network, (ii) all edges in G_t persist in $G_{t'}$ for $t' \geq t$, leading to cumulative and irreversible network growth over time, and (iii) At the time points after the first change, the networks are generated by a mixture of multiple graphon functions. These properties fundamentally distinguish our problem from classical change-point detection tasks, making it difficult to find suitable benchmark methods in simulations, and necessitate new methodological developments.

To address these challenges, we focus on subgraph counts, a well-established tool in network analysis (Bickel et al., 2011; Shao et al., 2022). A *key conceptual contribution* of our work, as established in Lemma 2.6, is that the expected counts of fixed subgraphs in GAN-M evolve as piecewise polynomial functions, where the order of the polynomial is determined by the size of the subgraph. This critical observation reframes the problem of learning structural changes as identifying shifts in the coefficients of a piecewise polynomial mean function, which represents a significant departure from classical approaches.

However, learning changes in piecewise polynomial trends presents a unique challenge: classical CUSUM or MOSUM statistics are designed for mean shifts in independent or weakly dependent data and do not account for the evolving polynomial structure. Moreover, subgraph counts across time points exhibit strong dependencies, as they are influenced by the cumulative growth of the network. This strong dependence invalidates the independence assumptions underpinning most existing methods, rendering them unsuitable for our context.

To overcome these obstacles, we propose a *novel extension* of the CUSUM framework tailored to polynomial trends in highly dependent data. This statistic, which we term **WESUM** (Weighted Sum), incorporates weights to effectively account for the polynomial nature of subgraph evolution and the dependencies across time points. A key strength of WESUM lies in its robustness to dependence, making it well suited for dynamic network settings.

We provide rigorous theoretical guarantees for WESUM, including detection bounds and consistency for change-point estimation, and establish rates that highlight its efficiency in dynamic network analysis. Extensive simulation studies validate the effectiveness of our approach, demonstrating its *superiority* in identifying structural changes in dynamic and evolving networks compared to classical methods. By bridging graphon theory with dynamic network modeling, our work introduces a fundamentally new class of change-point detection methodologies, providing powerful tools for understanding and analyzing complex network evolution.

The remainder of this paper is structured as follows. Section 2 outlines the problem setup, providing a detailed description of the graphon attachment network with change-points and its subgraph count properties. In Section 3, we introduce our proposed statistics and present the corresponding theoretical results. Section 4 presents the findings from our simulation studies. Section 6 concludes the paper and discusses potential directions for future research. Finally, detailed algorithms, simulation settings, and all technical proofs are provided in the appendix.

Notations: For a graph G, let $V(G)$ and $E(G)$ denote its set of nodes and edges, respectively. For a vector $x \in \mathbb{R}^p$, the norm is defined as $\|x\| = \left(\sum_{i=1}^p x_i^2\right)^{1/2}$. For a set S, $|S|$ denotes the number of elements in S. For a real number x, $\lceil x \rceil$ represents the smallest integer greater than or equal to x, and $\lfloor x \rfloor$ represents the largest integer smaller than or equal to x. Let $I(\cdot)$ denote the indicator function. For two positive real numbers a and b, we write $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. For two sequences of positive real numbers a_n and b_n, we say $a_n = O(b_n)$ if there exist positive constants N and C such that $a_n/b_n \leq C$ for $n > N$ and $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$.

2. Problem Setup

We introduce the Graphon Attachment Network Model (GAN-M), which combines the dynamic growth process of networks with the flexibility of graphon-based attachment probabilities.

Definition 2.1 (Graphon Attachment Network Model)**.** Let G_t denote the graph at time t, with A_t as its adjacency matrix, for $1 \leq t \leq T$. The model begins with an initial simple graph G_1 containing m nodes ($m \geq 1$), where each

node is assigned a random value $U_1, U_2, \ldots, U_m \sim U(0,1)$, drawn independently.

For $t > 1$, the graph G_t is formed recursively. Given the values $U_1, \ldots, U_{m+t-2}$ and the adjacency matrix A_{t-1}, the adjacency matrix A_t is updated by attachment as follows:

- Preserve all existing edges: for $1 \le i, j \le m+t-2$, set $A_t(i,j) = A_{t-1}(i,j)$.
- Assign $A_t(m+t-1, m+t-1) = 0$, ensuring no self-loops for the new node.
- Draw $U_{m+t-1} \sim U(0,1)$, independently of all random variables generated in the previous steps.
- Form edges between the new node $j := m+t-1$ and each existing node i ($1 \le i \le m+t-2$) independently, with probability:

$$\mathbb{P}(A_t(i,j) = 1 \mid U_i, U_j) = h_{T,t}(U_i, U_j),$$

where $h_{T,t} : [0,1]^2 \to [0,1]$ is a symmetric measurable function defining the attachment probability.

Definition 2.1 focuses on undirected simple graphs but can be readily extended to directed graphs with minimal modifications. The GAN-M framework models the attachment process such that new edges form only between the new node and existing nodes, reflecting a growth mechanism similar to the preferential attachment model. However, unlike the preferential attachment model, GAN-M incorporates a graphon-based approach, where the symmetric measurable function $h_{T,t}$, referred to as a *graphon*, governs the attachment probabilities. This integration allows GAN-M to capture complex and evolving network structures effectively.

The GAN-M framework generalizes several well-established network models, offering significant flexibility. Specifically, when $h_{T,t} = h$ for a fixed function h, GAN-M reduces to the traditional graphon network (Lovász, 2012). Furthermore, by setting $h_{T,t} = \rho_T h$, where $\rho_T \to 0$, the model generates a sparse graphon network. Additionally, the model encompasses L_p-graphons as introduced by Borgs et al. (2019; 2018). The flexibility of GAN-M is further enhanced by allowing the initial graph to be any simple finite graph, enabling the network to evolve from a wide range of starting configurations. This feature accommodates various real-world scenarios by incorporating pre-existing structures. A similar approach, where the network starts with a seeded graph, is also employed in the preferential attachment model (Bubeck et al., 2015; Hormozdiari et al., 2007).

Remark 2.2. The fitness model (Caldarelli et al., 2002) shares certain connections with the graphon model. In the fitness model, the link function can depend not only on the fitness values of nodes i and j but also on additional factors, such as the maximum fitness in the entire network or a threshold related to the network size. In this paper, we ground our model within graphon theory, as the graphon function can be viewed as a graph limit (Lovász, 2012). Moreover, our GAN-M extends the graphon model in two key aspects: (i) the graphon function is allowed to depend on T, t; and (ii) GAN-M allows the specification of an initial sample graph G_1.

A notable strength of the GAN-M lies in its triangular array structure, which allows $h_{T,t}$ to adapt to the final network size T. This adaptability enables GAN-M to generate networks with pronounced degree heterogeneity, where the maximum, average, and minimum node degrees differ substantially. Such heterogeneity reflects real-world network characteristics (Ke & Wang, 2024). For example, in Scenario 1 of Section A.2, we set $h_{T,t}(x,y) = T^{-0.6}(xy)^{-0.9} \wedge 1$, showcasing the model's ability to produce networks with diverse degree distributions. This versatility underscores the model's capacity to represent complex network dynamics and structural diversity.

A key innovation in GAN-M is the evolution of the graphon $h_{T,t}$ over time, allowing attachment probabilities to change as the network grows. This raises a fundamental question:

Do structural changes in the function $h_{T,t}$ result in shifts in the stochastic behavior of G_t and vice versa?

In real-world networks, such changes often signal shifts in underlying dynamics, such as variations in user behavior, modifications to system policies, or external influences on network growth. Learning and interpreting these changes is crucial for applications such as social network analysis, epidemic modeling, and communication networks, where the nature of connections evolves over time.

Thus, we can formalize the change detection problem as follows. Assume the existence of time points $\eta = \{\eta_0, \ldots, \eta_{K+1}\}$, where $1 = \eta_0 < \eta_1 < \cdots < \eta_K < \eta_{K+1} = T+1$, at which the graphon functions $h_{T,t}$ undergo changes. Specifically, $h_{T,t}$ differs from $h_{T,t-1}$ if and only if $t = \eta_k$ for some $1 \le k \le K$. The precise notion of "difference" between graphon functions is clarified in the following remark.

Remark 2.3 (Identifiability of a graphon). Identifiability issues are inherent in any graphon model. For instance, consider two graphons: $h_1(x,y) = xy$ and $h_2(x,y) = (1-x)(1-y)$. Although $h_1 \neq h_2$ for almost all $(x,y) \in [0,1]^2$, they are "essentially the same" because the random variables U_i and $1-U_i$ are interchangeable under measure-preserving transformations. More formally, let $\varphi : [0,1] \to [0,1]$ be a measure-preserving map, and define $h_1^{\varphi}(x,y) = h_1(\varphi(x), \varphi(y))$. Then, for every graph H,

the homomorphism densities $\hom(H, h_1^{\varphi})$ and $\hom(H, h_1)$ are identical, where $\hom(H, h_1)$ is the homomorphism density defined in (1) in Section 2.1. Consequently, the two graphons are indistinguishable.

Following Lovász (2012), we measure the difference between two graphons h_1 and h_2 using the cut norm:

$$\delta_{\square}(h_1, h_2) = \inf_{\varphi \in \Phi} \sup_{S,T \subset [0,1]} \left| \int_{S \times T} [h_1(x, y) - h_2(\varphi(x), \varphi(y))] \, dx \, dy \right|,$$

where Φ denotes the set of all invertible measure-preserving maps. We say that h_1 and h_2 differ if $\delta_{\square}(h_1, h_2) \neq 0$.

Thus, when we state that $h_{T,t}$ differs from $h_{T,t-1}$, we mean that $\delta_{\square}(h_{T,t}, h_{T,t-1}) \neq 0$.

Remark 2.4 (Networks observed within intervals). In the definition of the GAN-M, we assume that the time stamps of each network G_t are available; however, this may not always be the case in practice. In many real-world scenarios, networks are observed at fixed intervals, with multiple nodes added after each period. In such cases, we can treat these newly added nodes as belonging to the same time step t. We can then reorder these nodes arbitrarily, "unfold" them for analysis, and later "collapse" them back after learning the point of change. This approach is illustrated in Figure 1. Since the localization error in change-point detection cannot be smaller than $O_p(1)$, the theoretical results presented in this paper hold as long as the number of nodes entering the network at each time step (which may vary) remains bounded. We also note that a similar collapsing technique is employed in the preferential attachment model and its associated branching process; for further details, see Garavaglia & van der Hofstad (2018) and Garavaglia et al. (2022).

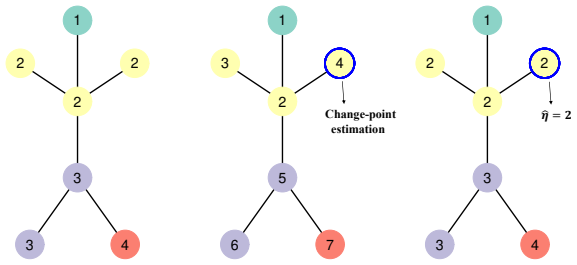

Figure 1. A diagram illustrating the process of the "unfold and collapse" method when multiple nodes enter the network at the same time.

2.1. Subgraph Counts

Our objective is to learn K, the number of change points, and η, their locations, from the final observation G_T and its adjacency matrix A_T. Intuitively, a change in G_t corresponds to a change in A_t, which manifests in observable features of the network at time t. Therefore, it is natural to consider various statistics of the network, and the particular statistic we focus on is the moments of the network, specifically subgraph counts (Bickel et al., 2011).

For a given graphon $h_{T,t}$ and a deterministic simple graph H (e.g., $H =$), we define the homomorphism density as:

$$\hom(H, h_{T,t}) = \int_{[0,1]^{|V(H)|}} \prod_{ij \in E(H)} h_{T,t}(x_i, x_j) \prod_{i \in V(H)} dx_i, \tag{1}$$

where $V(H)$ and $E(H)$ represent the node and edge sets of H, respectively.

Next, suppose that G_t is a graphon attachment network at time t (where $t \geq |V(H)|$). We define the subgraph count of H in G_t as:

$$N_H(G_t) = \sum_{\varphi: V(H) \to V(G_t)} \prod_{uv \in E(H)} I(\varphi(u), \varphi(v) \in E(G_t)),$$

where φ denotes an injective mapping from $V(H)$ to $V(G_t)$.

Remark 2.5 (Cut Norm and Homomorphism Density). Subgraph counts provide a natural and effective approach for learning changes in our model, as they are directly related to the cut norm and homomorphism density (Lovász, 2012, Corollary 10.34 and Lemma 10.32). Specifically, the following properties hold:

1. $\delta_{\square}(h_{T,1}, h_{T,2}) = 0 \iff \hom(H, h_{T,1}) = \hom(H, h_{T,2})$ for every simple graph H.
2. Let k be a positive integer. If for every simple graph H with k nodes, we have $|\hom(H, h_{T,1}) - \hom(H, h_{T,2})| \leq 2^{-k^2}$, then $\delta_{\square}(h_{T,1}, h_{T,2}) \leq \frac{50}{\sqrt{\log(k)}}$.

The subgraph counts $N_H(G_t)$ defined above satisfy the following key property, which is essential for the development of our main methodology:

Lemma 2.6. *In the context of this paper, suppose that $\eta_k \leq t < \eta_{k+1}$ for some $0 \leq k \leq K$. Then, we have*

$$\mathbb{E}[N_H(G_t)] = (t - \eta_k)^{|V(H)|} \hom(H, h_{T,\eta_{k+1}}) + Q_{|V(H)|-1}(t),$$

where $Q_{|V(H)|-1}(t)$ is a polynomial of degree $|V(H)| - 1$ in t.

This lemma is central to our approach, as it provides a clear and precise approximation for the expected subgraph counts at each time step t, accounting for potential structural changes in the network. The polynomial term $Q_{|V(H)|-1}(t)$ captures the gradual transition in network structure, while the leading term $(t - \eta_k)^{|V(H)|} \hom(H, h_{T,\eta_{k+1}})$ directly links the subgraph counts to the change in the graphon function $h_{T,t}$ at the change points. This relationship is pivotal in learning structural changes in the graph over time.

2.2. Why Not CUSUM?

CUSUM is a popular method for change-point detection. Let us first review the CUSUM statistic. Consider a sample X_r, where $r \in (s, e]$ is an integer indexing the time of X. To identify a change point within $(s, e]$, the CUSUM procedure constructs a statistic of the form

$$\tilde{X}^t_{s,e} = \frac{1}{\sqrt{e-s}} \sum_{r=s+1}^{t} \sqrt{\frac{e-t}{t-s}} X_r - \sum_{r=t+1}^{e} \sqrt{\frac{t-s}{e-t}} X_r. \tag{2}$$

The appeal of the CUSUM procedure lies in its ability to identify a time point $t \in (s, e]$ such that $|\mathbb{E}\tilde{X}^t_{s,e}|$ is maximized. Moreover, the method is highly visualizable, making it intuitive and easy to interpret in practical applications.

However, the CUSUM method is not applicable in the context of this paper. First, the size of G_t increases over time, which makes it impossible to directly use the adjacency matrix A_r as X_r in (2). Secondly, the CUSUM statistic is designed to detect changes in piecewise constant signals, while our subgraph count $N_H(G_r)$ has a piecewise polynomial expectation. Direct application of CUSUM would lead to failure, as the population version of the statistic at the true change-point location is no longer maximized. Therefore, it is necessary to develop a new statistic tailored for learning changes in piecewise polynomial signals, while retaining the structure and advantages of the CUSUM framework.

3. Learning Changes via WESUM

We introduce the **WESUM** (Weighted SUM) statistic, a novel extension of the classical CUSUM framework, specifically designed to address the challenges posed by graphon attachment networks. Unlike standard methods, WESUM utilizes subgraph counts $N_H(G_t)$ to learn structural changes in the dynamic and cumulative growth of networks. By handling issues such as evolving network sizes and strong dependencies, WESUM provides an interpretable and efficient solution, marking a significant advancement in the analysis of time-evolving graphons.

3.1. WESUM

The construction of WESUM is guided by two key principles. First, the coefficients of weights must be orthogonal to the polynomial vector up to order $|V(H)|$, ensuring that polynomial trends are removed. Second, the statistic should attain its maximum at the change point, allowing for precise localization. With these objectives in mind, we define the statistic as follows:

$$\tilde{X}^t_{s,e}(H) = \sqrt{\sum_{k=0}^{|V(H)|} \left(\tilde{X}^{t,k}_{s,e}(H)\right)^2}, \tag{3}$$

where

$$\tilde{X}^{t,k}_{s,e}(H) = \sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r) N_H(G_r), \quad 0 \le k \le |V(H)|, \tag{4}$$

and the coefficients $\{\tilde{w}^{t,k}_{s,e}(r)\}_{r=s+1}^{e}$ are determined through a projection procedure. For clarity, the subscripts for s, t, and e are omitted in the following:

$$a_k = (0, \cdots, 0, 1^k, 2^k, \cdots, (e-t)^k)^\top,$$

$$U_0 = \begin{pmatrix} 1 & s+1 & (s+1)^2 & \cdots & (s+1)^{|V(H)|} \\ 1 & s+2 & (s+2)^2 & \cdots & (s+2)^{|V(H)|} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & e & e^2 & \cdots & e^{|V(H)|} \end{pmatrix},$$

$$U_k = (U_{k-1}, a_{k-1}),$$

$$P_{U_k} = U_k (U_k^\top U_k)^{-1} U_k^\top,$$

$$W_k = (I - P_{U_k}) a_k.$$

If $W_k \neq 0$, we set

$$(\tilde{w}^{t,k}_{s,e}(s+1), \cdots, \tilde{w}^{t,k}_{s,e}(e))^\top = \frac{W_k}{\sqrt{W_k^\top W_k}}, \tag{5}$$

otherwise, the vector is set to zero.

To understand the connection between the statistic in (3) and the classical CUSUM statistic, we summarize some key properties below. Detailed proofs are provided in the appendix:

1. **Reduction to CUSUM**: When $|V(H)| = 0$, the statistic $\tilde{X}^t_{s,e}(H)$ reduces to the classical CUSUM statistic, which is designed to learn changes in piecewise constant signals.

2. **Orthogonality of Coefficients**: The coefficients $\tilde{w}^{t,k}_{s,e}(r)$ satisfy:

$$\sum_{r=s+1}^{e} \left(\tilde{w}^{t,k}_{s,e}(r)\right)^2 = 1, \quad \sum_{r=s+1}^{e} r^m \tilde{w}^{t,k}_{s,e}(r) = 0,$$

for all $0 \le m \le |V(H)|$. These conditions ensure that the coefficients are orthogonal to polynomial trends up to order $|V(H)|$, thereby maintaining the classical CUSUM property.

3. **Maximum at Change Point**: When $(s, e]$ contains exactly one change point η_k, the population version of the statistic (denoted $\tilde{f}^t_{s,e}(H)$) satisfies:

$$\tilde{f}^{\eta_k}_{s,e}(H) = \max_{t \in (s,e]} \tilde{f}^t_{s,e}(H),$$

ensuring that the statistic highlights the change point, and the location can be precisely localized using $\arg\max$ in Algorithm 2 in Section A.1.

In summary, the WESUM statistic retains the interpretability and effectiveness of the classical CUSUM procedure, while extending its utility to higher-order trends and complex network structures.

3.2. Algorithms

Building on the construction of WESUM, we adapt standard multiple change-point detection algorithms, originally designed for the CUSUM statistic, by replacing it with our proposed statistic. Specifically, we employ the random interval distillation procedure from Fan & Wu (2024), along with the seeded intervals method from Kovács et al. (2023), to enhance computational efficiency. Detailed descriptions of these algorithms can be found in Algorithm 1 and Algorithm 2, provided in Appendix A.1.

In summary, the proposed change-point detection algorithm proceeds as follows:

1. Line 2-12 in Algorithm 1: Construct an initial set of intervals $I = \{(s_m, e_m]\}_{m=1}^{M}$ based on seeded intervals.

2. Line 13-18 in Algorithm 1: For each interval $(s_m, e_m] \in I$, compute the maximum value of the statistic, $\max_{t:s_m<t\le e_m} |\tilde{X}^t_{s_m,e_m}|$, and retain those intervals where the computed value exceeds a predetermined threshold τ.

3. Line 19-33 in Algorithm 1: Apply the distillation method from Fan & Wu (2024) to refine the retained intervals, resulting in a set of disjoint intervals $S^* = \{[l_k, r_k]\}_{k=1}^{\hat{K}}$. The cardinality $|S^*|$ learns the total number of change-points, $\hat{K}$.

4. Algorithm 2: Expand each interval $[l_k, r_k]$ to $[\tilde{s}_k, \tilde{e}_k]$, and learn the location of the k-th change-point as $\hat{\eta}_k = \arg\max_{t:l_k<t\le r_k} |\tilde{X}^t_{\tilde{s}_k,\tilde{e}_k}|$.

The time complexity of Algorithm 1 and Algorithm 2 is $O(T\log(T))$, ensuring that the method is efficient and scalable for large networks.

3.3. Theoretical Results

Denote $\Delta = \min_{k=1,\cdots,K-1}\{\eta_{k+1} - \eta_k\} \wedge (\eta_1 - 1) \wedge (T + 1 - \eta_K)$ as the minimal spacing between change-points or boundary points. Let H be a deterministic finite simple graph, and let $\kappa_k = |\hom(H, h_{T,\eta_k}) - \hom(H, h_{T,\eta_{k-1}})|$ for $1 \le k \le K$ denote the jumping magnitude at each change-point. The key theoretical guarantee for the consistency of change-point detection and localization is stated in the following theorem:

Theorem 3.1. *Assume the following conditions hold:*

(i) There exists a constant $c_0 > 0$ such that

$$\min_k \kappa_k \Delta^{|V(H)|+1/2} \ge c_0 T^{|V(H)|-1/2}\log(T). \quad (6)$$

(ii) The threshold τ in Algorithm 1 satisfies $C_0 T^{|V(H)|-1/2}\log^{1/2}(T) < \tau < C_1 \min_k \kappa_k \Delta^{|V(H)|+1/2} - C_0 T^{|V(H)|-1/2}\log^{1/2}(T)$, where C_0 and C_1 are sufficiently large positive constants.

Under these conditions, the learned $\hat{K}$ and $\hat{\eta}_1, \ldots, \hat{\eta}_{\hat{K}}$ obtained from Algorithm 1 and Algorithm 2 satisfy

$$P\left(\hat{K} = K, \max_{1\le k\le K} |\hat{\eta}_k - \eta_k| \le \epsilon\right) \ge 1 - C_2 T^{-1},$$

where

$$\epsilon = C_3 \left(\frac{T^{|V(H)|-1/2}\sqrt{\log(T)}}{\min_k \kappa_k}\right)^{\frac{1}{|V(H)|+1/2}},$$

and $C_2, C_3 > 0$ are constants.

A few comments on this theorem are in order:

1. **Signal-to-Noise Ratio Condition**: Equation (6) establishes a signal-to-noise ratio condition, which links the minimal jump magnitude ($\min_k \kappa_k$) and spacing (Δ) to the sample size (T). This ensures that the signal is sufficiently distinguishable from noise. Notably, equation (6) does not necessitate the boundedness of any term in it; instead, it only requires that the combined terms collectively satisfy the signal-to-noise ratio condition.

2. **Correlations in Subgraph Counts**: Due to strong correlations between $N_H(G_t)$ and $N_H(G_{t-1})$, the term $T^{|V(H)|-1/2}$ appears in (6), reflecting the impact of temporal dependencies in the data.

3. **Minimal Spacing and Jump Magnitude**: When the number of change-points (K) is bounded and $\min_k \kappa_k$ is bounded away from zero, the minimal spacing Δ can be as small as $T^{(|V(H)|-1/2)/(|V(H)|+1/2)}\log(T)$. Conversely, if $\Delta \asymp T$, the jump magnitude ($\min_k \kappa_k$) can be as small as $\log(T)/T$. These results highlight the flexibility of our method in accommodating varying levels of sparsity and signal strength.

4. **Consistency and Localization Error**: The theorem guarantees that, with high probability, the learned number of change-points ($\hat{K}$) equals the true number (K). Additionally, the localization rate ϵ/Δ decreases as T increases, satisfying $\epsilon/\Delta = o(1)$ by (6). This ensures consistency, a fundamental property in change-point detection, as discussed in Fan & Wu (2024) and Yu et al. (2022).

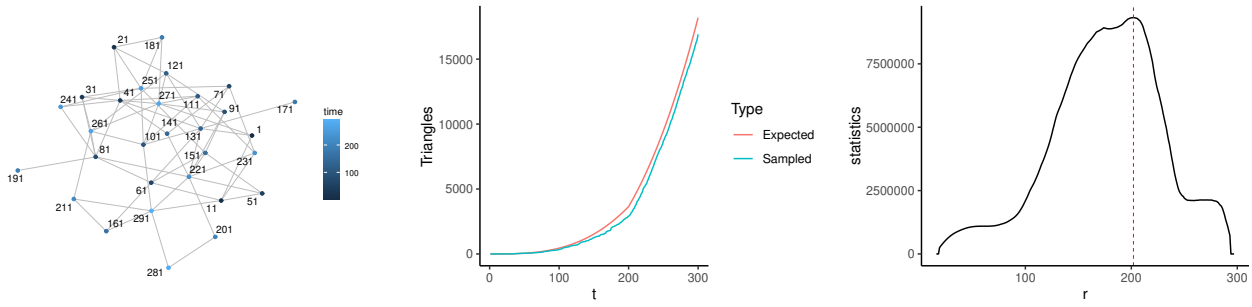

Figure 2. An illustrative example of the change-point detection procedure. Let $T = 300$ and $\eta = 200$, with $h_{T,t}(x, y) = (x - y)^2$ for $1 \le t \le 200$, and $h_{T,t}(x, y) = 1/6$ for $201 \le t \le 300$. The left panel shows a sample network, displaying only induced subgraphs at $10k + 1$ for $k = 0, \ldots, 29$. The middle panel plots the subgraph counts $N_H(G_t)$ for $H = \triangle$ (triangle) alongside their expected values. Using Algorithm 1 with $\tau = T^{2.5}\sqrt{\log T}$, we estimate $\hat{K} = 1$ and $S^* = \{[87, 225]\}$. Finally, applying Algorithm 2, we localize the change-point at $\hat{\eta} = 202$, close to the true value $\eta = 200$.

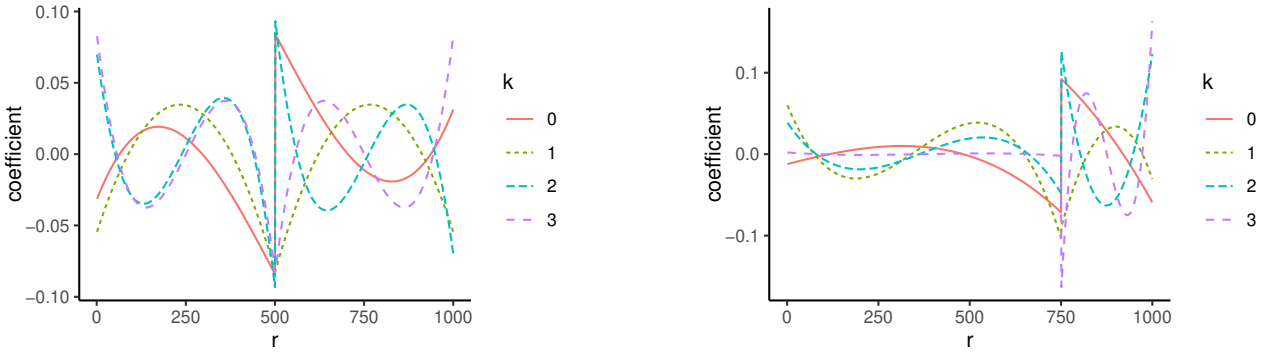

Figure 3. Plots of $\tilde{w}_{s,e}^{t,k}(r)$ for $s = 0, e = 1000$, and several values of t. Left figure: $t = 500$; Right figure: $t = 750$.

5. **Choice of H**: From (6), it is evident that a smaller size for H leads to a better detection bound. Therefore, we recommend using smaller subgraphs, such as  ,  , or  , for change-point detection, which aligns with the findings in our simulations.

Remark 3.2 (Utilizing multiple subgraphs). A natural question arises regarding which subgraphs are most suitable for learning changes. While determining the optimal choice of subgraphs may be an interesting avenue for further exploration, we can instead consider using multiple subgraphs simultaneously, such as $H =$  ,  , or  . In this case, we aggregate the information from these subgraphs by retaining an interval if the statistic corresponding to at least one of the subgraphs exceeds its pre-specified threshold. This approach enables the simultaneous use of different graph structures in the change-point detection process. The subsequent steps of the procedure, including the distillation method (line 19-33 in Algorithm 1) and change-point localization (Algorithm 2), remain unchanged. It is straightforward to verify that, under this aggregation of subgraphs, Theorem 3.1 still holds, ensuring both consistency and accurate localization of change-points.

4. Simulation Studies

In this section, we evaluate the performance of the proposed method through a series of simulation experiments. We use three commonly adopted evaluation metrics from the literature (e.g., Wang et al. (2021), Wang et al. (2020)):

- The difference between the learned and true number of change-points, $\hat{K} - K$, where $\hat{K}$ denotes the number of learned change-points and K represents the true number.
- The normalized Hausdorff distance $H(\hat{\eta}, \eta)/T$, where $H(\hat{\eta}, \eta)$ is defined as

$$H(\hat{\eta}, \eta) = \max\left\{\max_{x\in\eta}\min_{y\in\hat{\eta}}|x - y|, \max_{y\in\hat{\eta}}\min_{x\in\eta}|x - y|\right\},$$

  where a smaller value of $H(\hat{\eta}, \eta)/T$ indicates better performance.
- The Averaged Rand Index (ARI), as defined by Rand (1971) and Hubert & Arabie (1985), which ranges from

Table 1. Summary of results for Scenarios 1-4.

Scenario	Parameter(s)	Subgraph	$\hat{K}-K$			$H(\hat{\eta},\eta)/T$ ($\times 10^{-2}$)	ARI
			≤ -1	0	≥ 1		
1 (With Degree Heterogeneity)	$\alpha = 0.75$	⊶	0	92	8	9.40	0.85
		△	1	97	2	8.07	0.86
	0.65	⊶	0	95	5	4.64	0.92
		△	0	99	1	5.47	0.90
	0.5	⊶	0	98	2	1.93	0.96
		△	0	99	1	4.40	0.92
2 (Triangle Detection)	$T = 200$	△	5	95	0	9.94	0.87
	400	△	0	94	6	7.11	0.88
3 (SBM)	(C_1, C_2) =$(0.12, 0.04)$	⊶	0	100	0	0.26	1.00
		△	2	95	3	1.65	0.98
		□	0	90	10	18.81	0.87
	$(0.14, 0.04)$	⊶	0	100	0	0.24	1.00
		△	0	100	0	0.63	0.99
		□	0	97	3	17.94	0.87
4 (RDPG)	$(C_1, \ldots, C_4)$ $= (0.06, 0.02,$ $0.04, 0.01)$	⊶	0	91	9	1.58	0.98
		△	1	89	10	2.71	0.97
		□	19	81	0	6.77	0.92
	$(0.07, 0.02,$ $0.05, 0.01)$	⊶	0	97	3	0.57	0.99
		△	0	87	13	2.49	0.97
		□	13	85	2	6.13	0.94

0 to 1. A higher ARI value indicates a more accurate estimation.

For the subgraphs, we choose $H =$ ⊶, △, or □, as these are commonly used in the literature (e.g., Maugis et al. (2020), Shao et al. (2022)). The threshold τ in Algorithm 1 is chosen based on a data-driven approach inspired by Fan & Wu (2024). Specifically, we use

$$\tau = \max_{j=1,\ldots,T-h} \max_{j<t<j+h} |\tilde{X}^t_{j,j+h}(H)| \log(T),$$

where $h = \lfloor 3\log T \rfloor$ or $\lfloor 6\log T \rfloor$. When we suspect severe degree heterogeneity, we use $h = \lfloor 6\log T \rfloor$; otherwise, we use $h = \lfloor 3\log T \rfloor$.

Due to space constraints, the specific settings for the four simulations are provided in Appendix A.2. In each simulation setting, we repeat the experiment 100 times and report the mean values for $H(\hat{\eta},\eta)/T$, ARI, and the difference $\hat{K}-K$.

From the results in Tables 1, we observe that in Scenario 1, as α decreases (i.e., the magnitude of the jump increases), the detection accuracy improves for both ⊶ and △. A similar trend is evident in Scenario 2. It is important to note that in Scenario 2, the jump magnitude $\min_k \kappa_k$ is generally small, which necessitates a sufficiently large sample size for accurate estimation, even when the number of change-points K is known in advance. In Scenarios 3 and 4, larger signal levels enhance estimation accuracy. However, the convergence rate appears to be slower for larger subgraphs. Thus, we recommend using the smallest subgraphs, such as triangles, in practical applications. Overall, these findings highlight the effectiveness of our method across a wide range of network settings, including dense, sparse, and degree-heterogeneous networks.

5. A real data example

The email-Eu-core network data (Paranjape et al., 2017) was constructed using email communications from a large European research institution. This dataset captures all email interactions among members over a period of 803 days, involving 986 individuals (nodes). In our analysis, individuals are added to the network sequentially based on the timestamp of their first email interaction with colleagues. Two individuals are connected in the network if they have exchanged emails. The dataset is particularly suitable for the GAN-M model, as it exhibits a growing number of nodes, with incoming nodes having a likelihood of establishing connections with existing nodes.

The network exhibits substantial variability in node degrees. Specifically, the maximum degree is 345, the mean degree is 22, and the minimum degree is 1, highlighting the heterogeneous nature of the nodes. As recommended in Section 4, we set the threshold as $\tau = \max_{j=1,\ldots,T-h} \max_{j<t<j+h} |\tilde{X}^t_{j,j+h}(H)| \log(T)$, with $h = \lfloor 6\log(T) \rfloor$. We applied our algorithm using

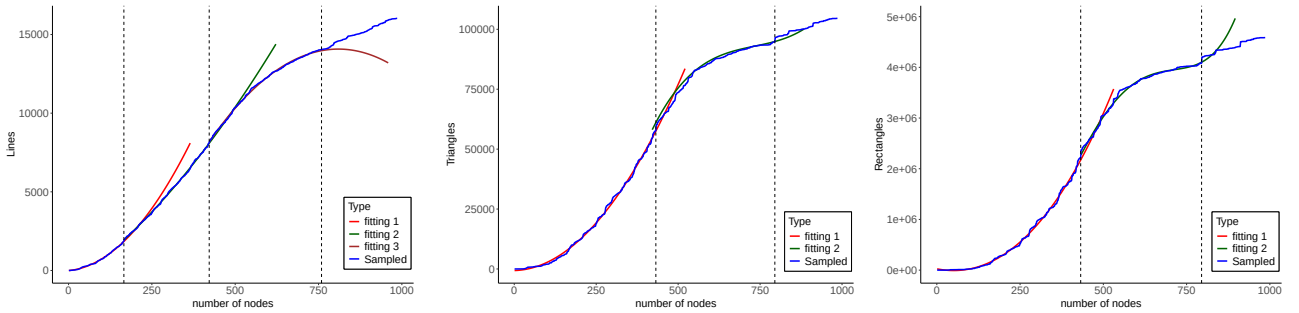

Figure 4. The counted subgraphs, along with the change-point detection results, are presented from left to right using line segments, triangles, and rectangles as the selected subgraph structures, respectively.

subgraph patterns, including line segments, triangles, and rectangles. For line segments, three change points were detected at nodes 166, 422, and 759. For triangles, two change points were identified at nodes 432 and 795. Similarly, for rectangles, two change points were detected at nodes 434 and 795. We visualize the counted subgraphs, including line segments, triangles, and rectangles, along with the change-point detection results in Figure 4.

Figure 4 provides valuable insights into the observed shifts. For triangles and rectangles, the rate of subgraph count growth before the first change point is notably higher than that after it. This likely reflects an initial surge in network activity, where frequent interactions and tasks lead to a denser connection structure in the early stages. The second change point, occurring at node 795, marks a significant increase in subgraph counts. This shift may correspond to the entry of a key figure into the network, altering connectivity dynamics and driving the observed structural change.

For line segment subgraphs, the last two detected change points closely align with those identified using triangles and rectangles, demonstrating the robustness of our method across different subgraph selections. However, an additional change point is detected at node 166, providing a more detailed view of early-stage connectivity changes. This observation supports the notion that subgraphs with fewer nodes can be more effective in capturing structural shifts.

6. Summary

In this paper, we introduce the graphon attachment network model, a novel and dynamic extension of traditional static networks, offering a robust framework for modeling and understanding the evolution of complex networks over time. By demonstrating that expected subgraph counts exhibit a piecewise polynomial structure, we provide a powerful new tool for learning structural changes in evolving networks. Building on this insight, we develop the WESUM statistic, a novel method specifically designed to learn dependent changes in piecewise polynomial signals, effectively addressing the long-standing challenge of change-point detection in dynamic networks.

Our approach is underpinned by rigorous theoretical guarantees, including consistency and convergence rates, which affirm the robustness of the method across a range of scenarios. Moreover, the performance of our technique is validated through extensive simulation studies, highlighting its superior accuracy and adaptability in diverse network settings.

This work opens up several exciting avenues for future research. Notably, the extension of our statistic to learn abrupt changes in graphons holds great promise for applications in more intricate dependency structures, such as locally stationary processes. Furthermore, the deeper exploration of graphon attachment networks and the development of generalized models for dynamic networks are essential steps toward advancing the field.

Acknowledgments

Weichi Wu is supported by NSFC No.12271287.

Impact Statement

This paper marks the first effort to integrate graphon and attachment models. The introduction of the WESUM statistic lays the foundation for future research in dynamic and evolving network modeling and data-driven change detection methods, with wide-ranging implications across fields such as epidemiology, finance, and the social sciences.

References

Aminikhanghahi, S. and Cook, D. J. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 2017.

Bai, T., Wang, L., Yin, D., Sun, K., Chen, Y., Li, W., and Li, D. Deep learning for change detection in remote sensing: a review. *Geo-spatial Information Science*, 26

(3):262–288, 2023.

Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

Bet, G., Bogerd, K., Castro, R. M., and van der Hofstad, R. Detecting a late changepoint in the preferential attachment model. *arXiv preprint arXiv:2310.02603*, 2023.

Bhamidi, S., Jin, J., and Nobel, A. Change point detection in network models: Preferential attachment and long range dependence. *The Annals of Applied Probability*, 28(1): 35–78, 2018.

Bickel, P. J., Chen, A., and Levina, E. The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280, 2011.

Borgs, C. and Chayes, J. Graphons: A nonparametric method to model, estimate, and design algorithms for massive networks. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 665–672, 2017.

Borgs, C., Chayes, J. T., Cohn, H., and Zhao, Y. An l^p theory of sparse graph convergence ii: Ld convergence, quotients and right convergence. *The Annals of Probability*, 46(1): 337–396, 2018.

Borgs, C., Chayes, J., Cohn, H., and Zhao, Y. An l^p theory of sparse graph convergence i: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 372(5):3019–3062, 2019.

Bubeck, S., Mossel, E., and Rácz, M. Z. On the influence of the seed graph in the preferential attachment model. *IEEE Transactions on Network Science and Engineering*, 2(1):30–39, 2015.

Caldarelli, G., Capocci, A., De Los Rios, P., and Munoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89(25):258702, 2002.

Chan, S. and Airoldi, E. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pp. 208–216. PMLR, 2014.

Chen, H. and Chu, L. Graph-based change-point analysis. *Annual Review of Statistics and Its Application*, 10(1): 475–499, 2023.

Chen, H., Qi, Z., and Shi, Z. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021.

Crespi, B. J. Vicious circles: positive feedback in major evolutionary and ecological transitions. *Trends in Ecology & Evolution*, 19(12):627–633, 2004.

De Ryck, T., De Vos, M., and Bertrand, A. Change point detection in time series data using autoencoders with a time-invariant representation. *IEEE Transactions on Signal Processing*, 69:3513–3524, 2021.

Duan, Y., Liu, J., Chen, S., Chen, L., and Wu, J. G-prompt: Graphon-based prompt tuning for graph classification. *Information Processing & Management*, 61(3):103639, 2024.

Fan, X. and Wu, W. Random interval distillation for detection of change-points in Markov chain Bernoulli networks. *arXiv preprint arXiv:2403.00600*, 2024.

Fan, X., Ma, F., Leng, C., and Wu, W. Low-rank approaches to graphon learning in networks. *arXiv preprint arXiv:2501.18785*, 2025.

Follain, B., Wang, T., and Samworth, R. J. High-dimensional changepoint estimation with heterogeneous missingness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):1023–1055, 2022.

Gallagher, S. and West, J. Reconceptualizing and expanding the positive feedback network effects model: A case study. *Journal of Engineering and Technology Management*, 26 (3):131–147, 2009.

Garavaglia, A. and van der Hofstad, R. From trees to graphs: collapsing continuous-time branching processes. *Journal of Applied Probability*, 55(3):900–919, 2018.

Garavaglia, A., Hazra, R. S., van der Hofstad, R., and Ray, R. Universality of the local limit of preferential attachment models. *arXiv preprint arXiv:2212.05551*, 2022.

Han, X., Jiang, Z., Liu, N., and Hu, X. G-mixup: Graph data augmentation for graph classification. In *International Conference on Machine Learning*, pp. 8230–8248. PMLR, 2022.

Hormozdiari, F., Berenbrink, P., Pržulj, N., and Sahinalp, S. C. Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution. *PLoS Computational Biology*, 3(7):e118, 2007.

Hornung, G. and Barkai, N. Noise propagation and signaling sensitivity in biological networks: a role for positive feedback. *PLoS Computational Biology*, 4(1):e8, 2008.

Hubert, L. and Arabie, P. Comparing partitions. *Journal of Classification*, 2:193–218, 1985.

Jiang, B., Li, J., and Yao, Q. Autoregressive networks. *Journal of Machine Learning Research*, 24(227):1–69, 2023.

Kaddouri, I., Naulet, Z., and Gassiat, É. On the impossibility of detecting a late change-point in the preferential attachment random graph model. *arXiv preprint arXiv:2407.18685*, 2024.

Ke, Z. T. and Wang, J. Optimal network membership estimation under severe degree heterogeneity. *Journal of the American Statistical Association*, pp. 1–15, 2024.

Kovács, S., Bühlmann, P., Li, H., and Munk, A. Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika*, 110(1):249–256, 2023.

Leskovec, J., Kleinberg, J., and Faloutsos, C. Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1 (1):2–es, 2007.

Lévy-Leduc, C. and Roueff, F. Detection and localization of change-points in high-dimensional network traffic data. *The Annals of Applied Statistics*, pp. 637–662, 2009.

Li, J., Fearnhead, P., Fryzlewicz, P., and Wang, T. Automatic change-point detection in time series via deep learning. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(2):273–285, 2024.

Lovász, L. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

Maugis, P.-A., Olhede, S., Priebe, C., and Wolfe, P. Testing for equivalence of network distribution using subgraph counts. *Journal of Computational and Graphical Statistics*, 29(3):455–465, 2020.

Newman, M. E. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):025102, 2001.

Padilla, O. H. M., Yu, Y., and Priebe, C. E. Change point localization in dependent dynamic nonparametric random dot product graphs. *The Journal of Machine Learning Research*, 23(1):10661–10719, 2022.

Paranjape, A., Benson, A. R., and Leskovec, J. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 601–610, 2017.

Peel, L. and Clauset, A. Detecting change points in the large-scale structure of evolving networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1), Feb. 2015.

Poncela, J., Gómez-Gardenes, J., Floría, L. M., Sánchez, A., and Moreno, Y. Complex cooperative networks from evolutionary preferential attachment. *PLoS One*, 3(6): e2449, 2008.

Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.

Rossi, R. A., Gallagher, B., Neville, J., and Henderson, K. Modeling dynamic behavior in large evolving graphs. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 667–676, 2013.

Ruiz, L., Chamon, L., and Ribeiro, A. Graphon neural networks and the transferability of graph neural networks. *Advances in Neural Information Processing Systems*, 33: 1702–1712, 2020.

Ruiz, L., Wang, Z., and Ribeiro, A. Graphon and graph neural network stability. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5255–5259. IEEE, 2021.

Shafique, A., Cao, G., Khan, Z., Asad, M., and Aslam, M. Deep learning-based change detection in remote sensing images: A review. *Remote Sensing*, 14(4):871, 2022.

Shao, M., Xia, D., Zhang, Y., Wu, Q., and Chen, S. Higher-order accurate two-sample network inference and network hashing. *arXiv preprint arXiv:2208.07573*, 2022.

Sun, J., Faloutsos, C., Papadimitriou, S., and Yu, P. S. Graphscope: parameter-free mining of large time-evolving graphs. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 687–696, 2007.

Truong, C., Oudre, L., and Vayatis, N. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

Wang, D., Yu, Y., and Rinaldo, A. Univariate mean change point detection: Penalization, cusum and optimality. *Electronic Journal of Statistics*, 14:1917–1961, 2020.

Wang, D., Yu, Y., and Rinaldo, A. Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics*, 49(1):203–232, 2021.

Wu, W. and Zhou, Z. Multiscale jump testing and estimation under complex temporal dynamics. *Bernoulli*, 30(3): 2372–2398, 2024.

Xu, J. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pp. 5433–5442, 2018.

Yu, Y., Chatterjee, S., and Xu, H. Localising change points in piecewise polynomials of general degrees. *Electronic Journal of Statistics*, 16(1):1855–1890, 2022.

Zaki, A., Attia, M., Hegazy, D., and Amin, S. Comprehensive survey on dynamic graph models. *International Journal of Advanced Computer Science and Applications*, 7(2), 2016.

Zhang, Y., Zhang, J., Sun, Y., and Wang, J. Change point detection in dynamic networks via regularized tensor decomposition. *Journal of Computational and Graphical Statistics*, 33(2):515–524, 2024.

Zhou, Y., Gao, S., Guo, D., Wei, X., Rokne, J., and Wang, H. A survey of change point detection in dynamic graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

A. Appendix

A.1. Algorithms

Algorithm 1 takes the subgraph counts $\{N_H(G_t)\}$ and threshold τ as inputs and outputs the learned number of change-points $\hat{K}$ and a set of intervals S^*. These intervals are then used to learn the locations of change-points via Algorithm 2. The algorithm consists of several steps:

In Lines 2-7, we utilize the original seeded intervals from Kovács et al. (2023). In Lines 8-12, these intervals are modified by inserting equally spaced points and partitioning them to ensure that each change-point is well-positioned within an interval, as described in Lemma A.1. This modification is primarily for technical proof purposes. The total length of all intervals in I is $O(T \log T)$, which ensures that the method is efficient in practice.

Lemma A.1. *Let* $\Delta = \min_{k=1,\ldots,K-1}\{\eta_{k+1} - \eta_k\} \wedge (\eta_1 - 1) \wedge (T + 1 - \eta_K)$ *represent the minimal spacing between change-points. If* $\Delta \to \infty$ *as* $T \to \infty$*, then for any* η_k*,* $1 \le k \le K$*, there exists an interval* $(s, e] \in I$ *such that* $\eta_k - 0.125\Delta < s < \eta_k - 0.025\Delta$ *and* $\eta_k + 0.025\Delta < e < \eta_k + 0.125\Delta$.

In Lines 13-18, we gather the intervals $(s_m, e_m] \in S$ where $\max_{t:s_m<t<e_m} |\tilde{X}^t_{s_m,e_m}| > \tau$, aiming to exclude intervals that contain no change-points. In Lines 19-33, we select endpoints from S to form new intervals. This step, which does not involve $N_H(G_t)$ or τ, is based on the general approach described in Fan & Wu (2024).

To localize the change-points, we take $\arg\max$ within the slightly lengthened intervals. Intuitively, as discussed in Section 3.1, in the population version, we have $\tilde{f}^{\eta_k}_{s,e}(H) = \max_t \tilde{f}^t_{s,e}(H)$ when the interval $(s, e]$ contains only one change-point η_k. Thus, taking the $\arg\max$ in the sample version provides an estimator $\hat{\eta}_k$ close to η_k. This approach is a natural extension of CUSUM methods.

A.2. Settings of Simulation Studies

Scenario 1: Model with Degree Heterogeneity. We set $T = 400$ and $\eta = 300$, and define the graphon function $h_{T,t}(x, y)$ as follows:

$$h_{T,t}(x,y) = \begin{cases} 1 \wedge T^{-0.6}(xy)^{-\alpha}, & t < 300, \\ 1 \wedge T^{-0.6}(xy)^{-0.9}, & t \ge 300, \end{cases} \tag{7}$$

where $\alpha = 0.75, 0.65, 0.5$. This graphon is designed to induce degree heterogeneity in the resulting network. To illustrate this, we summarize the degree quantiles from a sampled network generated by three different graphons: (I) $h_{T,t}(x, y) = 1 \wedge T^{-0.6}(xy)^{-0.9}$, (II) $h_{T,t}(x, y) = 1 \wedge T^{-0.6}(xy)^{-0.5}$, and (III) a mixture of the two graphons, corresponding to $\alpha = 0.5$ in (7), in Table 2. The table shows significant variability in the minimum, maximum, and average degrees, with a small subset of nodes displaying exceptionally high degrees. These results demonstrate the presence of degree heterogeneity, a feature commonly observed in real-world networks.

Table 2. Quantiles of node degrees in a sample network generated under Scenario 1.

Model	T	Min	25% Quantile	Mean	Median	75% Quantile	Max
(I)	400	17	35	65.81	46	76	378
	1000	31	60	126.03	81	132	999
	4000	55	99	240.94	147	242	3999
(II)	400	11	22.75	34.07	29	39	169
	1000	17	35.75	58.93	45	63.25	567
	4000	35	63	103.67	76	106	1901
(III)	400	24	47	91.82	68	112	399
	1000	55	86.75	182.59	116	204	999
	4000	107	161	376.79	220	367	3999

Scenario 2: Triangle Detection. We set $T = 200, 400$, and $\eta = 0.8T$, and define the graphon function $h_{T,t}(x, y)$ as follows:

$$h_{T,t}(x,y) = \begin{cases} \frac{(0.5+x)^2(0.5+y)^2}{1.8^4}, & t < \eta, \\ 0.1118, & t \ge \eta. \end{cases}$$

Algorithm 1 Random Interval Distillation for Piecewise Polynomial Signals

Input: $\{N_H(G_t)\}_{t=1}^T$, τ
For $1 \le k \le \lceil \log_{\sqrt{2}}(T) \rceil$, let
  $n_k = 2\lceil 2^{(k-1)/2} \rceil - 1$,
  $L_k = T2^{(k-1)/2}$,
  $J_k = \frac{T-L_k}{n_k-1}$,
  $I_k = \cup_{i=1}^{n_k} \{(\lfloor (i-1)J_k \rfloor, \lceil (i-1)J_k + L_k \rceil]\}$.
Let $I^* = \cup I_k$ and use the notation $I^* = \{(s_m^*, e_m^*)\}_{m=1}^{M^*}$. Let $I = \varnothing$.
while $1 \le m \le M^*$ **do**
  Let $b_m^{(i)} = \lfloor ((44-i)s_m^* + ie_m^*)/44 \rfloor$, for $i = 0, 1, \ldots, 44$.
  Add all intervals $(b_m^{(i)}, b_m^{(j)}]$ with $i < j$ into the set I.
end while
Use the notation $I = \{(s_m, e_m]\}_{m=1}^M$.
$S = \varnothing$.
while $1 \le m \le M$ **do**
  if $\max_{t: s_m < t < e_m} |\tilde{X}^t_{s_m, e_m}| > \tau$ **then**
    $S = S \cup \{(s_m, e_m]\}$.
  end if
end while
$\tilde{S} = S$, $i = 1$.
while $|\tilde{S}| > 0$ **do**
  $r_i = \min\{v : \exists u, (u, v] \in \tilde{S}\}$.
  $v^* = r_i$, $u^* = \max\{u : (u, v^*] \in \tilde{S}\}$.
  $\tilde{S} = \tilde{S} \setminus \{(u, v] \in \tilde{S} : (u, v] \cap (u^*, v^*) \neq \varnothing\}$.
  $i = i + 1$.
end while
$\hat{K}_0 = i - 1$, $\tilde{S} = S$, $i = 1$.
while $|\tilde{S}| > 0$ **do**
  $l_{\hat{K}_0 + 1 - i} = \max\{u : \exists v, (u, v] \in \tilde{S}\}$.
  $u^* = l_{\hat{K}_0 + 1 - i}$, $v^* = \min\{v : (u^*, v] \in \tilde{S}\}$.
  $\tilde{S} = \tilde{S} \setminus \{(u, v] \in \tilde{S} : (u, v] \cap (u^*, v^*) \neq \varnothing\}$.
  $i = i + 1$.
end while
$\hat{K} = i - 1$, $S^* = \{[l_j, r_j]\}_{j=1}^{\hat{K}}$.
Output: $\hat{K}$ and S^*.

Algorithm 2 Localization Procedure

Input: $\{N_H(G_t)\}_{t=1}^T$, S^* from Algorithm 1
$i = 1$, $\hat{K} = |S^*|$. Suppose $S^* = \{[l_k, r_k]\}_{k=1}^{\hat{K}}$.
Let $\hat{\Delta} = \min_{2 \le k \le \hat{K}} \left(\frac{l_k + r_k}{2} - \frac{l_{k-1} + r_{k-1}}{2} \right) \wedge \left(T + 1 - \frac{l_{\hat{K}} + r_{\hat{K}}}{2} \right) \wedge \left(\frac{l_1 + r_1}{2} - 1 \right)$
while $i \le \hat{K}$ **do**
  $s_i = \lfloor l_i - \hat{\Delta}/2 \rfloor$, $e_i = \lfloor r_i + \hat{\Delta}/2 \rfloor$
  $\hat{\eta}_i = \arg\max_{t: l_i < t \le r_i} |\tilde{X}^t_{s_i, e_i}|$
  $i = i + 1$
end while
Output: $\hat{\eta} = \{\hat{\eta}_1, \cdots, \hat{\eta}_{\hat{K}}\}$

In this scenario, we observe that $\hom(\multimap, h_{T,t})$ remains constant for all t. As a result, $\multimap$ is not suitable for learning the change point at η. Instead, we use Δ to learn the structural change in this case.

Scenario 3: Stochastic Block Model. We set $T = 450$ with change-points at $\eta = \{0.3T, 0.5T, 0.8T\}$, meaning the

change-points are unequally spaced. The graphon function $h_{T,t}(x,y)$ is defined as:

$$h_{T,t}(x,y) = \begin{cases} C_1(1+1.5I(x,y)), & \text{if } t < 0.3T \text{ or} \\ C_2(1+1.5I(x,y)), & 0.5T \le t < 0.8T \\ & \text{elsewhere} \end{cases}$$

where $I(x,y) = I(\lceil 3x \rceil = \lceil 3y \rceil)$. We set the parameter values $(C_1, C_2) = (0.12, 0.04)$ and $(0.14, 0.04)$.

Scenario 4: Random Dot Product Graph. We set $T = 450$ with change-points at $\eta = \{0.4T, 0.7T, 0.9T\}$, indicating unequally spaced change-points. The graphon function $h_{T,t}(x,y)$ is defined as:

$$h_{T,t}(x,y) = \begin{cases} C_1 + C_3 xy, & \text{if } t < 0.4T \text{ or} \\ & 0.7T \le t < 0.9T \\ C_2 + C_4 xy, & \text{elsewhere.} \end{cases}$$

We set the parameters as $(C_1, C_2, C_3, C_4) = (0.06, 0.02, 0.04, 0.01)$ for the first case, and $(C_1, C_2, C_3, C_4) = (0.07, 0.02, 0.05, 0.01)$ for the second. It is important to note that the graphs in this scenario are sparse.

A.3. Sensitivity

Table 3. Results for sensitivity analysis, Scenarios 3.

Scenario	Parameter(s)	Subgraph	$\hat{K} - K$			$H(\hat{\eta}, \eta)/T$ ($\times 10^{-2}$)	ARI
			≤ -1	0	≥ 1		
3 (SBM)	(C_1, C_2) =$(0.12, 0.04)$		0	100	0	0.26	1.00
			2	95	3	1.65	0.98
			0	100	0	0.31	1.00
			0	90	10	18.81	0.87
			0	90	10	19.01	0.87
	$(0.14, 0.04)$		0	100	0	0.24	1.00
			0	100	0	0.63	0.99
			0	100	0	0.24	1.00
			0	97	3	17.94	0.87
			0	96	4	18.76	0.87

To show that our algorithm is robust to different choices of subgraphs, we repeat the experiment in Scenario 3 (SBM model) and evaluate five types of subgraphs: line segments, triangles, connected triples, rectangles, and closed paths of length 4. The results indicate that triangles and connected triples yield similar outcomes, while rectangles and closed paths of length 4 also show comparable results. Additionally, triangles slightly outperform rectangles. This suggests that in application of our algorithm, the number of nodes is more significant than the specific shapes of the subgraphs.

A.4. Additional simulations with larger sizes

To demonstrate that our algorithm can handle large sample sizes, we repeat the experiment in Scenario 3 (SBM model) with a sample size of $T = 800$. We present the results of our algorithm under different threshold choices in Table 4. Recall that the threshold is defined as

$$\tau = \max_{j=1,\ldots,T-h} \max_{j<t<j+h} |\tilde{X}^t_{j,j+h}(H)| \log(T) \text{ with } h = \lceil k \log T \rceil. \tag{8}$$

We conduct experiments with $k = 4, 5, 6$. It turns out that our algorithm performs well.

A.5. Broader literature

In the literature, the graphon model has been widely employed in areas such as graph classification (Han et al., 2022; Duan et al., 2024), graph neural networks (Ruiz et al., 2020; 2021), and large-scale network learning (Chan & Airoldi, 2014; Xu, 2018; Fan et al., 2025). In addition, it has been used to explain phenomena such as the asymptotic normality of subgraph

Table 4. Results for larger network sizes with sensitivity analysis, Scenarios 3, $T = 800$.

Scenario	Threshold (see (8))	Parameter(s)	Subgraph	$\hat{K} - K$			$H(\hat{\eta}, \eta)/T$	ARI
				≤ -1	0	≥ 1	($\times 10^{-2}$)	
3 (SBM)	$k = 4$	(C_1, C_2) =$(0.12, 0.04)$		0	100	0	0.14	1.00
				0	97	3	0.32	1.00
		$(0.14, 0.04)$		0	100	0	0.14	1.00
				0	98	2	0.22	1.00
	$k = 5$	(C_1, C_2) =$(0.12, 0.04)$		0	100	0	0.13	1.00
				0	100	0	0.30	1.00
		$(0.14, 0.04)$		0	100	0	0.13	1.00
				0	100	0	0.23	1.00
	$k = 6$	(C_1, C_2) =$(0.12, 0.04)$		0	100	0	0.15	1.00
				7	93	0	1.75	0.99
		$(0.14, 0.04)$		0	100	0	0.14	1.00
				3	97	0	0.97	0.99

counts (Bickel et al., 2011) and to facilitate hypothesis testing for graph equivalence (Maugis et al., 2020). For an overview of the development of its theory, we refer to Lovász (2012); Borgs & Chayes (2017).

For change point detection methods, probabilistic approaches include Jiang et al. (2023); Padilla et al. (2022); Follain et al. (2022); Wu & Zhou (2024); Zhang et al. (2024), whereas deep learning-based approaches include De Ryck et al. (2021); Chen et al. (2021); Li et al. (2024) and those in Shafique et al. (2022); Bai et al. (2023). We refer to Truong et al. (2020); Zhou et al. (2024) for a review.

There are also works that investigated (heuristic) methods for detecting change points in dynamic networks, or explored the topological consequences of various network growth models (e.g., Leskovec et al. (2007); Sun et al. (2007); Rossi et al. (2013); Peel & Clauset (2015)).

A.6. The "positive feedback" mechanism and potential extensions of GAN-M

The positive feedback associated with node degrees, as suggested by a reviewer, is a compelling issue that represents an important mechanism beyond the graphon framework. Positive feedback implies that when a node forms a link to an existing node x, it increases the probability that subsequently added nodes will also form edges to x. Networks exhibiting positive feedback are prevalent in a variety of applications (Crespi, 2004; Hornung & Barkai, 2008; Gallagher & West, 2009). Current research on this topic largely focuses on preferential attachment models, see Newman (2001); Poncela et al. (2008) for a comprehensive review.

Our model has the potential to be extended to incorporate such a positive feedback mechanism, and we outline a possible implementation:

Remark A.2 (A hybrid model of Graphon Attachment Network and Preferential Attachment). Let G_t denote the graph at time t, with A_t as its adjacency matrix, for $1 \leq t \leq T$. Let $d_t(i)$ denote the degree of the node i in G_t, i.e., $d_t(i) = \sum_j A_t(i, j)$. The model begins with an initial simple graph G_1 containing m nodes ($m \geq 1$), where each node is assigned a random value $U_1, U_2, \ldots, U_m \sim U(0, 1)$, drawn independently.

For $t > 1$, the graph G_t is formed recursively. Given the values $U_1, \ldots, U_{m+t-2}$ and the adjacency matrix A_{t-1}, the adjacency matrix A_t is updated as follows:

1. Preserve all existing edges: for $1 \leq i, j \leq m+t-2$, set $A_t(i, j) = A_{t-1}(i, j)$. Assign $A_t(m+t-1, m+t-1) = 0$, ensuring no self-loops for the new node.

2. Draw $V_{m+t-1} \sim$ Bernoulli(p) independently of all random variables generated in the previous steps, with p being a pre-specified parameter.

3. If $V_{m+t-1} = 1$: Draw $U_{m+t-1} \sim U(0, 1)$, independently of all random variables generated in the previous steps. Form edges between the new node $j := m + t - 1$ and each existing node i ($1 \leq i \leq m + t - 2$) independently, with

probability:

$$\mathbb{P}(A_t(i,j)=1 \mid U_i, U_j) = h_{T,t}(U_i, U_j),$$

where $h_{T,t} : [0,1]^2 \to [0,1]$ is a symmetric measurable function defining the attachment probability.

4. If $V_{m+t-1} = 0$: Select a node i from the existing nodes according to the following rule: the probability of selecting each node is proportional to its degree. Form an edge between node i and the new node $m+t-1$.

The above model is a hybrid of the GAN-M and the preferential attachment model, retaining certain characteristics of the latter, which allows the model to incorporate a positive feedback mechanism. We leave further investigation into its dynamic properties and change-point detection for future work.

A.7. Proofs

Proof of Lemma 2.6. Since the size of the initial graph m is fixed, we can assume without loss of generality that $m = 1$. Additionally, assume without loss of generality that $V(G_t) = \{1, 2, \cdots, t\}$. With a slight abuse of notation, when we refer to $t_1, t_2 \in V(G_t)$, we are referring to the indices t_1, t_2 and time t_1, t_2 simultaneously, so $t_1 \vee t_2$ is well-defined. We have

$$\begin{aligned}\mathbb{E}(N_H(G_t) \mid U_1, \cdots, U_t) &= \sum_{\varphi: V(H)\to V(G_t)} \prod_{uv\in E(H)} \mathbb{P}(\varphi(u)\varphi(v) \in E(G_t) \mid U_1, \cdots, U_t) \\ &= \sum_{\varphi: V(H)\to V(G_t)} \prod_{uv\in E(H)} h_{T,\varphi(u)\vee\varphi(v)}(U_{\varphi(u)}, U_{\varphi(v)}).\end{aligned}$$

Consequently,

$$\begin{aligned}\mathbb{E}(N_H(G_t)) &= \sum_{\varphi: V(H)\to V(G_t)} \int_{[0,1]^{|V(H)|}} \prod_{uv\in E(H)} h_{T,\varphi(u)\vee\varphi(v)}(U_{\varphi(u)}, U_{\varphi(v)}) \prod_{u\in V(H)} dU_{\varphi(u)} \\ &= \sum_{m_0=0}^{|V(H)|} \sum_{\varphi: |\{u\in V(H): \varphi(u)\le \eta_k\}| = m_0} \int_{[0,1]^{|V(H)|}} \prod_{uv\in E(H)} h_{T,\varphi(u)\vee\varphi(v)}(U_{\varphi(u)}, U_{\varphi(v)}) \prod_{u\in V(H)} dU_{\varphi(u)}. \qquad (9)\end{aligned}$$

Note that

$$\begin{aligned}&\sum_{\varphi: |\{u\in V(H): \varphi(u)\le \eta_k\}| = 0} \int_{[0,1]^{|V(H)|}} \prod_{uv\in E(H)} h_{T,\varphi(u)\vee\varphi(v)}(U_{\varphi(u)}, U_{\varphi(v)}) \prod_{u\in V(H)} dU_{\varphi(u)} \\ &= \sum_{\varphi: |\{u\in V(H): \varphi(u)\le \eta_k\}| = 0} \int_{[0,1]^{|V(H)|}} \prod_{uv\in E(H)} h_{T,t}(U_{\varphi(u)}, U_{\varphi(v)}) \prod_{u\in V(H)} dU_{\varphi(u)} \\ &= \left\{ \prod_{m_1=0}^{|V(H)|-1} (t - \eta_k - m_1) \right\} \int_{[0,1]^{|V(H)|}} \prod_{uv\in E(H)} h_{T,\eta_{k+1}}(U_u, U_v) \prod_{u\in V(H)} dU_u \\ &= \left\{ \prod_{m_1=0}^{|V(H)|-1} (t - \eta_k - m_1) \right\} \hom(H, h_{T,\eta_{k+1}}) \\ &= (t-\eta_k)^{|V(H)|} \hom(H, h_{T,\eta_{k+1}}) + Q^0_{|V(H)|-1}(t),\end{aligned}$$

where $Q^0_{|V(H)|-1}(t)$ is a polynomial of order $|V(H)|-1$ with respect to t.

Similarly, for $m_0 \ge 1$,

$$\begin{aligned}&\sum_{\varphi: |\{u\in V(H): \varphi(u)\le \eta_k\}| = m_0} \int_{[0,1]^{|V(H)|}} \prod_{uv\in E(H)} h_{T,\varphi(u)\vee\varphi(v)}(U_{\varphi(u)}, U_{\varphi(v)}) \prod_{u\in V(H)} dU_{\varphi(u)} \\ &= Q^{m_0}_{|V(H)|-m_0}(t),\end{aligned}$$

where $Q^{m_0}_{|V(H)|-m_0}(t)$ is a polynomial of order $|V(H)|-m_0$ with respect to t. Then, by (9), we have

$$\mathbb{E}(N_H(G_t)) = (t-\eta_k)^{|V(H)|}\hom(H, h_{T,\eta_{k+1}}) + Q_{|V(H)|-1}(t),$$

where $Q_{|V(H)|-1}(t)$ is a polynomial of order $|V(H)|-1$ with respect to t. □

Proof of Lemma A.1. We follow the notations in Algorithm 1. Recall that Δ is the minimal spacing of change-points. By the proof of Theorem C.1 in Kovács et al. (2023), for each η_k, there exists an interval $(s^*_m, e^*_m] \in I^*$ such that $s^*_m \geq \eta_k - \Delta$, $e^*_m \leq \eta_k + \Delta$, $e^*_m - s^*_m \geq \Delta$, and

$$\left|\frac{s^*_m + e^*_m}{2} - \eta_k\right| \leq \frac{3}{8}(e^*_m - s^*_m).$$

Thus, we have $e^*_m \geq \eta_k + \frac{\Delta}{8}$ and $s^*_m \leq \eta_k - \frac{\Delta}{8}$.

Next, consider the set of 45 chosen points $\{b^{(i)}_m : i = 0, \ldots, 44\}$ from $(s^*_m, e^*_m]$. We have

$$|b^{(i)}_m - b^{(j)}_m| \leq \left\lceil \frac{e^*_m - s^*_m}{44} \right\rceil \leq \left\lceil \frac{\Delta}{22} \right\rceil$$

for all $i \neq j$. Therefore, the subinterval $[\eta_k + 0.025\Delta, \eta_k + \frac{\Delta}{8}]$ contains at least $\left\lceil \frac{0.1\Delta}{\lceil \frac{\Delta}{22} \rceil} \right\rceil - 1 \geq 1$ points from the set $\{b^{(i)}_m : i = 0, \ldots, 44\}$ for sufficiently large Δ. Similarly, there is at least 1 point in the set $\{b^{(i)}_m : i = 0, \ldots, 44\}$ that falls into the subinterval $[\eta_k - 0.125\Delta, \eta_k - 0.025\Delta]$.

Thus, there exist $b^{i^*}_m \in [\eta_k - 0.125\Delta, \eta_k - 0.025\Delta]$ and $b^{j^*}_m \in [\eta_k + 0.025\Delta, \eta_k + 0.125\Delta]$. □

Proof of Theorem 3.1. We divide the proof into two parts. In **Step 1**, we show that

$$\mathbb{P}\left(\{\hat{K} = K\}, \bigcap_{j=1}^{\hat{K}} \left\{[l_j, r_j] \text{ covers } \eta_j \text{ and } r_j - l_j \leq \frac{\Delta}{4}\right\}\right) \geq 1 - T^{-1}$$

where the intervals $[l_j, r_j]$ are obtained by Algorithm 1. In **Step 2**, we prove the final result.

Let $\tilde{f}^{t,k}_{s,e}(H) = \sum_{r=1}^{T} \tilde{w}^{t,k}_{s,e}(r)\mathbb{E}N_H(G_r)$ and $\tilde{f}^{t}_{s,e}(H) = \sqrt{\sum_{k=0}^{|V(H)|}(\tilde{f}^{t,k}_{s,e}(H))^2}$.

Step 1. By Lemma A.6, we have

$$\mathbb{P}(\mathcal{A}) \geq 1 - T^{-1}$$

where

$$\mathcal{A} = \left\{\sup_{s,t,e:s<t<e} \left|\tilde{X}^t_{s,e}(H) - \tilde{f}^t_{s,e}(H)\right| \leq CT^{|V(H)|-1/2}\sqrt{\log(T)}\right\}.$$

All the analysis in the rest of this proof is conducted on the event $\mathcal{A}$. For $(s, e] \in I$, when $(s, e]$ contains no change-point, by construction, $\tilde{f}^{t,k}_{s,e}(H) = 0$ for all $0 \leq k \leq |V(H)|$, so $\tilde{f}^{t}_{s,e}(H) = 0$. Then, $\mathcal{A}$ implies

$$\sup_{s<t<e} \tilde{X}^t_{s,e}(H) \leq CT^{|V(H)|-1/2}\sqrt{\log(T)}.$$

Thus, choosing $\tau > CT^{|V(H)|-1/2}\sqrt{\log(T)}$ excludes intervals without change-points.

For the change-point η_k, by Lemma A.1, there exists an interval $(s, e] \in I$ in Algorithm 1 such that

$$\eta_k - 0.125\Delta < s < \eta_k - 0.025\Delta, \quad \eta_k + 0.025\Delta < e < \eta_k + 0.125\Delta.$$

Moreover, by Lemma 2.6, $\mathbb{E}N_H(G_t)$ is a polynomial function of degree $|V(H)|$ in t, with the leading coefficient given by $\hom(H, h_{T,\eta_k})$ for $s < t < \eta_k$, and by $\hom(H, h_{T,\eta_{k+1}})$ for $\eta_k \leq t \leq e$. Then, by Lemma A.3, we have

$$\tilde{f}^{\eta_k}_{s,e}(H) \geq c_0|\hom(H, h_{T,\eta_k}) - \hom(H, h_{T,\eta_{k+1}})|\Delta^{|V(H)|+1/2}$$

where c_0 is a constant that depends only on $|V(H)|$. As a result, this interval can be retained as long as the event $\mathcal{A}$ occurs, and we choose a $\tau < c_0|\hom(H, h_{T,\eta_k}) - \hom(H, h_{T,\eta_{k+1}})|\Delta^{|V(H)|+1/2} - CT^{|V(H)|-1/2}\sqrt{\log(T)}$.

The remaining part of this step follows directly from the properties of random interval distillation described in Lines 19-32 of Algorithm 1. A detailed proof can be derived by following the argument presented in Theorem 3.1 of Fan & Wu (2024). Finally, we obtain

$$\mathbb{P}\left(\{\hat{K} = K\}, \bigcap_{j=1}^{\hat{K}} \left\{[l_j, r_j] \text{ covers } \eta_j \text{ and } r_j - l_j \le \frac{\Delta}{4}\right\}\right) \ge 1 - T^{-1}.$$

Step 2. By the arguments in **Step 1**, for η_k, with high probability, we have that $[l_k, r_k]$ covers η_k and $r_k - l_k \le \Delta/4$. Following the same argument as in Theorem 3.1 of Fan & Wu (2024), there exists an interval $(s, e] \in I$ such that $\eta_k \le r_k \le e$, $s \le l_k \le \eta_k$, $|e - \eta_k| \le \Delta/8$, and $|s - \eta_k| \le \Delta/8$. Thus, we have $r_k \le \eta_k + \Delta/8$ and $l_k \ge \eta_k - \Delta/8$. Recall the definition of $\hat{\Delta}$ in Line 3 of Algorithm 2. By construction, it is straightforward to verify that $7\Delta/8 \le \hat{\Delta} \le 9\Delta/8$. Therefore, we have $s_k = \lfloor l_k - \hat{\Delta}/2 \rfloor \in [\eta_k - 11\Delta/16, \eta_k - 7\Delta/16]$ and $e_k = \lfloor r_k + \hat{\Delta}/2 \rfloor \in [\eta_k + 7\Delta/16, \eta_k + 11\Delta/16]$.

Assume without loss of generality that $\hat{\eta}_k < \eta_k$. Similar to the proof of Lemma A.3, let $l = |V(H)|$, and let $\langle a, b\rangle = a^\top b$ for two vectors a and b. Let $u_0, u_1, \ldots, u_l$ be the orthogonalized and normalized vectors based on the columns of U_0 (defined in (5)). Let $a_m^b = (0, \ldots, 0, 1^m, \ldots, (e_k - b)^m)^\top$ for $s_k + 1 \le b \le e_k$, and let W_m^b be the coefficients from (5) based on a_m^b. Let $w_m^b = W_m^b / \sqrt{(W_m^b)^\top W_m^b}$.

By construction in (5), for any b, the vectors $u_0, u_1, \ldots, u_l, w_0^b, \ldots, w_l^b$ are mutually orthogonal. Let $X_r = N_H(G_r)$ for notational simplicity. Then, for $s_k < b \le e_k$, we have

$$(\tilde{X}^b_{s_k,e_k})^2 = \sum_{i=0}^{l} \langle X|_{(s_k+1):e_k}, w_i^b\rangle^2 = C_3 + \sum_{i=0}^{l} \langle X|_{(s_k+1):e_k}, u_i\rangle^2 + \sum_{i=0}^{l} \langle X|_{(s_k+1):e_k}, w_i^b\rangle^2,$$

where C_3 is a constant that depends on s_k, e_k, and $X|_{(s_k+1):e_k}$ is the vector $(X_{s_k+1}, \ldots, X_{e_k})^\top$. By the nature of orthogonal projection, we derive that

$$(\tilde{X}^b_{s_k,e_k})^2 = C_3 + \|P_V X|_{(s_k+1):e_k}\|^2 = C_3 + \|P_{V_1} X|_{(s_k+1):e_k}\|^2 = C_3 + \|P_{V_2^b} X|_{(s_k+1):b}\|^2 + \|P_{V_3^{e_k-b}} X|_{(b+1):e_k}\|^2,$$

where $V = (u_0, \ldots, u_l, w_0^b, \ldots, w_l^b)$,

$$V_1 = \begin{pmatrix} 1 & s_k+1 & (s_k+1)^2 & \cdots & (s_k+1)^l & 0 & 0 & \cdots & 0 \\ 1 & s_k+2 & (s_k+2)^2 & \cdots & (s_k+2)^l & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & b & b^2 & \cdots & b^l & 0 & 0 & \cdots & 0 \\ 1 & b+1 & (b+1)^2 & \cdots & (b+1)^l & 1 & 1^2 & \cdots & 1^l \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & e_k & e_k^2 & \cdots & e_k^l & 1 & e_k - b & \cdots & (e_k-b)^l \end{pmatrix},$$

$$V_2^b = \begin{pmatrix} 1 & s_k+1 & (s_k+1)^2 & \cdots & (s_k+1)^l \\ 1 & s_k+2 & (s_k+2)^2 & \cdots & (s_k+2)^l \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & b & b^2 & \cdots & b^l \end{pmatrix}, V_3^{e_k-b} = \begin{pmatrix} 1 & 1 & 1^2 & \cdots & 1^l \\ 1 & 2 & 2^2 & \cdots & 2^l \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (e_k-b) & (e_k-b)^2 & \cdots & (e_k-b)^l \end{pmatrix}.$$

Since $(\tilde{X}^{\hat{\eta}_k}_{s_k,e_k})^2 \ge (\tilde{X}^{\eta_k}_{s_k,e_k})^2$ by Algorithm 2, we derive that

$$\|P_{V_2^{\hat{\eta}_k}} X|_{(s_k+1):\hat{\eta}_k}\|^2 + \|P_{V_3^{e_k-\hat{\eta}_k}} X|_{\hat{\eta}_k+1:e}\|^2 \ge \|P_{V_2^{\eta_k}} X|_{(s_k+1):\eta_k}\|^2 + \|P_{V_3^{e_k-\eta_k}} X|_{\eta_k+1:e_k}\|^2. \tag{10}$$

Note that $\text{span}(V_2^b) = \text{span}(V_3^{b-s_k})$, and as a result, by subtracting the term (recall that we assume $\hat{\eta}_k < \eta_k$),

$$\|P_{V_3^{\hat{\eta}_k - s_k}} X|_{(s_k+1):\hat{\eta}_k}\|^2 + \|P_{V_3^{\eta_k-\hat{\eta}_k}} X|_{(\hat{\eta}_k+1):\eta_k}\|^2 + \|P_{V_3^{e_k-\eta_k}} X|_{(\eta_k+1):e_k}\|^2$$

from both sides of (10), we have

$$\begin{aligned}&\|P_{V_3^{e_k-\hat{\eta}_k}}X|_{(\hat{\eta}_k+1):e_k}\|^2-\|P_{V_3^{\eta_k-\hat{\eta}_k}}X|_{(\hat{\eta}_k+1):\eta_k}\|^2-\|P_{V_3^{e_k-\eta_k}}X|_{(\eta_k+1):e_k}\|^2\\&\geq\|P_{V_3^{\eta_k-s_k}}X|_{(s_k+1):\eta_k}\|^2-\|P_{V_3^{\hat{\eta}_k-s_k}}X|_{(s_k+1):\hat{\eta}_k}\|^2-\|P_{V_3^{\eta_k-\hat{\eta}_k}}X|_{(\hat{\eta}_k+1):\eta_k}\|^2.\end{aligned}\tag{11}$$

For the right-hand side, recall that we assume $\hat{\eta}_k<\eta_k$. Moreover, note that $\mathbb{E}X$ is a polynomial on $[u,v]$ for all $u<v<\eta_k$, implying $\mathbb{E}X|_{u:v}\in\text{span}(V_3^{v-u+1})$. Consequently, $(I-P_{V_3^{v-u+1}})\mathbb{E}X=0$. This gives

$$\begin{aligned}&\|P_{V_3^{\eta_k-s_k}}X|_{(s_k+1):\eta_k}\|^2-\|P_{V_3^{\hat{\eta}_k-s_k}}X|_{(s_k+1):\hat{\eta}_k}\|^2-\|P_{V_3^{\eta_k-\hat{\eta}_k}}X|_{(\hat{\eta}_k+1):\eta_k}\|^2\\&=\|X|_{(s_k+1):\eta_k}\|^2-\|(I-P_{V_3^{\eta_k-s_k}})X|_{(s_k+1):\eta_k}\|^2-\|X|_{(s_k+1):\hat{\eta}_k}\|^2\\&\quad+\|(I-P_{V_3^{\hat{\eta}_k-s_k}})X|_{(s_k+1):\hat{\eta}_k}\|^2-\|X|_{(\hat{\eta}_k+1):\eta_k}\|^2+\|(I-P_{V_3^{\eta_k-\hat{\eta}_k}})X|_{(\hat{\eta}_k+1):\eta_k}\|^2\\&=-\|(I-P_{V_3^{\eta_k-s_k}})(X-\mathbb{E}X)|_{(s_k+1):\eta_k}\|^2+\|(I-P_{V_3^{\hat{\eta}_k-s_k}})(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\|^2\\&\quad+\|(I-P_{V_3^{\eta_k-\hat{\eta}_k}})(X-\mathbb{E}X)|_{(\hat{\eta}_k+1):\eta_k}\|^2\\&=\|P_{V_3^{\eta_k-s_k}}(X-\mathbb{E}X)|_{(s_k+1):\eta_k}\|^2-\|P_{V_3^{\hat{\eta}_k-s_k}}(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\|^2\\&\quad-\|P_{V_3^{\eta_k-\hat{\eta}_k}}(X-\mathbb{E}X)|_{(\hat{\eta}_k+1):\eta_k}\|^2.\end{aligned}\tag{12}$$

For $\|P_{V_3^{\hat{\eta}_k-s_k}}(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\|^2$, note that with probability greater than $1-CT^{-1}$,

$$\begin{aligned}\|P_{V_3^{\hat{\eta}_k-s_k}}(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\|^2&=\|V_3^{\hat{\eta}_k-s_k}((V_3^{\hat{\eta}_k-s_k})^\top V_3^{\hat{\eta}_k-s_k})^{-1}(V_3^{\hat{\eta}_k-s_k})^\top(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\|^2\\&=\|((V_3^{\hat{\eta}_k-s_k})^\top V_3^{\hat{\eta}_k-s_k})^{-1/2}(V_3^{\hat{\eta}_k-s_k})^\top(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\|^2\\&=\sum_{i=1}^{l+1}\left(e_i^\top((V_3^{\hat{\eta}_k-s_k})^\top V_3^{\hat{\eta}_k-s_k})^{-1/2}(V_3^{\hat{\eta}_k-s_k})^\top(X-\mathbb{E}X)|_{(s_k+1):\hat{\eta}_k}\right)^2\\&\leq C_4T^{2|V(H)|-1}\log(T),\end{aligned}\tag{13}$$

where C_4 is a constant, $e_i=(0,\ldots,0,1,0,\ldots,0)^\top$, and the last inequality follows from Lemma A.7, noting that $\|e_i^\top((V_3^{\hat{\eta}_k-s_k})^\top V_3^{\hat{\eta}_k-s_k})^{-1/2}(V_3^{\hat{\eta}_k-s_k})^\top\|^2=1$ for all $1\leq i\leq l+1$. The other two terms in (12) can be handled similarly. Thus, we have

$$(12)\geq-2C_4T^{2|V(H)|-1}.\tag{14}$$

For the left-hand side of (11), by Lemma 6 in (Yu et al., 2022), we have

$$\begin{aligned}&\|P_{V_3^{e_k-\hat{\eta}_k}}X|_{(\hat{\eta}_k+1):e_k}\|^2-\|P_{V_3^{\eta_k-\hat{\eta}_k}}X|_{(\hat{\eta}_k+1):\eta_k}\|^2-\|P_{V_3^{e_k-\eta_k}}X|_{(\eta_k+1):e_k}\|^2\\&\leq-\Bigg|\sqrt{-\|P_{V_3^{e_k-\hat{\eta}_k}}\mathbb{E}X|_{(\hat{\eta}_k+1):e_k}\|^2+\|P_{V_3^{\eta_k-\hat{\eta}_k}}\mathbb{E}X|_{(\hat{\eta}_k+1):\eta_k}\|^2+\|P_{V_3^{e_k-\eta_k}}\mathbb{E}X|_{(\eta_k+1):e_k}\|^2}\\&-\sqrt{-\|P_{V_3^{e_k-\hat{\eta}_k}}(X-\mathbb{E}X)|_{(\hat{\eta}_k+1):e_k}\|^2+\|P_{V_3^{\eta_k-\hat{\eta}_k}}(X-\mathbb{E}X)|_{(\hat{\eta}_k+1):\eta_k}\|^2+\|P_{V_3^{e_k-\eta_k}}(X-\mathbb{E}X)|_{(\eta_k+1):e_k}\|^2}\Bigg|^2.\end{aligned}\tag{15}$$

Recall that $\kappa_k=|\hom(H,h_{T,\eta_k})-\hom(H,h_{T,\eta_{k-1}})|$ for $1\leq k\leq K$. By Proposition 8 in Yu et al. (2022), we have

$$\begin{aligned}&-\|P_{V_3^{e_k-\hat{\eta}_k}}\mathbb{E}X|_{(\hat{\eta}_k+1):e_k}\|^2+\|P_{V_3^{\eta_k-\hat{\eta}_k}}\mathbb{E}X|_{(\hat{\eta}_k+1):\eta_k}\|^2+\|P_{V_3^{e_k-\eta_k}}\mathbb{E}X|_{(\eta_k+1):e_k}\|^2\\&\geq c_l\kappa_k^2\left[|e_k-\eta_k|^{2l+1}\wedge|\hat{\eta}_k-\eta_k|^{2l+1}\right]\\&\geq c_l\kappa_k^2\left[|7\Delta/16|^{2l+1}\wedge|\hat{\eta}_k-\eta_k|^{2l+1}\right],\end{aligned}\tag{16}$$

where c_l is a constant that depends only on l. Furthermore, by following the arguments in (13), we have that with probability greater than $1-C/T$,

$$\begin{aligned}&-\|P_{V_3^{e_k-\hat{\eta}_k}}(X-\mathbb{E}X)|_{(\hat{\eta}_k+1):e_k}\|^2+\|P_{V_3^{\eta_k-\hat{\eta}_k}}(X-\mathbb{E}X)|_{(\hat{\eta}_k+1):\eta_k}\|^2+\|P_{V_3^{e_k-\eta_k}}(X-\mathbb{E}X)|_{(\eta_k+1):e_k}\|^2\\&\leq 2C_4T^{2|V(H)|-1}\log(T).\end{aligned}\tag{17}$$

Assume for the contradiction that for any constant $C_8 > 0$, there exists $T_0 \in \mathbb{N}$ such that for all $T > T_0$,

$$|\hat{\eta}_k - \eta_k| \geq C_8 \left(\frac{T^{2|V(H)|-1} \log(T)}{\min_k \kappa_k^2} \right)^{\frac{1}{2|V(H)|+1}}. \tag{18}$$

Then by (18), (16), and (6), we have

$$c_l \kappa_k^2 \left[|7\Delta/16|^{2l+1} \wedge |\hat{\eta}_k - \eta_k|^{2l+1} \right] \geq 8C_4 T^{2|V(H)|-1} \log(T),$$

where c_l is the constant in (16). This implies that

$$\begin{aligned}
&\|P_{V_3^{e_k - \hat{\eta}_k}} X|_{(\hat{\eta}_k+1):e_k}\|^2 - \|P_{V_3^{\eta_k - \hat{\eta}_k}} X|_{(\hat{\eta}_k+1):\eta_k}\|^2 - \|P_{V_3^{e_k - \eta_k}} X|_{(\eta_k+1):e_k}\|^2 \\
&\leq -\frac{1}{4} \left| \sqrt{-\|P_{V_3^{e_k - \hat{\eta}_k}} \mathbb{E}X|_{(\hat{\eta}_k+1):e_k}\|^2 + \|P_{V_3^{\eta_k - \hat{\eta}_k}} \mathbb{E}X|_{(\hat{\eta}_k+1):\eta_k}\|^2 + \|P_{V_3^{e_k - \eta_k}} \mathbb{E}X|_{(\eta_k+1):e_k}\|^2} \right|^2 \\
&\leq -\frac{1}{4} c_l \kappa_k^2 \left[|7\Delta/16|^{2l+1} \wedge |\hat{\eta}_k - \eta_k|^{2l+1} \right].
\end{aligned} \tag{19}$$

where we use the inequalities in (15), (16) and (17). Then by (11), (12), (14), and (19), we have

$$\kappa_k^2 \left[|7\Delta/16|^{2l+1} \wedge |\hat{\eta}_k - \eta_k|^{2l+1} \right] \leq \frac{8C_4}{c_l} T^{2|V(H)|-1} \log(T).$$

This gives

$$\kappa_k^2 \left[|7\Delta/16|^{2l+1} \wedge |\hat{\eta}_k - \eta_k|^{2l+1} \right] = 8C_4 c_l^{-1} T^{2|V(H)|-1} \log(T).$$

Note that $|\eta_k - \hat{\eta}_k| \leq |l_k - \eta_k| \leq \Delta/8 \leq 7\Delta/16$, therefore,

$$\begin{aligned}
\kappa_k^2 |\hat{\eta}_k - \eta_k|^{2l+1} &= \frac{8C_4}{c_l} T^{2|V(H)|-1} \log(T), \\
|\eta_k - \hat{\eta}_k| &\leq \left(\frac{8C_4}{c_l} \frac{T^{2|V(H)|-1} \log(T)}{\min_k \kappa_k^2} \right)^{\frac{1}{2l+1}}.
\end{aligned}$$

The above result contradicts (18). Recall that $l = |V(H)|$, and the result follows. □

A.8. Lemmas

Lemma A.3. *Let $f(t)$, $0 \leq t \leq T$, be a function such that*

$$f(t) = \sum_{j=0}^{l} a_{0j} t^j + \sum_{j=0}^{l} a_{1j} (t - \eta)^j I(t \geq \eta),$$

where a_{0j}, a_{1j} are coefficients satisfying $a_{1l} \neq 0$, and $l + 1 < \eta \leq T - l$ is a change-point. For an integer $t \in [1, T]$, let $\tilde{f}_{0,T}^{t,k} = \sum_{r=1}^{T} \tilde{w}_{0,T}^{t,k}(r) f(r)$ for $0 \leq k \leq l$, and define $\tilde{f}_{0,T}^{t} = \sqrt{\sum_{k=0}^{l} (\tilde{f}_{0,T}^{t,k})^2}$, where the coefficients $\tilde{w}_{0,T}^{t,k}(r)$ are defined in (5). Then the following hold:

1. *For every $1 \leq t \leq T$, we have $\tilde{f}_{0,T}^{\eta} \geq \tilde{f}_{0,T}^{t}$.*
2. *For every $\eta + l \leq t \leq T$, we have*

$$(\tilde{f}_{0,T}^{\eta})^2 - (\tilde{f}_{0,T}^{t})^2 \geq c_l a_{1l}^2 \left[\eta^{2l+1} \wedge (t - \eta)^{2l+1} \right],$$

   where $c_l > 0$ is a constant that depends only on l.

3. $$\tilde{f}_{0,T}^{\eta} \geq \sqrt{c_l} |a_{1l}| \Delta^{\frac{2l+1}{2}},$$

   where $\Delta = \eta \wedge (T - \eta)$ and $c_l > 0$ is a constant that depends only on l.

Remark A.4. For simplicity, we prove the result for $\tilde{f}^t_{0,T}$ with $0 < t \leq T$, but the same result holds for $\tilde{f}^t_{s,e}$ with $0 \leq s < t \leq e \leq T$.

Proof. Let $\langle a, b\rangle = a^\top b$ for two vectors a and b. Let $u_0, u_1, \cdots, u_l$ denote the orthogonalized and normalized vectors based on the columns of U_0 (defined in (5)) with $s = 0, e = T$. Let $a^b_k = (0, \cdots, 0, 1^k, \cdots, (T-b)^k)$ for $1 \leq b \leq T$, and let W^b_k be the coefficients from (5) based on a^b_k. Let $w^b_k = \frac{W^b_k}{\sqrt{(W^b_k)^\top W^b_k}}$. By the construction in (5), for any b, the vectors $u_0, u_1, \cdots, u_l, w^b_0, \cdots, w^b_l$ are mutually orthogonal. In the following proof, we use f as a shorthand for the vector $(f(1), f(2), \cdots, f(T))^\top$. Note that for any b,

$$\tilde{f}^b_{0,T} = \sqrt{\sum_{i=0}^{l} \langle f, w^b_i\rangle^2}.$$

Proof of the first result: Note that $f \in \text{span}\{u_0, u_1, \cdots, u_l, w^\eta_0, \cdots, w^\eta_l\}$. Then we have

$$\begin{aligned} f &= \sum_{i=0}^{l} \langle f, u_i\rangle u_i + \sum_{i=0}^{l} \langle f, w^\eta_i\rangle w^\eta_i, \\ \|f\|_2^2 &= \sum_{i=0}^{l} \langle f, u_i\rangle^2 + \sum_{i=0}^{l} \langle f, w^\eta_i\rangle^2. \end{aligned} \tag{20}$$

On the other hand, for $b \neq \eta$, f is not necessarily a linear combination of $u_0, u_1, \cdots, u_l, w^b_0, \cdots, w^b_l$. Therefore,

$$\|f\|_2^2 \geq \sum_{i=0}^{l} \langle f, u_i\rangle^2 + \sum_{i=0}^{l} \langle f, w^b_i\rangle^2.$$

As a result, $\sum_{i=0}^{l} \langle f, w^\eta_i\rangle^2 \geq \sum_{i=0}^{l} \langle f, w^b_i\rangle^2$, and the first result follows.

Proof of the second result: By (20), we have

$$\begin{aligned} \langle f, w^b_0\rangle &= \sum_{i=0}^{l} \langle f, w^\eta_i\rangle\langle w^\eta_i, w^b_0\rangle, \\ \langle f, w^b_1\rangle &= \sum_{i=0}^{l} \langle f, w^\eta_i\rangle\langle w^\eta_i, w^b_1\rangle, \\ &\vdots \\ \langle f, w^b_l\rangle &= \sum_{i=0}^{l} \langle f, w^\eta_i\rangle\langle w^\eta_i, w^b_l\rangle. \end{aligned}$$

Then

$$\begin{aligned} \langle f, w^b_0\rangle^2 &= \sum_{i=0}^{l} \langle f, w^\eta_i\rangle\langle f, w^b_0\rangle\langle w^\eta_i, w^b_0\rangle, \\ \langle f, w^b_1\rangle^2 &= \sum_{i=0}^{l} \langle f, w^\eta_i\rangle\langle f, w^b_1\rangle\langle w^\eta_i, w^b_1\rangle, \\ &\vdots \\ \langle f, w^b_l\rangle^2 &= \sum_{i=0}^{l} \langle f, w^\eta_i\rangle\langle f, w^b_l\rangle\langle w^\eta_i, w^b_l\rangle. \end{aligned}$$

Therefore,

$$
\begin{aligned}
&(\tilde{f}^{\eta}_{0,T})^2 - (\tilde{f}^{b}_{0,T})^2 \\
&= \sum_{i=0}^{l} \langle f, w_i^{\eta}\rangle^2 - \sum_{i=0}^{l} \langle f, w_i^{b}\rangle^2 \\
&= \sum_{i=0}^{l} \langle f, w_i^{\eta}\rangle^2 + \sum_{i=0}^{l} \langle f, w_i^{b}\rangle^2 - 2\sum_{i=0}^{l}\sum_{j=0}^{l} \langle f, w_i^{\eta}\rangle\langle f, w_j^{b}\rangle\langle w_i^{\eta}, w_j^{b}\rangle \\
&= \left\| \sum_{i=0}^{l} \langle f, w_i^{\eta}\rangle w_i^{\eta} - \sum_{i=0}^{l} \langle f, w_i^{b}\rangle w_i^{b} \right\|_2^2 \\
&= \left\| f - \sum_{i=0}^{l} \langle f, u_i\rangle u_i - \sum_{i=0}^{l} \langle f, w_i^{b}\rangle w_i^{b} \right\|_2^2 .
\end{aligned}
$$

Note that the above quantity is the residual of f projecting onto $u_0, \cdots, u_l, w_0^b, \cdots, w_l^b$, and consequently, onto $u_0, \cdots, u_l, a_0^b, \cdots, a_l^b$. Then for $b \geq \eta + l$, we have

$$
\begin{aligned}
&(\tilde{f}^{\eta}_{0,T})^2 - (\tilde{f}^{b}_{0,T})^2 \\
&= \min_{c_0,\cdots,c_l,c_{l+1},\cdots,c_{2l+1}} \left\| f - \sum_{i=0}^{l} c_{l+1+i} a_i^b - \sum_{i=0}^{l} c_i (1^i, \cdots, T^i)^\top \right\|_2^2 \\
&\geq \min_{c_0,\cdots,c_l} \left\| (f(1), \cdots, f(b))^\top - \sum_{i=0}^{l} c_i (1^i, \cdots, b^i)^\top \right\|_2^2 . \qquad (21)
\end{aligned}
$$

By (21), and since $\eta > l+1, b \geq \eta + l$, we apply Proposition 8 in Yu et al. (2022) to obtain

$$
\min_{c_0,\cdots,c_l} \left\| (f(1), \cdots, f(b))^\top - \sum_{i=0}^{l} c_i (1^i, \cdots, b^i)^\top \right\|_2^2 \geq \tilde{c}_l a_{1l}^2 \left[\eta^{2l+1} \wedge (b-\eta)^{2l+1} \right],
$$

where $\tilde{c}_l > 0$ is a constant depending on l. Thus, the result follows.

Proof of the third result: By taking $b = T$ in the second result and noting that $\tilde{f}^T_{0,T} = 0$ from (5), the result follows. □

Lemma A.5. *For $\tilde{w}^{t,k}_{s,e}$ defined in (5), we have for any $\varepsilon > 0$,*

$$
\mathbb{P}\left(\sup_{s,t,e,k} \left| \sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r) (N_H(G_r) - \mathbb{E}N_H(G_r)) \right| \geq \varepsilon \right) \leq C_{m1} T^3 \exp\left(-\frac{\varepsilon^2}{C_{m2} T^{2|V(H)|-1}} \right).
$$

Proof. Recall that m is the size of the initial graph. Define $A_T(\varphi(u), \varphi(v)) = I(\varphi(u)\varphi(v) \in E(G_r))$. For the triple (s,t,e), we have

$$
\begin{aligned}
\sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r) N_H(G_r) &= \sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r) \sum_{\varphi: V(H)\to V(G_r)} \prod_{uv\in E(H)} I(\varphi(u)\varphi(v) \in E(G_r)) \\
&= \sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r) \sum_{\varphi: V(H)\to V(G_r)} \prod_{uv\in E(H)} A_T(\varphi(u), \varphi(v)),
\end{aligned}
$$

where the sum is over the possible assignments of $\varphi : V(H) \to V(G_r)$. Let

$$
g_{s,t,e}(A_T(1,2), \ldots, A_T(T+m-2, T+m-1)) = \sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r) \sum_{\varphi: V(H)\to V(G_r)} \prod_{uv\in E(H)} A_T(\varphi(u), \varphi(v)).
$$

For $1 \le u \ne v \le T+m-1$, we derive that

$$\begin{aligned}
&|g_{s,t,e}(A_T(1,2),\cdots,A_T(u,v),\cdots,A_T(T+m-2,T+m-1)) \\
&- g_{s,t,e}(A_T(1,2),\cdots,|1-A_T(u,v)|,\cdots,A_T(T+m-2,T+m-1))| \\
&\le \sum_{r=s+1}^{e} |\tilde{w}_{s,e}^{t,k}(r)| \sum_{\varphi:V(H)\to V(G_r)} \sum_{i,j\in V(H)} (I(\varphi(i)\varphi(j)=uv)+I(\varphi(i)\varphi(j)=vu)) \\
&\le 2\sum_{r=s+1}^{e} |\tilde{w}_{s,e}^{t,k}(r)| \sum_{\varphi:V(H)\to V(G_r)} \sum_{i,j\in V(H)} I(\varphi(i)\varphi(j)=uv) I(u\le e, v\le e) \\
&\le 2\sum_{r=s+1}^{e} |\tilde{w}_{s,e}^{t,k}(r)| |V(G_r)|^{|V(H)|-2} I(u\le e, v\le e) \\
&\le 2\sqrt{\sum_{r=s+1}^{e} (\tilde{w}_{s,e}^{t,k}(r))^2}\sqrt{\sum_{r=s+1}^{e} \left(r^{|V(H)|-2}\right)^2} I(u\le e, v\le e) \\
&\le 2e^{|V(H)|-3/2} I(u\le e, v\le e). \qquad (22)
\end{aligned}$$

Here, we applied the Cauchy-Schwarz inequality and the fact that $\sum_{r=s+1}^{e}(\tilde{w}_{s,e}^{t,k}(r))^2 = 1$. Moreover, we have

$$\sum_{u=1}^{T+m-2}\sum_{v=u+1}^{T+m-1} \left(2e^{|V(H)|-3/2}\right)^2 I(u\le e, v\le e) \le 4(T+m-1)^{2|V(H)|-1}. \qquad (23)$$

By combining (22) and (23) and using McDiarmid's inequality, we get for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{r=s+1}^{e} \tilde{w}_{s,e}^{t,k}(r)(N_H(G_r)-\mathbb{E}N_H(G_r))\right| \ge \varepsilon \middle| U_1,\ldots,U_T\right) \le 2\exp\left(-\frac{\varepsilon^2}{2(T+m-1)^{2|V(H)|-1}}\right).$$

Consequently, we have

$$\mathbb{P}\left(\left|\sum_{r=s+1}^{e} \tilde{w}_{s,e}^{t,k}(r)(N_H(G_r)-\mathbb{E}N_H(G_r))\right| \ge \varepsilon\right) \le 2\exp\left(-\frac{\varepsilon^2}{2(T+m-1)^{2|V(H)|-1}}\right),$$

$$\mathbb{P}\left(\sup_{s,t,e,k}\left|\sum_{r=s+1}^{e} \tilde{w}_{s,e}^{t,k}(r)(N_H(G_r)-\mathbb{E}N_H(G_r))\right| \ge \varepsilon\right) \le 2|V(H)|(T+m-1)^3\exp\left(-\frac{\varepsilon^2}{2(T+m-1)^{2|V(H)|-1}}\right).$$

Since m is fixed, the result follows. □

Lemma A.6. *Recall $\tilde{X}_{s,e}^{t}(H)$ defined in (3), and let*

$$\tilde{f}_{s,e}^{t,k}(H) = \sum_{r=s+1}^{e} \tilde{w}_{s,e}^{t,k}(r)\mathbb{E}N_H(G_r), \quad \tilde{f}_{s,e}^{t}(H) = \sqrt{\sum_{k=0}^{|V(H)|} (\tilde{f}_{s,e}^{t,k}(H))^2}.$$

Then, for some large constant $C > 0$,

$$\mathbb{P}\left(\sup_{s,t,e:s<t<e}\left|\tilde{X}_{s,e}^{t}(H) - \tilde{f}_{s,e}^{t}(H)\right| \le CT^{|V(H)|-1/2}\log^{1/2}(T)\right) \ge 1-T^{-1}.$$

Proof. Let $\varepsilon_r = N_H(G_r) - \mathbb{E}N_H(G_r)$, $\tilde{\varepsilon}_{s,e}^{t,k}(H) = \tilde{X}_{s,e}^{t,k}(H) - \tilde{f}_{s,e}^{t,k}(H)$, and

$$\tilde{\varepsilon}_{s,e}^{t}(H) = \sqrt{\sum_{k=0}^{|V(H)|} (\tilde{\varepsilon}_{s,e}^{t,k}(H))^2}.$$

Thus, $\tilde{\varepsilon}^{t,k}_{s,e} = \sum_{r=s+1}^{e} \tilde{w}^{t,k}_{s,e}(r)\varepsilon_r$. By Lemma A.5, we have for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{s,t,e,k}\left|\sum_{r=s+1}^{e}\tilde{w}^{t,k}_{s,e}(r)(N_H(G_r)-\mathbb{E}N_H(G_r))\right|\geq\varepsilon\right)\leq C_{m1}T^3\exp\left(-\frac{\varepsilon^2}{C_{m2}T^{2|V(H)|-1}}\right).$$

Note that $|\tilde{X}^t_{s,e}(H) - \tilde{f}^t_{s,e}(H)| \leq \tilde{\varepsilon}^t_{s,e}(H)$. Therefore, we have

$$\begin{aligned}\mathbb{P}\left(\sup_{s<t<e}\left|\tilde{X}^t_{s,e}(H)-\tilde{f}^t_{s,e}(H)\right|\geq\epsilon\right)&\leq\mathbb{P}\left(\sup_{s,t,e:s<t<e}\left|\sum_{k=0}^{|V(H)|}(\tilde{\varepsilon}^{t,k}_{s,e}(H))^2\right|\geq\epsilon^2\right)\\&\leq\mathbb{P}\left(\sup_{s,t,e,k}\left|\tilde{\varepsilon}^{t,k}_{s,e}(H)\right|\geq c_1\epsilon\right)\\&\leq C_{m1}T^3\exp\left(-\frac{c_2\epsilon^2}{T^{2|V(H)|-1}}\right),\end{aligned}$$

where c_1, c_2, C_{m1} are positive constants. Finally, for

$$C_{m1}T^3\exp\left(-\frac{c_2\varepsilon^2}{T^{2|V(H)|-1}}\right)\leq T^{-1},$$

we can choose $\varepsilon = CT^{|V(H)|-1/2}\log^{1/2}(T)$ for some constant $C > 0$. □

Lemma A.7. *Suppose that $a_{s+1}, a_{s+2}, \ldots, a_e$ is a sequence of real numbers such that $\sum_{r=s+1}^{e} a_r^2 = 1$. Then, for any $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\left|\sum_{r=s+1}^{e}a_r\left(N_H(G_r)-\mathbb{E}N_H(G_r)\right)\right|\geq\varepsilon\right)\leq C_5\exp\left(-\frac{\varepsilon^2}{C_6T^{2|V(H)|-1}}\right),$$

where $C_5, C_6 > 0$ are constants. Consequently, we have

$$\mathbb{P}\left(\left|\sum_{r=s+1}^{e}a_r\left(N_H(G_r)-\mathbb{E}N_H(G_r)\right)\right|\geq C_7T^{|V(H)|-1/2}\sqrt{\log(T)}\right)\leq\frac{1}{T},$$

for some constant $C_7 > 0$.

Proof. The proof follows directly from the proof of Lemma A.5. □