Statistica Sinica **35** (2025), 1917-1946
doi:https://doi.org/10.5705/ss.202024.0048

MULTISCALE TESTS FOR POINT PROCESSES AND LONGITUDINAL NETWORKS

Youmeng Jiang and Min Xu*

Rutgers University

Abstract: We propose a new testing framework applicable to both the two-sample problem on point processes and the community detection problem on rectangular arrays of point processes, which we refer to as longitudinal networks; the latter problem is useful in situations where we observe interactions among a group of individuals over time. Our framework is based on a multiscale discretization scheme that considers not just the global null but also a collection of nulls local to small regions in the domain; in the two-sample problem, the local rejections tell us where the intensity functions differ and in the longitudinal network problem, the local rejections tell us when the community structure is most salient. We provide theoretical analysis for the two-sample problem and show that our method has minimax optimal power under a Holder continuity condition. We provide extensive simulation and real data analysis demonstrating the practicality of our proposed method.

Key words and phrases: Longitudinal network, multiscale, point process, two-sample test.

1. Introduction

In many applications involving network data, we observe just not a single static network but rather interactions over time. For example, in business applications, we may observe the timestamp of emails exchanged among employees in a company or transactions over time between people on an e-commerce website. In biology, animal behavioral researchers often use wearable devices to monitor physical interactions among a group of animals to understand their social dynamics (Gelardi et al., 2020).

In this paper, we study testing problems for interactions over time under the framework of *longitudinal networks*, also known as temporal networks. A longitudinal network A is a $n \times n$ array where each entry is an independent realization from a point process; for example, in an animal interaction network, the entry A_{jk} contains all the timestamps of when an interaction event between animal j and k was initiated. Based on the observed longitudinal network A, we ask whether the collection of intensity functions, one for each of the $\binom{n}{2}$ (or n^2 if network is directed) point processes, contain community structure. For static

*Corresponding author. E-mail: mx76@stat.rutgers.edu

networks, the problem of testing for community has been extensively studied, including Lei (2016) who proposed tests based on random matrix theory, and Gao and Lafferty (2017) and Jin, Ke and Luo (2021) who proposed tests based on subgraph count statistics. The longitudinal setting however introduces a new dimension to the problem: we may be interested in not just *whether* there is a community structure but also *when* the community structure is most apparent. For example, in an animal interaction network, the community structure may only be apparent during a specific time period, such as in the morning when the group is most active.

In this work, we propose a new multiscale testing framework based on discretization. In order to succinctly describe our framework, we first focus on the simpler problem of the two-sample test. Suppose we have two Poisson point processes over the same interval support $\mathcal{X} \subset \mathbb{R}$, with intensity functions λ_a and λ_b respectively. Our null hypothesis is that the two intensity functions are the same, i.e., $H_0 : \lambda_a = \lambda_b$. Our testing framework first partitions the ambient space into disjoint bins, which discretizes the Poisson process into a collection of independent Poisson random variables. The partition is chosen hierarchically at different scales to avoid the need to choose a smoothing parameter. In this way, we reduce the problem of testing Poisson processes to a hierarchical collection of tests on Poisson random variables, which we conduct by combining p-values obtained from Binomial exact tests and making the multiple testing adjustments via resampling under the null.

The advantage of this approach, aside from its computational simplicity, is that it can give granular local information: we can tell not just whether $\lambda_a \neq \lambda_b$ but where in support $\mathcal{X}$ that they differ significantly. We do this by testing not just the global null that $\lambda_a = \lambda_b$, but also a collection of local nulls that $\lambda_a|_I = \lambda_b|_I$ when restricted to a sub-region $I \subseteq \mathcal{X}$ in our hierarchical partition. To correct for sequential/multiple testing, we apply the adjustment method in Meinshausen (2008) to control the family-wise error rate. Somewhat surprisingly, the simultaneously valid tests for the local nulls can be done on top of the test for the global null *for free* in the sense that they can be done without any increase in computational complexity and without any decrease in statistical power.

This framework may be directly applied to longitudinal networks where we discretize the network into independent Poisson-weighted networks at different scales. This reduces the problem of testing for community in a longitudinal network to a hierarchical collection of tests for community in Poisson networks. To tackle the latter problem, we study tests based on the maximum eigenvalue as well as tests based on subgraph count statistics. We then combine the resulting p-values and make adjustments by resampling under the null. To generate samples under the null for networks, we propose a MCMC procedure based on a sampling algorithm for contingency tables. Although there are existing work on estimation for longitudinal networks and the related multi-layer network setting

(Zhang and Wang, 2025; Huang, Weng and Feng, 2023), we do not know of prior work focused on community structure testing for longitudinal networks.

One may ask whether a simple discretization scheme results in too much loss in power compared to existing tests on point processes based on say kernel smoothing (Fromont et al., 2013; Schrab et al., 2023) or wavelets (Taleb and Cohen, 2021). To that end, we analyze the power of our proposed framework theoretically in the two-sample testing problem and prove, under a Holder continuity condition, that when the dimension of the domain is small, our proposed test has optimal power in the sense that it attains minimax separation rate with respect to the distance $\int_I(\sqrt{\lambda_a} - \sqrt{\lambda_b})^2$ between the two intensity functions λ_a and λ_b. We also perform empirical studies validating that the discretization-based test has competitive power compared to existing approaches.

The remainder of the paper is organized in the following way. In Section 2, we define Poisson point process and the two-sample problem; we describe in detail our testing procedure in Section 2.2. In Section 3, we define the notion of longitudinal networks and testing for community structure; we describe our tests for three settings: symmetric networks with homogeneous baseline rate (Sec. 3.1), asymmetric networks with homogeneous baseline rate (Sec. S1.2), and degree-corrected networks with heterogeneous baseline rates (Sec. 3.2). In Section 4, we provide theorems characterizing the power of our proposed method for the two-sample test. Finally, in Section 5, we provide both simulation and real data experiments.

Notation: Given an integer K, we write $[K] := \{1, 2, \ldots, K\}$ and $[K]_0 := \{0, 1, 2, \ldots, K\}$. For a finite set L, we write $|L|$ to denote its cardinality. For a matrix A, we write $\lambda_1(A)$ to denote its maximum eigenvalue.

2. Tests on Point Processes

We first formally define a Poisson point process. Let the domain $\mathcal{X}$ be a compact subset of $\mathbb{R}^q$ with $\mathcal{B}(\mathcal{X})$ as the corresponding Borel σ-algebra. We say that $N : \mathcal{B}(\mathcal{X}) \to \mathbb{N}$ is a point process realization if it is a counting measure on $\mathcal{I}$ that is finite on every subset $I \in \mathcal{B}(\mathcal{X})$. We write $N(I) \in \mathbb{N}$ as the count of occurrences in $I \subset \mathcal{X}$, and write $N := N(\mathcal{X})$ as the total number of occurrences. We have $N < \infty$ since $\mathcal{X}$ is bounded. We write $X_1, X_2, \ldots, X_N \in \mathcal{I}$ as the locations of the occurrences.

For a finite measure $\Lambda(\cdot)$ on $\mathcal{X}$, we say that a random point process realization $N(\cdot)$ is generated by the inhomogeneous Poisson process $\mathrm{PP}(\Lambda)$ if, for all $k \in \mathbb{N}$, all disjoint subsets $A_1, A_2, \ldots A_k \in \mathcal{B}(\mathcal{I})$, and all $m_1, m_2, \ldots, m_k \in \mathbb{N}$, we have

$$\mathbb{P}\{N(A_1) = m_1, \ldots, N(A_k) = m_k\} = \prod_{i=1}^{k} \mathbb{P}\{N(A_i) = m_i\} = \prod_{i=1}^{k} \frac{e^{-\Lambda(A_i)}\Lambda(A_i)^{m_i}}{m_i!}.$$

We refer to $\Lambda(\cdot)$ as the intensity measure. In the case where $\Lambda(\cdot)$ has a density $\lambda(\cdot)$ (with respect to the Lebesgue measure), we also write $\mathrm{PP}(\lambda)$ as the same Poisson point process. We refer readers to Diggle (2013) and Kallenberg (2017) for additional details.

2.1. Two-sample test

Before considering the longitudinal network setting where we observe an array of point process realizations, we first study the two-sample setting where we observe two realizations: $N_a(\cdot) \sim \mathrm{PP}(\Lambda_a)$ and $N_b(\cdot) \sim \mathrm{PP}(\Lambda_b)$. Our goal is to test whether they have the same intensity measure, that is, we consider the null hypothesis

$$H_0 : \Lambda_a = \Lambda_b. \qquad \text{(Two-sample test)} \tag{2.1}$$

The null defined in (2.1) requires that $\Lambda_a = \Lambda_b$ everywhere on $\mathcal{X}$, so we refer to it as the *global null*. We can consider local tests where we ask whether $\Lambda_a = \Lambda_b$ when restricted to a sub-region. To formalize this, we define the notion of a hierarchical partitioning of $\mathcal{X}$.

Definition 1. Let $R \in \mathbb{N}$ be a resolution level. We say that $\boldsymbol{I} = \{\boldsymbol{I}^{(0)}, \boldsymbol{I}^{(1)}, \boldsymbol{I}^{(2)}, \ldots, \boldsymbol{I}^{(R)}\}$ is a hierarchical dyadic partition of $\mathcal{X}$ if $\boldsymbol{I}^{(0)} = \{I_1^{(0)}\}$ with $I_1^{(0)} = \mathcal{X}$ and

1. when $r = 1$, we let $\boldsymbol{I}^{(1)} = \{I_1^{(1)}, I_2^{(1)}\}$ be a partition of $\mathcal{X}$,
2. and for each $r > 1$, for each $\ell \in [2^r]$, let $I_\ell^{(r)}, I_{\ell+1}^{(r)}$ be a partition of $I_{(\ell+1)/2}^{(r-1)}$.

For each resolution level $r \in [R]_0$, the collection of intervals $\boldsymbol{I}^{(r)} = \{I_\ell^{(r)}\}_{\ell \in [2^r]}$ is a partition of $\mathcal{X}$. If we fix an interval $I_j^{(s)}$ where $s \in [R]_0$ and $j \in [2^s]$, then, defining

$$L(s, j, r) := \{2^{r-s}(j-1) + k : k = 1, 2, \ldots, 2^{r-s}\}, \text{ with } r > s,$$

we see that $\{I_j^{(s)}\}_{j \in L(s,j,r)}$ is a partition of $I_j^{(s)}$ at resolution level r. For example, we have $L(s, j, s+1) = \{2j-1, 2j\}$ so that $I_{2j-1}^{(s+1)}, I_{2j}^{(s+1)}$ is a dyadic partition of $I_j^{(s)}$ at one higher resolution level (Fig. 1).

To simplify notation, if $N(\cdot)$ is a point process realization on $\mathcal{X}$, we write $N^{(s,j)} := N(I_j^{(s)})$ as the number of occurrences in region $I_j^{(s)}$. When $\mathcal{X}$ is a one-dimensional interval, we can form the hierarchical partition $\mathbf{I}$ by recursively dividing each interval in halves. If $\mathcal{X}$ has dimension two or above, we can take any partitioning method that, in some sense, "evenly" divides each region. We also discuss how to construct $\boldsymbol{I}$ in Remark 3. For now, we assume that such a partition $\boldsymbol{I}$ is given and does not depend on the random realizations. Moreover, all of our discussions generalize to k-yadic partition in a straightforward way, but we will work with the dyadic version for simplicity of presentation.

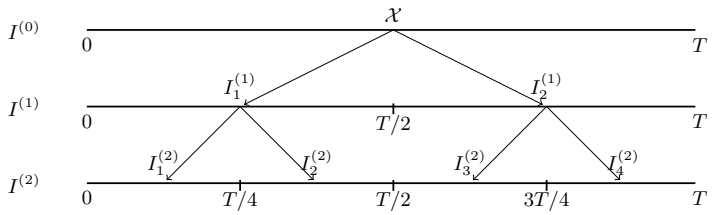


Figure 1.

For a given $\mathbf{I}$, we may then define the notion of a local null for the interval $I_j^{(s)}$:

$$H_0^{(s,j)} : \Lambda_a(\cdot) = \Lambda_b(\cdot) \text{ on } I_j^{(s)}. \qquad \text{(Local null)} \tag{2.2}$$

We note that, since $I_1^{(0)} = \mathcal{X}$, the null $H_0^{(0,1)}$ is exactly the global null. For reasons that will become clear, we also define a related notion of the local null, which we refer to as the *discretized local null*:

$$\bar{H}_0^{(s,j)} : \Lambda_a(I_j^{(s)}) = \Lambda_b(I_j^{(s)}). \qquad \text{(Discretized local null)} \tag{2.3}$$

We note that $H_0^{(s,j)}$ implies $\bar{H}_0^{(s,j)}$ but the two are generally not equivalent. They may be similar if $\Lambda_a(\cdot), \Lambda_b(\cdot)$ have a smooth density and if the region $I_j^{(s)}$ has a small diameter.

An important observation that underpins our testing procedure is the fact that the collection of local nulls $H_0^{(s,j)}$'s has a logical tree structure:

$$H_0^{(s,j)} \Rightarrow H_0^{(r,\ell)} \text{ for any } r \in \{s, s+1, \ldots, R\} \text{ and any } \ell \in L(s,j,r). \tag{2.4}$$

This holds because $\Lambda_a = \Lambda_b$ on the region $I_j^{(s)}$ implies that $\Lambda_a = \Lambda_b$ on every sub-region $I^{(r,\ell)} \subseteq I_j^{(s)}$. Because of this logical structure, if we do not reject $H_0^{(s,j)}$, then we should not reject $H_0^{(r,\ell)}$ for any sub-region $I^{(r,\ell)} \subseteq I^{(s,j)}$.

Remark 1. It is clear that $H_0^{(s,j)} \Rightarrow H_0^{(s+1,2j-1)} \cap H_0^{(s+1,2j)}$ (each local null implies its children). For the two sample test, we also have the reverse:

$$H_0^{s+1,2j-1} \cap H_0^{s+1,2j} \Rightarrow H_0^{(s,j)}, \tag{2.5}$$

that is, $H_0^{(s,j)}$ must be true if its two direct children are true. This allows us to obtain some small improvement in power when performing the multiple testing adjustment (Sec. 2.2.4). We note that (2.5) does not hold in the longitudinal network setting (cf. Rmk. 5).

Related work on testing for point processes. For the two-sample problem, Fromont et al. (2013) proposed tests for the global null using U-statistics based on kernel functions and proved their optimality; similar methods appear in Gretton et al. (2012). Methods based on scan statistics have been studied in

Kulldorff, Huang and Konty (2009), Walther (2010), and Picard, Reynaud-Bouret and Roquain (2018). In contrast to these work, our focus is on simultaneous testing of both the global and local nulls as well as on having a framework that easily extends to the longitudinal networks. There are also work on testing homogeneity (Fromont, Laurent and Reynaud-Bouret, 2011) and for testing whether the proportion of two intensity functions is a constant or increasing (Bovett and Saw, 1980; Deshpande, Mukhopadhyay and Naik-Nimbalkar, 1999).

2.2. Testing procedure for two-sample test

Our test procedure will produce valid p-values for H_0 and the entire family $H^{(s,j)}$ simultaneously in the following sense: we produce a collection of p-values $p^{(s,j)}$ such that if we reject

$$\mathcal{R}_\alpha = \{H_0^{(s,j)} : p^{(s,j)} \le \alpha \text{ and } p^{(s^*,j^*)} \le \alpha \text{ for all } I^{(s^*,j^*)} \text{ such that } I^{(s,j)} \subseteq I^{(s^*,j^*)}\}, \tag{2.6}$$

then we control family-wise error rate at level α, that is, $\mathcal{R}_\alpha$ contains no false positives with probability at least $\ge 1-\alpha$.

On a high level, our testing strategy is to approximate the local null $H_0^{(s,j)}$ by an intersection of discretized local nulls:

$$\check{H}_0^{(s,j)} := \bigcap_{r=s}^{R} \bigcap_{\ell \in L(s,j,r)} \bar{H}_0^{(r,\ell)}. \qquad \text{(Approximate local null)} \tag{2.7}$$

For a given $r \in \{s, s+1, \ldots, R\}$, we see that the intersection $\bigcap_{\ell \in L(s,j,r)} \bar{H}_0^{(r,\ell)}$ approximates $H_0^{(s,j)}$ at discretization/resolution level r. The additional intersection over r accounts for all the resolution levels. In the case where $s = 0$ and $j = 1$, we get an approximation of the global null for the two sample test (2.1):

$$\check{H}_0 \equiv \tilde{H}_0^{(0,1)} = \bigcap_{r=0}^{R} \bigcap_{\ell=1}^{2^r} \bar{H}_0^{(r,\ell)}. \tag{2.8}$$

Our testing procedure proceeds in four steps: (i) construct p-values $\bar{p}^{(r,\ell)}$'s for $\bar{H}_0^{(r,\ell)}$'s with exact tests, (ii) combine $p^{(r,\ell)}$ across ℓ at the same resolution level, (iii) combine p-value across the different resolution levels and use resampling to obtain individually valid p-values $\check{p}_F^{(s,j)}$ for each $H_0^{(s,j)}$, and (iv), apply sequential testing adjustment to obtain simultaneously valid p-values $p_F^{(s,j)}$. We explain each of the steps in detail below and give a concise description of the whole procedure in Algorithm 1. We also illustrate how the procedure works in an in-depth numerical example in Section 5.1.

2.2.1. Step 1: Compute p-value for each discretized local null

In the first step, we compute a p-value for each of the discretized local nulls, that is, for each $\bar{H}_0^{(r,\ell)}$ where $r \in [R], \ell \in [2^r]$. To test $\bar{H}_0^{(r,\ell)}$, we observe that the random counts $N_a^{(r,\ell)} := N_a(I_\ell^{(r)})$ and $N_b^{(r,\ell)} := N_b(I_\ell^{(r)})$ are Poisson random variables with means $\Lambda_a(I_\ell^{(r)}), \Lambda_b(I_\ell^{(r)}) \geq 0$ respectively. Define

$$N(\cdot) := N_a(\cdot) + N_b(\cdot)$$

as the aggregated realization. We then have that, under the null hypothesis and conditional on $N^{(r,\ell)}$, the random counts $N_a^{(r,\ell)}$ and $N_b^{(r,\ell)}$ have the binomial $\mathrm{Bin}(1/2, N^{(r,\ell)})$ distribution.

We take $N_a^{(r,\ell)}$ as the test statistic. Let $S_{\mathrm{Bin}(1/2,m)}(\cdot) := \mathbb{P}(|\mathrm{Bin}(1/2, m) - m/2| \geq \cdot)$ be the two-sided tail probability function, and write $\tilde{p}^{(r,\ell)} \equiv \tilde{p}^{(r,\ell)}(N_a(I_\ell^{(r)}), N^{(r,\ell)}) := S_{\mathrm{Bin}(1/2,N^{(r,\ell)})}(|N_a^{(r,\ell)} - N^{(r,\ell)}/2|)$ as the p-value. We may then reject the local null $\bar{H}_0^{(r,\ell)}$ at level $\alpha \in (0,1)$ if $\tilde{p}^{(r,\ell)} \leq \alpha$. However, under the null and conditional on $N^{(r,\ell)}$, $\tilde{p}^{(r,\ell)}$ has a discrete distribution. We may thus gain additional power by randomizing $\tilde{p}^{(r,\ell)}$ so that its distribution is continuous and uniform under the null. To that end, we generate an independent $U \sim \mathrm{Unif}[0,1]$, define $\tilde{S} := S_{\mathrm{Bin}(1/2,N^{(r,\ell)})}(|N_a^{(r,\ell)} - N^{(r,\ell)}/2| + 1)$, and define $\bar{p}^{(r,\ell)} = U\tilde{p}^{(r,\ell)} + (1-U)\tilde{S}$. The randomized p-value $\bar{p}^{(r,\ell)}$ has the $\mathrm{Unif}[0,1]$ distribution under $\bar{H}_0^{(r,\ell)}$. We also have that $\bar{p}^{(r,\ell)} \leq \tilde{p}^{(r,\ell)}$, so there is no loss in power (see proof in Prop. 3 in the Supplementary Material for completeness).

2.2.2. Step 2: Combining p-values of the same resolution level

In the second step, for each (s,j), we will consider each $r \in \{s, s+1, \ldots, R\}$ and combine the p-values $\{\bar{p}^{(r,\ell)} : \ell \in L(s,j,r)\}$. In the case where $s = 0$ and $j = 1$ so that $I_j^{(s)} = \mathcal{X}$, this amounts to combining the p-values $\{\bar{p}^{(r,\ell)}\}_{\ell \in [2^r]}$ for each $r \in [R]$. To simplify exposition, we describe the p-value combination method for when $s = 0$ and $j = 1$; the same method applies immediately for any (s,j).

We combine the p-values $\{\bar{p}^{(r,\ell)}\}_{\ell \in [2^r]}$ by specifying a function $f : [0,1]^{2^r} \to [0,1]$ and taking $f\big(\bar{p}^{(r,1)}, \ldots, \bar{p}^{(r,2^r)}\big)$. By choosing the combining function f carefully and using the fact that $\{\bar{p}^{(r,1)}, \ldots, \bar{p}^{(r,2^r)}\}$ are independent random variables uniform on $[0,1]$, we can guarantee that $f\big(\bar{p}^{(r,1)}, \ldots, \bar{p}^{(r,2^r)}\big)$ has the uniform distribution under the null. There are a number of reasonable choices for $f(\cdot)$ but we focus on two:

$$p_{\mathrm{F}}^{(r)} := S_{\chi^2_{2^r}}\left(-2 \sum_{\ell \in [2^r]} \log \bar{p}^{(r,\ell)}\right), \qquad \text{(Fisher combination)} \tag{2.9}$$

$$p_{\mathrm{M}}^{(r)} := S_{\beta_{2^r}}\left(\min_{\ell \in [2^r]} \bar{p}^{(r,\ell)}\right), \qquad \text{(Minimum combination)} \tag{2.10}$$

where $S_{\chi^2_{2^r}}(\cdot) := \mathbb{P}(\chi^2_{2^r} \geq \cdot)$ is the right tail probability function for the χ^2 distribution with 2^r degree of freedom and $S_{\beta_{2^r}}(\cdot) := \mathbb{P}(\mathrm{Beta}_{1,2^r} \leq \cdot)$ is the left

tail probability of the Beta distribution with parameter $(1, 2^r)$. The fact that $p_F^{(r)}$ and $p_M^{(r)}$ have the uniform distribution under the null follows from the fact that the negative sum of logarithm of independent Unif$[0,1]$ random variables has the χ^2 distribution and that the minimum of independent Unif$[0,1]$ random variables has the Beta distribution. For the remainder of this paper, we use Fisher combination by default but it would be trivial to use minimum combination instead.

We can follow the same procedure to compute, for any (s,j) and $r \in \{s, s+1, \ldots, R\}$, the p-value $p_F^{(s,j,r)}$ which combines $\{\bar{p}^{(r,\ell)}\}_{\ell \in L(s,j,r)}$. Moreover, we derive a dynamic program that, in the process of computing $p_F^{(1)}, \ldots, p_F^{(R)}$ for the global null case $(s = 0, j = 1)$, can simultaneously and without any additional computational burden, compute the whole collection $\{\{p_F^{(s,j,r)}\}_{r=s}^R\}_{s \in [R], j \in [2^r]}$. The dynamic program uses an iterative bottom-up approach and runs in time $O(N)$. We give the details in Algorithm 2.

Remark 2. The question of which combination method has more power depends on what the alternative is. We show through our theoretical analysis in Section 4 that when the integrated difference $\int\{(\lambda_a - \lambda_b)/\lambda\}^2$ is large, then Fisher combination has higher power. On the other hand, if $|\lambda_a - \lambda_b|$ is large only on a small region and 0 elsewhere, then the minimum combination method has higher power.

2.2.3. Step 3: Combining across different resolution levels

Finally, to obtain the p-value for $H_0^{(s,j)}$, we combine $\{p_F^{(s,j,r)}\}_{r=s}^R$ across resolution levels $\{s, s+1, \ldots, R\}$. There are again a number of choices, but we propose

$$\tilde{p}_F^{(s,j)} := \min\big\{p_F^{(s,j,r)} \,:\, r \in \{s, s+1, \ldots, R\}\big\}.$$

Since the random variables $\{p_F^{(s,j,r)}\}_{r=s}^R$ are not independent, the distribution of $\tilde{p}_F^{(s,j)}$ under H_0 is difficult to characterize exactly. Instead, we make adjustments to $\tilde{p}_F^{(s,j)}$. One straightforward way is to make the Bonferroni adjustment, where we let

$$\check{p}_F^{(s,j)} = (R - s + 1) \cdot \tilde{p}_F. \tag{2.11}$$

We may also adjust $\tilde{p}_F$ by resampling. We note that the realizations $N_a(\cdot)$ and $N_b(\cdot)$ can be equivalently characterized by two sequences of random variables $X_1, X_2, \ldots, X_N$ taking value on $\mathcal{X}$ and $M_1, M_2, \ldots, M_N$ taking value on $\{-1, 1\}$ where $N = N_a + N_b$ is the random length of the sequence. The occurrences of $N_a(\cdot)$ comprise of all X_i where $M_i = -1$ and the occurrences of $N_b(\cdot)$ comprise of those points for which $M_i = 1$.

Under H_0, $M_1, \ldots, M_N$ would be Rademacher random variables, that is, $\mathbb{P}_0(M_i = 1) = 1/2$, and independent of $X_1, \ldots, X_N$. Hence, to resample B samples from H_0, we do the following, for $b^* = 1, 2, \ldots, B$:

1. Generate $M_1^{(b^*)}, \ldots, M_N^{(b^*)} \sim$ Rademacher independently.
2. Take $N_a^{(b^*)}(\cdot) = \{X_i : M_i^{(b^*)} = -1\}$ and $N_b^{(b^*)}(\cdot) = \{X_i : M_i^{(b^*)} = 1\}$.

On each sample $N_a^{(b^*)}(\cdot), N_b^{(b^*)}(\cdot)$, we then repeat steps 1, 2, and the first part of step 3 to compute $\tilde{p}_{F,\,b^*}^{(s,j)}$ for each $s \in [R]$ and $j \in [2^s]$. We may then define

$$\check{p}_F^{(s,j)} = \frac{1}{B} \sum_{b^*=1}^{B} \mathbb{1}\big(\tilde{p}_{F,b^*}^{(s,j)} \leq \tilde{p}_F^{(s,j)}\big). \tag{2.12}$$

2.2.4. Step 4: Sequential/multiple testing adjustment

The p-values $\check{p}_F^{(s,j)}$'s produced from step 3 are individually valid in that, under $H_0^{(s,j)}$, we have that $\check{p}_F^{(s,j)} \leq \alpha$ with probability at most α. To account for sequential/multiple testing, we use the adjustment method proposed by Meinshausen (2008). For each (s, j) where $s \in [R]$ and $j \in [2^s]$, define

$$\mathcal{L}(s,j) = \begin{cases} \#\{\text{terminal nodes emanating from } (s,j) \text{ in } \boldsymbol{I}\} & \text{if } (s,j) \text{ is not a terminal node} \\ 2 & \text{if } (s,j) \text{ is a terminal node,} \end{cases}$$

where a terminal node in $\boldsymbol{I}$ is a region $I_\ell^{(r)}$ with no sub-region. If $\boldsymbol{I}$ is a full binary tree with R resolution levels, then $\mathcal{L}(s,j) = 2^{R-s}$ if $s < R$ and $\mathcal{L}(R,j) = 2$. The total number of terminal regions is $\mathcal{L}(0,1)$ which, in the case of a full binary tree, is 2^R. We then define the final adjusted p-value:

$$p_F^{(s,j)} = \check{p}_F^{(s,j)} \cdot \frac{\mathcal{L}(0,1)}{\mathcal{L}(s,j)} = \check{p}_F^{(s,j)} \cdot 2^{s \wedge (R-1)}. \tag{2.13}$$

We note in particular that the p-value $p_F^{(0,1)}$ for the global null does not receive any adjustment so that any rejections we make of the local nulls $H_0^{(s,j)}$ comes "for free" on top of our test for the global null. In other words, conducting the tests for the local nulls does not decrease our power for the global null.

Using the fact that under $H_0^{(s,j)}$, we have $\mathbb{P}(\check{p}^{(s,j)} \leq \alpha) \leq \alpha$, and Theorem 2 in Meinshausen (2008), the following FWER guarantee immediately follows:

Theorem 1. *The rejection set $\mathcal{R}_\alpha$ formed via* (2.6) *with p-values $\{p_F^{(s,j)}\}_{s\in[R], j\in[2^s]}$, as defined in* (2.13)*, has family-wise error rate (FWER) at most α.*

The overall procedure (Alg. 1) is computationally efficient. The whole collection $\{p_F^{(s,j,r)}\}$ can be computed in $O(N)$ time using Algorithm 2 so that the overall procedure has runtime complexity $O(NB)$ where B is the number of resampling repetitions.

Remark 3. Recall that we can equivalently describe the two realizations $N_a(\cdot)$ and $N_b(\cdot)$ through a sequence (of random length) of positions $X_1, \ldots, X_N \in \mathcal{X}$

Algorithm 1 Computing simultaneously valid p-values $p_F^{(s,j)}$ for all $H_0^{(s,j)}$.

Input: Poisson process realizations $N_a(\cdot)$ and $N_b(\cdot)$ and a hierarchical partitioning $\boldsymbol{I} = \{I_\ell^{(r)}\}_{r\in[R],\ell\in[2^r]}$ of the domain.

Output: $H_0^{(s,j)}$.

1: **for** each $r \in [R]$ **do**
2: **for** each $\ell \in [2^r]$ **do**
3: Set $\tilde{p}^{(r,\ell)} = S_{\mathrm{Bin}(1/2,N^{(r,\ell)})}(|N_a^{(r,\ell)} - N^{(r,\ell)}/2|)$.
4: Use randomization described in Section 2.2.1 to obtain $\bar{p}^{(r,\ell)}$.
5: **end for**
6: **end for**
7: Apply Algorithm 2 on $\{\bar{p}^{(r,\ell)}\}_{r\in[R],\ell\in[2^r]}$ to obtain $\{\{p_F^{(s,j,r)}\}_{r=s}^R\}_{s\in[R],j\in 2^s}$.
8: Compute $\tilde{p}_F^{(s,j)} := \min\{p_F^{(s,j,r)} : r \in \{s, s+1, \ldots, R\}\}$.
9: **for** $b^* \in \{1, 2, \ldots, B\}$ **do**:
10: Generate $M_1^{(b^*)}, \ldots, M_N^{(b^*)} \sim$ Rademacher independently.
11: Take $N_a^{(b^*)}(\cdot) = \{X_i : M_i^{(b^*)} = -1\}$ and $N_b^{(b^*)}(\cdot) = \{X_i : M_i^{(b^*)} = 1\}$.
12: Repeat lines 1 to 8 on $N_a^{(b^*)}(\cdot)$ and $N_b^{(b^*)}(\cdot)$ to obtain $\tilde{p}_{F,b^*}^{(s,j)}$.
13: **end for**
14: Compute the raw p-values $\check{p}_F^{(s,j)} := (1/B)\sum_{b^*=1}^B \mathbb{1}\{\tilde{p}_{F,b^*}^{(s,j)} \le \tilde{p}_F^{(s,j)}\}$.
15: Compute the adjusted p-values $p_F^{(s,j)} = \check{p}_F^{(s,j)} 2^{s\wedge(R-1)}$.

Algorithm 2 Dynamic program for computing the collection of p-values $\{\{p_F^{(s,j,r)}\}_{r=s}^R\}_{s\in[R],j\in[2^r]}$.

Input: a collection $\{\bar{p}^{(r,\ell)}\}_{r\in[R],\ell\in[2^r]}$.

Output: The collection $\{\{p_F^{(s,j,r)}\}_{r=s}^R\}_{s\in[R],j\in[2^r]}$.

1: For every $r \in [R], \ell \in [2^r]$, set $m_0^{r,\ell} = -2\log\bar{p}^{(r,\ell)}$
2: **for** $k \in \{1, 2, \ldots, R-1\}$ **do**:
3: For every $r \in [R-k]$ and $\ell \in [2^r]$, set $m_k^{(r,\ell)} = m_{k-1}^{r+1,2\ell-1} + m_{k-1}^{r+1,2\ell}$.
4: **end for**
5: **for** $s \in [R], j \in [2^s], r \in \{s, s+1, \ldots, R\}$ **do**:
6: Set $p_F^{(s,j,r)} = S_{\chi^2_{2^{r-s}}}(m_{r-s}^{r,j})$.
7: **end for**

of the union of $N_a(\cdot)$ and $N_b(\cdot)$, and a set of markers $M_1, \ldots, M_N \in \{-1, +1\}$, where $M_i = -1$ implies that X_i belong to $N_a(\cdot)$. We can then see that the type I error guarantee holds conditional on the aggregate positions $X_1, \ldots, X_N$ of the union of the two realizations. This is because the p-values $\bar{p}^{(r,\ell)}$ produced in step 1 for the discretized local nulls are valid conditional on the positions; we only use the fact that the random markers $M_1, \ldots, M_N$ are, conditionally on N, independent Rademacher random variables. An important implication of this fact is that the hierarchical partition $\boldsymbol{I}$ can depend on the aggregate positions $\{X_1, \ldots, X_N\}$ so long as it does not depend on $\{M_1, \ldots, M_n\}$; in particular, we can split each region such that each sub-region has equal number of "unmarked"

points. In practice, we recommend choosing the hierarchical partition in this way and setting the maximum resolution level $R = O(\log n)$ such that each bin at level R contains a constant number of points, say 10 or 20.

Remark 4. The adjustment method in Meinshausen (2008) can, in fact, be improved by looking at the test sequentially and removing any previously rejected hypotheses from the set of terminal nodes under consideration. This is analogous to how Holm's method improves upon the Bonferroni method. We refer readers to the excellent paper by Goeman and Solari (2010) for more details.

3. Tests on Longitudinal Networks

In this section, we consider interaction processes among a group of n individuals. For each pair of individuals $u, v \in [n]$, we write $N_{uv}(\cdot)$ as the realization that captures the interaction events between u and v over time. The collection $\{N_{uv}(\cdot) : u, v \in [n]\}$ is therefore an array of point process realizations which we refer to as a longitudinal network. Here, we take the network to be symmetric/undirected in that $N_{uv}(\cdot) = N_{vu}(\cdot)$; we study the directed/asymmetric setting in Section S1.2. Suppose $N_{uv}(\cdot) \sim PP(\Lambda_{uv})$ for $\binom{n}{2}$ intensity measures $\{\Lambda_{uv}\}$. We aim to test whether the intensity measures Λ_{uv}'s are all identical. When n is large however, the space of alternative hypothesis is enormous so that it is important for us to designate a plausible alternative with which to test against.

We therefore assume that there is an underlying block/community structure. More precisely, suppose each individual u belongs to one of K communities and write $\boldsymbol{\sigma}(u) \in [K]$ as the community membership of u, where $\boldsymbol{\sigma} : [n] \to [K]$ is the community assignment function. We assume that the probability distribution of the interactions between u, v depends only on the community memberships of u and v. More precisely, for each pairs of communities $s, t \in [K]$, let $\Gamma_{st}(\cdot)$ be an intensity measure and suppose

$$\Lambda_{uv} = \Gamma_{\boldsymbol{\sigma}(u)\boldsymbol{\sigma}(v)} \quad \text{for individuals } u, v \in [n].$$

We then define the null hypothesis to be $H_0 : K = 1$ and the alternative to be $H_1 : K > 1$. More precisely, we define the null hypothesis

$$H_0 : N_{uv}(\cdot) \sim \mathrm{PP}(\Gamma), \text{ for some } \Gamma, \text{ for all } u, v \in [n]. \qquad \text{(Symmetric array test)} \tag{3.1}$$

This is the generalization of the two-sample test to the array case. In many applications however, individuals may have different baseline rates of interactions. To capture potential rate heterogeneity, we propose to augment the block model with a vector $\boldsymbol{\theta} \in [0, \infty)^n$ of non-negative scalars and let $N_{uv}(\cdot) \sim \mathrm{PP}(\Gamma_{\boldsymbol{\sigma}(u)\boldsymbol{\sigma}(v)}\theta_u\theta_v)$.

We may then consider the same test of whether there exists a community structure in the interactions. An equivalent formulation is to define

$$H_0 : N_{uv}(\cdot) \sim \mathrm{PP}(\Gamma\theta_u\theta_v), \text{ for some } \Gamma \text{ and } \boldsymbol{\theta}, \text{ for all } u, v \in [n]. \tag{3.2}$$

We refer to (3.2) as the *degree-corrected array test*. In the next section, we focus on the symmetric array test and consider the degree-corrected array test in Section 3.2.

Related work on longitudinal networks. Recently, there has been increased attention on the modeling of dynamic networks (Holme and Saramäki, 2012). For example, Xu and Hero (2013) employ a state space model to describe temporal changes at the level of the connectivity pattern, DuBois, Butts and Smyth (2013) introduced a family of relational event models that capture the heterogeneity in the underlying interaction dynamics of network data over time. Modelling temporal interaction between two nodes by Poisson processes is also considered in Corneli, Latouche and Rossi (2016) and Matias, Rebafka and Villers (2018), where they provide likelihood-based algorithms for membership estimation. Zhang and Wang (2025) study longitudinal networks from a tensor factorization perspective; they discretize the time into bins and propose an adaptive merging method to ensure that the discretized network is not too sparse. We refer readers to the introduction in Zhang and Wang (2025) for a more extensive review of estimation methods for longitudinal networks. Unlike estimation, testing for longitudinal networks has not received much attention, which is where our work enters the picture.

3.1. Testing procedure for longitudinal networks

We now consider the test of interaction processes among a group of individuals, defined in (3.1). As with the two-sample test, we construct our testing procedure based on a hierarchical partitioning $\boldsymbol{I} = \{I_\ell^{(r)}\}$ of the support $\mathcal{X}$ of the interactions, as described in Definition 1. The test follows the same steps as that described in Section 2.1. The only differences are in step 1, where we specify different test statistics for the discretized local null, and in step 4, where we specify different resampling algorithms.

Following the two-sample test described in Section 2.1, we define, for a resolution level $r \in [R]$ and $\ell \in [2^r]$, the local null

$$H_0^{(r,\ell)} : \Lambda_{uv} = \Gamma, \text{ on } I_\ell^{(r)} \text{ for some common } \Gamma, \text{ for all } u, v \in [n].$$

We also define discretized local null

$$\bar{H}_0^{(r,\ell)} : \Lambda_{uv}(I_\ell^{(r)}) = \gamma, \text{ for some common } \gamma \geq 0, \text{ for all } u, v \in [n].$$

Our testing procedure follows the same steps as the two-sample test.

3.1.1. Step 1: Compute p-values for each discretized local null.

In the first step, we test the discretized local null $\bar{H}_0^{(r,\ell)}$ for some $r \in [R]$ and $\ell \in [2^r]$. Define $N_{uv}^{(r,\ell)} := N_{uv}(I_\ell^{(r)})$ and observe that $N_{uv}^{(r,\ell)} \sim \text{Poisson}\{\Lambda_{uv}(I_\ell^{(r)})\}$. To motivate our test, define an integer matrix $A^{(r,\ell)} \in \mathbb{N}^{n\times n}$ where

$$A_{uv}^{(r,\ell)} = \begin{cases} N_{uv}^{(r,\ell)}, & u \neq v \\ 0, & u = v \end{cases}. \tag{3.3}$$

We view $A^{(r,\ell)}$ as the adjacency matrix of a weighted network. If the intensity measures $\{\Lambda_{uv}\}$ has a block structure in that $\Lambda_{uv} = \Gamma_{\boldsymbol{\sigma}(u)\boldsymbol{\sigma}(v)}$ where $\boldsymbol{\sigma}(u), \boldsymbol{\sigma}(v) \in [K]$ are the cluster membership of u and v (cf. Sec. 3), then $A^{(r,\ell)}$ is a random matrix that follows a Poisson Stochastic Block Model (SBM). To be precise, define a matrix $\boldsymbol{\gamma} \in \mathbb{R}^{K\times K}$ where, for $s, t \in [K]$, we have $\boldsymbol{\gamma}_{st} = \Gamma_{st}(I_\ell^{(r)})$. For $u \neq v$, we then have $A_{uv}^{(r,\ell)} \sim \text{Poisson}(\boldsymbol{\gamma}_{\boldsymbol{\sigma}(u)\boldsymbol{\sigma}(v)})$.

Without loss of generality, we may assume that no clusters are empty and that the rows of $\boldsymbol{\gamma}$ are distinct. Now let $\boldsymbol{P}$ be a $n \times n$ matrix given by

$$\boldsymbol{P}_{uv} = \boldsymbol{\gamma}_{\sigma(u)\sigma(v)}. \tag{3.4}$$

We can then see that $\mathbb{E}_{\boldsymbol{\gamma},\sigma}[A] = \boldsymbol{P} - \text{diag}(\boldsymbol{P})$ and now for each $r \in [R]$ and $\ell \in [2^r]$, we can restate the *discretized local null* as

$$\bar{H}_0^{(r,\ell)} : \boldsymbol{P} = \gamma^{(r,\ell)} \mathbf{1}_n \mathbf{1}_n^\mathsf{T} \quad \text{for some constant } \gamma^{(r,\ell)} > 0, \tag{3.5}$$

where $\mathbf{1}$ a vector of all ones of length n.

Given an observed adjacency matrix $A^{(r,\ell)}$, an intuitive idea for the goodness-of-fit test is to remove the signal using an estimate of the true mean $\gamma^{(r,\ell)}$ and test whether the residual matrix is a noise matrix. Let $\hat{\gamma}^{(r,\ell)} = \{2/(n^2 - n)\} \sum_{u<v} A_{uv}^{(r,\ell)}$ be an estimator of the true Poisson mean, we denote the empirically centered and re-scaled adjacency matrix by $\tilde{A}^{(r,\ell)}$

$$\tilde{A}_{uv}^{(r,\ell)} := \begin{cases} (A_{uv}^{(r,\ell)} - \hat{\gamma}^{(r,\ell)})/\sqrt{(n-1)\hat{\gamma}^{(r,\ell)}}, & u \neq v, \\ 0, & u = v. \end{cases} \tag{3.6}$$

The asymptotic distribution of the extreme eigenvalues of the empirically centered and re-scaled adjacency matrix has been studied in Bickel and Sarkar (2013) and Lei (2016) while the entries are Bernoulli. We extend their result to the case with Poisson distributed edges.

Theorem 2. *For each $r \in [R], \ell \in [2^r]$, Let $A^{(r,\ell)}$ be the adjacency matrix generated from a Poisson Stochastic Block Model and $\tilde{A}^{(r,\ell)}$ be defined as in* (3.6)*, then under the local null hypotheses* (3.5) *and as $n \to \infty$, we have $n^{2/3}\{\lambda_1(\tilde{A}^{(r,\ell)}) - 2\} \xrightarrow{d} TW_1$, where TW_1 is the Tracy–Widom law with $\beta = 1$.*

We relegate the proof of Theorem 2 to Section S1.5.1 of the Supplementary Material. In Section S1.5.2, we also characterize the behavior of $\lambda_1(\tilde{A}^{(r,\ell)})$ under the alternative setting where $K > 1$, showing that $\lambda_1(\tilde{A}^{(r,\ell)})$ diverges as n increases.

Theorem 2 shows that, if we take $\lambda_1(A^{(r,\ell)})$ to be the test statistics for the local null $\bar{H}_0^{(r,\ell)}$, we may obtain the asymptotically valid p-value using the Tracy–Widom distribution. Denote $F_{\mathrm{TW1}}(\cdot)$ as the CDF of the Tracy–Widom law with $\beta = 1$ and define the local p-value

$$\begin{aligned}\bar{p}^{(r,\ell)} &\equiv \bar{p}^{(r,\ell)}\big(A^{(r,\ell)}\big)\\ &:= 2\min\Big(F_{\mathrm{TW1}}[n^{2/3}\{\lambda_1(A^{(r,\ell)})\} - 2], 1 - F_{\mathrm{TW1}}[n^{2/3}\{\lambda_1(A^{(r,\ell)})\} - 2]\Big).\end{aligned}$$

In the same manner, we could then reject the discretized local null $\bar{H}_0^{(r,\ell)}$ at level $\alpha \in (0,1)$ if $p_0^{(r,\ell)} \le \alpha$. The Tracy–Widom law is asymptotic, but we will perform resampling in step 3 so that the final p-values that we produce are still valid for any finite n.

We note that when the resolution level is very high, the discretized networks may be very sparse and thus any tests on them may have low power. This presents no intrinsic difficulty since we aggregate results from multiple resolution levels. In practice, we recommend setting R so that all the discretized networks are connected.

3.1.2. Step 2: Combining p-values of the same resolution level

We follow exactly the same procedure described in Section 2.2.2 to obtain $p_F^{(s,j,r)}$ (or $p_M^{(s,j,r)}$) for every $s \in [R]$, $j \in [2^s]$, and $r \in \{s, s+1, \ldots, R\}$.

3.1.3. Step 3: Combining across different resolution levels

We again define $\tilde{p}_F^{(s,j)} = \min\{p_F^{(s,j,r)} : r \in \{s, s+1, \ldots, R\}\}$. We can make a Bonferroni adjustment just as before, but we propose adjusting with resampling.

Unlike the two-sample test setting, where we only have two realizations, here we have $\binom{n}{2}$ realizations, denoted as $\{N_{uv}(\cdot) : u, v \in [n], u < v\}$. They could also be equivalently characterized by two sequences of random variables $X_1, X_2, \ldots, X_N \in \mathcal{X}$ and random tuples $M_1, M_2, \ldots, M_N$ taking value on $\binom{[n]}{2} := \{(u,v) : u, v \in [n], u < v\}$, where $N = \sum_{u<v} N_{uv}$ is the random length of the sequence. Under the global null H_0, $M_1, M_2, \ldots, M_N$ are multinomial random tuples distributed uniformly over the set $\binom{[n]}{2}$, that is, $\mathbb{P}_0\{M_i = (u,v)\} = 2/\{n(n-1)\}$ for each $i \in [N]$ and $(u,v) \in \binom{[n]}{2}$. Thus we could resample B samples, $b^* = 1, 2, \ldots, B$, from H_0 in the following way:

1. Generate $M_1^{(b^*)}, \ldots, M_N^{(b^*)} \sim \mathrm{Uniform}\binom{[n]}{2}$ independently.
2. Take $N_{uv}^{(b^*)}(\cdot) = \{X_i : M_i^{(b^*)} = (u,v)\}$ for each pair of $(u,v) \in \binom{[n]}{2}$.

On each sample collection $\{N_{uv}^{(b^*)}(\cdot) : u, v \in [n], u < v\}$, we could then compute the simulated unadjusted p-value $\tilde{p}_{F,b^*}^{(s,j)}$ as in Section 2.2.3 and output $\check{p}_F^{(s,j)} := (1/B)\sum_{b^*=1}^{B} \mathbb{1}\big(\tilde{p}_{F,b^*}^{(s,j)} \le \tilde{p}_F^{(s,j)}\big)$ as the individually valid p-value for $H_0^{(s,j)}$.

3.1.4. Step 4: Sequential/multiple testing adjustment

Following Section 2.2.4, we define $p_F^{(s,j)} = \check{p}_F^{(s,j)} \cdot 2^{s \wedge (R-1)}$.

3.2. Degree-corrected symmetric array test

In this setting, we define the local null at $r \in [R]$ and $\ell \in [2^r]$ as

$$H_0^{(r,\ell)} : \Lambda_{uv} = \Gamma\theta_u\theta_v, \text{ on } I_\ell^{(r)} \text{ for some } \Gamma, \boldsymbol{\theta}, \text{ for all } u, v \in [n],$$

where $\boldsymbol{\theta} \in [0,\infty)^n$ is a vector of baseline rate for each of the n individuals. We define the discretized local null as $\bar{H}_0^{(r,\ell)} : \Lambda_{uv}(I_\ell^{(r)}) = \gamma\theta_u\theta_v$, for some $\gamma, \boldsymbol{\theta}$, for all $u, v \in [n]$.

Remark 5. Note that in our definition, we do not require two local nulls $H_0^{(r,\ell)}$ and $H_0^{(r',\ell')}$ to have the same degree correction parameter $\boldsymbol{\theta}$. This means that, although we have the logical implication that $H_0^{(r,\ell)} \Rightarrow H_0^{(r+1,2\ell-1)} \cap H_0^{(r+1,2\ell)}$ (each local null implies its children), we do not have the reverse: if $H_0^{(r,\ell)}$ is false, it may still be that both $H_0^{(r+1,2\ell-1)}$ and $H_0^{(r+1,2\ell)}$ are true. This implies that we may have no power against certain alternatives. This issue is difficult to overcome completely but it can be ameliorated by performing the test with different choices of the hierarchical partition.

3.2.1. Step 1: Compute p-value for each discretized local nulls

For $r \in [R]$ and $\ell \in [2^r]$, define $N_{uv}^{(r,\ell)} := N_{uv}(I_\ell^{(r)})$ as with Section 3.1 and define the adjacency matrix $A_{uv}^{(r,\ell)} = N_{uv}^{(r,\ell)}$ for $u \neq v$ and $A_{uv}^{(r,\ell)} = 0$ if $u = v$.

Suppose the intensity measures $\{\Lambda_{uv}\}$ has a block structure so that $\Lambda_{uv} = \Gamma_{\sigma(u)\sigma(v)}$, where $\sigma(u), \sigma(v) \in [K]$ are the cluster memberships of individual u and v. In the setting where each individual $u \in [n]$ has its own baseline rate of interactions $\theta_u > 0$, we have that $A_{uv}^{(r,\ell)} \sim \text{Poisson}(\boldsymbol{\gamma}_{\sigma(u)\sigma(v)}\theta_u\theta_v)$, where $\boldsymbol{\gamma}_{st} := \Gamma_{st}(I_\ell^{(r)})$. This model is similar to the so-called degree-corrected stochastic block model (DCSBM), where the edges are Bernoulli-distributed binary random variables instead of Poisson random integers as we have in our setting.

To test each discretized local null, we apply the Signed Triangle (SgnT) and the Signed Quadrilateral (SgnQ) statistics introduced and analyzed by Jin, Ke and Luo (2021). To define the SgnT and SgnQ statistic, first define a vector $\hat{\eta}^{(r,\ell)}$ and a scalar $V^{(r,\ell)}$ as

$$\hat{\eta}^{(r,\ell)} = \left(\frac{1}{V^{(r,\ell)}}\right)^{1/2} A^{(r,\ell)}\mathbf{1}_n, \qquad \text{where } V^{(r,\ell)} = \mathbf{1}_n' A^{(r,\ell)}\mathbf{1}_n. \tag{3.7}$$

The Signed-Polygon(SgnT) statistic T_n is then defined as

$$T^{(r,\ell)} \equiv T(I_\ell^{(r)}) = \sum_{\substack{u_1,u_2,u_3\in[n]\\ u_1\neq u_2\neq u_3}} \big(A^{(r,\ell)}_{u_1u_2} - \hat\eta^{(r,\ell)}_{u_1}\hat\eta^{(r,\ell)}_{u_2}\big)\cdot\big(A^{(r,\ell)}_{u_2u_3} - \hat\eta^{(r,\ell)}_{u_2}\hat\eta^{(r,\ell)}_{u_3}\big) \cdot \big(A^{(r,\ell)}_{u_3u_1} - \hat\eta^{(r,\ell)}_{u_3}\hat\eta^{(r,\ell)}_{u_1}\big). \tag{3.8}$$

In a similar manner, we define the Signed-Quadrilateral(SgnQ) statistic as

$$Q^{(r,\ell)} \equiv Q(I_\ell^{(r)}) = \sum_{\substack{u_1,u_2,u_3,u4\in[n]\\ u_1\neq u_2\neq u_3\neq u_4}} \big(A^{(r,\ell)}_{u_1u_2} - \hat\eta^{(r,\ell)}_{u_1}\hat\eta^{(r,\ell)}_{u_2}\big)\big(A^{(r,\ell)}_{u_2u_3} - \hat\eta^{(r,\ell)}_{u_2}\hat\eta^{(r,\ell)}_{u_3}\big) \cdot \big(A^{(r,\ell)}_{u_3u_4} - \hat\eta^{(r,\ell)}_{u_3}\hat\eta^{(r,\ell)}_{u_4}\big)\big(A^{(r,\ell)}_{u_4u_1} - \hat\eta^{(r,\ell)}_{u_4}\hat\eta^{(r,\ell)}_{u_1}\big). \tag{3.9}$$

The intuition behind the SgnT and SgnQ test statistics is that a network with community structure tends to have more triangles and quadrilaterals than a network with a similar number of edges but without community structure. We refer readers to Gao and Lafferty (2017) and Jin, Ke and Luo (2018, 2021) for a more detailed discussion.

Theorems 2.1 and 2.2 in Jin, Ke and Luo (2021) prove asymptotic normality for $T^{(r,\ell)}$ and $Q^{(r,\ell)}$ for degree-corrected stochastic block model, where the edges are Bernoulli (they actually prove it for the more general mixed membership model). However, a careful examination of their proof shows that their result applies, without modification, to the setting where the edges are Poisson. We restate their result below; see Section 5.3.1 for experimental validation.

Proposition 1 (follows from Thms. 2.1 and 2.2 in Jin, Ke and Luo, 2021**).** *Suppose* $\|\theta\|_2\to\infty, \|\theta\|_\infty\to 0, (\|\theta\|_2^2/\|\theta\|_1)\sqrt{\log\|\theta\|_1}\to 0$ *as* $n\to\infty$*, then*

$$\frac{T^{(r,\ell)}}{\sqrt{6}(\|\hat\eta^{(r,\ell)}\|^2-1)^{3/2}} \xrightarrow{d} N(0,1), \quad \textit{and} \quad \frac{Q^{(r,\ell)}-2(\|\hat\eta^{(r,\ell)}\|^2-1)^2}{\sqrt{8}(\|\hat\eta^{(r,\ell)}\|^2-1)^2} \xrightarrow{d} N(0,1). \tag{3.10}$$

Based on Proposition 1, we could take $T^{(r,\ell)}$ or $Q^{(r,\ell)}$ as our test statistic for the discretized local null $\bar H_0^{(r,l)}$, define the corresponding local p-value as $\bar p^{(r,\ell)} \equiv \bar p^{(r,\ell)}\big(T^{(r,\ell)}\big) := 2[1-\Phi(|T^{(r,\ell)}/\sqrt{6}(\|\hat\eta^{(r,\ell)}\|^2-1)^{3/2}|)]$ or as $\bar p^{(r,\ell)} \equiv \bar p^{(r,\ell)}\big(Q^{(r,\ell)}\big) := 2[1-\Phi(|\{Q^{(r,\ell)}-2(\|\hat\eta^{(r,\ell)}\|^2-1)^2\}/\sqrt{8}(\|\hat\eta^{(r,\ell)}\|^2-1)^2|)]$.

3.2.2. Step 2: Computing p-values of the same resolution level

We follow exactly the same procedure described in Section 2.2.2 to obtain $p_F^{(s,j,r)}$ (or $p_M^{(s,j,r)}$) for every $s\in[R]$, $j\in[2^s]$, and $r\in\{s,s+1,\dots,R\}$.

3.2.3. Step 3: Combine across different resolution levels

We again define $\tilde p_F^{(s,j)} = \min\{p_F^{(s,j,r)} : r\in\{s,s+1,\dots,R\}\}$. We can make a Bonferroni adjustment just as before, but we propose to adjust with resampling.

The added challenge here, however, is that we do not observe the degree correction parameter vector $\boldsymbol{\theta}$.

To describe the resampling procedure, we again characterize the $\binom{n}{2}$ realizations $\{N_{uv}(\cdot) : u, v \in [n], u < v\}$ as a random sequence $X_1, \ldots, X_N \in \mathcal{X}$ and $M_1, \ldots, M_N$ taking value on $\binom{[n]}{2}$. Then, under H_0, we have

$$\mathbb{P}_0\{M_i = (u, v)\} = \frac{\theta_u \theta_v}{\sum_{u'<v'} \theta_{u'} \theta_{v'}}. \tag{3.11}$$

We cannot directly use (3.11) to resample $M_1, \ldots, M_N$ since we do not observe $\boldsymbol{\theta}$. To overcome this problem, we condition on the degree of all the individuals, which is a sufficient statistic for $\boldsymbol{\theta}$. To be precise, we write $A^{(0)}_{uv} := N_{uv}(\mathcal{X})$ as the total number of interactions between u and v, and define $D_u \equiv D_u(A^{(0)}) := \sum_{v \neq u} A^{(0)}_{uv}$ as the degree of individual u. Equivalently, we may express D_u as a function of $M_1, \ldots, M_N$ via the equation $D_u(M) = \sum_{i=1}^N \mathbb{1}\{u \in M_i\}$. Write $\boldsymbol{m} := \{(u_1, v_1), \ldots, (u_N, v_N)\}$ as a possible outcome for $M_1, \ldots, M_N$, and write $\boldsymbol{D}(\boldsymbol{m}) := \{D_u(\boldsymbol{m})\}_{u \in [n]}$ as the corresponding vector of all the degrees, then

$$\mathbb{P}_0\big(\{M_1, \ldots, M_N\} = \boldsymbol{m}\big) = \frac{\prod_{i=1}^N \theta_{u_i} \theta_{v_i}}{(\sum_{u'<v'} \theta_{u'} \theta_{v'})^N} = \frac{\prod_{u \in [n]} (\theta_u)^{D_u(\boldsymbol{m})}}{(\sum_{u'<v'} \theta_{u'} \theta_{v'})^N}.$$

For a vector $\boldsymbol{d} \in \mathbb{N}^N$, define $\mathcal{M}_{\boldsymbol{d}} := \big\{\boldsymbol{m} = \big((u_1, v_1), \ldots, (u_N, v_N)\big) : \boldsymbol{D}(\boldsymbol{m}) = \boldsymbol{d}\big\}$ as the set of all possible outcomes of $M_1, \ldots, M_N$ that result in degree vector $\boldsymbol{d}$. Then, we have

$$\begin{aligned}\mathbb{P}_0\big(\{M_1, \ldots, M_N\} = \boldsymbol{m} \,|\, \boldsymbol{D}(\boldsymbol{m}) = \boldsymbol{d}\big) &= \frac{\mathbb{P}_0\big(\{M_1, \ldots, M_N\} = \boldsymbol{m}\big)}{\sum_{\boldsymbol{m}' \in \mathcal{M}_{\boldsymbol{d}}} \mathbb{P}_0\big(\{M_1, \ldots, M_N\} = \boldsymbol{m}\big)} \\ &= \frac{1}{|\mathcal{M}_{\boldsymbol{d}}|}.\end{aligned}$$

Importantly, conditional on the degree, the distribution of $M_1, \ldots, M_N$ does not depend on $\boldsymbol{\theta}$. Therefore, writing $\boldsymbol{D}^{\text{obs}}$ as the observed degree vector, we propose to generate $b^* = 1, \ldots, B$ Monte Carlo samples $M_1^{(b*)}, \ldots, M_N^{(b*)}$ from the conditional distribution

$$\mathbb{P}_0\big(\{M_1, \ldots, M_N\} = \cdot \,|\, \boldsymbol{D}(\cdot) = \boldsymbol{D}^{\text{obs}}\big). \tag{3.12}$$

To generate from (3.12), we use a Metropolis–Hastings algorithm. Given current state $\boldsymbol{m} = \{(u_1, v_1), \ldots, (u_N, v_N)\}$, we generate a proposal $\boldsymbol{m}'$ by choosing a pair $i, j \in [n]$ where $i < j$. Denote $m_i = (u_i, v_i)$ and $m_j = (u_j, v_j)$. We then generate m_i', m_j' by drawing uniformly from each of the following five outcomes:

$$\begin{pmatrix} m_i' \\ m_j' \end{pmatrix} \sim \text{Unif}\left\{ \begin{pmatrix} u_j & v_j \\ u_i & v_i \end{pmatrix}, \begin{pmatrix} u_i & v_j \\ u_j & v_i \end{pmatrix}, \begin{pmatrix} u_j & v_i \\ u_i & v_j \end{pmatrix}, \begin{pmatrix} u_i & u_j \\ v_i & v_j \end{pmatrix}, \begin{pmatrix} v_i & v_j \\ u_i & u_j \end{pmatrix} \right\}. \tag{3.13}$$

We complete the proposal by letting $m'_k = m_k$ for every $k \neq i, j$. If m'_i or m'_j contain multi-edge, that is, both end-points of m'_i or m'_j refer to the same node, then reject m'. Otherwise, it is straightforward to verify that the Metropolis–Hastings ratio is exactly 1 and we accept the proposal m'.

Proposition 2. *The Markov Chain specified via* (3.13) *and the acceptance rule given below* (3.13) *is ergodic and has stationary distribution that is uniform on* $\mathcal{M}_{\boldsymbol{d}}$.

The simpler version of Proposition 2 for contingency tables is well known (Diaconis and Sturmfels, 1998). We give a proof for the longitudinal network setting in Section S1.4 of the Supplementary Material.

We may thus generate $M_1^{(b*)}, \ldots, M_N^{(b*)}$ by taking some steps of the Metropolis–Hastings algorithm and obtain our resampled point process realizations $\{N_{uv}^{(b^*)}(\cdot)\}_{u,v}$. For each b^*, we may then obtain $\tilde{p}_{F,b^*}^{(s,j)}$ and construct the final p-value as $\check{p}_F^{(s,j)} = (1/B)\sum_{b^*=1}^{B} \mathbb{1}\{\tilde{p}_{F,b^*}^{(s,j)} \leq \tilde{p}_F^{(s,j)}\}$.

Step 4: Sequential/multiple testing adjustment.

We make the sequential/multiple testing adjustment on $\check{p}_F^{(s,j)}$ to obtain simultaneously valid $p_F^{(s,j)}$ just as in Section 2.2.4, with one key difference because we do not have the logical implication that $H_0^{(s+1,2j-1)} \cap H_0^{(s+1,2j)} \Rightarrow H_0^{(s,j)}$, that is, if a local null is false, it could still be that the two children are true. As a consequence, we must redefine $\mathcal{L}(s, j)$ as

$$\mathcal{L}(s,j) = \begin{cases} \#\{\text{terminal nodes emanating from } (s,j) \text{ in } \boldsymbol{I}\} & \text{if } (s,j) \text{ is not a terminal node} \\ 1 & \text{if } (s,j) \text{ is a terminal node.} \end{cases}$$

In the case where the hierarchical partitions $\boldsymbol{I}$ is a full binary tree, we have that $\mathcal{L}(s, j) = 2^{R-s}$ and $\mathcal{L}(0, 1) = 2^R$. We then define the simultaneously valid p-value as $p_F^{(s,j)} = \check{p}_F^{(s,j)}\{\mathcal{L}(0,1)/\mathcal{L}(s,j)\} = 2^s \check{p}_F^{(s,j)}$. If we form our rejections via (2.6) using $\{p_F^{(s,j)}\}$, then our FWER is controlled at α. We succinctly summarize all the steps in Algorithm 4.

4. Theoretical Analysis

In this section, we analyze the power of our proposed test under the two-sample test setting where we have two intensities functions λ_a and λ_b; we write $\lambda = \lambda_a + \lambda_b$. Under the global null, we suppose that $\lambda_a = \lambda_b = \lambda/2$ on the support $\mathcal{X}$. We state our results in terms of power against the global null but they are applicable to local nulls as well. Throughout this section, we let ν be the base measure with respect to which the λ_a, λ_b are defined; one can assume ν is the Lebesgue measure for simplicity. We also take the partitions $\boldsymbol{I}$ to be fixed

and state our results in terms of deterministic conditions on $\boldsymbol{I}$.

Our results are of the following form: under an alternative hypothesis where λ_a and λ_b are sufficiently separated according to some notion of distance, our proposed tests at level α will have power at least β. More precisely, writing p as the p-value that we output, we have that $\mathbb{P}_0(p \le \alpha) \le \alpha$ and $\mathbb{P}(p \le \alpha) \ge 1 - \beta$. Our first two results consider the Fisher combination test. For a given hierarchical partition $\boldsymbol{I}$ and a resolution level r, it measures the separation between λ_a and λ_b in terms of a discretized L_2 divergence.

Theorem 3. *Let $\alpha, \beta \in (0,1)$ and let C be a universal constant greater than 1 whose value is specified in the proof. Assume $\int_{I_\ell^{(R)}} \lambda d\nu \ge 2$ for all $\ell \in [2^R]$. Assume there exists $r \in [R]$ such that*

$$\frac{1}{4}\sum_{\ell=1}^{2^r}\left(\frac{\int_{I_\ell^{(r)}} \lambda_a - \lambda_b \, d\nu}{\int_{I_\ell^{(r)}} \lambda \, d\nu}\right)^2 \int_{I_\ell^{(r)}} \lambda d\nu \ge 2^{r/2}\left(\frac{C^{1/2}}{\beta} + 2\log^{1/2}\frac{R}{\alpha}\right) + 2\log\frac{R}{\alpha}. \quad (4.1)$$

Then, we have that $\mathbb{P}(p_F^{(0,1)} \le \alpha) \ge 1 - 2\beta$.

We prove Theorem 3 in Section S2.1 of the Supplementary Material. To better understand the implications of Theorem 3, we next take the support $\mathcal{X}$ to be a compact subset of $\mathbb{R}^q$, and the hierarchical partition $\boldsymbol{I}$ to be any partition such that each split divides a region into two sub-regions, whose volume is halved and whose diameter is reduced by a factor of $O(2^{-1/q})$. Since there is no fixed sample size for Poisson processes, we write $n := \int_{\mathcal{X}} \lambda d\nu$ so that it is the "expected sample size"; note then that λ/n is a probability measure.

Theorem 4. *Let $\mathcal{X} \subset \mathbb{R}^q$ and let $\lambda = \lambda_a + \lambda_b$. Suppose $0 < c_{\min} := \inf_{x\in\mathcal{X}} \lambda(x) \le \sup_{x\in\mathcal{X}} \lambda(x) =: c_{\max} < \infty$. Let $n := \int_{\mathcal{X}} \lambda \, d\nu$ and suppose $(\lambda_a - \lambda_b)/\lambda$ is γ-Holder continuous:*

$$\left|\frac{\lambda_a(x) - \lambda_b(x)}{\lambda(x)} - \frac{\lambda_a(y) - \lambda_b(y)}{\lambda(y)}\right| \le C_H \|x - y\|_2^\gamma \qquad \text{for all } x, y \in \mathcal{X}.$$

Let $R = \lfloor \log_2(n/2) - \log_2(c_{\max}/c_{\min}) \rfloor$ and $\{I_l^{(r)}\}_{r\in[R], l\in[2^r]}$ be a dyadic partition of $\mathcal{X}$ such that for all $r \in [R]$ and $l \in [2^r]$, $\nu(I_l^{(r)}) = \nu(I)/2^r$ and $diam(I_l^{(r)}) \le C_d 2^{-r/q}$. Let $\alpha, \beta \in (0,1)$ and suppose

$$\int_{\mathcal{X}}\left(\frac{\lambda_a - \lambda_b}{\lambda}\right)^2 \frac{\lambda}{n} \, d\nu \ge \begin{cases} C_1 n^{-4\gamma/(q+4\gamma)}\{\beta^{-1} + \log(\log n/\alpha)\} & \text{if } \gamma/q \ge 1/4 \\ C_1 n^{-2\gamma/q}(\beta^{-1} + \log\{\log n/\alpha\}) & \text{if } \gamma/q \le 1/4 \end{cases}, \quad (4.2)$$

where $C_1 > 0$ depends only on $c_{\max}/c_{\min}$, C_H, and C_d. Then, $\mathbb{P}(p_F^{(0,1)} \le \alpha) \ge 1 - 2\beta$.

We prove Theorem 4 in Section S2.2 of the Supplementary Material. It is important to note that Theorem 4 only requires the difference $(\lambda_a - \lambda_b)/\lambda$ to be

Holder continuous; no smooth assumptions are made on the individual intensity functions themselves. Moreover, because we use a hierarchical partition instead of a fixed resolution level, our test is able to adapt to the unknown smoothness γ of the underlying function $(\lambda_a - \lambda_b)/\lambda$, that is, it attains the separation rate (4.2) without knowledge of γ.

Remark 6. As a direct consequence of Theorem 4, we see that when $\mathcal{X} \subseteq \mathbb{R}$ (so that $q = 1$) and when $(\lambda_a - \lambda_b)/\lambda$ is Lipschitz (so that $\gamma = 1$), then our test has nontrivial power when the separation $\int\{(\lambda_a - \lambda_b)/\lambda\}^2(\lambda/n)d\nu$ (which is the squared $L_2(\lambda/n)$ distance between λ_a/λ and λ_b/λ) is of order $n^{-4/5}$. This matches, up to log-factor, the lower bound on the minimax separation rate given by Fromont, Laurent and Reynaud-Bouret (2011) (Sec. 2), showing that our test has minimax optimal power up to log-factors in this setting. We conjecture that when $(\lambda_a - \lambda_b)/\lambda$ is Lipschitz, our test is minimax optimal when $q \leq 4$ and suboptimal when $q > 4$.

Next, we consider the testing procedure that combines all the p-values of the same resolution level by taking the minimum instead of using the Fisher combination function. The next result is the analog of Theorem 3. However, the separation strength between λ_a and λ_b is measured by taking the maximum among the regions rather than taking the sum.

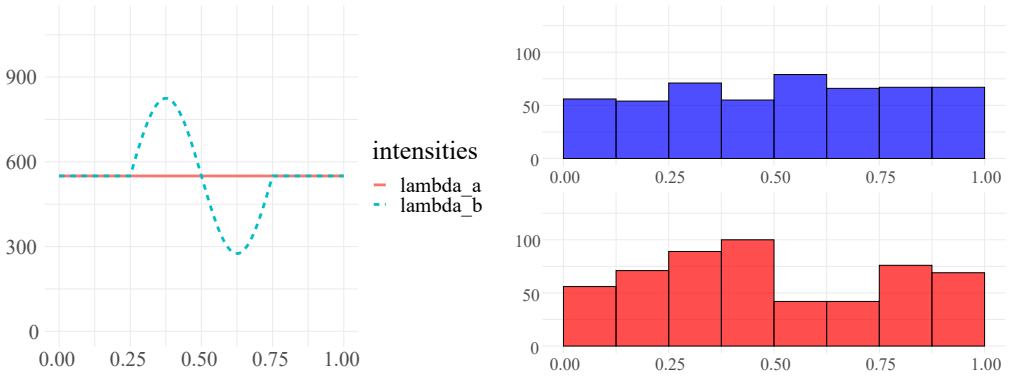
Theorem 5. *Let $\alpha, \beta \in (0,1)$ and let C be a universal constant whose value is specified in the proof. Assume that $\int_{I_l^{(R)}} \lambda d\nu \geq 2$ for all $l \in [2^R]$. Assume there exists $r \in [R]$ such that*

$$\frac{1}{4}\max_{l\in[2^r]}\left(\frac{\int_{I_l^{(r)}} \lambda_a - \lambda_b d\nu}{\int_{I_l^{(r)}} \lambda d\nu}\right)^2 \int_{I_l^{(r)}} \lambda d\nu \geq 2r + \frac{C^{1/2}}{\beta} + 2\log\frac{R}{\alpha}.$$

Then, we have that $\mathbb{P}(p_M \leq \alpha) \geq 1 - 2\beta$.

We give the proof of Theorem 5 in Section S2.3 of the Supplementary Material. To more clearly see the implication of Theorem 5, we consider a setting where $|\lambda_a - \lambda_b|$ is non-zero possibly only on a small region S. We show that, so long as the hierarchical partition is chosen so that each split divides the volume equally, the minimum combination method will have non-trivial power. The following result is an analog of Theorem 4.

Theorem 6. *Let $\mathcal{X} \subset \mathbb{R}$ be an interval and let $S \subset \mathcal{X}$ be a sub-interval. Let $\{I_l^{(r)}\}_{r\in[R]}$ be a dyadic partition of $\mathcal{X}$ such that $R = \lfloor \log_2(n/2) - \log_2(\lambda_{max}/\lambda_{min})\rfloor$ and that for all $r \in [R]$ and $l \in [2^r]$, $\nu(I_l^{(r)}) = \nu(\mathcal{X})2^{-r}$. Suppose that $0 < c_{\min} \leq \lambda \leq c_{\max} < \infty$ and that $\int_I \lambda d\nu = n$. Suppose that $\left|\{\lambda_a(x) - \lambda_b(x)\}/2\lambda(x)\right| \geq \delta_S > 0$ for all $x \in S$ and also that $\nu(S)/\nu(\mathcal{X}) \geq (c_{\max}/c_{\min})(8/n)$.*



(a) Intensity functions λ_a, λ_b (b) Realizations $N_a(\cdot), N_b(\cdot)$

Figure 2. Example of a two-sample test setting.

For any $\alpha, \beta \in (0,1)$*, if*

$$\delta_S^2 \frac{\nu(S)}{\nu(\mathcal{X})} \geq \frac{C_2}{n}\left(\log n + \beta^{-1} + \log\frac{1}{\alpha}\right) \tag{4.3}$$

for $C_2 > 0$ *dependent only on* $c_{\max}/c_{\min}$*, then we have that* $\mathbb{P}(p_M \leq \alpha) \geq 1 - 2\beta$.

We give the proof of Theorem 6 in Section S2.4 of the Supplementary Material. We note that Theorem 6 does not require any smoothness on the difference $\lambda_a - \lambda_b$; it only requires that the difference is at least δ_S in magnitude on the region S.

Remark 7. Although our theoretical results are developed for the two-sample test, the high level proof strategy may be applicable for analyzing the power of longitudinal network tests as well. The main difficulty however is that, when analyzing all the $l \in [2^r]$ bins at high resolution level r, many of the bins may have very small signal strength but nevertheless cannot be ignored because on aggregate, they contribute to the overall separation from the null hypothesis. In Theorem 3, we handle the low signal bins through a careful analysis of the tail probability of the binomial distribution (Lems. 6 and 7 in Sec. S2.5 of the Supplementary Material). Extending this step to the longitudinal network setting is a nontrivial problem, which we leave for future work.

5. Experimental Studies

5.1. An illustrative example

We begin with a single example for the two-sample test to give a concrete illustration of our testing procedure. We let the two intensity functions be $\lambda_a(x) = 550\mathbb{1}_{[0,1]}$ and $\lambda_b(x) = 550 + 275\sin\{4\pi(x-1/4)\}\mathbb{1}_{[1/4,3/4]}$; see Figure 2(a). We take our hierarchical partition $\boldsymbol{I}$ to have $R = 3$ resolution level by dyadic splitting so

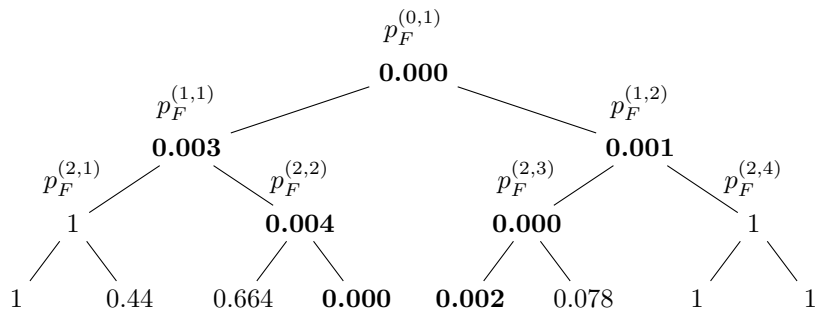


Figure 3. Output simultaneously valid p-values. Bold indicates rejection at $\alpha = 0.05$.

that $I_1^{(1)} = [0, 1/2]$ and $I_2^{(1)} = [1/2, 1]$. We see then that $H_0^{(2,1)}$ and $H_0^{(2,4)}$ are true since $\lambda_a = \lambda_b$ on the regions $[0, 1/4]$ and $[3/4, 1]$. We generate two point process realizations and show them in Figure 2(b).

We then perform our testing procedure with $B = 1000$ resampling draws and display the final simultaneously valid p-values $\{p_F^{(s,j)}\}$ in Figure 3. We see that, at level $\alpha = 0.05$, we correctly reject the global null $H_0^{(0,1)}$. Moreover, we correctly reject the local nulls $H_0^{(1,0)}, H_0^{(1,1)}$ at resolution level $r = 1$ and the local nulls $H_0^{(2,2)}, H_0^{(2,3)}$ at resolution level $r = 2$. At resolution level $r = 3$, we reject $H_0^{(3,4)}, H_0^{(3,5)}$ but missed $H_0^{(3,3)}, H_0^{(3,6)}$. This is not unexpected since the multiple testing adjustment makes it rather difficult to reject hypotheses at the most granular resolution level.

5.2. Simulation for the two-sample test

Next, we provide simulations results of our proposed test for the two-sample problem. We Let $N_a(\cdot)$ and $N_b(\cdot)$ be two Poisson point processes on $\mathcal{X} = [0, 1]$, with intensity functions $\lambda_a(\cdot)$ and $\lambda_b(\cdot)$ respectively. In Section S3.1 of the Supplementary Material, we verify that our tests have the desired type I error. Here, we study the power of our proposed test. For comparison with our proposed test, we also present simulations results of other two-sample testing procedures. The first is the kernel-based test proposed in Gretton et al. (2012) and Fromont et al. (2013). Recall that we can characterize $N_a = \{X_i : M_i = 1\}$ and $N_b = \{X_i = -1\}$, then for any symmetric kernel function $K : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$, the test statistic of the kernel-based test is given by $T_{\text{kernel}} = \sum_{i \neq j \in [N]} K(X_i, X_j) M_i M_j$. There are many choices of the Kernel but in this simulation we use the Gaussian kernel $K(X_i, X_j) := \exp\{-(X_i - X_j)/2\sigma^2\}$ which is shown to have good performance in practice by Fromont et al. (2013). We also consider the conditional Kolmogorov–Smirnov test for performance comparison, which is commonly used for two-sample problems. We again apply Monte–Carlo method to approximate the exact p-value of these two tests. In total, we consider 5 different tests denoted by MF, MM, KN_1, KN_2, KS, where

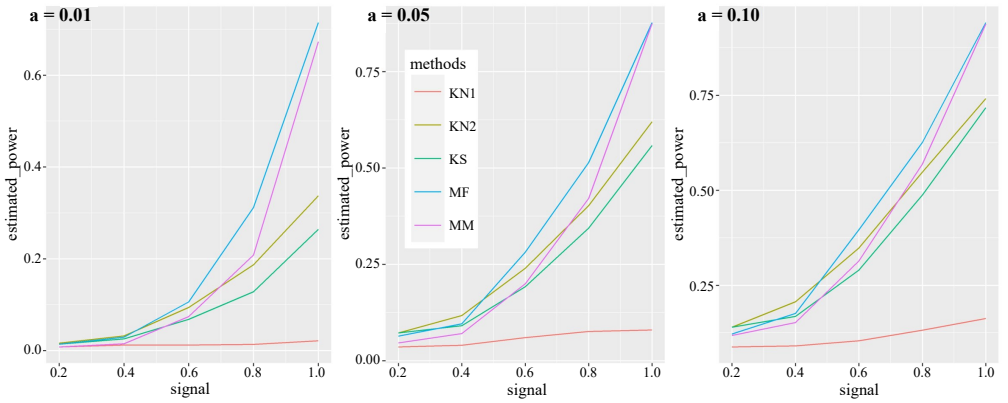


Figure 4. The rejection proportions of the five tests under different signal strength with 1,000 simulations. (Left) level $\alpha = 0.01$. (Center) level $\alpha = 0.05$. (Right) level $\alpha = 0.10$.

MF and MM represent our *multi-scale binning* test with Fisher combination and minimum combination, respectively. The tests KN_1, KN_2 represent the Gaussian kernel test with parameter $\sigma = 0.5$ and $\sigma = 0.1$ of the kernel function. The test KS represents the conditional Kolmogorov–Smirnov test.

We let $\lambda_a(\cdot)$ be a constant function and $\lambda_b(\cdot)$ be a piecewise constant function given below

$$\lambda_a(x) = 50 \cdot \mathbf{1}_{[0,1]}(x),$$
$$\lambda_b(x) = 50\{(1-p) \cdot \mathbf{1}_{[1,1/4]}(x) + (1+p) \cdot \mathbf{1}_{(1/4,1/2]}(x) + \mathbf{1}_{(1/2,1]}(x)\}.$$

For each $p \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$, we perform 1,000 repetitions and use 500 resampling draws in each repetition. In Figure 4, we plot the empirical power, i.e., the proportion of rejections out of the 1,000 simulations for each of the five tests in previous experiments, at three different levels $\alpha = 0.01, 0.05, 0.10$.

At all three levels, we can see that the kernel-based test with large bandwidth has very bad performance, which is not surprising since the large bandwidth smooths the difference between the two Poisson realizations. The other four tests have similar number of rejections when the signal is relatively weak, while as the value of p becomes larger, the MF and MM test, i.e., our *multi-scale binning* test with Fisher and minimum combination, dominates the K-S test and kernel-based tests.

5.3. Simulation study of array test

Next we study tests for degree-corrected longitudinal networks described in Section 3.2. We provide the results on the non-degree-corrected setting in Section S3.2 of the Supplementary Material.

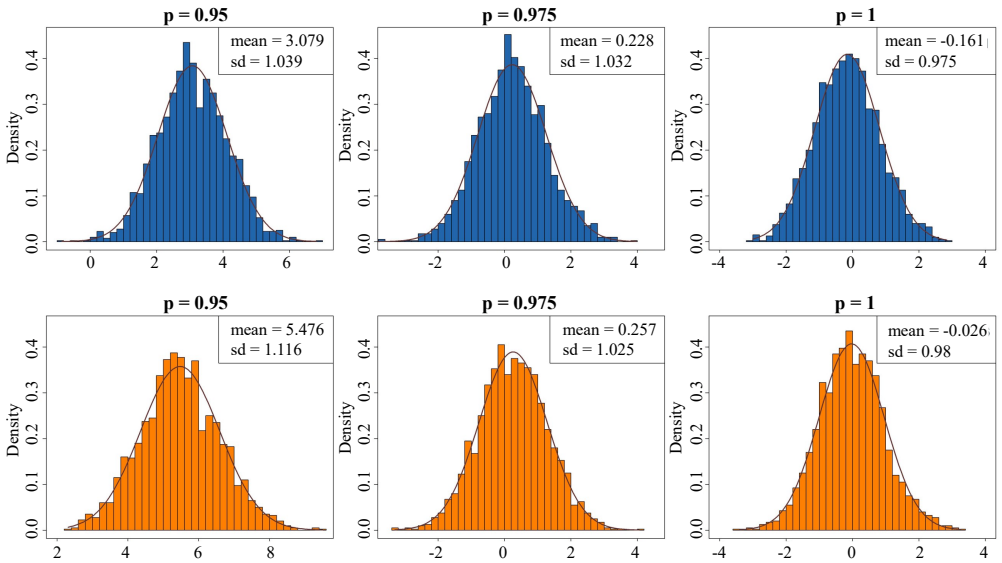


Figure 5. Histogram of Signed-Polygon test statistics with different values of p.

5.3.1. Testing degree-corrected longitudinal networks

We generate a degree corrected longitudinal network with $K = 2$ communities and n nodes; we generate community membership $\boldsymbol{\sigma}$ so that the first $n/2$ nodes are in community 1 and the second half are in community 2. To generate the degree correction parameter $\boldsymbol{\theta}$, we first create unnormalized $\tilde{\boldsymbol{\theta}} \in [0, \infty)^n$ and then set $\theta_u = s \cdot (\tilde{\theta}_u / \|\tilde{\boldsymbol{\theta}}\|)$ for a fixed sparsity parameter s so that $\|\boldsymbol{\theta}\| = s$. We generate the interactions between u, v as

$$N_{uv}(\cdot) \sim \begin{cases} \mathrm{PP}(\lambda_a \cdot \theta_u \theta_v), & \sigma_u = \sigma_v, \\ \mathrm{PP}(\lambda_b \cdot \theta_u \theta_v), & \sigma_u \neq \sigma_v, \end{cases}$$

where λ_a is the within-community intensity function and λ_b is the between-community intensity function.

We first empirically verify the null distribution of $T^{(r,\ell)}$ and $Q^{(r,\ell)}$ – the SgnT (3.8) and SgnQ (3.9) test statistics for each discretized local null. We set $n = 1000$ number of nodes, $s = 100$ sparsity level, and let $\tilde{\theta}_u = u$ for each $u \in [n]$. We set $\lambda_a = \mathbf{1}_{[0,1]}$ and $\lambda_b = p\mathbf{1}_{[0,1]}$ where we let $p \in \{0.95, 0.975, 1\}$ so that the null is either true or close to being true. For each value of p, we generate 2,000 longitudinal networks and reduce each to a single static weighted network A such that $A_{uv} = N_{uv}([0, 1])$. We then compute the centered, scaled SgnT and SgnQ test statistics for A and plot their empirical distribution respectively in Figure 5.

Next we study the power of our proposed test. We set $n = 100$, $s = 12$, and consider two levels of degree heterogeneity: (1) **moderate degree heterogeneity** with *i.i.d.* $\tilde{\theta}_u \sim \mathrm{Unif}(2, 3)$, and (2) **severe degree heterogeneity** with fixed

Table 1. The proportion of rejections of the proposed array test under moderate degree heterogeneity level, where $\tilde{\theta}_u \sim \text{Unif}(2,3)$.

	SgnT					SgnQ				
p	0.6	0.7	0.8	0.9	1.0	0.6	0.7	0.8	0.9	1.0
$\alpha = 0.01$	0.98	0.15	0.02	0	0	0.84	0.77	0.11	0.02	0.01
$\alpha = 0.05$	0.99	0.41	0.07	0.05	0.03	1	0.95	0.15	0.07	0.04
$\alpha = 0.10$	0.99	0.59	0.11	0.10	0.06	1	0.97	0.22	0.09	0.05
$\alpha = 0.25$	0.99	0.84	0.27	0.23	0.22	1	0.99	0.35	0.19	0.23

Table 2. The proportion of rejections of the proposed array test under severe degree heterogeneity level with fixed $\tilde{\theta}_u = \sqrt{u}$.

	SgnT					SgnQ				
p	0.6	0.7	0.8	0.9	1.0	0.6	0.7	0.8	0.9	1.0
$\alpha = 0.01$	0.37	0.03	0.01	0	0	0.95	0.19	0.02	0.02	0
$\alpha = 0.05$	0.64	0.15	0.08	0.05	0.06	0.96	0.35	0.06	0.05	0.04
$\alpha = 0.10$	0.80	0.29	0.14	0.09	0.12	0.98	0.43	0.13	0.11	0.08
$\alpha = 0.25$	0.95	0.52	0.29	0.26	0.22	1	0.61	0.23	0.22	0.25

$\tilde{\theta}_u = \sqrt{u}$. We let

$$\lambda_a(x) = s\frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx}\mathbf{1}_{[0,1]}(x), \quad \lambda_b(x) = p \cdot s\frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx}\mathbf{1}_{[0,1]}(x).$$

We note that $\int_0^1 \lambda_a = s$ and $\int_0^1 \lambda_b = ps \leq s$ so the expected total number of interaction events in the network is at most $s\|\boldsymbol{\theta}\|^2 = s^3 = 1728$, which implies that the network is relatively sparse. To study the power of our proposed test at different signal strength, we let p vary within the set $\{0.6, 0.7, 0.8, 0.9, 1.0\}$. We let the number of resolution levels R be equal to 4 and let $I^{(r,\ell)} := [(\ell-1)/2^r, \ell/2^r) \subset \mathcal{X}$ be the discretized intervals for $l \in [2^r], r \in \{1,2,3,4\}$. For each experiment, we create $B = 400$ resampling draws using the Metropolis–Hastings algorithm described in Section 3.2.3 to derive the adjusted p-values. We perform the proposed multiscale binning test for the global null using both SgnT and SgnQ test statistics and summarize the empirical proportion of rejections out of the 200 experiments under the two degree heterogeneity levels in Table 1 and Table 2 respectively.

We observe that tests based on either the SgnT or the SgnQ test statistic have the desired type I error control. The SgnQ test however has higher power in all the settings that we have studied. We also see that the severity of degree heterogeneity negatively affects the power of our testing procedure.

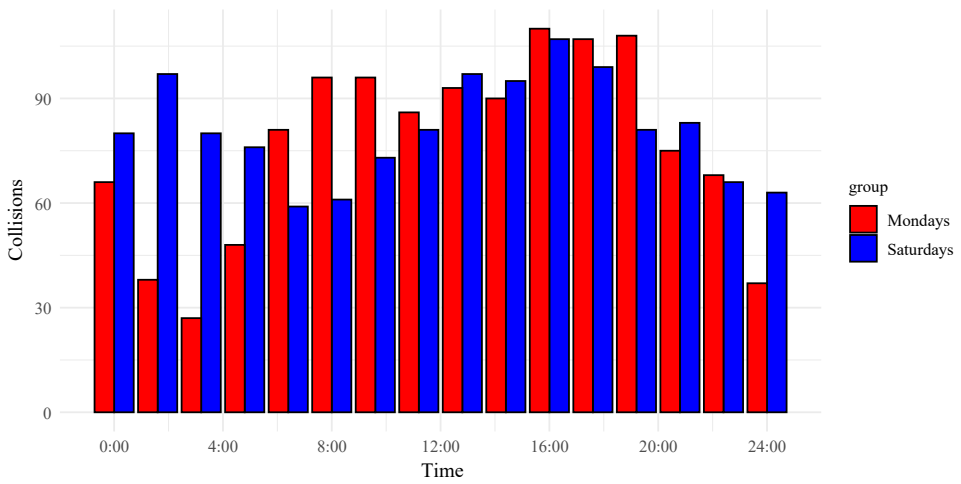


Figure 6. Number of NYC collisions on Mondays and Saturdays at different times of day.

5.4. Traffic collision data

We conduct the two-sample test using data derived from motor vehicle collision events in New York City, recorded by the New York Police Department during January 2022. We test whether there exists a significant difference between the occurrences of crash events on Mondays and Saturdays during this period. To achieve this, we treat the time-stamps of the collisions on Mondays and Saturdays as two realizations of Poisson Processes. Let $\boldsymbol{I}$ be the time interval between 00:00 and 24:00 and with the number of resolution level $R = 4$, we evenly partition $\boldsymbol{I}$ by dyadic splitting such that $I_1^{(1)} = [00:00, 12:00)$, $I_1^{(2)} = [12:00, 24:00)$ and so on. We provide a barplot of the number of collision events within each discretized interval at the most granular resolution level (Fig. 6). We can see that the distribution of collisions are very similar on Mondays and Saturdays from noon to midnight. However, in the morning, there are many more collisions happening on Monday on Saturday. On the other hand, in late night, collisions are more frequent on Saturday than on Sunday.

We perform the proposed testing procedure with 1,000 bootstrap samples and use the Fisher combination method to combine p-values at the same resolution level. We provide the tree of valid p-values in Figure 7. At level $\alpha = 0.05$, we reject the global null $H_0^{(0,1)}$, which make sense as the two occurrences data do have different patterns. Notice that we also successfully rejected the local nulls $H_0^{(4,2)}$, $H_0^{(4,3)}$ and $H_0^{(4,6)}$, where the difference in number of occurrences are among the largest at the resolution level $r = 4$. Although we see from Figure 6 that the number of collisions are different around midnight, we are not able to reject the corresponding local null $H_0^{(4,16)}$ due to the sequential testing adjustment.

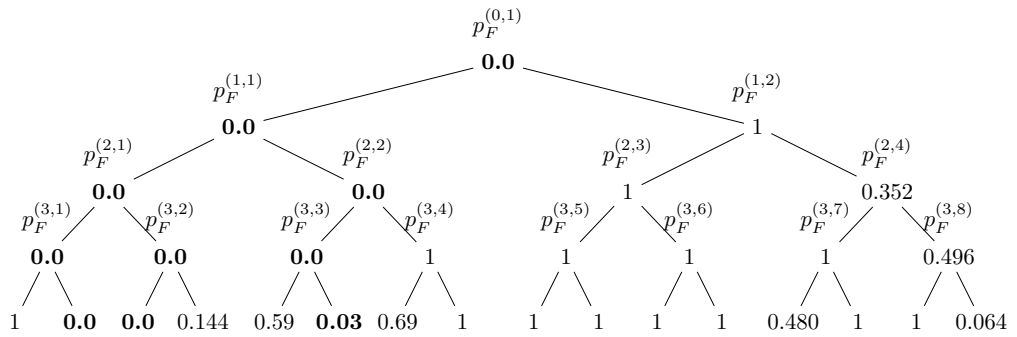


Figure 7. Simultaneously valid p-values for testing NYC collision occurrences on Mondays and Saturdays. Bold indicates rejection at level $\alpha = 0.05$.

5.5. Primates interaction data

In this study, we analyze a network of pairwise interactions within a group of 13 Guinea baboons residing in an enclosure at a Primate Center in France, starting from June 13th, 2019 (Gelardi et al., 2020). The dataset was gathered using wearable proximity sensors attached to leather collars worn on the front side of the 13 baboons. These sensors utilized low-power radio communication, exchanging packets when the distance between two baboons was approximately less than 1.5 meters. The collected data was aggregated with a temporal resolution of 20 seconds, defining interaction between two individuals if their sensors exchanged at least one packet during a 20-second interval. We consider here mainly three days of data collected between July 8th and July 10th, 2019, capturing a total of 6,458 interactions among the 13 baboons. To better analyze how the network evolves over the course of a typical day, we aggregate the interactions across all the days so that each interaction has only the hour/minute/second information. We notice that the degrees are quite heterogeneous and, thus, believe that a degree-corrected model is best suited for the analysis.

We perform the multi-binning test using SgnQ statistics with a bootstrap sample size of $B = 600$ and a resolution number of $R = 4$. The intervals are evenly discretized between 5:30 AM and 10:00 PM. At the most granular level, each bin corresponds to approximately 1 hour. We provide the the tree of simultaneous valid p-values in Figure 9. At level $\alpha = 0.05$, our testing procedure reject the global null and all the local nulls at resolution level $r = 1, 2, 3$. Notice that there are three local nulls, $H_0^{(4,10)}, H_0^{(4,11)}$, and $H_0^{(4,15)}$ at resolution level $r = 4$ that our test failed to reject. We display the discretized networks that correspond to these two of these intervals in Figure 8. We also display another instance where we do reject the null as a comparison. We see that the results make sense based on the visualization since there appears to be 3 distinct clusters ({Violette, Mosphere, Harlem}, {Petoulette, Pipo, Kali}, and others) for the morning interaction graph.

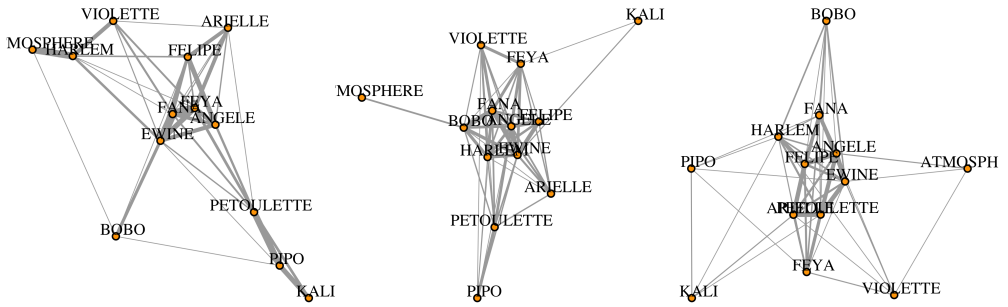


Figure 8. Subgraphs of baboon interaction network at different times of day. (Left) Interactions between 6:48 AM and 7:48 AM in the morning. (Center) Interactions between 3:53 PM and 4:53 PM in the afternoon. (Right) Interactions between 7:55 PM and 8:55 PM at night.

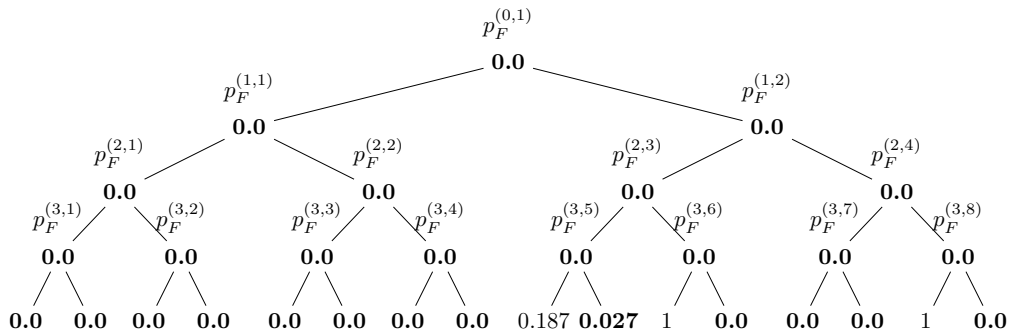


Figure 9. Simultaneously valid p-values for testing baboons interaction network. Bold indicates rejection at level $\alpha = 0.05$.

Supplementary Material

The proofs of all the major results and additional experimental studies are included in the online Supplementary Material of this paper.

Acknowkedgments

This research is supported by the United States National Science Foundation grant DMS-2113671 and the National Institute of Health grant 1R01GM157610-01.

References

Bickel, P. J. and Sarkar, P. (2013). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **78**, 253–273.

Bovett, J. M. and Saw, J. G. (1980). On comparing two Poisson intensity functions. *Communications in Statistics - Theory and Methods* **9**, 943–948.

Corneli, M., Latouche, P. and Rossi, F. (2016). Block modelling in dynamic networks with non-homogeneous Poisson processes and exact ICL. *Social Network Analysis and Mining* **6**, Article no. 55.

Deshpande, J., Mukhopadhyay, M. and Naik-Nimbalkar, U. (1999). Testing of two sample proportional intensity assumption for non-homogeneous Poisson processes. *Journal of Statistical Planning and Inference* **81**, 237–251.

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics* **26**, 363–397.

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC Press.

DuBois, C., Butts, C. and Smyth, P. (2013). Stochastic blockmodeling of relational event dynamics. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics* **PMLR 31**, 238–246.

Fromont, M., Laurent, B. and Reynaud-Bouret, P. (2011). Adaptive tests of homogeneity for a Poisson process. *Annales de l'IHP - Probabilités et Statistiques* **47**, 176–213.

Fromont, M., Laurent, B., Reynaud-Bouret, P. et al. (2013). The two-sample problem for Poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics* **41**, 1431–1461.

Gao, C. and Lafferty, J. (2017). Testing network structure using relations between small subgraph probabilities. *arXiv:1704.06742*.

Gelardi, V., Godard, J., Paleressompoulle, D., Claidiere, N. and Barrat, A. (2020). Measuring social networks in primates: wearable sensors versus direct observations. In *Proceedings of the Royal Society A* **476**, 20190737.

Goeman, J. J. and Solari, A. (2010). The sequential rejection principle of familywise error control. *The Annals of Statistics*, 3782–3810.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research* **13**, 723–773.

Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics Reports* **519**, 97–125.

Huang, S., Weng, H. and Feng, Y. (2023). Spectral clustering via adaptive layer aggregation for multi-layer networks. *Journal of Computational and Graphical Statistics* **32**, 1170–1184.

Jin, J., Ke, Z. T. and Luo, S. (2018). Network global testing by counting graphlets. In *Proceedings of the 35th International Conference on Machine Learning* **PMLR 80**, 2333–2341.

Jin, J., Ke, Z. T. and Luo, S. (2021). Optimal adaptivity of signed-polygon statistics for network testing. *The Annals of Statistics* **49**, 3408–3433.

Kallenberg, O. (2017). *Random Measures, Theory and Applications*. Springer.

Kulldorff, M., Huang, L. and Konty, K. (2009). A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics* **8**, Article no. 58.

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics* **44**, 401–424.

Matias, C., Rebafka, T. and Villers, F. (2018). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika* **105**, 665–680.

Meinshausen, N. (2008). Hierarchical testing of variable importance. *Biometrika* **95**, 265–278.

Picard, F., Reynaud-Bouret, P. and Roquain, E. (2018). Continuous testing for Poisson process intensities: A new perspective on scanning statistics. *Biometrika* **105**, 931–944.

Schrab, A., Kim, I., Albert, M., Laurent, B., Guedj, B. and Gretton, A. (2023). MMD aggregated two-sample test. *Journal of Machine Learning Research* **24**, Article no. 194.

Taleb, Y. and Cohen, E. A. (2021). Multiresolution analysis of point processes and statistical thresholding for Haar wavelet-based intensity estimation. *Annals of the Institute of Statistical Mathematics* **73**, 395–423.

Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *The Annals of Statistics* **38**, 1010–1033.

Xu, K. S. and Hero, A. O. (2013). Dynamic stochastic blockmodels: Statistical models for time-evolving networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, 201–210. Springer.

Zhang, H. and Wang, J. (2025). Efficient estimation for longitudinal network via adaptive merging. *Journal of the American Statistical Association.* DOI: 10.1080/01621459.2025.2455202 .

(Received February 2024; accepted April 2024)