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 consider 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.

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 traditional static networks, the problem of testing for community structure has been extensively studied, including Lei (2016) who proposed tests based on random matrix theory and Gao and Lafferty (2017) and Jin et al. (2021) who proposed tests based on subgraph count statistics. The new longitudinal setting however introduces new dimensions 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. This framework is motivated by longitudinal networks but it also provides a new approach to the two-sample test on point processes, outside of the network context. Indeed, to succinctly describe our framework, it is simplest to first consider the two-sample test problem. 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 $\lambda_a\big|_I = \lambda_b\big|_I$ 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 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.

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., 2021) 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 and review relevant literature in Section 2.3. 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 (Section 3.1), asymmetric networks with homogeneous baseline rate (Section 3.2), and degree-corrected networks with heterogeneous

baseline rates (Section 3.3). 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 define the notion of Poisson point process formally. 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 the readers to Diggle (2013) and Kallenberg et al. (2017) for additional details on the properties of Poisson point processes.

2.1 Two-sample test

For two point process realizations $N_a(\cdot) \sim \mathrm{PP}(\Lambda_a)$ and $N_b(\cdot) \sim \mathrm{PP}(\Lambda_b)$, we test whether they have the same intensity measure, that is, we consider the null hypothesis

$$H_0 \; : \; \Lambda_a = \Lambda_b. \qquad \text{(Two-sample test)} \qquad (21)$$

4

The null defined in (21) requires that $\Lambda_a = \Lambda_b$ everywhere on the domain $\mathcal{X}$. We thus refer to it as the *global null*. We can consider local tests where we ask whether Λ_a and Λ_b are equal when restricted to a sub-domain. To formalize this, we first define the notion of a hierarchical partitioning of the domain $\mathcal{I}$.

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 $I_{(\ell+1)/2}^{(r-1)}$.

It is clear that 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 $I^{(s,j)}$ 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. See Figure 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.

For a given hierarchical partition $\mathbf{I}$, we may then define the notion of a local null for the
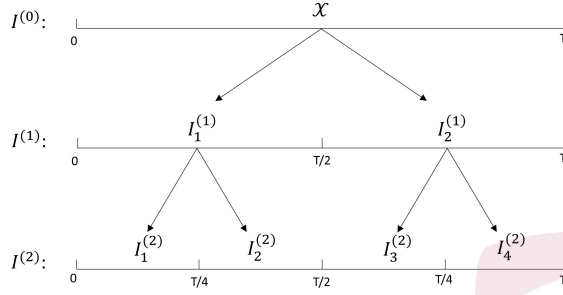
Figure 1

interval $I_j^{(s)}$:

$$H_0^{(s,j)} \ : \ \Lambda_a(\cdot) = \Lambda_b(\cdot) \text{ on } I_j^{(s)}. \qquad \text{(Local null)} \qquad (22)$$

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*:

$$\bar{H}_0^{(s,j)} \ : \ \Lambda_a(I_j^{(s)}) = \Lambda_b(I_j^{(s)}). \qquad \text{(Discretized local null)} \qquad (23)$$

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:

$$H_0^{(s,j)} \Rightarrow H_0^{(r,\ell)} \text{ for any } r \in \{s, s+1, \dots, R\} \text{ and any } \ell \in L(s,j,r). \qquad (24)$$

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 in fact also have the reverse

$$H_0^{s+1,2j-1} \cap H_0^{s+1,2j} \Rightarrow H_0^{(s,j)}, \qquad (25)$$

6

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 (see Section 2.2.4). We note that (25) does not hold in the longitudinal network setting (c.f. Remark 5).

2.2   Testing procedure for two-sample test

Our test procedure will produce simultaneously 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

$$\mathcal{R}_\alpha = \{H_0^{(s,j)} : p^{(s,j)} \leq \alpha \text{ and } p^{(s^*,j^*)} \leq \alpha \text{ for all } I^{(s^*,j^*)} \text{ such that } I^{(s,j)} \subseteq I^{(s^*,j^*)}\}, \quad (26)$$

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

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

$$\check{H}_0^{(s,j)} := \bigcap_{r=s}^{R} \bigcap_{\ell \in L(s,j,r)} \bar{H}_0^{(r,\ell)}. \qquad \text{(Approximate local null)} \qquad (27)$$

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 (21):

$$\check{H}_0 \equiv \tilde{H}_0^{(0,1)} = \bigcap_{r=0}^{R} \bigcap_{\ell=1}^{2^r} \bar{H}_0^{(r,\ell)}. \qquad (28)$$

Our testing procedure proceeds in four steps.

Step 1: construct p-values $\bar{p}^{(r,\ell)}$'s for $\bar{H}_0^{(r,\ell)}$'s with exact tests.

Step 2: Combine $p^{(r,\ell)}$ across ℓ at the same resolution level.

Step 3: Combine p-value across the different resolution levels and use resampling to obtain individually valid p-values $\check{p}_F^{(s,j)}$ for each $H_0^{(s,j)}$.

Step 4: 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 $\text{Bin}(\frac{1}{2}, N^{(r,\ell)})$ distribution.

Thus, we may take $N_a^{(r,\ell)}$ to be the test statistic. Write $S_{\text{Bin}(\frac{1}{2},m)}(\cdot) := \mathbb{P}(|\text{Bin}(\frac{1}{2},m) - \frac{m}{2}| \geq \cdot)$ as the two-sided tail probability function, and write

$$\tilde{p}^{(r,\ell)} \equiv \tilde{p}^{(r,\ell)}\big(N_a(I_\ell^{(r)}), N^{(r,\ell)}\big) := S_{\text{Bin}(\frac{1}{2},N^{(r,\ell)})}\left(\left|N_a^{(r,\ell)} - \frac{N^{(r,\ell)}}{2}\right|\right)$$

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, let $U \sim \text{Unif}[0,1]$ be an independent random variable, define $\tilde{S} := S_{\text{Bin}(\frac{1}{2},N^{(r,\ell)})}\big(\big|N_a^{(r,\ell)} - \frac{N^{(r,\ell)}}{2}\big| + 1\big)$, 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 $\text{Unif}[0,1]$ 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 (we provide a proof in Proposition 3 in the appendix for completeness).

8

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:

$$p_{\mathrm{F}}^{(r)} := S_{\chi_{2^r}^2}\left(-2 \sum_{\ell \in [2^r]} \log \bar{p}^{(r,\ell)}\right), \qquad \text{(Fisher combination)} \tag{29}$$

$$p_{\mathrm{M}}^{(r)} := S_{\beta_{2^r}}\left(\min_{\ell \in [2^r]} \bar{p}^{(r,\ell)}\right), \qquad \text{(Minimum combination)} \tag{210}$$

where $S_{\chi_{2^r}^2}(\cdot) := \mathbb{P}(\chi_{2^r}^2 \geq \cdot)$ 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 $\mathrm{Unif}[0, 1]$ random variables has the χ^2 distribution and that the minimum of independent $\mathrm{Unif}[0, 1]$ random variables has the Beta distribution. For the remainder of this paper, we use Fisher combination by default but it is trivial to replace it with minimum combination method 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

$\big\{\,\{p_F^{(s,j,r)}\}_{r=s}^R\,\big\}_{s\in[R],j\in[2^r]}$. 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 $\int(\frac{\lambda_a-\lambda_b}{\lambda})^2$ 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

$$\check{p}_F^{(s,j)} = (R - s + 1) \cdot \tilde{p}_F. \tag{211}$$

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

$$\check{p}_F^{(s,j)} = \frac{1}{B} \sum_{b^*=1}^{B} \mathbb{1}\big(\tilde{p}_{F,b^*}^{(s,j)} \leq \tilde{p}_F^{(s,j)}\big). \tag{212}$$

2.2.4   Step 4: sequential/multiple testing adjustment

The unadjusted 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

$$\mathcal{L}(s,j) = \begin{cases} \#\{\text{terminal nodes emanating from } (s,j) \text{ in } \boldsymbol{I}\} & \text{if } (s,j) \text{ is not a terminal node} \\ 2 & \text{if } (s,j) \text{ is a terminal node} \end{cases} \tag{213}$$

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:

$$p_F^{(s,j)} = \check{p}_F^{(s,j)} \cdot \frac{\mathcal{L}(0,1)}{\mathcal{L}(s,j)} = \check{p}_F^{(s,j)} \cdot 2^{s \wedge (R-1)} \tag{214}$$

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 (26) with p-values $\{p_F^{(s,j)}\}_{s \in [R], j \in [2^s]}$, as defined in (214), has family-wise error rate (FWER) at most α.*

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 $X_1, \ldots, X_N \in \mathcal{X}$ 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 $N_a(\cdot)$. Under this characterization, we 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 $M_1, \ldots, M_N \in \{-1, +1\}$ 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 in a way such that each sub-region has equal number of "unmarked" points.

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 Bonferroni method. We refer the readers to the excellent paper by Goeman and Solari (2010) for more detail.

2.3   Related work on point processes

If we assume the two Poisson processes are homogeneous, then the two-sample problem is equivalent to testing equality of the means of two Poisson population. This problem has been widely studied by many related works, see Cox (1953), Gail (1974) and Krishnamoorthy and Thomson (2004). Some numerical comparisons of different test statistics and testing procedures are also provided, for example, by Chiu (2010) and Ng et al. (2007). For testing non-homogeneous Poisson process, problem arises since count of occurrences within a certain observation period does not contain detailed spatial or temporal information of the processes. Some work has been done with respect to testing whether the proportion of intensity functions is a constant or increasing over time (Bovett and Saw, 1980; Deshpande et al., 1999). In the case of testing the intensity functions themselves, Fromont et al. (2013) constructed testing procedures using U-statistics

---

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: Simultaneously valid p-values $p_F^{(s,j)}$ for each $H_0^{(s,j)}$.

1: **for** each $r \in [R]$ **do**

2:      **for** each $\ell \in [2^r]$ **do**

3:          Set $\tilde{p}^{(r,\ell)} = S_{\mathrm{Bin}(\frac{1}{2}, N^{(r,\ell)})}\big(|N_a^{(r,\ell)} - N^{(r,\ell)}/2|\big)$

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

15: Compute the adjusted p-values $p_F^{(s,j)} = \check{p}_F^{(s,j)} 2^{s\wedge(R-1)}$.

13

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

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

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

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**

---

based on both single and multiple kernel functions; see Gretton et al. (2012) also. On the other hand, the scanning method is also commonly used to detect clusters or compare samples of observations modeled by Poisson process, see Kulldorff et al. (2009) and Walther et al. (2010). Picard et al. (2018) proposed a continuous testing procedure based on scan statistics for both the test of homogeneity of a single Poisson process and two-sample testing of equality of intensity functions, whose test statistic is the minimum among a discrete p-value process.

3. Tests on longitudinal networks

In this section, we can 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 3.2.

Suppose $N_{uv}(\cdot) \sim PP(\Lambda_{uv})$ for $\binom{n}{2}$ intensity measures $\{\Lambda_{uv}\}$. We aim to test whether the

14

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

$$H_0 : N_{uv}(\cdot) \sim \text{PP}(\Gamma), \text{ for some } \Gamma, \text{ for all } u, v \in [n]. \qquad \text{(Symmetric array test)} \qquad (315)$$

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

$$N_{uv}(\cdot) \sim \text{PP}(\Gamma_{\boldsymbol{\sigma}(u)\boldsymbol{\sigma}(v)}\theta_u\theta_v).$$

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

$$H_0 : N_{uv}(\cdot) \sim \text{PP}(\Gamma\theta_u\theta_v), \text{ for some } \Gamma \text{ and } \boldsymbol{\theta}, \text{ for all } u, v \in [n]. \qquad (316)$$

We refer to (316) 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 3.3.

3.1   Testing procedure for longitudinal networks

We now consider the test of interaction processes among a group of individuals, defined in (315). 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

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

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 Γ function. 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.

Our testing procedure follows the same 4 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

$$A_{uv}^{(r,\ell)} = \begin{cases} N_{uv}^{(r,\ell)}, & u \neq v \\ 0, & u = v \end{cases}. \tag{317}$$

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 (c.f. Section 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

$$\boldsymbol{P}_{uv} = \boldsymbol{\gamma}_{\sigma(u)\sigma(v)} \tag{318}$$

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

$$\bar{H}_0^{(r,\ell)} : \boldsymbol{P} = \gamma^{(r,\ell)} \mathbf{1}_n \mathbf{1}_n^\mathsf{T} \quad \text{for some constant } \gamma^{(r,\ell)} > 0 \tag{319}$$

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 $\hat{\gamma}^{(r,\ell)} = \frac{2}{n^2-n} \sum_{u<v} A_{uv}^{(r,\ell)}$ be an estimator of the true Poisson mean, we denote the empirically centered and re-scaled adjacency matrix by $\tilde{A}^{(r,\ell)}$

$$\tilde{A}_{uv}^{(r,\ell)} := \begin{cases} \dfrac{A_{uv}^{(r,\ell)} - \hat{\gamma}^{(r,\ell)}}{\sqrt{(n-1)\hat{\gamma}^{(r,\ell)}}}, & u \neq v, \\ 0, & u = v. \end{cases} \tag{320}$$

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 distributed. Here 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 (320), then under the local null hypotheses (319) and as $n \to \infty$, we have*

$$n^{2/3}(\lambda_1(\tilde{A}^{(r,\ell)}) - 2) \xrightarrow{d} TW_1. \tag{321}$$

17

where TW_1 is the Tracy–Widom law with $\beta = 1$, and $\lambda_1(A^{(r,\ell)})$ is the largest eigenvalue of $\tilde{A}^{(r,\ell)}$.

We relegate the proof of Theorem 2 to Section S1.4.1 of the Appendix. In Section S1.4.2, 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. These results combined suggests that $\lambda_1(\tilde{A}^{(r,\ell)})$ is a good test statistic.

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 distribution with $\beta = 1$ and define the local p-value

$$\bar{p}^{(r,\ell)} \equiv \bar{p}^{(r,\ell)}\big(A^{(r,\ell)}\big) := 2\min\left(F_{\mathrm{TW1}}\Big(n^{2/3}\big(\lambda_1(A^{(r,\ell)})\big) - 2\Big) \right), 1 - F_{\mathrm{TW1}}\Big(n^{2/3}\big(\lambda_1(A^{(r,\ell)})\big) - 2\Big) \right)$$

In the same manner, we could then reject the discretized local null $\bar{H}_0^{(r,\ell)}$ at level $\alpha \in (0,1)$ if $p_0^{(r,\ell)} \leq \alpha$, and since the Tracy-Widom distribution is continuous, we don't need to transform the p-value as we did for the two-sample test. The Tracy-Widom distribution is asymptotic as $n \to \infty$, so for any finite n, the quantity $\bar{p}^{(r,\ell)}$ may not be an exact p-value. However, because we perform adjustment by resampling in step 3, the final p-values that we produce are still valid for any finite n.

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 to adjust 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, $\mathbb{P}_0\{M_i = (u,v)\} = \frac{2}{n(n-1)}$ 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 $M_1^{(b^*)}, \ldots, M_N^{(b^*)} \sim \text{Uniform}\binom{[n]}{2}$ 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 $\check{p}_F^{(s,j)} := \frac{1}{B}\sum_{b^*=1}^{B} \mathbb{1}\big(\tilde{p}_{F,b^*}^{(s,j)} \leq \tilde{p}_F^{(s,j)}\big)$ as the individually valid p-value for $H_0^{(s,j)}$.

3.1.4   Step 4: sequential/multiple testing adjustment

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

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

(322)

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

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

If we form our rejections via (26) using $\{p_F^{(s,j)}\}$, then our FWER is controlled at level α.

3.2   Asymmetric arrays of interactions

So far we've considered testing symmetric interaction processes among a single group of individuals, we can also extend this to the problem of testing the asymmetric interactions between two possibly different groups of individuals. Let V_1 and V_2 be two sets of individuals and suppose

$|V_1| = m$ and $|V_2| = n$. Now we let $N_{uv}(\cdot)$ represents the temporal interactions events between individual $u \in V_1$ and individual $v \in V_2$, resulting in a collection of asymmetric point process realizations $\{N_{uv}(\cdot) : u \in V_1, \ v \in V_2\}$ where

$$N_{uv}(\cdot) \sim PP(\Lambda_{uv}), \quad \text{for intensity measure } \Lambda_{uv}.$$

Now we have $m \times n$ realizations in total and similarly we can reduce the the dimensionality of this problem by assuming community structures in both groups. Suppose there are K_1 and K_2 communities respectively in groups V_1 and V_2, we again assume the intensity function of the realization between two individuals only depends on their community memberships. More precisely, let $\boldsymbol{\sigma}_1 : [m] \to [K_1]$ and $\boldsymbol{\sigma}_2 : [n] \to [K_2]$ be two clustering function on groups V_1 and V_2, then we assume

$$\Lambda_{uv} = \Gamma_{\boldsymbol{\sigma}_1(u)\boldsymbol{\sigma}_2(v)}, \text{ for any } u \in V_1 \text{ and } v \in V_2$$

where $\{\Gamma_{st}\}_{s,t\in[K_1]\times[K_2]}$ is a collection of $K_1 K_2$ intensity measures. We can again consider the goodness-of-fit test of the community structure with null hypothesis

$$H_0 : K_1 = K_2 = 1 \quad vs. \quad H_0 : K_1 \cdot K_2 > 1 \tag{323}$$

and with a partition of $\boldsymbol{I}$ of the support $\mathcal{X}$, we can define each discretized local null as

$$\bar{H}_0^{(r,\ell)} : \Lambda_{uv}(I_\ell^{(r)}) = \gamma^{(r,\ell)}, \text{ for some common } \gamma^{(r,\ell)} \geq 0 \text{ and for all } u \in V_1, v \in V_2. \tag{324}$$

Similar to the local adjacency matrix $A^{(r,\ell)}$ defined in previous section, we let $B^{(r,\ell)}$ be a $m \times n$ matrix with entries being counts of interactions between any two individuals from groups V_1 and V_2 respectively, within interval $I_\ell^{(r)}$

$$B_{uv}^{(r,\ell)} = N_{uv}(I_\ell^{(r)}), \text{ for any } u \in V_1 \text{ and } v \in V_2$$

To test each discretized local null $\bar{H}_0^{(r,\ell)}$ given observed matrix $B^{(r,\ell)}$, we again remove the mean effect and check whether the residual matrix looks like random noise. We define

$$\hat{\gamma}^{(r,\ell)} = \sum_{u \in V_1} \sum_{v \in V_2} \frac{B_{uv}^{(r,\ell)}}{mn}. \tag{325}$$

20

Moreover, we define

$$\tilde{B}^{(r,l)} = \frac{B^{(r,\ell)} - \hat{\gamma}^{(r,\ell)}}{\sqrt{m \cdot \hat{\gamma}^{(r,\ell)}}} \in \mathbb{R}^{n \times m} \tag{326}$$

as the empirically scaled and centered counterpart of $B^{(r,\ell)}$, with $\hat{\gamma}^{(r,\ell)}$ defined as in (325) and let $\tilde{W}^{(r,\ell)} = (\tilde{B}^{(r,\ell)})^\mathsf{T} \tilde{B}^{(r,\ell)}$. Then we have the following limiting distribution of the largest eigenvalues of $\tilde{W}^{(r,\ell)}$.

Theorem 3. *Let $\lambda_1(\tilde{W}^{(r,\ell)})$ be the largest eigenvalue of matrix $\tilde{W}^{(r,\ell)}$ and suppose $\lim_{n \to \infty} n/m \in (0, \infty)$. Then for each $r \in [R]$ and $\ell \in [2^r]$, under the discretized local null hypothesis $\bar{H}_0^{(r,\ell)}$ given in (324), we have*

$$\frac{m \cdot \lambda_1(\tilde{W}^{(r,\ell)}) - (\sqrt{n} + \sqrt{m})^2}{(\sqrt{n} + \sqrt{m})(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})^{1/3}} \xrightarrow{d} TW_1. \tag{327}$$

as $n, m \to \infty$.

We relegate the proof of Theorem 3 to Section S1.4.3 of the Appendix.

Using Theorem 3, we can let $\lambda_1(\tilde{W}^{(r,\ell)})$ be the test statistics for the local test (324), and derive the p-value for the discretized local null as

$$\bar{p}^{(r,\ell)} \equiv p^{(r,\ell)}(\tilde{W}^{(r,\ell)})$$

$$:= 2\min\left(F_{\mathrm{TW1}}\left(\frac{m \cdot \lambda_1(\tilde{W}^{(r,\ell)}) - (\sqrt{n} + \sqrt{m})^2}{(\sqrt{n} + \sqrt{m})(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})^{1/3}} \right), 1 - F_{\mathrm{TW1}}\left(\frac{m \cdot \lambda_1(\tilde{W}^{(r,\ell)}) - (\sqrt{n} + \sqrt{m})^2}{(\sqrt{n} + \sqrt{m})(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})^{1/3}} \right) \right)$$

Steps 2, 3, and 4 proceed in the same way as the symmetric case, except that the resampling method changes slightly. To generate samples under the null in this scenario, we can just change the distribution of the random marks to be $\mathbb{P}\big(M_i = (u, v)\big) = \frac{1}{mn}$, $\forall i \in [N]$, $u \in V_1$ and $v \in V_2$. Then we generate sequences of random marks $\{M_1^{(b^*)}, \dots, M_N^{(b^*)}\}$ under the aforementioned distribution and let the collection $\{N_{uv}^{(b^*)}(\cdot) : u \in V_1, v \in V_2\}$ be a resample of the observed asymmetric array.

3.3   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. In our definition, we do not require any two local nulls $H_0^{(r,\ell)}$ and $H_0^{(r',\ell')}$ to have the same baseline rate or the same Γ. Therefore, 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) but not the reverse (c.f. Remark 5).

We define the discretized local null as

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

3.3.1   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 et al. (2021). To define the SgnT and SgnQ statistic, first define a vector $\hat{\eta}^{(r,\ell)}$ and a scalar $V^{(r,\ell)}$ as

$$\hat{\eta}^{(r,\ell)} = \left(1/V^{(r,\ell)}\right)^{\frac{1}{2}} A^{(r,\ell)} \mathbf{1}_n, \qquad \text{where } V^{(r,\ell)} = \mathbf{1}_n' A^{(r,\ell)} \mathbf{1}_n. \tag{328}$$

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

$$
\begin{aligned}
T^{(r,\ell)} \equiv T(I_\ell^{(r)}) = \sum_{\substack{u_1,u_2,u_3 \in [n] \\ u_1 \neq u_2 \neq u_3}} & \left(A_{u_1 u_2}^{(r,\ell)} - \hat{\eta}_{u_1}^{(r,\ell)} \hat{\eta}_{u_2}^{(r,\ell)}\right) \cdot \left(A_{u_2 u_3}^{(r,\ell)} - \hat{\eta}_{u_2}^{(r,\ell)} \hat{\eta}_{u_3}^{(r,\ell)}\right) \\
& \cdot \left(A_{u_3 u_1}^{(r,\ell)} - \hat{\eta}_{u_3}^{(r,\ell)} \hat{\eta}_{u_1}^{(r,\ell)}\right)
\end{aligned}
\tag{329}
$$

22

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

$$Q^{(r,\ell)} \equiv Q(I_\ell^{(r)}) = \sum_{\substack{u_1,u_2,u_3,u4\in[n] \\ u_1\neq u_2\neq u_3\neq u_4}} \left(A_{u_1 u_2}^{(r,\ell)} - \hat{\eta}_{u_1}^{(r,\ell)}\hat{\eta}_{u_2}^{(r,\ell)}\right)\left(A_{u_2 u_3}^{(r,\ell)} - \hat{\eta}_{u_2}^{(r,\ell)}\hat{\eta}_{u_3}^{(r,\ell)}\right)$$

$$\cdot \left(A_{u_3 u_4}^{(r,\ell)} - \hat{\eta}_{u_3}^{(r,\ell)}\hat{\eta}_{u_4}^{(r,\ell)}\right)\left(A_{u_4 u_1}^{(r,\ell)} - \hat{\eta}_{u_4}^{(r,\ell)}\hat{\eta}_{u_1}^{(r,\ell)}\right) \tag{330}$$

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

Theorem 2.1 and 2.2 in Jin et al. (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 network edges are Poisson. We restate their result below; see also Section 5.3.2 for experimental validation.

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

$$\frac{T^{(r,\ell)}}{\sqrt{6}(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^{3/2}} \xrightarrow{d} N(0,1) \tag{331}$$

and

$$\frac{Q^{(r,\ell)} - 2(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^2}{\sqrt{8}(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^2} \xrightarrow{d} N(0,1) \tag{332}$$

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

$$\bar{p}^{(r,\ell)} \equiv \bar{p}^{(r,\ell)}\left(T^{(r,\ell)}\right) := 2\left[1 - \Phi\left(\left|\frac{T^{(r,\ell)}}{\sqrt{6}(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^{3/2}}\right|\right)\right]$$

and

$$\bar{p}^{(r,\ell)} \equiv \bar{p}^{(r,\ell)}\left(Q^{(r,\ell)}\right) := 2\left[1 - \Phi\left(\left|\frac{Q^{(r,\ell)} - 2(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^2}{\sqrt{8}(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^2}\right|\right)\right].$$

3.3.2   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 $r \in \{s, s+1, \ldots, R\}$.

3.3.3   Step 3: combine across different resolution levels

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

$$\mathbb{P}_0\{M_i = (u,v)\} = \frac{\theta_u \theta_v}{\sum_{u' < v'} \theta_{u'} \theta_{v'}} \tag{333}$$

We cannot directly use (333) 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 $A_{uv}^{(0)} := N_{uv}(\mathcal{X})$ as the total number of interactions between u and v and define

$$D_u \equiv D_u(A^{(0)}) := \sum_{v \neq u} A_{uv}^{(0)}$$

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

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

Importantly, conditional on the degree, the distribution of $M_1, \ldots, M_N$ does not depend on $\boldsymbol{\theta}$. Therefore, writing $\boldsymbol{D}^{\mathrm{obs}}$ 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

$$\mathbb{P}_0\big(\{M_1, \ldots, M_N\} = \cdot \mid \boldsymbol{D}(\cdot) = \boldsymbol{D}^{\mathrm{obs}}\big). \tag{334}$$

To generate from (334), 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:

$$\binom{m_i'}{m_j'} \sim \text{Unif} \left\{ \left(\begin{smallmatrix} u_j & v_j \\ u_i & v_i \end{smallmatrix} \right), \left(\begin{smallmatrix} u_i & v_j \\ u_j & v_i \end{smallmatrix} \right), \left(\begin{smallmatrix} u_j & v_i \\ u_i & v_j \end{smallmatrix} \right), \left(\begin{smallmatrix} u_i & u_j \\ v_i & v_j \end{smallmatrix} \right), \left(\begin{smallmatrix} v_i & v_j \\ u_i & u_j \end{smallmatrix} \right) \right\}. \tag{335}$$

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 $\boldsymbol{m}'$. Otherwise, it is straightforward to verify that the Metropolis–Hastings ratio is exactly 1 and we accept the proposal $\boldsymbol{m}'$.

Proposition 2. *The Markov Chain specified via* (335) *and the acceptance rule given below* (335) *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 (see e.g. Diaconis and Sturmfels (1998)). We give a proof for the longitudinal network setting in Section S1.3 of the appendix.

Thus, we may generate $M_1^{(b*)}, \ldots, M_N^{(b*)}$ by taking some number of 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 $\check{p}_F^{(s,j)} = \frac{1}{B} \sum_{b^*=1}^{B} \mathbb{1}\{\tilde{p}_{F,b^*}^{(s,j)} \leq \tilde{p}_F^{(s,j)}\}$.

Step 4: sequential/multiple testing adjustment. Following Section 3.1.4, we define $p_F^{(s,j)} = \check{p}_F^{(s,j)} \cdot 2^{-s}$. We succinctly summarize all the steps in Algorithm 4.

3.4   Related work on longitudinal networks

Recently, there has been increased attention on the modeling of dynamic networks, from both the physics community (Holme and Saramäki, 2012) as well as statistics. For example, Xu and Hero (2013) employ a state space model to describe temporal changes at the level of the connectivity pattern, DuBois et al. (2013) introduced a family of relational event models that captures the heterogeneity in underlying interaction dynamics of network data over time. Modelling temporal

interaction between two nodes by Poisson processes is also considered in Corneli et al. (2016) and Matias et al. (2018), where they provide likelihood-based algorithms for membership estimation. Zhang and Wang (2023) studies 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 the readers to the introduction in Zhang and Wang (2023) for a more extensive review of estimation methods for longitudinal networks.

4.   Theoretical analysis

In this section, we theoretically analyze the power of our proposed test under the two-sample test setting where we have two intensities functions λ_a and λ_b and write $\lambda = \lambda_a + \lambda_b$. Under the global null hypothesis, we suppose that $\lambda_a = \lambda_b = \frac{\lambda}{2}$ 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 $\mathbb{P}_0(p \leq \alpha) \leq \alpha$ and $\mathbb{P}(p \leq \alpha) \geq 1 - \beta$.

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 4. *Let $\alpha, \beta \in (0,1)$ and let C be a universal constant greater than 1 whose value is specified in the proof.*

Assume $\int_{I_\ell^{(R)}} \lambda d\nu \geq 2$ for all $\ell \in [2^R]$. Assume there exists $r \in [R]$ such that

$$\frac{1}{4}\sum_{\ell=1}^{2^r}\left(\frac{\int_{I_\ell^{(r)}} \lambda_a - \lambda_b \, d\nu}{\int_{I_\ell^{(r)}} \lambda \, d\nu}\right)^2 \int_{I_\ell^{(r)}} \lambda d\nu \geq 2^{r/2}\left(\frac{C^{1/2}}{\beta} + 2\log^{1/2}\frac{R}{\alpha}\right) + 2\log\frac{R}{\alpha}. \qquad (436)$$

26

Then, we have that

$$\mathbb{P}(p_F^{(0,1)} \leq \alpha) \geq 1 - 2\beta.$$

We prove Theorem 4 in Section S2.1 of the Appendix. To better understand the implications of Theorem 4, we next take the support $\mathcal{X}$ to be a compact subset of $\mathbb{R}^q$ and take the hierarchical partition $\boldsymbol{I}$ to be any partition such that each split divides a region into two sub-regions whose volume is halved (measured with respect to ν) and whose diameter is reduced by a factor of $O(2^{-\frac{1}{q}})$. Since there is no fixed sample size for Poisson processes, we write $n := \int_{\mathcal{X}} \lambda \, d\nu$ so that it is in some sense an "expected sample size"; we note then that $\frac{\lambda}{n}$ is a probability measure on $\mathcal{X}$.

Theorem 5. *Let $\mathcal{X} \subset \mathbb{R}^q$ and let $\lambda = \lambda_a + \lambda_b$. Suppose $0 < c_{\min} := \inf_{x \in \mathcal{X}} \lambda(x) \leq \sup_{x \in \mathcal{X}} \lambda(x) =: c_{\max} < \infty$. Let $n := \int_{\mathcal{X}} \lambda \, d\nu$ and suppose that $\frac{\lambda_a - \lambda_b}{\lambda}$ is γ-Holder continuous on $\mathcal{X}$ in that*

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

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

$$\int_{\mathcal{X}} \left(\frac{\lambda_a - \lambda_b}{\lambda} \right)^2 \frac{\lambda}{n} \, d\nu \geq \begin{cases} C_1 n^{-\frac{4\gamma}{q+4\gamma}} \left(\beta^{-1} + \log \frac{\log n}{\alpha} \right) & \text{if } \gamma/q \geq 1/4 \\ C_1 n^{-\frac{2\gamma}{q}} \left(\beta^{-1} + \log \frac{\log n}{\alpha} \right) & \text{if } \gamma/q \leq 1/4 \end{cases}, \qquad (437)$$

where $C_1 > 0$ depends only on $c_{\max}/c_{\min}$, C_H, and C_d. Then,

$$\mathbb{P}(p_F^{(0,1)} \leq \alpha) \geq 1 - 2\beta.$$

We prove Theorem 5 in Section S2.2 of the Appendix. It is important to note that Theorem 5 only requires the difference $\frac{\lambda_a - \lambda_b}{\lambda}$ to be Holder continuous; no smooth assumptions are made on the individual intensity functions themselves.

Remark 6. As a direct consequence of Theorem 5, we see that when $\mathcal{X} \subseteq \mathbb{R}$ (so that $q = 1$) and when $\frac{\lambda_a - \lambda_b}{\lambda}$ is Lipschitz (so that $\gamma = 1$), then our test has nontrivial power when the separation $\int (\frac{\lambda_a - \lambda_b}{\lambda})^2 \frac{\lambda}{n} d\nu$ (which is the squared $L_2(\lambda/n)$ distance between λ_a/λ and λ_b/λ) is

27

of order $n^{-\frac{4}{5}}$. This matches, up to log-factor, the lower bound on the minimax separation rate given by Fromont et al. (2011) (see Section 2 within), showing that our test has minimax optimal power up to log-factors in this setting. We conjecture that when $\frac{\lambda_a - \lambda_b}{\lambda}$ 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 4. 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 6. *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 6 in Section S2.3 of the Appendix. To more clearly see the implication of Theorem 6, 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 5.

Theorem 7. *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 $R = \lfloor \log_2 \frac{n}{2} - \log_2 \frac{\lambda_{max}}{\lambda_{min}} \rfloor$ 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 $\left| \frac{\lambda_a(x) - \lambda_b(x)}{2\lambda(x)} \right| \geq \delta_S > 0$ for all $x \in S$ and also that $\frac{\nu(S)}{\nu(\mathcal{X})} \geq \frac{c_{\max}}{c_{\min}} \frac{8}{n}$.*

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

$$\delta_S^2 \frac{\nu(S)}{\nu(\mathcal{X})} \geq \frac{C_2}{n} \left(\log n + \beta^{-1} + \log \frac{1}{\alpha} \right) \tag{438}$$

28

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 7 in Section S2.4 of the Appendix. We note that Theorem 7 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.

5. Experimental studies

5.1 An illustrative example

We begin with a single simulation example for the two-sample test to give a concrete illustration of our testing procedure. Here, we let the two intensity functions be $\lambda_a(x) = 550 \mathbb{1}_{[0,1]}$

$$\lambda_a(x) = 550 \qquad \text{for } x \in [0,1],$$

$$\lambda_b(x) = \begin{cases} 550 & \text{if } x \in [0,1/4] \text{ or } x \in [3/4,1] \\ 275 \cdot \sin(4\pi(x - 1/4)) + 550 & \text{if } x \in [1/4,3/4] \end{cases}$$

and they are zero outside of $[0,1]$; see Figure 2a. We take our hierarchical partition $\boldsymbol{I}$ to have $R = 3$ resolution level by dyadic splitting so 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 2b.

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 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.

We note that the p-values $p_F^{(s,j)}$ are very different from the discretized null p-values $\bar{p}^{(s,j)}$ computed in step 1. Indeed, in this example, we have that $\bar{p}^{(0,1)} = 0.37$ since $\int_0^1 \lambda_a = \int_0^1 \lambda_b$ so

29

(a) Intensity functions λ_a, λ_b

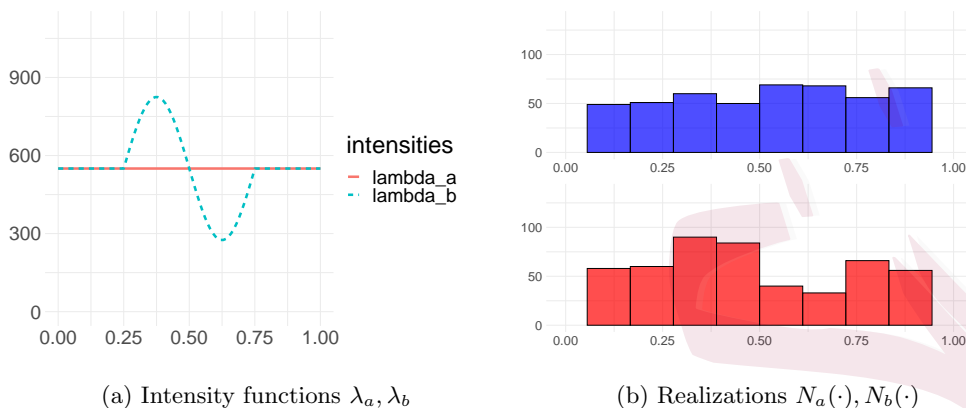(b) Realizations $N_a(\cdot), N_b(\cdot)$

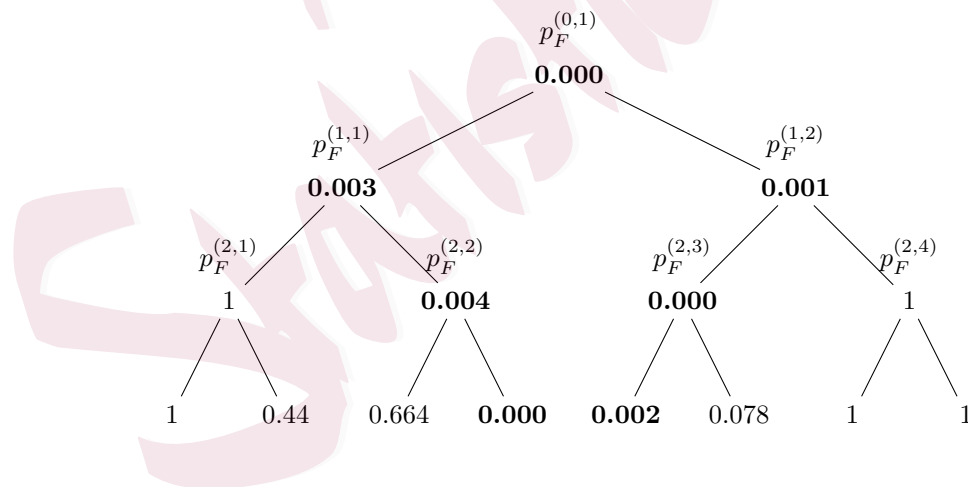Figure 2: Example of a two-sample test setting

Figure 3: Simultaneously valid p-values produced by our testing procedure. Bold indicates rejection at level $\alpha = 0.05$.

that the discretized local null $\bar{H}_0^{(0,1)}$ is actually true.

5.2   Simulation for the two-sample test

Next we provide simulations results of our proposed test for the two-sample test. We Let $N_a(\cdot)$ and $N_b(\cdot)$ be two Poisson point process on support $\mathcal{X} = [0,1]$, with densities $\lambda_a(\cdot)$ and $\lambda_b(\cdot)$ on $\mathcal{X}$ with respect to the Lebesgue measure ν on $\mathcal{X}$. In Section S3.1 of the appendix, 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

$$K(X_i, X_j) := \exp\left\{-\frac{(X_i - X_j)}{2\sigma^2}\right\} \tag{539}$$

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 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 $\sigma = 0.2$ of the kernel function (539). 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

$$\lambda_a(x) = 50 \cdot \mathbf{1}_{[0,1]}(x)$$

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

31

For each $p \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$, we realize 1000 pair of simulations of two Poisson process with intensities λ_a and λ_b respectively in $[0, 1]$ and again generate 500 bootstrap resamples for each simulated pair of Poisson realization. We plot the empirical power, i.e., the proportion of rejections out of the 1000 simulations for each of the five tests in previous experiments, at three different levels $\alpha = 0.01, 0.05, 0.10$. Note that in this particular experiment, we let the bandwidth of the two kernel functions in test KS1 and KS2 to be $\sigma_1 = 0.5$ and $\sigma_2 = 0.1$ respectively.
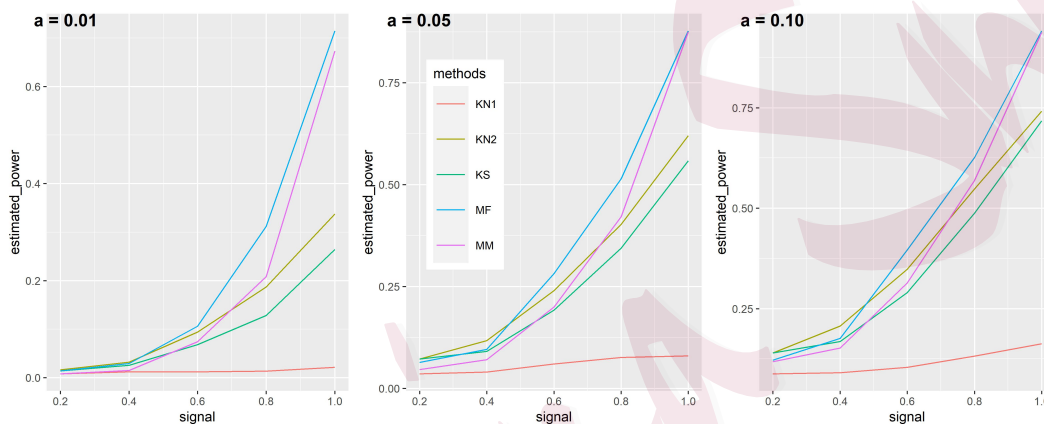

Figure 4: The proportion of rejections of the five tests out of 1000 simulated samples under different signal strength. Left: level $\alpha = 0.01$. Center: level $\alpha = 0.05$. Right: level $\alpha = 0.10$.

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. We conjecture that our *multi-binning* tests have better performance than the KS test in this case is because that the expected difference of empirical CDFs of the two Poisson realizations reaches its maximum at $x = 1/4$ and will decrease as x gets larger, while the *multi-binning* tests will continue collecting information, especially for the test using Fisher's combination.

5.3 Simulation study of array test

5.3.1 testing rate homogeneous array

As an empirical verification of Theorem 2, in Figure 11, we plot the finite sample distributions of the largest eigenvalue of the adjacency matrix $A^{(r,\ell)}$ under the null hypothesis. We give the details of the experimental set-up in Section S3.2 of the appendix; in that section, we also discuss the bootstrap correction method proposed by Bickel and Sarkar (2016) to improve the Tracy-Widom approximation.
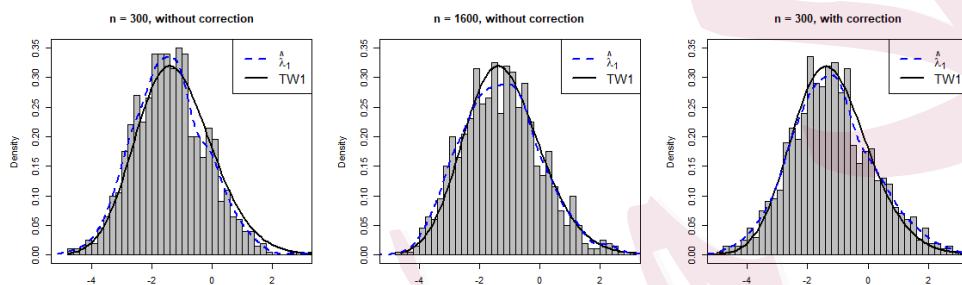

Figure 5: The empirical distribution of 1000 simulated samples of centered and scaled largest eigenvalues of $\tilde{A}$, compared with the Tracy-Widom distribution.

Next, we consider two alternative Poisson SBM models with $K = 2$ and $K = 3$ equally sized communities respectively. We let the probability distribution of the interactions between two nodes u, v only depends on whether they are in the same community and we denote the intensity function of realizations between individuals within the same community as $\lambda_{\text{same}}(\cdot)$ and from different communities as $\lambda_{\text{diff}}(\cdot)$. We then define

$$\lambda_{\text{same}}(x) = s \cdot \mathbf{1}_{[0,1]}(x), \quad \lambda_{\text{diff}}(x) = s \cdot \frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx}\mathbf{1}_{[0,1]}(x)$$

for both the two alternative Poisson SBM models, where s is a parameter that controls the sparsity levels of the networks in this experiments. We again have $n = 200$ and either $K = 2$ or $K = 3$ equally sized communities. We then generate 200 sample collections of realizations on the same support for each of the two models and for each value of $s \in \{0.1, 0.175, 0.25, 0.5, 1\}$ and conduct our proposed test on these samples where the bootstrap sample size and partitioning of

the support are exactly the same as in the preceding experiment. The proportions of rejections for the two SBM models under different sparsity levels are recorded in Table 1.

	$K = 2$					$K = 3$				
s	1.0	0.5	0.25	0.175	0.1	1.0	0.5	0.25	0.175	0.1
$\alpha = 0.01$	1	1	0.98	0.41	0.055	1	1	1	0.785	0.05
$\alpha = 0.05$	1	1	0.99	0.575	0.115	1	1	1	0.905	0.145
$\alpha = 0.10$	1	1	0.995	0.63	0.165	1	1	1	0.97	0.24
$\alpha = 0.25$	1	1	1	0.74	0.36	1	1	1	1	0.38

Table 1: The proportion of rejections of the proposed array test out of 200 simulated samples of networks of 200 nodes at different sparsity and confidence levels $\alpha \in \{0.01, 0.05, 0.10, 0.25\}$.

5.3.2 testing rate heterogeneous array

Now we investigate the empirical performance of the *multi-binning* test on testing array of processes with heterogeneous baseline rate. We first empirically verify the null distribution of $T^{(r,\ell)}$ and $Q^{(r,\ell)}$, the SgnT (329) and SgnQ (330) test statistics for each discretized local nulls. For an arbitrary $I \in \mathcal{X}$, we consider a Poisson DCSBM with n individuals and $K = 2$ communities, where the edges $N_{uv}(I)$ are just Poisson random integers. We let $\theta \in [0, \infty)^n$ be the vector representing the baseline rate of interactions for each individual. Then for some $p \in [0, 1]$,

$$N_{uv}(I) \sim \begin{cases} \text{Poisson}(\theta_u \theta_v), & \boldsymbol{\sigma}_u = \boldsymbol{\sigma}_v, \\ \text{Poisson}(\theta_u \theta_v \cdot p), & \boldsymbol{\sigma}_u \neq \boldsymbol{\sigma}_v. \end{cases}$$

In order to have a control of the sparsity level of the network, we introduce a intermediate vector $\tilde{\theta} \in [0, \infty)^n$ and a scalar $s > 0$. Then we let $\theta_u = s \cdot \frac{\tilde{\theta}_u}{\|\tilde{\theta}\|}$ such that $\|\theta\| = s$.

We let the number of nodes $n = 1000$, $s = 100$ and let $\tilde{\theta}_u = u$ for each $u \in [n]$. We consider two equally sized communities. Then for each $p \in \{0.95, 0.975, 1\}$, we generate 2000 samples from the above network. We then derive the centered, scaled SgnT and SgnQ test statistics for

34

each sample and plot their empirical distribution respectively in Figure 6.

Figure 6: Histogram of Signed Polygon test statistics with different values of p.
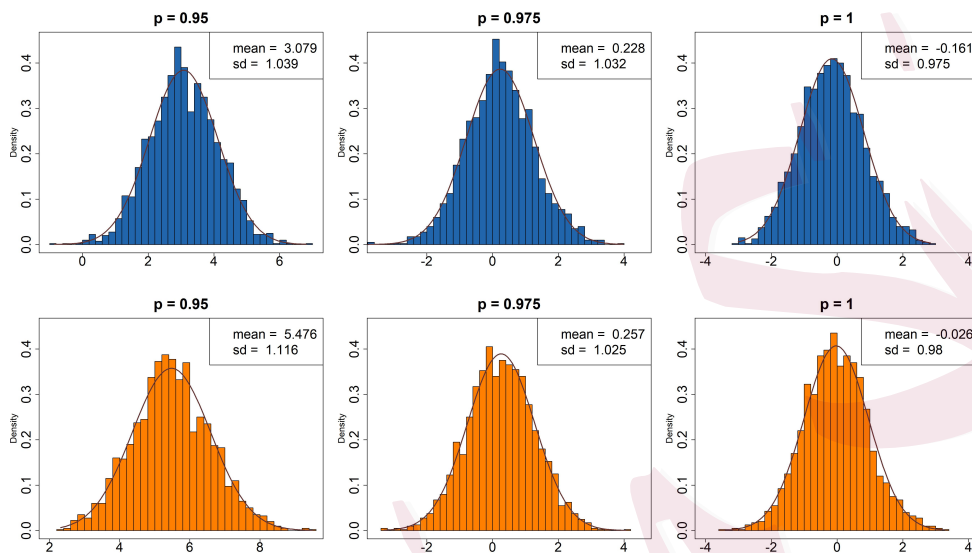
Next we study the empirical performance of the proposed test on array of Poisson process realizations with heterogeneous baseline rate, where the p-value for each local null $\bar{H}_0^{(r,\ell)}$ is derived from the null distribution of the two Signed Polygon statistics. We consider a model with $n = 100$ individuals, number of communities $K = 2$ and again let the community membership to be such that $\boldsymbol{\sigma}_u = 1$ if and only if $u \leq \frac{n}{2}$. With a fixed sparsity level where we let $s = 12$ and $\theta_u = s \cdot \frac{\tilde{\theta}_u}{\|\tilde{\theta}\|}$, we consider two levels of degree heterogeneity: (1) **moderate degree heterogeneity** with iid $\tilde{\theta}_u \sim \text{Unif}(2,3)$ and (2) **severe degree heterogeneity** with fixed $\tilde{\theta}_u = \sqrt{u}$.

To study the performance of the proposed test at different signal strength, we let p, the variable that controls the off-diagonal connection probability, to vary within the set $\{0.6, 0.7, 0.8, 0.9, 1.0\}$. Then for each distribution of the degree vector θ and each value of p, we generate 200 sample collections of Poisson realizations $\{N_{uv}(\cdot) : u, v \in [n]\}$ on support $\mathcal{X} = [0,1]$ where the within community and between community densities are

$$\lambda_{\text{in}}(x) = s \cdot \frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx} \mathbf{1}_{[0,1]}(x), \quad \lambda_{\text{out}}(x) = p \cdot s \cdot \frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx} \mathbf{1}_{[0,1]}(x)$$

respectively. In this experiment we let the number of resolution levels $R = 4$ and $I^{(r,\ell)} :=$ $[\frac{\ell-1}{2^r}, \frac{\ell}{2^r}) \subset \mathcal{X}$ be the discretized intervals for all $l \in [2^r], r \in \{1, 2, 3, 4\}$. Also for each generated observation we again produce $B = 400$ bootstrap samples using the Metropolis-Hastings algorithm described in Section 3.3.3 to derive the adjusted p-values. Then we perform the proposed *multi-binning* test using both SgnT and SgnQ test statistics on each sample at the two heterogeneity level and for different values of p. We summarize the proportion of rejections out of the 200 samples under the two degree heterogeneity levels in Table 2 and Table 3 respectively.

	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	0.6	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 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
$\alpha = 0.01$	0.37	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	0.10	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

Table 3: The proportion of rejections of the proposed array test under severe degree heterogeneity level with fixed $\tilde{\theta}_u = \sqrt{u}$.

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 police department(NYPD) during January 2022. We believe that these collision events can be effectively modeled by a Poisson Point process. We focus on testing whether there exists a significant difference between the occurrences of crash events at Mondays and Saturdays during this period. To achieve this, we treat the occurrence times of collisions on Mondays and Saturdays as two realizations of Poisson Processes. Let I be the time interval between 00:00 and 24:00 and with the number of resolution level $R = 4$, we evenly partition 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, see Figure 7. We can see that the distribution of collisions are very similar on Mondays and Saturdays from noon to midnight. However there are much more collisions happening in Monday mornings than in Saturday morning and on the other hand, collisions are more frequent at Saturday pre-dawn period.
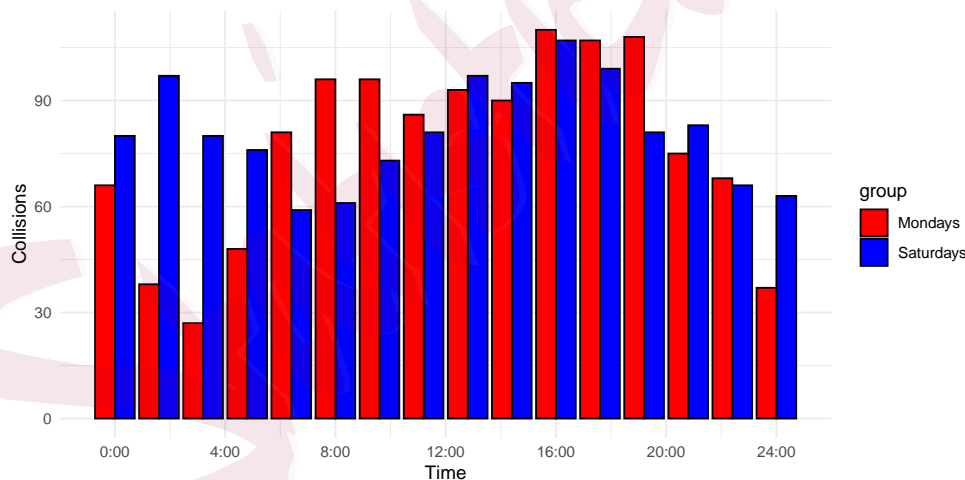

Figure 7: Number of NYC collisions on Mondays and Saturdays at different time of day.

Then we perform the proposed testing procedure with 1000 bootstrap samples and use the fisher combination method to combine p-values at the same resolution level. We provide the tree
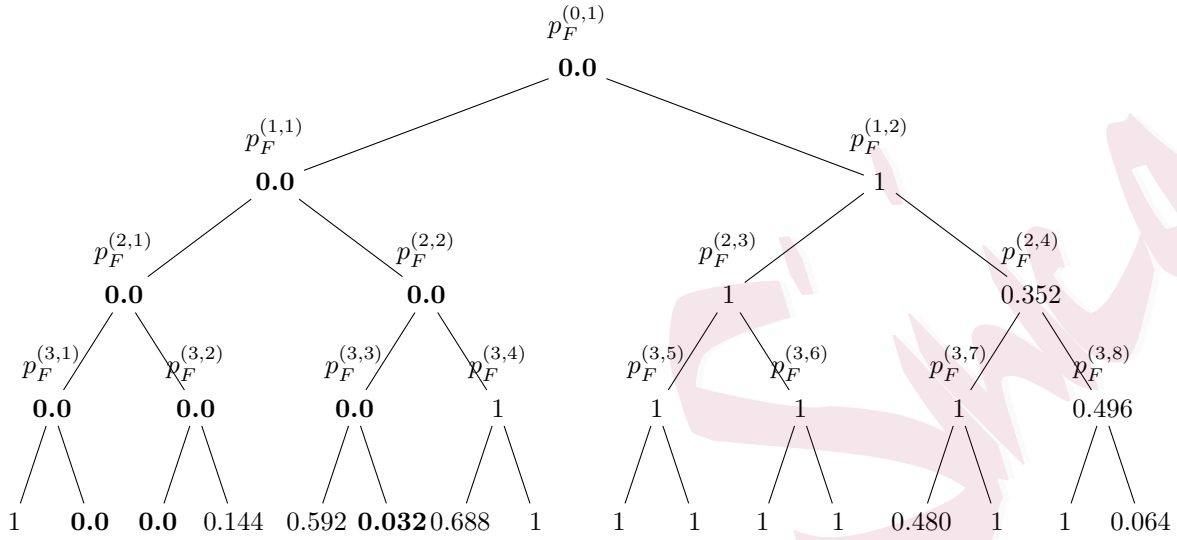
Figure 8: Simultaneous valid p-values for testing NYC collision occurrences at Mondays and Saturdays. Bold indicates rejection at level $\alpha = 0.05$

of valid p-values in Figure 8. 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 $r = 4$. However, we can see from Figure 7 that the number of collisions are remarkably different right before midnight while the corresponding local null $H_0^{(4,16)}$ is rejected due to the sequential testing adjustment.

5.5   Primates interaction data

In this study, we analyze a network of dyadic 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 individuals. 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 6458 interactions among the 13 baboons. To better analyze how the network evolves during the day, we aggregate all the interactions considering only the hour and second each interaction happens while ignoring the actual date it was recorded. We notice the degrees of each individuals are quite different, we believe a degree corrected model is best suited for the analysis.
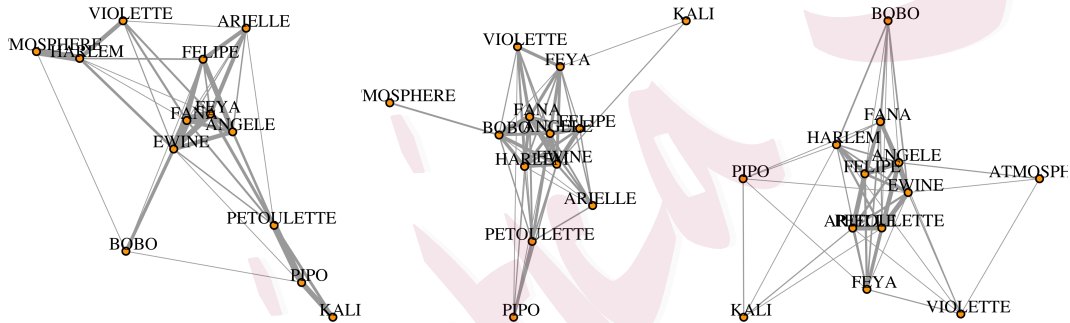

Figure 9: Subgraphs of baboon interaction network at different times of day. **Left:** Interactions in the morning between 6:48AM and 7:48AM. **Center:** In the afternoon between 15:53AM and 16:53AM. **Right:** At night between 7:55PM and 8:55PM.

We perform the multi-bining test using SgnQ statistics with boot strap sample size $B = 600$ and number of resolution number $R = 4$ and the intervals are evenly discretized between 5:30AM and 10:00PM. 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 $\alpha = 0.01$, our testing procedure rejected the global null and all the local nulls at resolution level $r = 1, 2, 3$. Notice that there are two local nulls $H_0^{(4,11)}$ and $H_0^{(4,15)}$ at resolution level $r = 4$ that our test failed to reject. We display the discretized network in these two intervals in Figure 9. We also display
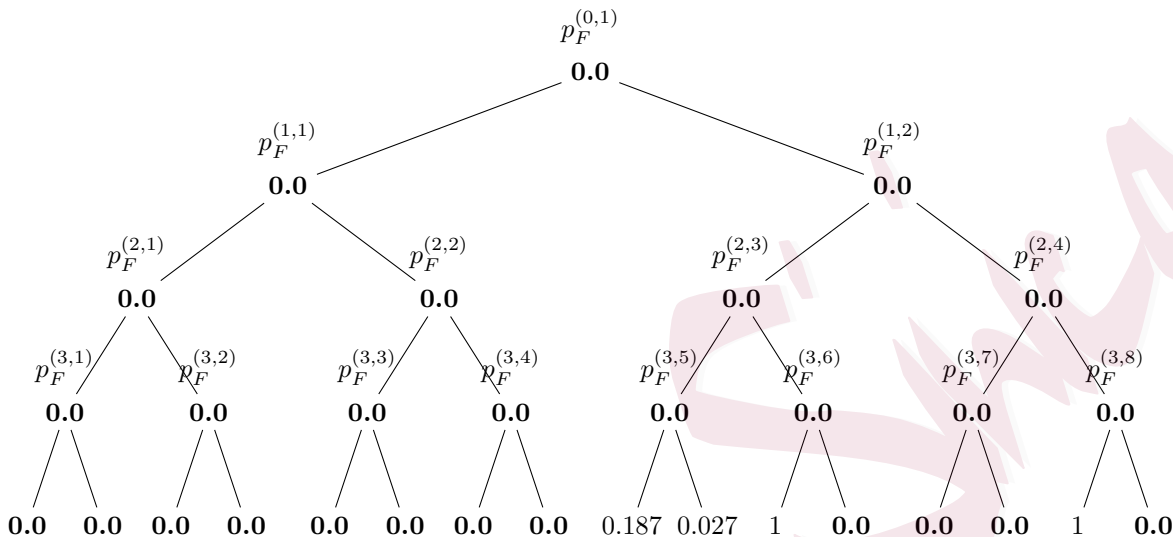
39

Figure 10: Simultaneous valid p-values for testing baboons interaction network. Bold indicates rejection at level $\alpha = 0.01$

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 for the morning interaction graph.

References

Alex, B., Erdős, L., Knowles, A., Yau, H.-T., Yin, J., et al. (2014). Isotropic local laws for sample covariance and generalized wigner matrices. *Electronic Journal of Probability*, 19.

Bickel, P. J. and Sarkar, P. (2013). Hypothesis testing for automated community detection in networks. *arXiv preprint arXiv:1311.2694*.

Bickel, P. J. and Sarkar, P. (2016). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):253–273.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

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

Chiu, S. N. (2010). Parametric bootstrap and approximate tests for two poisson variates. *Journal of Statistical Computation and Simulation*, 80(3):263–271.

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(1):55.

Cox, D. (1953). Some simple approximate tests for poisson variates. *Biometrika*, 40(3/4):354–360.

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(2):237–251.

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of statistics*, 26(1):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 *Artificial Intelligence and Statistics*, pages 238–246.

Erdős, L., Knowles, A., Yau, H.-T., and Yin, J. (2013). The local semicircle law for a general class of random matrices. *Electronic Journal of Probability*, 18:1–58.

Erdős, L., Yau, H.-T., and Yin, J. (2012). Rigidity of eigenvalues of generalized wigner matrices. *Advances in Mathematics*, 229(3):1435–1515.

Fromont, M., Laurent, B., and Reynaud-Bouret, P. (2011). Adaptive tests of homogeneity for a poisson process. In *Annales de l'IHP Probabilités et statistiques*, volume 47, pages 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(3):1431–1461.

Gail, M. (1974). Power computations for designing comparative poisson trials. *Biometrics*, pages 231–237.

Gao, C. and Lafferty, J. (2017). Testing network structure using relations between small subgraph probabilities. *arXiv preprint 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. *Proceedings of the Royal Society A*, 476(2236):20190737.

Goeman, J. J. and Solari, A. (2010). The sequential rejection principle of familywise error control. *The Annals of Statistics*, pages 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(1):723–773.

Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.

Jin, J., Ke, Z. T., and Luo, S. (2018). Network global testing by counting graphlets. *arXiv preprint arXiv:1807.08440*.

Jin, J., Ke, Z. T., and Luo, S. (2021). Optimal adaptivity of signed-polygon statistics for network testing. *The Annals of Statistics*, 49(6):3408–3433.

Kallenberg, O. et al. (2017). *Random measures, theory and applications*, volume 1. Springer.

Krishnamoorthy, K. and Thomson, J. (2004). A more powerful test for comparing two poisson means. *Journal of Statistical Planning and Inference*, 119(1):23–35.

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(1):1–9.

Latała, R. (2005). Some estimates of norms of random matrices. *Proceedings of the American Mathematical Society*, 133(5):1273–1282.

Lee, J. O., Yin, J., et al. (2014). A necessary and sufficient condition for edge universality of wigner matrices. *Duke Mathematical Journal*, 163(1):117–173.

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, 44(1):401–424.

Matias, C., Rebafka, T., and Villers, F. (2018). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika*, 105(3):665–680.

Meinshausen, N. (2008). Hierarchical testing of variable importance. *Biometrika*, 95(2):265–278.

Ng, H., Gu, K., and Tang, M.-L. (2007). A comparative study of tests for the difference of two poisson means. *Computational statistics & data analysis*, 51(6):3085–3099.

Picard, F., Reynaud-Bouret, P., and Roquain, E. (2018). Continuous testing for poisson process intensities: a new perspective on scanning statistics. *Biometrika*, 105(4):931–944.

Pillai, N. S., Yin, J., et al. (2014). Universality of covariance matrices. *The Annals of Applied Probability*, 24(3):935–1001.

Schrab, A., Kim, I., Albert, M., Laurent, B., Guedj, B., and Gretton, A. (2021). Mmd aggregated two-sample test. *arXiv preprint arXiv:2110.15073*.

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. et al. (2010). Optimal and fast detection of spatial clusters with scan statistics. *The Annals of Statistics*, 38(2):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*, pages 201–210. Springer.

REFERENCES

Zhang, H. and Wang, J. (2023). Efficient estimation for longitudinal network via adaptive merging. *Preprint*.

Youmeng Jiang

Statistics Department

Rutgers University

yj228stat.rutgers.edu

Min Xu

Statistics Department

Rutgers University

mx76stat.rutgers.edu

Supplementary material to "Multiscale Tests for Point

Processes and Longitudinal Networks"

S1. Supplementary material for Sections 2 and 3

S1.1 Concise description of algorithms for longitudinal networks

S1.2 Randomizing p-value

Let X be a discrete random variable taking value on $\{x_1, \ldots, x_m\} \subset \mathbb{R}$ where we have the ordering $x_1 \leq x_2 \leq \ldots x_m$. Define $S(x) = \mathbb{P}(X \geq x)$ and

$$q_1 := S(x_1) = 1, \ q_2 := S(x_2), \ldots, q_m := S(x_m), \ q_{m+1} := 0,$$

so that $S(X)$ takes value on $\{q_1, q_2, \ldots, q_m\}$. We define random variable $\tilde{S}$ such that if $S(X) = q_i$, then $\tilde{S} = q_{i+1}$.

Proposition 3. *Let $U \sim Unif[0,1]$ be independent of X. Define*

$$Z := U \cdot S(X) + (1 - U) \cdot \tilde{S}. \tag{S1.1}$$

Then, we have that $Z \leq S(X)$ and that $Z \sim Unif[0,1]$.

Proof. Since $\tilde{S} < S(X)$ by definition, it is clear that $Z \leq S(X)$ as well. To show that Z has the Unif$[0,1]$ distribution, fix $t \in (0,1)$. Then there exists $i \in [m]$ such that $q_i \geq t > q_{i+1}$. We then have that

$$\mathbb{P}(Z \leq t) = \mathbb{P}(S(X) \leq q_{i+1}) + \mathbb{P}(Z \leq t, S(X) = q_i, \tilde{S} = q_{i+1})$$

$$= q_{i+1} + \mathbb{P}\left(U \leq \frac{t - q_{i+1}}{q_i - q_{i+1}}\right)\mathbb{P}(S(X) = q_i) = t,$$

where the last inequality follows because $\mathbb{P}(S(X) = q_i) = \mathbb{P}(X = x_i) = S(x_i) - S(x_{i+1}) = q_i - q_{i+1}$. The Proposition follows as desired.

$\square$

Algorithm 3 Computing simultaneously valid p-values for $H_0^{(s,j)}$ in the symmetric array case.

INPUT: Poisson process realizations $\{N_{uv}(\cdot), u < v \in [n]\}$ and a hierarchical partitioning $\boldsymbol{I} = \{I_\ell^{(r)}\}_{r \in [R], \ell \in [2^r]}$ of the domain.

OUTPUT: p-values $\{p_F^{(s,j)}\}$.

1: **for** each $r \in [R]$ **do**

2:      **for** each $\ell \in [2^r]$ **do**

3:          Define integer matrix $A^{(r,\ell)}$ as (317).

4:          Set
$$\bar{p}^{(r,\ell)} = 2\min\left(F_{\text{TW1}}\left(n^{2/3}\big(\lambda_1(A^{(r,\ell)})\big) - 2\right), 1 - F_{\text{TW1}}\left(n^{2/3}\big(\lambda_1(A^{(r,\ell)})\big) - 2\right)\right).$$

5:      **end for**

6: **end for**

7: Apply Algorithm 2 on $\{\bar{p}^{(r,\ell)}\}$ 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: Run Metropolis–Hastings described in Section 3.1.3 to generate $M_1^{(b^*)}, \ldots, M_N^{(b^*)}$ for $b^* \in [B]$ and let $\{N_{uv}^{(b^*)}(\cdot)\}$ be the corresponding point process realizations.

10: **for** $b^* \in \{1, 2, \ldots, B\}$ **do:**

11:      For each $r \in [R]$ and $\ell \in [2^r]$, construct $A_{uv}^{(r,\ell)\,(b^*)}$ from $\{N_{uv}^{(b^*)}(\cdot)\}$.

12:      Repeat lines 1 to 8 to obtain $\tilde{p}_{F,b^*}^{(s,j)}$.

13: **end for**

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

15: Compute $p_F^{(s,j)} = \check{p}_F^{(s,j)} 2^s$.

Algorithm 4 Computing simultaneously valid p-values for $H_0^{(s,j)}$ in the degree-corrected setting.

INPUT: Poisson process realizations $\{N_{uv}(\cdot), u < v \in [n]\}$ and a hierarchical partitioning $\boldsymbol{I} = \{I_\ell^{(r)}\}_{r \in [R], \ell \in [2^r]}$ of the domain.

OUTPUT: p-values $\{p_F^{(s,j)}\}$

1: **for** each $r \in [R]$ **do**

2:     **for** each $\ell \in [2^r]$ **do**

3:         Set $p^{(r,\ell)} = 2\left[1 - \Phi\left(\left|\frac{T^{(r,\ell)}}{\sqrt{6}(\|\hat{\eta}^{(r,\ell)}\|^2 - 1)^{3/2}}\right|\right)\right].$

4:     **end for**

5: **end for**

6: Apply Algorithm 2 on $\{\bar{p}^{(r,\ell)}\}$ to obtain $\{\{p_F^{(s,j,r)}\}_{r=s}^R\}_{s \in [R], j \in [2^s]}.$

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

8: Run Metropolis–Hastings algorithm described in Section 3.3.3 to generate $\boldsymbol{m}^{(b^*)}$ for $b^* \in [B]$ and let $\{N_{uv}^{(b^*)}(\cdot)\}$ be the corresponding point process realizations.

9: **for** $b^* \in \{1, 2, \ldots, B\}$ **do**:

10:     For each $r \in [R]$ and $\ell \in [2^r]$, construct $A_{uv}^{(r,\ell)\,(b^*)}$ from $\{N_{uv}^{(b^*)}(\cdot)\}.$

11:     Repeat lines 1 to 8 to obtain $\tilde{p}_F^{(b^*)}.$

12: **end for**

13: Compute the adjusted p-value $\check{p}_F^{(s,j)} := \frac{1}{B}\sum_{b^*=1}^B \mathbb{1}\{\tilde{p}_{F,(b^*)}^{(s,j)} \leq \tilde{p}_F^{(s,j)}\}.$

14: Compute $p^{(s,j)} = \check{p}_F^{(s,j)} \cdot 2^{-s}.$

S1.3   Proof of Proposition 2

Proof. Let $T(\cdot|\cdot)$ be the transition probability of the Markov Chain specified via (335), we first verify that the Metropolis-Hastings ratio is 1 by showing that $T(\cdot|\cdot)$ is a symmetric distribution, i.e., for any two sample vector $\boldsymbol{m}^{(1)} \neq \boldsymbol{m}^{(2)} \in \mathcal{M}_{\boldsymbol{d}}$, we have $T(\boldsymbol{m}^{(1)}|\boldsymbol{m}^{(2)}) = T(\boldsymbol{m}^{(2)}|\boldsymbol{m}^{(1)})$. It is obvious that the necessary condition for $T(\boldsymbol{m}^{(1)}|\boldsymbol{m}^{(2)})$ to be positive, is that there must exist exactly two indices $i \neq j \in [N]$ such that $m_i^{(1)} \neq m_i^{(2)}$, $m_j^{(1)} \neq m_j^{(2)}$ while the other elements are all the same for the two vectors. We can see that $T(\boldsymbol{m}^{(1)}|\boldsymbol{m}^{(2)}) = T(\boldsymbol{m}^{(2)}|\boldsymbol{m}^{(1)}) = \frac{1}{5\binom{[N]}{2}}$ regardless of the values of $m_i^{(1)}, m_j^{(1)}, m_i^{(2)}, m_j^{(2)}$.

Next, we show that the Markov Chain is irreducible on the support $\mathcal{M}_{\boldsymbol{d}}$. By definition it suffice to show that for any two sample $\boldsymbol{m}^{(1)} \neq \boldsymbol{m}^{(2)} \in \mathcal{M}_{\boldsymbol{d}}$, there exist a finite steps path $\boldsymbol{m}^{(1)} \longrightarrow \boldsymbol{m}^{(2)}$. Since the vector of all degrees $\boldsymbol{D}(\boldsymbol{m})$ are identical for all $\boldsymbol{m} \in \mathcal{M}_{\boldsymbol{d}}$, we can show there exist a path $\boldsymbol{m}^{(1)} \longrightarrow \boldsymbol{m}^{(2)}$ which sequentially matching each element of $\boldsymbol{m}^{(1)}$ to be the same as in $\boldsymbol{m}^{(2)}$. To be specific, we denote the j_{th} elements of vectors $\boldsymbol{m}^{(1)}$ and $\boldsymbol{m}^{(2)}$ as $\boldsymbol{m}_j^{(1)} = (u_j, v_j)$ and $\boldsymbol{m}_j^{(2)} = (u_j', v_j')$ respectively. Let $i = \min\left\{ j \in [N] : \boldsymbol{m}_j^{(1)} \neq \boldsymbol{m}_j^{(2)} \right\}$, we first show that we can go from $\boldsymbol{m}^{(1)}$ to an intermediate state $\boldsymbol{m}^{(1,i)} \in \mathcal{M}_{\boldsymbol{d}}$ in finite steps, such that $\boldsymbol{m}_j^{(1,i)} = \boldsymbol{m}_j^{(2)}$ for all $j \leq i$. We can easily see for $m_i^{(1)} \neq m_i^{(2)}$, there could only be two cases

1. $u_i = u_i', v_i \neq v_i'$ or $u_i \neq u_i', v_i = v_i'$

2. $u_i \neq v_i \neq u_i' \neq v_i'$

For the first case, suppose $u_i = u_i', v_i \neq v_i'$, then by the fact $\boldsymbol{D}(\boldsymbol{m}^{(1)}) = \boldsymbol{D}(\boldsymbol{m}^{(2)})$ there must exist $s > i$ such that $u_s = v_i'$ or $v_s = v_i'$. Then by (335) we can easily check that we can go from $\boldsymbol{m}^{(1)}$ to $\boldsymbol{m}^{(1,i)}$ in one step with $T(\boldsymbol{m}^{(1,i)}|\boldsymbol{m}^{(1)}) = \frac{1}{5\binom{[N]}{2}}$. For the second case, we can go from $\boldsymbol{m}^{(1)}$ to $\boldsymbol{m}^{(1,i)}$ in two steps where in the first step we move to a state $\boldsymbol{m}^{(1,i)\prime}$ such that $\boldsymbol{m}_i^{(1,i)\prime} = (u_i', v_j)$ which is just the state in the first case, so by the same reason we can move to $\boldsymbol{m}^{(1,i)}$ in the second step. Notice that the above paths does not depend on the index i, thus there exist an integer $t \leq 2N$ such that we can sequentially move from $\boldsymbol{m}^{(1)} \longrightarrow \boldsymbol{m}^{(1,i)} \longrightarrow \cdots \longrightarrow \boldsymbol{m}^{(1,N)} = \boldsymbol{m}^{(2)}$ in t steps.

Given any state $\boldsymbol{m}$, if not all edges are the same, i.e., $m_i = (u, v)$ for some $u, v \in [n]$ and all $i \in [N]$, then we can always find $m_i = (u_i, v_i)$ and $m_j = (u_j, v_j)$ with $i < j \in [N]$ such that

48

the nodes u_i, v_i, u_j, v_j satisfy one of the two cases listed above. For case 1, we can see that the five outcomes contains multi-edges, so the Markov Chain can stay at current state with positive probability and thus the period for this state is 1. For 2, we can easily check the state can return in t steps for any $t \geq 2$, thus the period for this state is also 1. Then by irreducibility, we can conclude that the Markov Chain is aperiodic.

With the Markov Chain being irreducible and aperiodic, it converges to its unique stationary distribution. Then by the construction of the Metropolis-Hastings algorithm below (335), it is guaranteed that the stationary distribution is the target distribution, i.e., the uniform distribution on $\mathcal{M}_{\boldsymbol{d}}$.

$\square$

S1.4   Supplementary material for Section 3.1

Recall that we let $A^{(r,\ell)}$, defined in (317), be the adjacency matrix of a undirected Poisson Stochastic Block Model with K communities. We denote $\boldsymbol{\sigma}$ as the membership vector and $\boldsymbol{\gamma}$ as the connection intensity between different communities, as discussed in Section 3.1.1. Without loss of generality, we omit all the superscripts of $A^{(r,\ell)}$ that represents the partition of the support $\mathcal{X}$ and just use A to denote adjacency matrix generated from Poisson Stochastic Block Model in all subsequent analysis in Section S1.4 for simplicity. For the same reason, we also omit the subscripts of $B^{(r,\ell)}$, the matrix with entries being counts of interactions between two groups of individuals, in the proof of theorem 3.

S1.4.1   Proof of Theorem 2

Proof. Under the null hypothesis, we have that $\boldsymbol{P} = \gamma \mathbf{1}_n \mathbf{1}_n^\mathsf{T}$, for some constant $\gamma > 0$.

Let $\tilde{A}'$ be a $n \times n$ matrix such that

$$
\tilde{A}'_{uv} = \begin{cases} (A_{uv} - \hat{\gamma})/\sqrt{(n-1)\gamma}, & u \neq v \\ (\gamma - \hat{\gamma})/\sqrt{(n-1)\gamma}, & u = v \end{cases}
$$

Where $\hat{\gamma} = \frac{2}{n^2 - n} \sum_{u < v} A_{uv}$ is an estimator of γ. Let $C_n = n(\gamma - \hat{\gamma})/\sqrt{(n-1)\gamma}$ and matrix $\tilde{A}^*$ be as defined in (S1.4). Then by definition we have that $\tilde{A}' = \tilde{A}^* + \Delta'$, where $\Delta' =$

$(\gamma - \hat{\gamma})\mathbf{1}_n\mathbf{1}_n^{\mathsf{T}}/\sqrt{(n-1)\gamma} = C_n\mathbf{1}_n\mathbf{1}_n^{\mathsf{T}}/n$. Note that $\hat{\gamma}$ is the sample mean of $n(n-1)/2$ $i.i.d$ Poisson random variables with mean γ, we can apply the Poisson tail bound (S1.5) again and get

$$\mathbb{P}\left(|\gamma - \hat{\gamma}| > s\right) \le 2\exp\left\{-\frac{n(n-1)s^2}{4(\gamma + s)}\right\}$$

And thus we have $|\gamma - \hat{\gamma}| = o_p(\log n/n)$ and that $C_n = o_p(\log n/\sqrt{n})$.

Let μ_i^* be the eigenvector of $\tilde{A}^*$ corresponding to its i_{th} largest eigenvalue. Then by Lemma 2, we have a lower bound on the largest eigenvalue of $\tilde{A}'$:

$$\begin{aligned}
\lambda_1(\tilde{A}') &\ge (\mu_1^*)^{\mathsf{T}}\tilde{A}'\mu_1^* \\
&= \lambda_1(\tilde{A}^*) + (\mu_1^*)^{\mathsf{T}}\Delta'\mu_1^* \\
&= \lambda_1(\tilde{A}^*) + C_n(\mu_1^*)^{\mathsf{T}}\mathbf{1}_n\mathbf{1}_n^{\mathsf{T}}\mu_1^*/n \\
&= \lambda_1(\tilde{A}^*) + \tilde{O}_p(1/n) \cdot o_p(\log n/\sqrt{n}) \\
&\ge \lambda_1(\tilde{A}^*) - o_p(n^{-2/3})
\end{aligned}$$

To derive the upper bound of $\lambda_1(\tilde{A}')$, we denote μ_1' as the eigenvector corresponding to the largest eigenvalue of $\tilde{A}'$. Let $\{a_1, \ldots, a_n\}$ be the coordinates of the vector μ_1' with respect to the basis $\{\mu_1^*, \ldots, \mu_n^*\}$, i.e., $\mu_1' = \sum_{i=1}^n a_i\mu_i^*$. Define $\mathcal{S}_{C_n} \subset [n] := \{i \in [n] : \lambda_i(\tilde{A}^*) > (\lambda_1(\tilde{A}^*) - |C_n|)\}$ as the set of indices of those eigenvalues of $\tilde{A}^*$ that lies in the interval $(\lambda_1(\tilde{A}^*) -$

$|C_n|, \lambda_1(\tilde{A}^*)]$. Then By Lemma 2 and the fact that $|C_n|$ is the largest eigenvalue of Δ', we have

$$
\begin{aligned}
\lambda_1(\tilde{A}') &= (\mu'_1)^{\mathsf{T}} \tilde{A}' \mu'_1 \\
&= (\mu'_1)^{\mathsf{T}} \tilde{A}^* \mu'_1 + (\mu'_1)^{\mathsf{T}} \Delta' \mu'_1 \\
&\leq \sum_{i=1}^{n} a_i^2 \lambda_i(\tilde{A}^*) + \big(\sum_{i \in \mathcal{S}_{Cn}} a_i (\mu_i^*)^{\mathsf{T}} \big) |\Delta'| \big(\sum_{i \in \mathcal{S}_{Cn}} a_i (\mu_i^*) \big) \\
&\quad + \big(\sum_{i \in ([n]/\mathcal{S}_{Cn})} a_i (\mu_i^*)^{\mathsf{T}} \big) |\Delta'| \big(\sum_{i \in ([n]/\mathcal{S}_{Cn})} a_i \mu_i^* \big) \\
&\leq \lambda_1(\tilde{A}^*) \sum_{i \in \mathcal{S}_{Cn}}^{m} a_i^2 + \big(\lambda_1(\tilde{A}^*) - |C_n| \big) \sum_{i \in ([n]/\mathcal{S}_{Cn})} a_i^2 \\
&\quad + |\mathcal{S}_{Cn}| \cdot \sum_{i \in ([n]/\mathcal{S}_{Cn})}^{m} a_i^2 (\mu_i^*)^{\mathsf{T}} |\Delta'| \mu_i^* \quad + |C_n| \sum_{i \in ([n]/\mathcal{S}_{Cn})}^{m} a_i^2 \\
&\leq \lambda_1(\tilde{A}^*) + |\mathcal{S}_{Cn}| |C_n| \cdot \big(\sum_{i \in ([n]/\mathcal{S}_{Cn})}^{m} a_i^2 \cdot \tilde{O}_p(1/n) \big) \\
&= \lambda_1(\tilde{A}^*) + |\mathcal{S}_{Cn}| \cdot \tilde{O}_p(1/n) \cdot o_p(\log n/\sqrt{n})
\end{aligned}
$$

Then we could bound the size of $\mathcal{S}_{Cn}$ by using the results from Erdős et al. (2012) and Bickel and Sarkar (2016), where the main idea is that the empirical counting of the eigenvalues is close to the semicircle counting functions.

Let $N(a, b)$ be the number of eigenvalues of $\tilde{A}^*$ lying in interval $(a, b]$, and define $N_{sc}(a, b) := n \int_a^b \rho_{sc}(x) dx$, where $\rho_{sc} = (1/2\pi)((4 - x^2)_+)^{1/2}$ denote the the density of the semicircle law discussed in Erdős et al. (2012). Following Theorem 2.2 in Erdős et al. (2012) and the discussion in Bickel and Sarkar (2016), there exist constant $A_0 > 1$, C, c and $d < 1$, such that for any L satisfying the following:

$$
A_0 \log\log n \leq L \leq \log(10n)/\log\log n
$$

and for $|a|, |b| < 5$, we have :

$$
\begin{aligned}
&\mathbb{P}\left(|N(a, b) - N_{sc}(a, b)| \geq 2(\log n)^L \right) \\
&\leq \mathbb{P}\left(|N(-\infty, b) - N_{sc}(\infty, b)| \geq (\log n)^L \right) + \mathbb{P}\left(|N(-\infty, a) - N_{sc}(\infty, a)| \geq (\log n)^L \right) \\
&\leq 2C \exp\{-c(\log n)^{(-dL)}\}
\end{aligned}
$$

51

Notice that $\mathcal{S}_{C_n} = N\big(\lambda_1(\tilde{A}^*) - |C_n|, \lambda_1(\tilde{A}^*)\big)$, and from the above inequality we have that:

$$\mathcal{S}_{C_n} = N\big(\lambda_1(\tilde{A}^*) - |C_n|, \lambda_1(\tilde{A}^*)\big) = N_{sc}\big(\lambda_1(\tilde{A}^*) - |C_n|, \lambda_1(\tilde{A}^*)\big)\big) + O_p(\log n)^L \qquad \text{(S1.2)}$$

And

$$
\begin{aligned}
N_{sc}\big(\lambda_1(\tilde{A}^*) - |C_n|, \lambda_1(\tilde{A}^*)\big)\big) &= n \int_{\lambda_1(\tilde{A}^*) - |C_n|}^{\lambda_1(\tilde{A}^*)} (\frac{1}{2\pi}((4 - x^2)_+)^{1/2} dx \\
&\leq n \int_{\lambda_1(\tilde{A}^*) - |C_n|}^{2} (\frac{1}{2\pi}((4 - x^2)_+)^{1/2} dx \\
&= O(n|C_n|^{3/2}) \\
&= o_p\big(n^{1/4}(\log n)^{3/2}\big)
\end{aligned}
$$

Where the second to last equality holds by using the area of a rectangle with side length $(2 - |C_n|)$ and $\sqrt{4 - (2 - |C_n|)^2}$ to cover the actual size of the integral.

Now we can see that

$$
\begin{aligned}
\lambda_1(\tilde{A}') &\leq \lambda_1(\tilde{A}^*) + |\mathcal{S}_{C_n}| \cdot \tilde{O}_p(1/n) \cdot o_p(\log n/\sqrt{n}) \\
&= \lambda_1(\tilde{A}^*) + o_p\big(n^{1/4}(\log n)^{3/2}\big) \cdot \tilde{O}_p(1/n) \cdot o_p(\log n/\sqrt{n}) \\
&= \lambda_1(\tilde{A}^*) + \tilde{O}_p\big((\log n)^{5/2} n^{-5/4}\big) \\
&\leq \lambda_1(\tilde{A}^*) + o_p(n^{-2/3})
\end{aligned}
$$

And combining the lower and upper bound we have that

$$\lambda_1(\tilde{A}') = \lambda_1(\tilde{A}^*) + o_p(n^{-2/3}) \qquad \text{(S1.3)}$$

Now let's get back to the target matrix $\tilde{A} = \sqrt{\frac{\hat{\gamma}}{\gamma}}\big(\tilde{A}' - \frac{C_n}{n} I_n\big)$. By triangle inequality of matrix norm we have

$$\big\|\tilde{A}'\big\| - \big\|\frac{C_n}{n} I_n\big\| \leq \big\|\tilde{A}' - \frac{C_n}{n} I_n\big\| \leq \big\|\tilde{A}'\big\| + \big\|\frac{C_n}{n} I_n\big\|$$

And since $\|\frac{C_n}{n} I_n\| = |C_n/n| = o_p(\log n \cdot n^{-3/2})$, we could easily see that

$$
\begin{aligned}
\lambda_1(\tilde{A}) &= \sqrt{\hat{\gamma}/\gamma} \cdot \big\|\tilde{A}' - \frac{C_n}{n} I_n\big\| \\
&= \big(1 + o_p(\log n/n)\big)\big(\lambda_1(\tilde{A}') + o_p(\log n \cdot n^{-3/2})\big) \\
&= \lambda_1(\tilde{A}') + o_p(n^{-2/3}) \\
&= \lambda_1(\tilde{A}^*) + o_p(n^{-2/3})
\end{aligned}
$$

52

Finally by Lemma 1 and Slutsky's lemma, we have

$$n^{2/3}\big(\lambda_1(\tilde{A}) - 2\big) \xrightarrow{d} \mathrm{TW}_1.$$

□

Lemma 1. *Let $\boldsymbol{P}$ be defined in* (318) *and $\tilde{A}^*$ be a matrix such that*

$$\tilde{A}^*_{uv} = \begin{cases} (A_{uv} - \boldsymbol{P}_{uv})/\sqrt{(n-1)\boldsymbol{P}_{uv}}, & u \neq v \\ \\ 0, & u = v \end{cases} \tag{S1.4}$$

Then we have

$$n^{2/3}(\lambda_1(\tilde{A}^*) - 2) \xrightarrow{d} TW_1.$$

Proof. Consider a $n \times n$ real symmetric Wigner matrix

$$G^*_{uv} = \frac{1}{\sqrt{n-1}} x_{uv}, \quad 1 \leq u, v \leq n$$

Where the off-diagonal elements are *i.i.d.* standard normal distributed random variables and the diagonal elements are zeros. Theorem 1.2. in Lee et al. (2014) implies that the largest eigenvalue of G^* weakly converges to the Tracy-Widom distribution.

Next, by tail bound of Poisson random variables, for any $s > 0$ and $1 \leq u, v \leq n$ we have

$$\mathbb{P}\left(\left| \tilde{A}^*_{uv} \right| > \frac{s}{\sqrt{n-1}} \right) \leq 2\exp(-\frac{\boldsymbol{P}_{uv} s^2}{2(\boldsymbol{P}_{uv} + \sqrt{\boldsymbol{P}_{uv}}s)}) \tag{S1.5}$$

And thus there exist a constant v independent of n, such that for any $s \geq 1$ we have

$$\mathbb{P}\left(\left| \tilde{A}^*_{uv} \right| > \frac{s}{\sqrt{n-1}} \right) \leq v^{-1}\exp(-s^v)$$

The above inequality shows that the entries of $\tilde{A}^*$ have a uniformly subexponential decay, and thus by Theorem 2.4 in Erdős et al. (2012), we have that $n^{2/3}(\lambda_1(\tilde{A}^*) - 2)$ converges to $n^{2/3}(\lambda_1(G^*) - 2)$ in distribution. □

Lemma 2. *For each $1 \leq i \leq n$, let μ_i^* be the eigenvector of $\tilde{A}^*$ corresponding to the i_{th} largest eigenvalue $\lambda_i(\tilde{A}^*)$. Then for any deterministic unit vector $\mathbf{v}$, we have*

$$((\mu_i^*)^\mathsf{T}\mathbf{v})^2 = \tilde{O}_p(1/n), \quad \text{uniformly for all } i \in [n] \tag{S1.6}$$

Where we define $a_n = \tilde{O}_p(b_n)$, if for any $\epsilon > 0$ and $D > 0$, there exists $n_0 = n_0(\epsilon, D)$ such that

$$\mathbb{P}(a_n \geq n^\epsilon b_n) \leq n^{-D} \quad \text{for all } n \geq n_0.$$

Lemma 2 is a direct application of the eigenvector delocalization theorem proposed in Alex et al. (2014). Note that the conditions of Theorem 2.16 in Alex et al. (2014) does not apply to our configuration of $\tilde{A}^*$ since the diagonal entries are made to be all zeros while the original condition requires that all elements of the matrix should have positive variance. However, Erdős et al. (2013) provides a local semicircle law(Theorem 2.3) which holds even when some entries of a generalized Wigner matrix have zero variance, and as a result of the local semicircle law, the eigenvector delocalization theorem still holds in our setting. See also discussions in Bickel and Sarkar (2016) and Lei (2016).

S1.4.2   Maximum eigenvalue test statistic under an alternative

We consider the limiting distribution of $\lambda_1(\tilde{A}^{(r,\ell)})$ under some alternative cases. When the adjacency matrix is generated from a Stochastic Block Model with $K > 1$ communities and Bernoulli entries, Bickel and Sarkar (2013) shows that the largest eigenvalue of the scaled and centered adjacency matrix is $O(\sqrt{n})$, given that the community probability matrix ψ is diagonally dominant. Lei (2016) provided a more general result which requires that each community has size at least proportional to n/K. The following proposition is a direct extension of Theorem 3.3 in Lei (2016) to the Poisson network.

Proposition 4. *Let $A^{(r,\ell)}$ be an adjacency matrix from Poisson stochastic model with K communities and let $\mathcal{G}_k = \{u \in [n] : \sigma(u) = k\}$ be the set of vertices that belong to group k for $k \in [K]$.*

Assume there exist a constant $C_K > 0$ such that for all n we have

$$\min_{k \in [K]} |\mathcal{G}_k| \geq C_K \cdot n \tag{S1.7}$$

Then for any $r \in [R]$ and $\ell \in [2^r]$, if $K > 1$ we have

$$\lambda_1(\tilde{A}^{(r,\ell)}) \geq \frac{\sqrt{n}\delta C_K - O_p(1)}{\left(\|\gamma\|_{max} + o_p(\log n/n)\right)^{1/2}} \tag{S1.8}$$

where δ is the minimum ℓ_∞ distance between any two distinct rows of γ.

54

Proof. Let $\hat{\boldsymbol{P}} = \hat{\gamma}\mathbf{1}_n\mathbf{1}_n^\mathsf{T}$, we have

$$
\begin{aligned}
\big\|\tilde{A}\big\| &= \big((n-1)\hat{\gamma}\big)^{-1/2}\big\|A - \big(\hat{\boldsymbol{P}} - \mathrm{diag}(\hat{\boldsymbol{P}})\big)\big\| \\
&\geq \big((n-1)\hat{\gamma}\big)^{-1/2}\Big(\big\|\boldsymbol{P} - \hat{\boldsymbol{P}} - \mathrm{diag}(\boldsymbol{P} - \hat{\boldsymbol{P}})\big\| - \big\|A - \big(\boldsymbol{P} - \mathrm{diag}(\boldsymbol{P})\big)\big\|\Big)
\end{aligned}
$$

We can see that the matrix $A - \big(\boldsymbol{P} - \mathrm{diag}(\boldsymbol{P})\big)$ has off-diagonal entries being independent, centered Poisson random variables and diagonal entries being all zeros. By Theorem 2 in Latała (2005), we have that there exist some $C' > 0$ such that

$$
\mathbb{E}\big\|A - \big(\boldsymbol{P} - \mathrm{diag}(\boldsymbol{P})\big)\big\| \leq C'\sqrt{n} \tag{S1.9}
$$

and thus $\big\|A - \big(\boldsymbol{P} - \mathrm{diag}(\boldsymbol{P})\big)\big\| \leq O_p(\sqrt{n})$.

To derive an upper bound of $\big\|\boldsymbol{P} - \hat{\boldsymbol{P}} - \mathrm{diag}(\boldsymbol{P} - \hat{\boldsymbol{P}})\big\|$, we notice that since $K > 1$, there exist two community $k_1 \neq k_2$. Let $\boldsymbol{g}_{k_1} = \{u \in [n] : \sigma(u) = k_1\}$ and $\boldsymbol{g}_{k_2} = \{u \in [n] : \sigma(u) = k_2\}$ be the set of vertices that belong to k_1 and k_2 respectively. Since we assume the matrix $\boldsymbol{\gamma}$ have pairwise distinct rows, there must exist a group $k_3 \in [K]$ such that $\boldsymbol{\gamma}_{k_1 k_3} \neq \boldsymbol{\gamma}_{k_2 k_3}$, and we can choose

$$
k_3 = \operatorname*{argmin}_{k' \in [K]} |\boldsymbol{\gamma}_{k_1 k'} - \boldsymbol{\gamma}_{k_2 k'}|.
$$

Note that k_3 can be equal to k_1 or k_2. Now let $\boldsymbol{D}$ be a submatrix of $\boldsymbol{P} - \hat{\boldsymbol{P}} - \mathrm{diag}(\boldsymbol{P} - \hat{\boldsymbol{P}})$, which only consist the rows in $k_1 \cup k_2$ and columns in k_3. We can see that when $k_1 \neq k_2 \neq k_3$, after some row permutaions D could be seen as:

$$
\boldsymbol{D} = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}
$$

where D_1 is a $|k_1| \times |k_3|$ matrix with all entries equal to $\boldsymbol{\gamma}_{k_1 k_3} - \hat{\gamma}$ and D_2 is a $|k_2| \times |k_3|$ matrix with all entries equal to $\boldsymbol{\gamma}_{k_2 k_3} - \hat{\gamma}$. Then we have

$$
\begin{aligned}
\big\|\boldsymbol{D}\big\| &\geq \max\Big((\boldsymbol{\gamma}_{k_1 k_3} - \hat{\gamma})\sqrt{|k_1| \cdot |k_3|}, (\boldsymbol{\gamma}_{k_2 k_3} - \hat{\gamma})\sqrt{|k_2| \cdot |k_3|}\Big) \\
&\geq n\delta C_K
\end{aligned}
$$

When $k_3 = k_1$ or $k_3 = k_2$, we can see $\boldsymbol{D}$ could still be permuted into a block matrix with blocks D_1 and D_2. However in this case, one of the blocks have all diagonal entries being zeros,

since we do not allow self-loops. Without loss of generality, we assume that $k_3 = k_1$, and we still have the same lower bound of $\|\boldsymbol{D}\|$ by

$$\|\boldsymbol{D}\| \geq \max\Big((\boldsymbol{\gamma}_{k_1 k_1} - \hat{\gamma})(|k_1| - 1), (\boldsymbol{\gamma}_{k_2 k_1} - \hat{\gamma})\sqrt{|k_2| \cdot |k_1|}\Big)$$

$$\geq n\delta C_K - O_p(1)$$

Finally we have

$$\|\tilde{A}\| \geq \big((n-1)\hat{\gamma}\big)^{-1/2}\big(n\delta C_K - O_p(\sqrt{n})\big)$$

$$\geq \frac{\sqrt{n}\delta C_K - O_p(1)}{\big(\|\boldsymbol{\gamma}\|_{\max} + o_p(\log n/n)\big)^{1/2}}$$

$\square$

S1.4.3   Proof of Theorem 3

Proof. Under the null hypothesis, we have that $B_{uv} \sim \text{Poisson}(\gamma)$ for some $\gamma > 0$ and for any $u \in V_1, v \in V_2$. Recall that we denote $\hat{\gamma}$ (325) as an estimator of γ, matrix $\tilde{B}$ as the empirically centered and scaled counterpart of B and $\tilde{W} = \tilde{B}^\mathsf{T}\tilde{B}$.

Let $\tilde{B}^*$ be as defined in (S1.13) and $\tilde{B}'$ be a $m \times n$ matrix with entries $\tilde{B}'_{uv} = (B_{uv} - \hat{\gamma})/\sqrt{m\gamma}$. Then we have $\tilde{B}' = \tilde{B}^* + \alpha\Delta$, where $\Delta = \mathbf{1}_m \mathbf{1}_n^\mathsf{T}$ and $\alpha = \frac{\gamma - \hat{\gamma}}{\sqrt{m\gamma}}$. Denote $\tilde{W}^* = \tilde{B}^{*\mathsf{T}}\tilde{B}^*$ and $\tilde{W}' = \tilde{B}'^\mathsf{T}\tilde{B}'$. Then we let $\big(\lambda_i(\tilde{W}^*), \mu_i^*\big)_{i=1}^n$ be the pairs of eigenvalue and eigenvector of matrix $\tilde{W}^*$ with the eigenvalues in a non-increasing order, namely $\lambda_1(\tilde{W}^*) \geq \lambda_2(\tilde{W}^*) \cdots \geq \lambda_n(\tilde{W}^*)$. Similarly, we let $(\lambda_i(\tilde{W}'), \mu_i')_{i=1}^n$ be the pairs of eigenvalue and eigenvectors of $\tilde{W}'$, where the eigenvalues are in non-increasing order as well.

First let us derive the a lower bound on $\lambda_1(\tilde{W}')$, the largest eigenvalue of $\tilde{W}' = \tilde{B}'^\mathsf{T}\tilde{B}'$:

$$\lambda_1(\tilde{W}') \geq (\mu_1^*)^\mathsf{T}\tilde{B}'^\mathsf{T}\tilde{B}'\mu_1^*$$

$$= (\mu_1^*)^\mathsf{T}\tilde{B}^{*\mathsf{T}}\tilde{B}^*\mu_1^* + \alpha(\mu_1^*)^\mathsf{T}(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^\mathsf{T}\tilde{B}^* + \alpha\Delta^\mathsf{T}\Delta)\mu_1^*$$

$$= \lambda_1(\tilde{W}^*) + \alpha(\mu_1^*)^\mathsf{T}(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^\mathsf{T}\tilde{B}^* + \alpha\Delta^\mathsf{T}\Delta)\mu_1^*$$

$$\geq \lambda_1(\tilde{W}^*) - |\alpha(\mu_1^*)^\mathsf{T}(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^\mathsf{T}\tilde{B}^* + \alpha\Delta^\mathsf{T}\Delta)\mu_1^*|$$

Let $\tilde{B}^* = \sum_{i=1}^n \sqrt{\lambda_i(\tilde{W}^*)} s_i^* \mu_i^{*\mathsf{T}}$ be the singular value decomposition of $\tilde{B}^*$. Then we have:

$$\mu_i^{*\mathsf{T}} \tilde{B}^{*\mathsf{T}} = \sum_{i=1}^n \sqrt{\lambda_i(\tilde{W}^*)} \mu_1^{*\mathsf{T}} \mu_i^* s_i^{*\mathsf{T}} = \sqrt{\lambda_i(\tilde{W}^*)} s_i^{*\mathsf{T}}$$

Notice that s_i^*, μ_i^* are the eigenvectors of the matrix $\tilde{B}^* \tilde{B}^{*\mathsf{T}}, \tilde{B}^{*\mathsf{T}} \tilde{B}^*$ respectively, and we can easily check the conditions in Lemma 4 hold for both matrix $\tilde{B}^*$ and its transpose, thus we have

$$\alpha \mu_1^{*\mathsf{T}} \tilde{B}^{*\mathsf{T}} \Delta \mu_1^* = \alpha \sqrt{\lambda_1(\tilde{W}^*)} s_i^{*\mathsf{T}} \mathbf{1}_m \mathbf{1}_n^{\mathsf{T}} \mu_1^*$$
$$= o_p(\log n/n^{3/2}) \cdot \tilde{O}_p(1) \cdot \lambda_1(\tilde{W}^*)$$
$$= \tilde{O}_p(\log n/n^{3/2}) \cdot O_p(n^{-1/3})$$

Where the $o_p(\log n/n^{3/2})$ term is derived by noticing that $\hat{\gamma}$ is the sample mean of $m \times n$ independent Poisson random variables, and again by the Poisson tail bound (S1.5) we get $\alpha = \frac{\gamma - \hat{\gamma}}{\sqrt{m\gamma}} = o_p(\log n/n^{3/2})$. On the other hand it is easily seen that $\alpha^2 \mu_1^{*\mathsf{T}} \Delta^{\mathsf{T}} \Delta \mu_1^* = \tilde{O}_p(1) \cdot o_p((\log n)^2/n^2)$, which indicates that:

$$\lambda_1(\tilde{W}') \geq \lambda_1(\tilde{W}^*) - |\alpha(\mu_1^*)^{\mathsf{T}} (\tilde{B}^{*\mathsf{T}} \Delta + \Delta^{\mathsf{T}} \tilde{B}^* + \alpha \Delta^{\mathsf{T}} \Delta) \mu_1^*|$$
$$\geq \lambda_1(\tilde{W}^*) - \tilde{O}_p(\log n/n^{3/2}) \cdot O_p(n^{-1/3}) - \tilde{O}_p((\log n)^2/n^2)$$
$$= \lambda_1(\tilde{W}^*) - \tilde{O}_P(\log n \cdot n^{-11/6}) \tag{S1.10}$$

Next, we derive an upper bound for the largest eigenvalue of $\tilde{W}'$. Let $\{a_1, \ldots, a_n\}$ be the coordinates of μ_1', the eigenvector of $\tilde{W}'$ associated with its largest eigenvalue, with respect to the basis consisting of eigenvector of $\tilde{W}^*$, i.e., $\mu_1' = \sum_{i=1}^n a_i \mu_i^*$. Let $\mathcal{S} = \{i \in [n] : \lambda_i(\tilde{W}^*) > \lambda_1(\tilde{W}^*) - 2\|\alpha(\tilde{B}^{*\mathsf{T}} \Delta + \Delta^{\mathsf{T}} \tilde{B}^* + \alpha \Delta^{\mathsf{T}} \Delta)\|\}$, such that $|\mathcal{S}|$ is the number of $\lambda_i(\tilde{W}^*)$'s in the interval $(\lambda_1(\tilde{W}^*) - 2\|\alpha(\tilde{B}^{*\mathsf{T}} \Delta + \Delta^{\mathsf{T}} \tilde{B}^* + \alpha \Delta^{\mathsf{T}} \Delta)\|, \lambda_1(\tilde{W}^*))$. Let $\mathbf{v}_1 = \sum_{i=1}^m a_i \mu_i^*$ and $\mathbf{v}_2 = \sum_{i=m+1}^n a_i \mu_i^*$

so that $\mu_1' = \mathbf{v}_1 + \mathbf{v}_2$. we have:

$$\lambda_1(\tilde{W}') = (\mu_1')^{\mathsf{T}} \tilde{W}' \mu_1'$$

$$= (\mu_1')^{\mathsf{T}} \tilde{B}^{*\mathsf{T}} \tilde{B}^* \mu_1' + \alpha(\mu_1')^{\mathsf{T}} (\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)\mu_1'$$

$$\leq \lambda_1(\tilde{W}^*) \cdot \sum_{j \in \mathcal{S}} a_j^2 + \left(\lambda_1(\tilde{W}^*) - 2\|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)\|\right) \cdot \sum_{j \in ([n]/\mathcal{S})} a_j^2$$

$$+ 2\mathbf{v}_1^{\mathsf{T}}|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)|\mathbf{v}_1 + 2\mathbf{v}_2^{\mathsf{T}}|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)|\mathbf{v}_2$$

$$\leq \lambda_1(\tilde{W}^*) - 2\|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)\| \sum_{j \in ([n]/\mathcal{S})} a_j^2$$

$$+ 2m \sum_{j \in \mathcal{S}} a_j^2 (\mu_j^*)^{\mathsf{T}}|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)|\mu_j^*$$

$$+ 2\|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^{\mathsf{T}}\tilde{B}^* + \alpha\Delta^{\mathsf{T}}\Delta)\| \sum_{j \in ([n]/\mathcal{S})} a_j^2$$

$$\leq \lambda_1(\tilde{W}^*) + 2|\mathcal{S}| \sum_{j \in \mathcal{S}} a_j^2 (\mu_j^*)^{\mathsf{T}} \left(|2\alpha(\tilde{B}^{*\mathsf{T}}\Delta| + |\alpha^2\Delta^{\mathsf{T}}\Delta)|\right)\mu_j^*$$

$$\leq \lambda_1(\tilde{W}^*) + 2|\mathcal{S}| \sum_{j \in \mathcal{S}} a_j^2 \left(\lambda_j(\tilde{W}^*) \cdot \tilde{O}_p(\log n/n^{3/2}) + \tilde{O}_p((\log n)^2/n^2)\right)$$

$$\leq \lambda_1(\tilde{W}^*) + 2|\mathcal{S}| \left(O_p(n^{-1/3}) \cdot \tilde{O}_p(\log n/n^{3/2}) + \tilde{O}_p((\log n)^2/n^2)\right)$$

$$\leq \lambda_1(\tilde{W}^*) + 2|\mathcal{S}| \cdot \tilde{O}_p(\log n \cdot n^{-11/6})$$

Now let $a = \lambda_1(\tilde{W}^*) - 2\|\alpha(\tilde{B}'\Delta + \Delta'\tilde{B} + \alpha\Delta'\Delta)\|$ and $b = \lambda_1(\tilde{W}^*)$. We can see that $|\mathcal{S}| = \mathcal{N}(a) - \mathcal{N}(b)$. Noticing that $|(\mathcal{N}(a) - \mathcal{N}(b)) - (\mathcal{N}_m(a) - \mathcal{N}_m(b))| = |(\mathcal{N}(a) - \mathcal{N}_m(a)) - (\mathcal{N}(b) - \mathcal{N}_m(b))|$ and together with Lemma 5 we have that for any $\varepsilon > 0$

$$\mathbb{P}(|(\mathcal{N}(a) - \mathcal{N}(b)) - (\mathcal{N}_m(a) - \mathcal{N}_m(b))| \geq 2n^{-1}\log(n)^{C_\varepsilon \log\log(n)})$$

$$\leq \mathbb{P}[(\mathcal{N}(a) - \mathcal{N}_m(a)) \geq n^{-1}\log(n)^{C_\varepsilon \log\log(n)}] + \mathbb{P}[(\mathcal{N}(b) - \mathcal{N}_m(b)) \geq n^{-1}\log(n)^{C_\varepsilon \log\log(n)}]$$

$$\leq 2n^{C_\varepsilon} \exp(-\log(n)^{\varepsilon \log\log(n)})$$

and which indicates that $|\mathcal{S}| = |\mathcal{N}_m(a) - \mathcal{N}_m(b)| + O_p\left(n^{-1}\log(n)^{C_\varepsilon \log\log(N)}\right)$. Since $\mathcal{N}_m(a) - \mathcal{N}_m(b) = n\int_a^b \varrho_m(x)dx$, and it is easily seen by simple calculus that $\varrho_m(x)$ achieves it's local maximum at $x = \frac{(1-n/m)^2}{1+n/m}$, and $\varrho_m(\frac{(1-n/m)^2}{1+n/m}) = \frac{1}{\pi|1-n/m|\sqrt{n/m}}$. Thus we could have the

58

following bound on the size of $|\mathcal{N}_m(a) - \mathcal{N}_m(b)|$:

$$|\mathcal{N}_m(a) - \mathcal{N}_m(b)| \leq 2n\|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^\mathsf{T}\tilde{B}^* + \alpha\Delta^\mathsf{T}\Delta)\|\frac{1}{\pi|1 - \frac{n}{m}|\sqrt{\frac{n}{m}}}$$

And we could see that

$$\left\|\alpha(\tilde{B}^{*\mathsf{T}}\Delta + \Delta^\mathsf{T}\tilde{B}^* + \alpha\Delta^\mathsf{T}\Delta)\right\| \leq |\alpha|\left(\left\|\tilde{B} * \mathsf{T}\Delta\right\| + \left\|\Delta'\tilde{B}^*\right\| + |\alpha| \cdot \left\|\Delta'\Delta\right\|\right)$$

$$\leq |\alpha|\left(2\left\|\tilde{B}^*\right\| \cdot \|\Delta\|_F + |\alpha| \cdot \|\Delta\|_F^2\right)$$

$$= o_p(\log n \cdot n^{-3/2})\left(2\sqrt{mn\lambda_1(\tilde{W}^*)} + mn \cdot o_p(\log n \cdot n^{-3/2})\right)$$

$$= o_p(\log n \cdot n^{-5/6})$$

Thus we have that $|\mathcal{S}| = |\mathcal{N}_m(a) - \mathcal{N}_m(b)| + O_p\left(n^{-1}\log(n)^{C_\varepsilon \log\log(n)}\right) \leq o_p(\log n \cdot n^{1/6})$ and

$$\lambda_1(\tilde{W}') \leq \lambda_1(\tilde{W}^*) + |\mathcal{S}| \cdot \tilde{O}_P(\log n \cdot n^{-11/6})$$

$$\leq \lambda_1(\tilde{W}^*) + \tilde{O}_p\left((\log n)^2 \cdot n^{-17/6}\right) \tag{S1.11}$$

Combine (S1.10) and (S1.11) we conclude that

$$\lambda_1(\tilde{W}') = \lambda_1(\tilde{W}^*) + o_p(n^{-2/3}) \tag{S1.12}$$

Finally we can look at the target matrix $\tilde{B}$. We see that $\tilde{B} = \tilde{B}' \times \sqrt{\frac{\gamma}{\hat{\gamma}}}$ and we can derive from the Poisson tail bond (S1.5) that $\sqrt{\frac{\gamma}{\hat{\gamma}}} = o_p(\log n \cdot n^{-1/2})$. Thus we can also have

$$\lambda_1(\tilde{W}') = \sqrt{\frac{\gamma}{\hat{\gamma}}}\lambda_1(\tilde{W}')$$

$$= \left(1 + o_p(\log n \cdot n^{-1/2})\right)\left(\lambda_1(\tilde{W}^*) + o_p(n^{-2/3})\right)$$

$$= \lambda_1(\tilde{W}^*) + o_p(n^{-2/3})$$

Then by Lemma 3 and Slutsky's theorem, we get the result of Theorem 3. $\qquad\square$

Lemma 3. *Let $\tilde{B}_{uv}^*$ be a matrix with entries:*

$$\tilde{B}_{uv}^* := \frac{B_{uv} - \gamma}{\sqrt{m\gamma}}, \ \forall u \in [m], v \in [n] \tag{S1.13}$$

and let $\tilde{W}^* = \tilde{B}^{*\mathsf{T}}\tilde{B}^*$ with $\lambda_1(\tilde{W}^*)$ being its largest eigenvalue. Suppose that $\lim_{n\to\infty} n/m \in (0, \infty)$, then we have, as $m, n \to \infty$:

$$\frac{m\lambda_1(\tilde{W}^*) - (\sqrt{n} + \sqrt{m})^2}{(\sqrt{m} + \sqrt{n})(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})^{1/3}} \xrightarrow{d} TW_1, \tag{S1.14}$$

Lemma 4 (Theorem 2.8 in Alex et al. (2014)). *Let G be an $m \times n$ random matrix with independent entries satisfying*

$$\mathbb{E}G_{uv} = 0, \quad \mathbb{E}|G_{uv}|^2 = \frac{1}{\sqrt{nm}}.$$

Assume that m and n satisfy the bounds $n^{1/C} \leq m \leq n^C$ for some $C > 0$. Suppose for all $p \in \mathbb{N}$, there exist C_p such that

$$\mathbb{E}|(mn)^{1/4}G_{uv}|^p \leq C_p$$

Let $\boldsymbol{\mu}_i$ be the eigenvalue of $G^\mathsf{T}G$ associated with its i_{th} largest eigenvalue. Then for any $\varepsilon > 0$ we have

$$|\boldsymbol{\mu}_i^\mathsf{T}\mathbf{v}|^2 = \tilde{O}_P(1/n)$$

uniformly for all $i \leq (1 - \varepsilon)\min(m, n)$ and any deterministic unit vector $\mathbf{v} \in \mathbb{R}^n$.

Lemma 5 (Theorem 3.3 in Pillai et al. (2014)). *Let $\xi_\pm = (1 \pm \sqrt{\frac{n}{m}})^2$, and denote the Marchenko-Pastur law by ϱ_m, which is given by*

$$\varrho_m(x) = \frac{m}{2\pi n}\sqrt{\frac{[(\xi_+ - x)(x - \xi_-)]_+}{x^2}}$$

Let $\beta \in \mathbb{R}$, define the empirical spectral distribution of $(\tilde{B}^*)^\mathsf{T}\tilde{B}^*$ by:

$$\mathcal{N}(\beta) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{[\beta,\infty)}\big(\lambda_i(\tilde{W}^*)\big)$$

And the distribution given by the Marchenko-Pastur law:

$$\mathcal{N}_m(\beta) := \int_\beta^\infty \varrho_m(x)dx$$

If $\lim_{n\to\infty} \frac{n}{m} \in (0,\infty) \setminus \{1\}$, then for any $\varepsilon > 0$, there exists a constant C_ε such that:

$$\mathbb{P}(|\mathcal{N}(\beta) - \mathcal{N}_m(\beta)| \geq n^{-1}\log(n)^{C_\varepsilon \log\log(n)}) \leq n^{C_\varepsilon}\exp(-\log(n)^{\varepsilon \log\log(n)})$$

S2. Supplementary material for Section 4

For $r \in [R]$ and $\ell \in [2^r]$, define

$$\delta^{(r,\ell)} := \frac{\int_{I_\ell^{(r)}} \lambda_a - \lambda_b \, d\nu}{\int_{I_\ell^{(r)}} \lambda_a + \lambda_b \, d\nu}, \tag{S2.15}$$

so that $\frac{\int_{I_\ell^{(r)}} \lambda_a\, d\nu}{\int_{I_\ell^{(r)}} \lambda\, d\nu} = \frac{1}{2} + \delta^{(r,\ell)}$ and $\frac{\int_{I_\ell^{(r)}} \lambda_b\, d\nu}{\int_{I_\ell^{(r)}} \lambda\, d\nu} = \frac{1}{2} - \delta^{(r,\ell)}$.

Proposition 5. *For every $r \in [R]$ and $l \in [2^r]$, $N_a^{(r,\ell)} \sim Bin\big(\frac{1}{2} + \delta^{(r,\ell)}, N^{(r,\ell)}\big)$ conditional on $N^{(r,\ell)}$ and if $\delta^{(r,\ell)} = 0$, then $p^{(r,\ell)} \sim Unif[0,1]$.*

Moreover, for every $r \in [R]$, conditional on $\{N^{(r,\ell)}\}_{\ell \in [2^r]}$, the collection of random variables $\{N_a^{(r,\ell)}\}_{l \in [2^r]}$ are mutually independent.

Proof. Since $N_a^{(r,\ell)} = N_a(I_\ell^{(r)})$, we have that $N_a^{(r,\ell)}$ has the Poisson distribution with mean $\int_{I_l^{(r)}} \lambda_a d\nu$. Since $N^{(r,\ell)} - N_a^{(r,\ell)} = N_b^{(r,\ell)}$ has the Poisson distribution with mean $\int_{I_l^{(r)}} \lambda_b d\nu$, and is independent of $N_a^{(r,\ell)}$, we have that, for any $s, t \in \mathbb{N}$ where $s \leq t$,

$$
\begin{aligned}
\mathbb{P}(N_a^{(r,\ell)} = s \mid N^{(r,\ell)} = t) &= \frac{\mathbb{P}(N_a^{(r,\ell)} = s, N_b^{(r,\ell)} = t - s)}{\mathbb{P}(N^{(r,\ell)} = t)} \\
&= \frac{\frac{1}{s!} e^{-\int_{I_l^{(r)}} \lambda_a d\nu} \left\{ \int_{I_l^{(r)}} \lambda_a d\nu \right\}^s \frac{1}{(t-s)!} e^{-\int_{I_l^{(r)}} \lambda_b d\nu} \left\{ \int_{I_l^{(r)}} \lambda_b d\nu \right\}^{t-s}}{\frac{1}{t!} e^{-\int_{I_l^{(r)}} \lambda_a + \lambda_b d\nu} \left\{ \int_{I_l^{(r)}} \lambda_a + \lambda_b d\nu \right\}^t} \\
&= \binom{t}{s} \left(\frac{1}{2} + \delta^{(r,\ell)} \right)^s \left(\frac{1}{2} - \delta^{(r,\ell)} \right)^{t-s},
\end{aligned}
$$

and the first claim follows directly. If $\delta_l^{(k)} = 0$, then $\hat{p}_l^{(k)}$ is uniform by Proposition 3.

The second claim follows from the independent increment property of a Poisson process. $\square$

S2.1   Proof of Theorem 4

Proof. (of Theorem 4)

Let $r^* \in [R]$ denote the resolution level that satisfies (436). Recalling that $\delta_l^{(r^*)}$ is defined as (S2.15), we define the event

$$
\mathcal{E}_{r^*} := \left\{ \sum_{l=1}^{2^{r^*}} (N^{(r^*,l)} - 1) \delta_l^{(r^*)2} \geq 2^{r^*/2} \left(\frac{C^{1/2}}{\beta^{1/2}} + 2\log^{1/2} \frac{R}{\alpha} \right) + 2\log \frac{R}{\alpha} \right\},
$$

where C is the universal constant specified in Theorem S1 which we may assume to be greater than 1. Then, by Theorem S1,

$$
\begin{aligned}
\mathbb{P}\big(p_F \leq \alpha\big) &\leq \mathbb{P}\big(p_F^{(r^*)} \leq \frac{\alpha}{R}\big) \\
&\leq \mathbb{P}\left(\left\{ p_F^{(r^*)} \leq \frac{\alpha}{R} \right\} \cap \mathcal{E}_{r^*} \right) + \mathbb{P}(\mathcal{E}_{r^*}^c) \leq \beta + \mathbb{P}(\mathcal{E}_{r^*}^c).
\end{aligned}
$$

In order to upper bound the probability of $\mathcal{E}^c_{r*}$, we observe, by our assumption that $\mathbb{E}N^{(r^*,l)} = \int_{I_l^{(r^*)}} \lambda d\nu \geq 2$ for all $l \in [2^{r^*}]$ and the fact that $|\delta_l^{(r^*)2}| \leq \frac{1}{2}$, that

$$\mathbb{E}\sum_{l=1}^{2^{r^*}}(N^{(r^*,l)} - 1)\delta_l^{(r^*)2} \geq \frac{1}{2}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2} \quad \text{and} \tag{S2.16}$$

$$\mathrm{Var}\sum_{l=1}^{2^{r^*}}(N^{(r^*,l)} - 1)\delta_l^{(r^*)2} \leq \frac{1}{4}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2}. \tag{S2.17}$$

As a short hand, we write

$$W \equiv \sum_{l=1}^{2^{r^*}}(N^{(r^*,l)} - 1)\delta_l^{(r^*)2}, \quad \text{and}$$

$$T_{r^*,R,\alpha,\beta} \equiv 2^{r^*/2}\Big(\frac{C^{1/2}}{\beta^{1/2}} + 2\log^{1/2}\frac{R}{\alpha}\Big) + 2\log\frac{R}{\alpha}.$$

We note that by (436), we have $\mathbb{E}W \geq \frac{1}{2}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big) \geq T_{r^*,R,\alpha,\beta}$. By this, Chebyshev' inequality, and (S2.16) and (S2.17), we have

$$\mathbb{P}(\mathcal{E}^c_{r*}) = \mathbb{P}\Big\{\sum_{l=1}^{2^{r^*}}(N^{(r^*,l)} - 1)\delta_l^{(r^*)2} \leq T_{r^*,R,\alpha,\beta}\Big\}$$

$$= \mathbb{P}\big\{W \leq T_{r^*,R,\alpha,\beta}\big\}$$

$$= \mathbb{P}\big\{W - \mathbb{E}W \leq T_{r^*,R,\alpha,\beta} - \mathbb{E}W\big\}$$

$$\leq \mathrm{Var}(W) \cdot \big\{\mathbb{E}W - T_{r^*,R,\alpha,\beta}\big\}^{-2}$$

$$\leq \mathrm{Var}\Big\{\sum_{l=1}^{2^{r^*}}(N^{(r^*,l)} - 1)\delta_l^{(r^*)2}\Big\}\Big\{\mathbb{E}\Big(\sum_{l=1}^{2^{r^*}}(N^{(r^*,l)} - 1)\delta_l^{(r^*)2}\Big) - T_{r^*,R,\alpha,\beta}\Big\}^{-2}$$

$$\leq \Big\{\frac{1}{4}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2}\Big\}\Big\{\frac{1}{2}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2} - T_{r^*,R,\alpha,\beta}\Big\}^{-2}$$

$$\leq \Big\{\frac{1}{4}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2}\Big\}\Big\{\frac{1}{4}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2}\Big\}^{-2}$$

$$\leq \Big\{\frac{1}{4}\sum_{l=1}^{2^{r^*}}\Big(\int_{I_l^{(r^*)}} \lambda d\nu\Big)\delta_l^{(r^*)2}\Big\}^{-1} \leq \beta,$$

where the penultimate inequality follows from (436) and the fact that $C \geq 1$. The Theorem follows as desired.

$\square$

S2.2 Proof of Theorem 5

Proof. (of Theorem 5)

We first claim that, writing C as the universal constant specified in Theorem 4,

$$\frac{1}{4n}\int_I \left(\frac{\lambda_a - \lambda_b}{\lambda}\right)^2 \lambda\, d\nu \geq \min_{r\in[R]} \frac{2^{r/2}}{n}\left(\frac{C^{1/2}}{\beta} + 2\log^{1/2}\frac{R}{\alpha}\right) + \frac{2}{n}\log\frac{R}{\alpha} + \frac{C_H C_d}{2}2^{-\frac{r\gamma}{q}}. \quad \text{(S2.18)}$$

To see that this claim is true, define

$$\tilde{r} = \min\left(R, \left\lfloor \frac{\log_2 \frac{n}{2}}{\frac{1}{2} + \frac{2\gamma}{q}} - \log_2\frac{c_{\max}}{c_{\min}} \right\rfloor\right),$$

or equivalently,

$$\tilde{r} = \begin{cases} \left\lfloor \frac{\log_2 \frac{n}{2}}{\frac{1}{2} + \frac{2\gamma}{q}} - \log_2\frac{c_{\max}}{c_{\min}} \right\rfloor & \text{if } \gamma/q \geq 1/4 \\ R & \text{if } \gamma/q \leq 1/4 \end{cases}.$$

Then, using the fact that $\log_2 \frac{n}{2} - \log_2\frac{c_{\max}}{c_{\min}} - 1 \leq R \leq \log_2 \frac{n}{2} - \log_2\frac{c_{\max}}{c_{\min}}$, we have

$$\frac{2^{\tilde{r}/2}}{n}\left(\frac{C^{1/2}}{\beta} + 2\log^{1/2}\frac{R}{\alpha}\right) + \frac{2}{n}\log\frac{R}{\alpha} + \frac{C_H C_d}{2}2^{-\frac{2\tilde{r}\gamma}{q}}$$

$$\leq \begin{cases} \frac{1}{4}C_1 n^{-\frac{4\gamma}{q+4\gamma}}\left(\beta^{-1} + \log\frac{\log n}{\alpha}\right) & \text{if } \gamma/q \geq 1/4 \\ \frac{1}{4}C_1 n^{-\frac{2\gamma}{q}}\left(\beta^{-1} + \log\frac{\log n}{\alpha}\right) & \text{if } \gamma/q \leq 1/4 \end{cases}$$

for some $C_1 > 0$ whose value depends only on $\frac{c_{\max}}{c_{\min}}$, C_H, and C_d. Therefore, we have from assumption (437) that claim (S2.18) holds. Then, by Lemma 8, we have that for every $r \in [R]$,

$$\frac{1}{n}\sum_{l=1}^{2^r}\left(\frac{\int_{I_l^{(r)}}\lambda_a - \lambda_b d\nu}{\int_{I_l^{(r)}}\lambda d\nu}\right)^2 \int_{I_l^{(r)}}\lambda d\nu$$

$$\geq \frac{1}{n}\int_I\left(\frac{\lambda_a - \lambda_b}{\lambda}\right)^2 \lambda\, d\nu - \frac{C_H C_d}{2}2^{-\frac{2\gamma r}{q}}.$$

Thus, using (S2.18), we may conclude that there exists a $r \in [R]$ such that

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

From the hypothesis of the theorem, we also have that for all $l \in [2^R]$,

$$\int_{I_l^{(R)}}\lambda d\nu = n\frac{\int_{I_l^{(R)}}\lambda\,\nu}{\int_I \lambda\, d\nu} \geq n\frac{c_{\min}}{c_{\max}}\frac{\nu(I_l^{(R)})}{\nu(I)} \geq n\frac{c_{\min}}{c_{\max}}2^{-R} \geq 2,$$

63

where, in the final inequality, we use the assumption that $R \leq \log_2 \frac{n}{2} - \log_2 \frac{c_{\max}}{c_{\min}}$.

Then, from Theorem 4, it holds that $\mathbb{P}(p_F \leq \alpha) \geq 1 - 2\beta$ and the conclusion of the Theorem

follows as desired.

$\square$

S2.3   Proof of Theorem 6

Proof. (of Theorem 6)

The proof is similar to that of Theorem 4. Let $r^* \in [R]$ and $l^* \in [2^{r^*}]$ denote the resolution

level and bin such that

$$\frac{1}{4}\left(\frac{\int_{I_{l^*}^{(r^*)}} \lambda_a - \lambda_b d\nu}{2\int_{I_{l^*}^{(r^*)}} \lambda d\nu}\right)^2 \int_{I_{l^*}^{(r^*)}} \lambda d\nu \geq 2r^* + \frac{C^{1/2}}{\beta} + 2\log\frac{K}{\alpha}.$$

Define the event

$$\mathcal{E}_{r^*l^*} := \left\{(m_{l^*}^{(r^*)} - 1)\delta_{l^*}^{(r^*)2} \geq 2r^* + \frac{C^{1/2}}{\beta^{1/2}} + 2\log\frac{K}{\alpha}\right\}.$$

By Theorem S2, we have that

$$\mathbb{P}(\hat{p}_{\min} \geq \frac{\alpha}{K}) \leq \mathbb{P}(\hat{p}^{(k)} \geq \frac{\alpha}{K})$$

$$\leq \mathbb{P}\left(\left\{\hat{p}^{(k)} \geq \frac{\alpha}{K}\right\} \cap \mathcal{E}_{r^*l^*}\right) + \mathbb{P}(\mathcal{E}_{r^*l^*}^c)$$

$$\leq \beta + \mathbb{P}(\mathcal{E}_{r^*l^*}^c).$$

To bound $\mathbb{P}(\mathcal{E}_{r^*l^*}^c)$, we use our assumption that $\mathbb{E}m_{l^*}^{(r^*)} = \int_{I_{l^*}^{(r^*)}} \lambda d\nu \geq 2$ and the fact that

$\delta_{l^*}^{(r^*)} \leq 1$ to obtain

$$\mathbb{E}(m_{l^*}^{(r^*)} - 1)\delta_{l^*}^{(r^*)2} \geq \frac{1}{2}\left(\int_{I_{l^*}^{(r^*)}} \lambda d\nu\right)\delta_{l^*}^{(r^*)2} \qquad \text{and}$$

$$\mathrm{Var}(m_{l^*}^{(r^*)} - 1)\delta_{l^*}^{(r^*)2} \leq \left(\int_{I_{l^*}^{(r^*)}} \lambda d\nu\right)\delta_{l^*}^{(r^*)2}.$$

We have then

$$
\mathbb{P}(\mathcal{E}^c_{l^* r^*}) = \mathbb{P}\left((m^{(r^*)}_{l^*} - 1)\delta^{(r^*)2}_{l^*} \leq \frac{C^{1/2}}{\beta^{1/2}} + 2r^* + 2\log \frac{K}{\alpha} \right)
$$

$$
\leq \{\mathrm{Var}(m^{(r^*)}_{l^*} - 1)\delta^{(r^*)2}_{l^*}\}\left\{ \mathbb{E}(m^{(r^*)}_{l^*} - 1)\delta^{(r^*)2}_{l^*} - \left(2r^* + \frac{C^{1/2}}{\beta^{1/2}} + 2\log \frac{K}{\alpha} \right) \right\}^{-2}
$$

$$
\leq \left\{ \frac{1}{4}\left(\int_{I^{(r^*)}_{l^*}} \lambda d\nu \right) \delta^{(r^*)2}_{l^*} \right\}^{-1} \leq \beta.
$$

□

S2.4   Proof of Theorem 7

Proof. Let $r := \lceil \log_2 \frac{\nu(I)}{\nu(S)} \rceil$ so that

$$
\frac{\nu(I)}{2^{r-1}} \geq \nu(S) \geq \frac{\nu(I)}{2^r}.
$$

We observe that since $\frac{\nu(S)}{\nu(I)} \geq \frac{c_{\max}}{c_{\min}}\frac{8}{n}$ by assumption,

$$
r \leq \lceil \log_2 \frac{\nu(I)}{\nu(S)} \rceil < \lfloor \log_2 \frac{n}{2} - \log_2 \frac{c_{\max}}{c_{\min}} \rfloor = R.
$$

Hence, $\{I^{(r+1)}_l\}$ exists in our dyadic partitioning and there exists $l^* \in [2^{r+1}]$ such that the interval $I^{(r+1)}_{l^*} \subset S$. Let C be the universal constant specified in Theorem 6 and let $C_2 := \frac{32c_{\max}}{c_{\min}}C^{1/2}$. From (438), we have that

$$
\frac{1}{4} \max_{l \in [2^{r+1}]} \left(\frac{\int_{I^{(r+1)}_l} \lambda_a - \lambda_b d\nu}{\int_{I^{(r+1)}_l} \lambda d\nu} \right)^2 \int_{I^{(r+1)}_l} \lambda d\nu
$$

$$
\geq \frac{1}{4} \left(\frac{\int_{I^{(r+1)}_{l^*}} \frac{\lambda_a - \lambda_b}{\lambda} \lambda d\nu}{\int_{I^{(r+1)}_{l^*}} \lambda d\nu} \right)^2 \int_{I^{(r+1)}_{l^*}} \lambda d\nu
$$

$$
\geq \frac{1}{4}\delta^2_S \int_{I^{(r+1)}_{l^*}} \lambda d\nu \overset{(a)}{\geq} \frac{1}{4}\delta^2_S n \frac{c_{\min}}{c_{\max}} 2^{-(r+1)}
$$

$$
\geq \delta^2 n 2^{-(r-1)} \frac{c_{\min}}{16c_{\max}} \geq \delta^2_S n \frac{\nu(S)}{\nu(I)} \frac{c_{\min}}{16c_{\max}} \geq n\delta^2_S \frac{\nu(S)}{\nu(I)} \frac{2C^{1/2}}{C_2}
$$

$$
\geq 2\log n + \frac{C^{1/2}}{\beta} + 2\log \frac{1}{\alpha} \geq 2r + \frac{C^{1/2}}{\beta} + 2\log \frac{R}{\alpha},
$$

where inequality (a) follows from the fact that

$$
\int_{I^{(r+1)}_{l^*}} \lambda d\nu = \frac{\int_{I^{(r+1)}_{l^*}} \lambda d\nu}{\int_I \lambda d\nu} \geq n \frac{c_{min}}{c_{max}} 2^{-(r+1)}.
$$

The conclusion of the theorem follows from Theorem 6.

□

S2.5   Auxiliary results

Recall that, for a positive integer m, we define

$$S_{\text{Bin}(\frac{1}{2},m)}(t) := \mathbb{P}(|\text{Bin}(\frac{1}{2},m) - \frac{m}{2}| \geq t) \tag{S2.19}$$

$$S_{\chi^2_m}(t) := \mathbb{P}(\chi^2_m \geq t). \tag{S2.20}$$

Moreover, define

$$\mathcal{M}_m := \left\{ -\frac{m}{2}, -\frac{m}{2} + 1, \ldots, \frac{m}{2} - 1, \frac{m}{2} \right\} \tag{S2.21}$$

$$\mathcal{M}_m^+ := \begin{cases} \left\{ 0, 1, \ldots, \frac{m}{2} \right\} & \text{if } m \text{ is even,} \\[2mm] \left\{ \frac{1}{2}, \frac{3}{2}, \ldots, \frac{m}{2} \right\} & \text{if } m \text{ is odd.} \end{cases} \tag{S2.22}$$

Theorem S1. *Let d be a positive integer. For each $l \in [d]$, let $m_l \in \mathbb{N}$, $\delta_l \in [0, \frac{1}{2}]$, and let $A_1, \ldots, A_d$ be independent random variables where $A_l \sim Bin(\frac{1}{2} + \delta_l, m_l)$,*

Let $U_1, \ldots, U_d$ be independent random variables distributed uniform on $[0,1]$ and independent of $A_1, \ldots, A_d$. Define $p_l := U_l \cdot S_{Bin(\frac{1}{2},m_l)}(|A_l - \frac{m_l}{2}|) + (1 - U_l)S_{Bin(\frac{1}{2},m_l)}(|A_l - \frac{m_l}{2}| + 1)$, define the set $L := \{l \in [d] : m_l \geq 2\}$, and define $p := S_{\chi^2_{2|L|}}\left(\sum_{l \in L} -2\log p_l\right)$.

Then, there exists a universal constant $C > 0$ such that, for any $\alpha, \beta \in (0,1)$, if

$$\sum_{l \in L}(m_l - 1)\delta_l^2 \geq |L|^{1/2}\left(\frac{C^{1/2}}{\beta^{1/2}} + 2\log^{1/2}\frac{1}{\alpha}\right) + 2\log\frac{1}{\alpha}, \tag{S2.23}$$

then $\mathbb{P}(p \leq \alpha) \geq 1 - \beta$.

Proof. Define

$$L_1 := \{l \in L : (m_l - 1)\delta_l^2 \geq 2\} \quad \text{and} \quad L_2 := \{l \in L : (m_l - 1)\delta_l^2 < 2\}.$$

For simplicity of presentation, we write $Z_l := -2\log p_l$ and $\tilde{Z}_l := \frac{4}{m_l}\left(A_l - \frac{m_l}{2}\right)^2$ for $l \in L$. By Hoeffding's inequality, it holds that $S_{\text{Bin}(\frac{1}{2},m)}(t) \leq 2\exp\{-2\frac{t^2}{m}\}$. Therefore, we have that

$$Z_l = -2\log\left\{U_l \cdot S_{\text{Bin}(\frac{1}{2},m)}(|A_l - \frac{m_l}{2}|) + (1 - U_l)S_{\text{Bin}(\frac{1}{2},m)}\left(|A_l - \frac{m_l}{2}| + 1\right)\right\}$$

$$\geq -2\log\left\{S_{\text{Bin}(\frac{1}{2},m)}(|A_l - \frac{m_l}{2}|)\right\} \geq \frac{4}{m_l}\left(A_l - \frac{m_l}{2}\right)^2 - 2\log 2 = \tilde{Z}_l - 2\log 2.$$

66

By Lemma 9, we have

$$S_{\chi^2_{2|L|}}\left(2|L| + \sqrt{8}|L|^{1/2}\log^{1/2}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\right) \leq \alpha, \qquad (S2.24)$$

By (S2.24), the fact that $S_{\chi^2_{2|L|}}(\cdot)$ is monotone decreasing, and the fact that $2\log 2 \leq 2$, we have

$$
\begin{aligned}
\mathbb{P}(p \geq \alpha) &= \mathbb{P}\left\{S_{\chi^2_{2|L|}}\left(\sum_{l\in L} -2\log p_l\right) \geq \alpha\right\}\\
&\leq \mathbb{P}\left\{\sum_{l\in L} -2\log p_l \leq 2|L| + \sqrt{8}|L|^{\frac{1}{2}}\log^{\frac{1}{2}}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\right\}\\
&\leq \mathbb{P}\left\{\sum_{l\in L_1}(\tilde{Z}_l - 2\log 2) + \sum_{l\in L_2} Z_l \leq 2|L| + \sqrt{8}|L|^{\frac{1}{2}}\log^{\frac{1}{2}}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\right\}\\
&\leq \mathbb{P}\left\{\sum_{l\in L_1}(\tilde{Z}_l - \mathbb{E}\tilde{Z}_l) + \sum_{l\in L_2}(Z_l - \mathbb{E}Z_l) \leq \sum_{l\in L_1}(4 - \mathbb{E}\tilde{Z}_l)\right.\\
&\qquad\qquad\left. + \sum_{l\in L_2}(2 - \mathbb{E}Z_l) + 2|L|^{\frac{1}{2}}\log^{\frac{1}{2}}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\right\}. \quad (\star)
\end{aligned}
$$

We now observe that by Lemma 7,

$$
\begin{aligned}
&\sum_{l\in L_1}(4 - \mathbb{E}\tilde{Z}_l) + \sum_{l\in L_2}(2 - \mathbb{E}Z_l) + 2|L|^{1/2}\log^{1/2}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\\
&\leq -\sum_{l\in L} 2(m_l - 1)\delta_l^2 + 2|L|^{1/2}\log^{1/2}\frac{1}{\alpha} + 2\log\frac{1}{\alpha} \leq 0,
\end{aligned}
$$

where the final inequality follows by our assumption (S2.23). Therefore, returning to $(\star)$, we may apply Chebyshev's inequality to obtain

$$
\begin{aligned}
(\star) &\leq \frac{\sum_{l\in L_1}\mathrm{Var}\tilde{Z}_l + \sum_{l\in L_2}\mathrm{Var}Z_l}{\left\{-\sum_{l\in L}2(m_l-1)\delta_l^2 + 2|L|^{\frac{1}{2}}\log^{\frac{1}{2}}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\right\}^2}\\
&\leq Cd\left\{-2\sum_{l\in L}(m_l-1)\delta_l^2 + 2|L|^{\frac{1}{2}}\log^{\frac{1}{2}}\frac{1}{\alpha} + 2\log\frac{1}{\alpha}\right\}^{-2} \leq \beta.
\end{aligned}
$$

where, in the penultimate inequality, we used Lemmas 6 and where $C > 0$ is the universal constant specified in Lemma 6. The conclusion of the Theorem follows as desired.

$\square$

Theorem S2. *Let d be a positive integer. For each $l \in [d]$, let $m_l \in \mathbb{N}$, $\delta_l \in [0, 1/2]$, and let $A_1, \ldots, A_d$ be independent random variables where $A_l \sim Bin\left(\frac{1}{2} + \delta_l, m_l\right)$.*

Let $U_1, \ldots, U_d$ be independent random variables distributed uniform on $[0, 1]$ and independent of $A_1, \ldots, A_d$. Define $p_l := U_l \cdot S_{Bin(\frac{1}{2}, m_l)}(|A_l - \frac{m_l}{2}|) + (1 - U_l)S_{Bin(\frac{1}{2}, m_l)}(|A_l - \frac{m_l}{2}| + 1)$, define

the set $L := \{l \in [d] \ : \ m_l \geq 2\}$, *and define* $p_{min} := F_{Beta,|L|}\big(\min_{l \in L} p_l\big)$ *where* $F_{Beta,|L|}(x) :=$
$\mathbb{P}(Beta(1, |L| + 1) \leq x)$ *for any* $x \in \mathbb{R}$.

There exists universal constants $C > 0$ *such that for any* $\alpha, \beta \in (0, 1)$, *if*

$$\max_{l \in [d]} (m_l - 1)\delta_l^2 \geq \frac{C^{1/2}}{\beta^{1/2}} + 2\log\frac{|L|}{\alpha},$$

then $\mathbb{P}(p_{min} \leq \alpha) \geq 1 - \beta$.

Proof. Let C be the maximum of 4 and the universal constant specified in Lemma 6. By assumption, there exists $l^* \in [d]$ be such that

$$(m_{l^*} - 1)\delta_{l^*}^2 \geq \frac{C^{1/2}}{\beta^{1/2}} + 2\log\frac{|L|}{\alpha} \geq 2, \tag{S2.25}$$

where the last inequality follows since $\beta, \alpha \in (0, 1)$.

By Hoeffding's inequality, we have that

$$-2\log p_{l^*} \geq -2\log S_{\mathrm{Bin}(\frac{1}{2}, m_{l^*})}(|A_{l^*} - \frac{m_{l^*}}{2}|)$$
$$\geq \frac{4}{m_{l^*}}\left(A_{l^*} - \frac{m_{l^*}}{2}\right)^2 - 2\log 2.$$

We write $\tilde{Z}_{l^*} := \frac{4}{m_{l^*}}\big(A_{l^*} - \frac{m_{l^*}}{2}\big)^2$ so that $-2\log p_{l^*} \geq \tilde{Z}_{l^*} - 2\log 2 \geq \tilde{Z}_{l^*} - 4$.

For any $\alpha, \beta \in (0, 1)$, we may use the fact that $F_{\mathrm{Beta},|L|}(x) \leq |L|x$ to show that

$$\mathbb{P}(p_{\min} \geq \alpha) \leq \mathbb{P}\big(\min_{l \in L} p_l \geq \frac{\alpha}{|L|}\big) \leq \mathbb{P}\big(p_{l^*} \geq \frac{\alpha}{|L|}\big)$$
$$= \mathbb{P}\big(-2\log p_{l^*} \leq 2\log\frac{|L|}{\alpha}\big)$$
$$\leq \mathbb{P}\left(\tilde{Z}_{l^*} - \mathbb{E}\tilde{Z}_{l^*} \leq (4 - \mathbb{E}\tilde{Z}_{l^*}) + 2\log\frac{|L|}{\alpha}\right). \quad (\star)$$

By Lemma 7 and (S2.25), we have that

$$(4 - \mathbb{E}\tilde{Z}_{l^*}) + 2\log\frac{|L|}{\alpha} \leq -2(m_{l^*} - 1)\delta_{l^*}^2 + 2\log\frac{|L|}{\alpha} \leq 0.$$

Therefore, continuing on from $(\star)$, we have by Chebyshev inequality and Lemma 6 that

$$(\star) \leq \frac{\mathrm{Var}(\tilde{Z}_{l^*})}{\big\{-(m_{l^*} - 1)\delta_{l^*}^2 + 2\log\frac{|L|}{\alpha}\big\}^2} \leq C\left\{-(m_{l^*} - 1)\delta_{l^*}^2 + 2\log\frac{|L|}{\alpha}\right\}^{-2} \leq \beta.$$

The conclusion of the Theorem follows as desired.

$\square$

Lemma 6. *Let m be a positive integer and let $\delta \in [0, 1/2]$. Let $A \sim Bin(\frac{1}{2} + \delta, m)$ and let $U \sim$ Unif$[0, 1]$ be independent of A. Define $Z := -2\log\big\{U \cdot S_{Bin(\frac{1}{2}, m)}(|A - \frac{m}{2}|) + (1 - U)S_{Bin(\frac{1}{2}, m)}(|A - \frac{m}{2}| + 1)\big\}$ and $\tilde{Z} := \frac{4}{m}(A - \frac{m}{2})^2$.*

There exists a universal constant $C > 0$ such that

1. *if $(m - 1)\delta^2 < 2$, then $Var(Z) \leq C$,*

2. *and $Var(\tilde{Z}) \leq C$.*

Proof. First assume that $(m - 1)\delta^2 < 2$. By increasing the value of the universal constant C if necessary, we may assume without the loss of generality that $m \geq 17$.

Define $\mathcal{M}_m$ as (S2.21). Let P, Q be probability measures on $\mathcal{M}_m$ such that P is the distribution of $|A - \frac{m}{2}|$ and Q is the distribution of $|Bin(\frac{1}{2}, m) - \frac{m}{2}|$.

Then, let $\tilde{S}_0(\cdot)$ be defined as in Lemma 10, we have by the same lemma that

$$
\begin{aligned}
\mathrm{Var}\, Z &\leq \mathbb{E}Z^2 \\
&= \int_0^1 \sum_{s \in \mathcal{M}_m} 4 \log^2 \tilde{S}_0(|s| + u) \cdot \frac{P(s)}{Q(s)} Q(s)\, du \\
&\leq \underbrace{\left\{ \sum_{s \in \mathcal{M}_m} \left(\frac{P(s)}{Q(s)}\right)^2 Q(s) \right\}^{1/2}}_{\text{Term 1}} \tag{S2.26} \\
&\quad + \underbrace{\left\{ \int_0^1 \sum_{s \in \mathcal{M}_m} \left(4 \log^2 \{\tilde{S}_0(|s| + u)\}\right)^2 Q(s)\, du \right\}^{1/2}}_{\text{Term 2}}. \tag{S2.27}
\end{aligned}
$$

Term 2 of (S2.27) is equal to $16 \cdot \mathbb{E} \log^4 \tilde{S}_0(|Bin(\frac{1}{2}, m) - \frac{m}{2}| + U)$. Since the random variable $\tilde{S}_0(|Bin(\frac{1}{2}, m) - \frac{m}{2}| + U)$ is uniformly distributed on $[0, 1]$, we have that Term 2 is upper bounded by a universal constant.

For Term 1, we define $r := \frac{1 + 2\delta}{1 - 2\delta}$ and observe that for any $s \in \mathcal{M}_m$,

$$
\begin{aligned}
\frac{P(s)}{Q(s)} &= \frac{1}{2} 2^m \left\{ \left(\frac{1}{2} + \delta\right)^{\frac{m}{2} + s} \left(\frac{1}{2} - \delta\right)^{\frac{m}{2} - s} + \left(\frac{1}{2} + \delta\right)^{\frac{m}{2} - s} \left(\frac{1}{2} - \delta\right)^{\frac{m}{2} + s} \right\} \\
&= \frac{1}{2}(1 - 4\delta^2)^{\frac{m}{2}} (r^s + r^{-s}) \leq r^s.
\end{aligned}
$$

Since we assume $(m - 1)\delta^2 \leq 2$ and since we assume that $m \geq 17$, we have that $\delta^2 \leq \frac{1}{8}$ and

thus $0 \le \log r \le 8\delta$. Let W be a random variable distributed as $|\mathrm{Bin}(\frac{1}{2}, m) - \frac{m}{2}|$. Then

$$
\begin{aligned}
\sum_{s \in \mathcal{M}} r^{2s} Q(s) = \mathbb{E} r^{2W} &= \int_1^\infty \mathbb{P}(r^{2W} \ge t) \, dt \\
&= \int_1^\infty \mathbb{P}\left(W \ge \frac{\log t}{2 \log r}\right) dt \\
&\le \int_1^\infty \exp\left(-\frac{\log^2 t}{4m \log^2 r}\right) dt \\
&\le \int_1^\infty t^{-\frac{\log t}{2^{12}}} \, dt \le C,
\end{aligned}
$$

where $C > 0$ is a universal constant.

Now we turn to the second claim. Write $A = \sum_{i=1}^m \epsilon_i$ where $\epsilon_1, \dots, \epsilon_m$ are independent and identically distributed $\mathrm{Ber}(\frac{1}{2} + \delta)$ random variables.

For any $i \in [m]$, we have

$$
\mathrm{Var}._{\,|\, \{\epsilon_{-i}\}} \left[\left(A - \frac{m}{2}\right)^2\right] \le \sup_{z \in [-\frac{m}{2}, \frac{m}{2}]} \mathrm{Var}\left[(z + \epsilon_i)^2\right] \le m.
$$

Thus, by the Efron–Stein inequality,

$$
\begin{aligned}
\mathrm{Var}\tilde{Z} = \frac{16}{m^2} \mathrm{Var}\left[\left(A - \frac{m}{2}\right)^2\right] \\
\le \frac{16}{m^2} \mathbb{E} \sum_{i=1}^m \mathrm{Var}._{\,|\, \{\epsilon_{-i}\}}\left[\left(A - \frac{m}{2}\right)^2\right] \le 16.
\end{aligned}
$$

The conclusion of the lemma follows as desired.

$\square$

Lemma 7. *Let m be a positive integer and let $\delta \in [0, 1/2]$. Let $A \sim \mathrm{Bin}(m, \frac{1}{2} + \delta)$ and let $U \sim \mathrm{Unif}[0,1]$ be independent of A. Define $Z := -2 \log\{U \cdot S_{\mathrm{Bin}(\frac{1}{2}, m)}(|A - \frac{m}{2}|) + (1-U) S_{\mathrm{Bin}(\frac{1}{2}, m)}(|A - \frac{m}{2}| + 1)\}$ and $\tilde{Z} := \frac{4}{m}(A - \frac{m}{2})^2$.*

We have that

1. *$\mathbb{E} Z - 2 \ge 8(m-1)\delta^2$,*

2. *and if $(m-1)\delta^2 \ge 2$, then $\mathbb{E}\tilde{Z} - 4 \ge 2(m-1)\delta^2$.*

Proof. Define $\mathcal{M}_m$ as (S2.21) and note that $\mathcal{M}_m = -\mathcal{M}_m$. For $s \in \mathcal{M}_m$, write $P_m(s, \delta) = \binom{m}{\frac{m}{2}+s}(\frac{1}{2} + \delta)^{\frac{m}{2}+s}(\frac{1}{2} - \delta)^{\frac{m}{2}-s}$ as the probability that $\mathrm{Bin}(m, \frac{1}{2} + \delta)$ random variable is equal to $\frac{m}{2} + s$ and $Q_m(s) := \binom{m}{\frac{m}{2}+s}(\frac{1}{2})^m = P_m(s, 0)$.

Define $W = |A - \frac{m}{2}| + U$. We also define

$$
\begin{aligned}
F_m(\delta) = \mathbb{E}Z &= \mathbb{E}\big[-2\log \tilde{S}_0(W)\big] \\
&= \sum_{s \in \mathcal{M}_m} P_m(s, \delta) \int_0^1 \big\{-2\log \tilde{S}_0(|s| + u)\big\}\, du,
\end{aligned}
$$

where the definition of $\tilde{S}_0(\cdot)$ and the second equality follow from Lemma 10. We note then that

$$
\mathbb{E}Z - 2 \geq F_m(\delta) - F_m(0).
$$

Moreover, since the function $\delta \mapsto P_m(s, \delta)$ is equal to its Taylor series expansion for all $\delta \in (-\frac{1}{2}, \frac{1}{2})$, the same holds for $F_m(\delta)$, that is,

$$
F_m(\delta) = F_m(0) + \sum_{j=1}^{\infty} F_m^{(j)}(0)\frac{\delta^j}{j!}, \quad \text{for all } \delta \in (-\frac{1}{2}, \frac{1}{2}).
$$

By symmetry, $P_m(s, \delta) = P_m(-s, -\delta)$ and thus, $F_m(\delta) = F_m(-\delta)$ and $F_m^{(j)}(0) = 0$ when j is an odd integer. When j is an even integer, we have that, by Lemma 11,

$$
F_m^{(j)}(0) = \sum_{s\mathcal{M}_m} \left(\partial_\delta^{(j)} P_m(s, \delta)\big|_{\delta=0}\right) \int_0^1 \big\{-2\log \tilde{S}_0(|s| + u)\big\}\, du \geq 0.
$$

We now claim that $F_m^{(2)}(0) \geq 8(m-1)$. To see this, first observe that, by Hoeffding's inequality, it holds that $-2\log\big\{S_{\mathrm{Bin}(\frac{1}{2}, m)}(|s|)\big\} \geq \frac{4}{m}s^2 - 2\log 2$. Moreover, since $\sum_{s \in \mathcal{M}_m} P_m(s, \delta) = 1$, writing $P_m^{(2)}(s, \delta)$ as second derivative of $P_m(s, \delta)$ with respect to δ, we have $\sum_{s \in \mathcal{M}_m} P_m^{(2)}(s, \delta) = 0$. Thus, using the fact that $P_m^{(2)}(s, 0) \geq 0$ for all $s \in \mathcal{M}_m$ (by Lemma 11), we have that, for any $\delta \in (-1/2, 1/2)$,

$$
\begin{aligned}
F_m^{(2)}(\delta) &= \sum_{s \in \mathcal{M}_m} P_m^{(2)}(s, \delta) \int_0^1 \big\{-2\log \tilde{S}_0(|s| + u)\big\}\, du \\
&\geq \sum_{s \in \mathcal{M}_m} P_m^{(2)}(s, \delta)\big(-2\log\big\{S_{\mathrm{Bin}(\frac{1}{2}, m)}(|s|)\big\}\big) \\
&\geq \sum_{s \in \mathcal{M}_m} P_m^{(2)}(s, \delta)\frac{4}{m}s^2 \\
&\geq \frac{4}{m}\left(\frac{d}{d\delta}\right)^2 \underbrace{\left\{\sum_{s \in \mathcal{M}_m} P_m(s, \delta)s^2\right\}}_{\mathbb{E}(A - m/2)^2 \text{ where } A \sim \mathrm{Bin}(1/2 + \delta, m)} \\
&= \frac{4}{m}\left(\frac{d}{d\delta}\right)^2\left\{\frac{m}{4}(1 - 4\delta^2) + m^2\delta^2\right\} = 1 + 8(m-1).
\end{aligned}
$$

We may conclude then that $F_m^{(2)}(0) \geq 8(m-1)$ as desired. Therefore, we have that

$$\mathbb{E}Z - 2 = F_m(\delta) - F_m(0) \geq 8(m-1)\delta^2.$$

For the second claim of the Lemma, we note that $A \sim \mathrm{Bin}(\frac{1}{2} + \delta, m)$. Therefore, assuming $(m-1)\delta^2 \geq 2$, we have that

$$
\begin{aligned}
\mathbb{E}\tilde{Z} - 4 &= \frac{4}{m}\mathbb{E}\big(A - \frac{m}{2}\big)^2 - 4 \\
&= \frac{4}{m}\Big\{\mathbb{E}\big(A - \frac{m}{2} - m\delta\big)^2 + m^2\delta^2\Big\} - 4 \\
&= (1 - 4\delta^2) + 4m\delta^2 - 4 \geq 2m\delta^2
\end{aligned}
$$

as desired. The conclusion of the lemma thus follows.   □

Lemma 8. *Let $I \subset \mathbb{R}^q$ and let $I_1, \ldots, I_L$ be a partition of I such that $\mathrm{diam}(I_l) \leq C_d L^{-1/q}$ for all $l \in [L]$ for some $C_d > 0$. Write $\delta := \frac{\lambda_a - \lambda_b}{\lambda}$ and suppose that δ is γ-Holder continuous for $\gamma \in (0,1]$, i.e., $|\delta(x) - \delta(y)| \leq C_H\|x - y\|_2^\gamma$ for all $x, y \in I$, for some $C_H > 0$.*

Then, we have that

$$0 \leq \int_I \Big(\frac{\lambda_a - \lambda_b}{\lambda}\Big)^2 \lambda \, d\nu - \sum_{l=1}^{L}\left\{\Big(\frac{\int_{I_l} \lambda_a - \lambda_b \, d\nu}{\int_{I_l} \lambda \, d\nu}\Big)^2 \int_{I_l} \lambda \, d\nu\right\} \leq 2C_H C_d d^{-\frac{2\gamma}{q}}\Big(\int_I \lambda \, d\nu\Big).$$

Proof. Fix an arbitrary $l \in [L]$ and define $\mathbb{E}^{(l)}[\cdot]$ as expectation with respect to the probability measure with density $\frac{\lambda}{\int_{I_l} \lambda d\nu}$. We then have that

$$\frac{\int_{I_l} \big(\frac{\lambda_a - \lambda_b}{\lambda}\big)^2 \lambda \, d\nu}{\int_{I_l} \lambda \, d\nu} = \mathbb{E}^{(l)}[\delta^2] \geq \{\mathbb{E}^{(l)}\delta\}^2 = \left\{\frac{\int_{I_l} \lambda_a - \lambda_b d\nu}{\int_{I_l} \lambda d\nu}\right\}^2.$$

For the other direction, we observe that

$$
\begin{aligned}
&\frac{\int_{I_l} \big(\frac{\lambda_a - \lambda_b}{\lambda}\big)^2 \lambda \, d\nu}{\int_{I_l} \lambda \, d\nu} - \left\{\frac{\int_{I_l} \lambda_a - \lambda_b d\nu}{\int_{I_l} \lambda d\nu}\right\}^2 \\
&\qquad = \mathbb{E}^{(l)}[\delta^2] - \{\mathbb{E}^{(l)}\delta\}^2 = \mathrm{Var}^{(l)}(\delta) \\
&\qquad \overset{(a)}{=} \frac{1}{2}\mathbb{E}^{(l)}\big[(\delta(X) - \delta(Y))^2\big] \\
&\qquad \leq \frac{1}{2}C_H\mathbb{E}^{(l)}\|X - Y\|_2^{2\gamma} \leq \frac{1}{2}C_H \sup_{x,y \in I_l}\|x - y\|_2^{2\gamma} \overset{(b)}{\leq} \frac{C_H C_d}{2}L^{-2\gamma/q}.
\end{aligned}
$$

72

where in inequality (a), the random variables X, Y are independent and distributed with density $\frac{\lambda}{\int_{I_l} \lambda \, d\nu}$ and where in inequality (b), we use the assumption that $\mathrm{diam}(I_l) \leq C_d L^{-1/q}$.

In summary, we have that, for each $l \in [L]$,

$$0 \leq \int_{I_l} \left(\frac{\lambda_a - \lambda_b}{\lambda} \right)^2 \lambda \, d\nu - \left(\frac{\int_{I_l} \lambda_a - \lambda_b d\nu}{\int_{I_l} \lambda \, d\nu} \right)^2 \int_{I_l} \lambda \, d\nu \leq \frac{C_H C_d}{2} L^{-2\gamma/q} \int_{I_l} \lambda \, d\nu.$$

By summing over $l \in [L]$, the conclusion of the theorem follows as desired.

$\square$

S2.6   Technical lemmas

Lemma 9. *Let $X \sim \chi^2_{2k}$. Then, we have that for all $t > 0$,*

$$\mathbb{P}(X \geq 2k + 2\sqrt{2kt} + 2t) \leq e^{-t}.$$

Proof. If $X \sim \chi^2_{2k}$, then $\mathbb{E}X = 2k$ and X is also a $\mathrm{Gamma}(k, 2)$ random variable and hence its moment generation function is bounded by $\mathbb{E}e^{\lambda(X - \mathbb{E}X)} \leq \frac{4k\lambda^2}{2(1 - 2\lambda)}$ for all $\lambda \in (0, \frac{1}{c})$ by Boucheron et al. (2013, Section 2.4). Then, it holds by Boucheron et al. (2013, Theorem 2.3) that $\mathbb{P}(X - \mathbb{E}X \geq \sqrt{8kt} + ct) \leq e^{-t}$ for all $t > 0$. The Lemma immediately follows. $\square$

Lemma 10. *Let $m \in \mathbb{N}$ and let $\tilde{A} \sim Bin(\frac{1}{2}, m)$. Define $\tilde{W} = |\tilde{A} - \frac{m}{2}| + U$ where $U \sim Unif[0, 1]$ is independent of $\tilde{A}$. We let $\mathcal{M}^+_m$ be as defined in* (S2.22).

Write $\tilde{S}_0(z) := \mathbb{P}(\tilde{W} \geq z)$. We have that, for any $z \in \mathbb{R}$,

$$\tilde{S}_0(z) = \begin{cases} (1 - (z - k_1))S_{Bin(\frac{1}{2}, m)}(k_1) + (z - k_1)S_{Bin(\frac{1}{2}, m)}(k_1 + 1) & \text{if } z \in \left[\min \mathcal{M}^+_m, 1 + \max \mathcal{M}^+_m \right) \\ 1 & \text{if } z < \min \mathcal{M}^+_m \\ 0 & \text{if } z \geq 1 + \max \mathcal{M}^+_m, \end{cases}$$

(S2.28)

where in the first case, k_1 is defined as $k_1 := \max\{k \in \mathcal{M}^+_m : k \leq z\}$.

Moreover, we have that

$$\tilde{S}'_0(z) = \begin{cases} -\mathbb{P}\left(|\tilde{A} - \frac{m}{2}| = k_1 \right) & \text{if } z \in \left[\min \mathcal{M}^+_m, 1 + \max \mathcal{M}^+_m \right) \\ 0 & \text{else .} \end{cases}$$

Finally, let $A \sim Bin(\frac{1}{2}+\delta, m)$ and let $W = |A - \frac{m}{2}| + U$ where $U \sim Unif[0,1]$ is independent

of A, we have that

$$\tilde{S}_0(W) \stackrel{d}{=} (1-U)S_{Bin(\frac{1}{2}, m)}\left(|A - \frac{m}{2}|\right) + U \cdot S_{Bin(\frac{1}{2}, m)}\left(|A - \frac{m}{2}| + 1\right).$$

Proof. To establish the first claim, let $z \in \left[\min \mathcal{M}_m^+, 1 + \max \mathcal{M}_m^+\right)$ and let $k_1 := \max\{k \in \mathcal{M}_m^+ :$

$k \leq z\}$. Define the event

$$\mathcal{E}_{k_1} = \left\{|\tilde{A} - \frac{m}{2}| = k_1\right\}.$$

Then, we have that

$$\tilde{S}_0(z) = \mathbb{P}(\tilde{W} \geq z)$$

$$= \mathbb{P}(\mathcal{E}_{k_1})\mathbb{P}(\tilde{W} \geq z \,|\, \mathcal{E}_{k_1}) + \mathbb{P}\left(|\tilde{A} - \frac{m}{2}| > k_1\right).$$

$$= \left\{S_{Bin(\frac{1}{2}, m)}(k_1) - S_{Bin(\frac{1}{2}, m)}(k_1 + 1)\right\}(1 - (z - k_1)) + S_{Bin(\frac{1}{2}, m)}(k_1 + 1).$$

The first claim (S2.28) follows immediately.

The second claim follows by direct differentiation, and the third claim follows directly from

the first claim. The whole Lemma thus follows as desired. □

Lemma 11. *For $m \in \mathbb{N}$, $s \in \mathcal{M}_m$ (defined as (S2.21)), and $\delta \in (-\frac{1}{2}, \frac{1}{2})$, define $P_m(s, \delta) = \binom{m}{\frac{m}{2}+s}(\frac{1}{2}+\delta)^{\frac{m}{2}+s}(\frac{1}{2}-\delta)^{\frac{m}{2}-s}$. We then have that, for any integer $j \geq 1$,*

$$\frac{\partial^{(2j)}}{\partial \delta^{(2j)}} P_m(s, \delta)\Bigg|_{\delta=0} \geq 0.$$

Proof. First suppose $s \geq 0$. Since $|2\delta| < 1$, we have that,

$$P_m(s, \delta) = \binom{m}{\frac{m}{2}+s} 2^{-m} \left(\frac{1 + 2\delta}{1 - 2\delta}\right)^s$$

$$= \binom{m}{\frac{m}{2}+s} 2^{-m}(1 + 2\delta)^s \left(1 + \sum_{k=1}^{\infty} (2\delta)^k\right)^s.$$

It is thus clear that in Taylor series expansion of $\delta \mapsto P_m(s, \delta)$, all the coefficients are non-negative

and and thus, $\frac{\partial^{(2j)}}{\partial \delta^{(2j)}} P_m(s, \delta) \geq 0$.

If $s \leq 0$ on the other hand, the same claim follows by the fact that

$$P_m(-s, \delta) = P_m(s, -\delta).$$

The lemma thus immediately follows. □

S3. Supplementary material for Section 5

S3.1 Two sample test Type I error

We consider the following three intensities

1. $\lambda_a(x) = \lambda_b(x) = 40 \cdot \mathbf{1}_{[0,1]}(x)$

2. $\lambda_a(x) = \lambda_b(x) = 40 \cdot \big(\sin(2\pi x) + 1\big)$

3. $\lambda_a(x) = \lambda_b(x) = 40 \cdot \frac{x(1-x)^4}{\int_0^1 x(1-x)^4 dx} \mathbf{1}_{[0,1]}(x)$

The first function is uniform, while the other two are not, indicating the intensities changes on the support. Note that the third function is the scaled beta density function with parameters (2,5). For each of the three cases under the null hypothesis, we conduct 2000 simulations of two independent Poisson processes with the intensities functions given in the corresponding case and present the proportions of rejections out of all simulations based on the adjusted p-value of each test. We generate 500 bootstrap resamples of each of the 2000 pairs of Poisson processes conditional on the total number of observations of the pooled process $N = N_a + N_b$, and use the same resamples to derive adjusted p-values for all tests. We provide the percentage of rejections at level $\alpha = 0.05, 0.1$ and 0.25 for the five test procedures under 3 different intensities, the results are given in Table 4. We can see from the results that these five tests all attains the corresponding nominal levels, which is not a surprise due to the Monte Carlo Approximation of the exact rejection threshold.

S3.2   Empirical verification of Tracy-Widom approximation and bootstrap correction

Table 4: The empirical level (% of rejections) of different tests under the null

Test	case 1			case 2			case 3		
	5%	10%	25%	5%	10%	25%	5%	10%	25%
MF	4.9	9.8	25.7	5.5	8.9	23.7	5.1	10.2	24.5
MM	4.6	8.9	22.4	5.4	10.4	23.9	4.9	10.4	26.0
KN_1	4.8	9.6	23.5	5.1	9.5	24.2	6.1	11.3	27.2
KN_2	4.7	9.7	25.9	4.4	9.3	24.1	5.9	10.9	25.8
KS	5.1	9.9	24.9	4.5	9.1	24.6	6.2	11.4	25.9

S3.2   Empirical verification of Tracy-Widom approximation and bootstrap correction

To see how fast the largest eigenvalues converge to the limiting distribution, we consider two cases with the numbers of nodes $n = 300$ and $n = 1600$ respectively. For each case we simulate 1000 adjacency matrix A whose entries $\{A_{ij} : i \neq j \leq n\}$ are independent and identically distributed Poisson random variables with mean equals to 20. Then we plot the sample distribution of the test statistics, i.e., $n^{2/3}(\lambda_1(\tilde{A}) - 2)$ against the Tracy-Widom distribution, where $\tilde{A}$ is the empirically centered and scaled version of A.

We can see from the first two graphs in Figure 11 that when $n = 300$ the sample distribution deviates in location compared with the target distribution and when $n = 1600$ the location is corrected but there still is some difference in scale. Though there are some differences in location and scale, we can see the sample distribution does have similar shape with the Tracy-Widom distribution even when the number of nodes is as small as 300. In similar experiments where adjacency matrices have Bernoulli distributed entries, Bickel and Sarkar (2016) proposed to apply bootstrap correction to the largest eigenvalue, where they generate parametric bootstrap samples of the adjacency matrices and use the bootstrapped mean and variance of the largest eigenvalues to shift and scale the test statistics to have a better match with the Tracy-Widom distribution.

S3.2   Empirical verification of Tracy-Widom approximation and bootstrap correction

Here we adapted the same bootstrap correction technique to the eigenvalues of adjacency matrix with $n = 300$ nodes, where we generate 50 bootstrap samples for each sample adjacency matrix. We plot the empirical distribution of the test statistics after bootstrap correction as the third graph in Figure 11. We can see that even with just 50 bootstrap samples, the sample distributions of the test statistics looks much closer to the target distribution.
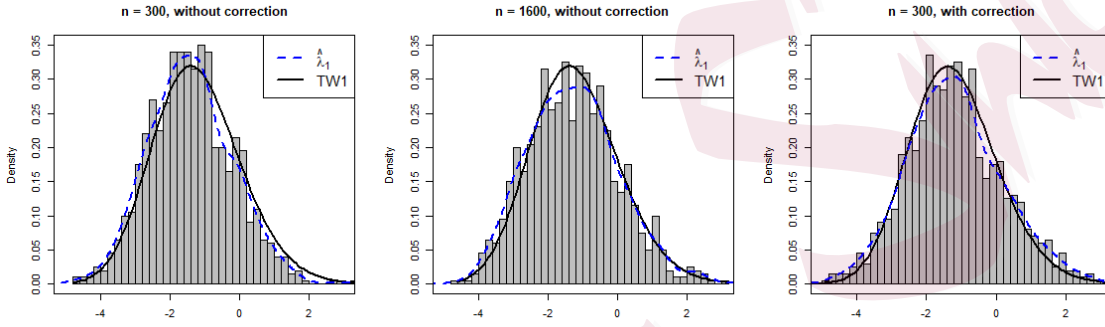

Figure 11: The empirical distribution of 1000 simulated samples of centered and scaled largest eigenvalues of $\tilde{A}$, compared with the Tracy-Widom distribution.

Remember that here we are using the Tracy-widom distribution to compute p-values for every local null hypotheses $\bar{H}_0^{(r,\ell)}$ and we generate bootstrap samples to estimate the exact critical threshold for the global null hypothesis $\bar{H}_0$, thus we could simply use the same bootstrap samples generated for testing the global null to correct the location and scale of the largest eigenvalue of each local adjacency matrix $\tilde{A}^{(r,\ell)}$. Given observed collection of Poisson process realizations $\{N_{uv}(\cdot) : u < v \in [n]\}$, we describe the procedure to derive the local p-values with bootstrap correction of the location and scale of the largest eigenvalue in the following steps:

1. For $b^* = 1, 2, \ldots, B$, generate bootstrap sample collections $\{N_{uv}^{b^*}(\cdot) : u < v \in [n]\}$ as described in Section 3.1.3.

2. For the observed realization, estimate the Poisson mean $\hat{\lambda}^{(r,\ell)}$ of each discretized interval as $\hat{\lambda}^{(r,\ell)} = \frac{1}{n^2-n} \sum_{u \neq v} N_{uv}^{(r,\ell)}$ and let $\tilde{A}^{(r,\ell)}$ be the centered and re-scaled adjacency matrix

77

S3.2 Empirical verification of Tracy-Widom approximation and bootstrap correction

for interval $I^{(r,\ell)}$

$$
\tilde{A}_{uv}^{(r,\ell)} := \begin{cases} \frac{N_{uv}^{(r,\ell)} - \hat{\lambda}^{(r,\ell)}}{\sqrt{(n-1)\hat{\lambda}^{(r,\ell)}}}, & u \neq v, \\[2mm] 0, & u = v. \end{cases}
$$

and let $\lambda_1(\tilde{A}^{(r,\ell)})$ be the largest eigenvalue of adjacency matrix $\tilde{A}^{(r,\ell)}$.

3. Do step 2 for every bootstrap resamples to derive their largest eigenvalues $\lambda_1(\tilde{A}^{(r,\ell,b^*)})$ at every discretized interval. Then we calculate the sample mean and standard deviation of $\left\{\lambda_1(\tilde{A}^{(r,\ell,b^*)}) : b^* \in \{1,2,\ldots,B\}\right\}$ for each $r \in [R], \ell \in [2^r]$ and denote them as $\hat{\mu}_1^{(r,\ell)}, \hat{s}_1^{(r,\ell)}$ respectively.

4. Denote μ_{tw} and s_{tw} as the mean and standard deviation of Tracy-Widom distribution with $\beta = 1$ and let

$$
\lambda_{bc}^{(r,\ell)} = \mu_{\text{tw}} + s_{\text{tw}} \frac{\lambda_1(\tilde{A}^{(r,\ell)}) - \hat{\mu}_1^{(r,\ell)}}{\hat{s}_1^{(r,\ell)}}
$$

be the test statistic after bootstrap correction.

5. Finally we compute the p-value for the discretized local null $\bar{H}_)^{(r,\ell)}$ as

$$
p^{(r,\ell)} \equiv p^{(r,\ell)}\big(\lambda_{bc}^{(r,\ell)}\big) := 2\min\Big(F_{\text{TW1}}\Big(\lambda_{bc}^{(r,\ell)}\Big), 1 - F_{\text{TW1}}\Big(\lambda_{bc}^{(r,\ell)}\Big)\Big)
$$