Proofs of Statements and Numerical Experiments

EC.1. Upper bound results

In this section, we provide the Lemmas that control the type I and II errors of the tests for problems (P1) and (P2).

EC.1.1. Testing at a given point

LEMMA EC.1. *For any $\alpha \in (0,1)$ the type I error of test (9) is bounded by α.*

Proof. We have $\alpha(\psi^{\tau}_{n,T}) = \sup_{(\Theta_0,\Pi)\in S_n(\omega_n)} \mathrm{P}_{(\Theta_0,\Pi)}\Big\{\|Z_T(\tau)\|_{2\to 2} > H_{\alpha,n}\Big\}$. Recall that $Z_T(t) = -\mu_T(t)\Pi\odot\Delta\Theta^{\tau} + \xi(t)$, where

$$\xi(t) = \sqrt{\frac{t(T-t)}{T}}\left(\frac{1}{t}\sum_{s=1}^{t} W^s - \frac{1}{T-t}\sum_{s=t+1}^{T} W^s\right)$$

and that under the null hypothesis $\Delta\Theta^{\tau} = 0$ and $W^s = (W^s_{ij}) \in [-1,1]^{n\times n}$ are independent centered Bernoulli matrices with independent entries taking values in $\{1-\Pi_{ij}\Theta^0_{ij}, -\Pi_{ij}\Theta^0_{ij}\}$ with the success probability $\Pi_{ij}\Theta^0_{ij}$. Consequently, applying Lemma EC.9 with $\|\Pi\odot\Theta^0\|_{1,\infty} \le \omega_n$ to bound

$$\alpha(\psi^{\tau}_{n,T}) = \sup_{(\Theta_0,\Pi)\in S_n(\omega_n)} \mathrm{P}_{(\Theta_0,\Pi)}\Big\{\|\xi(\tau)\|_{2\to 2} > 2\sqrt{2}(1+\epsilon)\sqrt{\omega}_n + C_\epsilon \log\Big(\frac{2n}{\alpha}\Big)\Big\}$$

we get $\alpha(\psi_{n,T}) \le \alpha$. □

LEMMA EC.2. *Let $\alpha,\beta \in (0,1)$ and $H_{\alpha,n}$ be given by* (10). *Suppose that*

$$\mathcal{R}_{n,\tau} \ge 4\sqrt{2}(1+\epsilon)\sqrt{\frac{\omega_n}{T}} + \frac{C_\epsilon}{\sqrt{T}}\Big(\log\Big(\frac{2n}{\alpha}\Big) + \log\Big(\frac{2n}{\beta}\Big)\Big). \tag{EC.1}$$

Then the type II error of test (9) is bounded by β.

Proof. For ease of notation we denote

$$(\boldsymbol{\Theta}, \Delta\boldsymbol{\Theta}^{\tau}) = \{(\Theta + \Delta\Theta^{\tau}, \Pi), (\Theta, \Pi)\}\,.$$

By definition, the type II error is $\beta(\psi^{\tau}_{n,T}, \mathcal{R}_{n,\tau}) = \sup_{(\boldsymbol{\Theta},\Delta\boldsymbol{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}(\omega_n,\mathcal{R}_{n,\tau})} \mathrm{P}_{(\boldsymbol{\Theta},\Delta\boldsymbol{\Theta}^{\tau})}\left\{\|Z_T(\tau)\|_{2\to 2} \le H_{\alpha,n}\right\}$. Using the triangle inequality, we can show that

$$\mathrm{P}_{(\boldsymbol{\Theta},\Delta\boldsymbol{\Theta}^{\tau})}\left\{\|Z_T(\tau)\|_{2\to 2} \le H_{\alpha,n}\right\} \le \mathrm{P}_{(\boldsymbol{\Theta},\Delta\boldsymbol{\Theta}^{\tau})}\Big\{\|\xi(\tau)\|_{2\to 2} \ge \mu_T(\tau)\|\Delta\Theta^{\tau}\|_{2\to 2} - H_{\alpha,n}\Big\}.$$

The choice of $H_{\alpha,n}$ and the fact that $\mu_T(\tau)=\sqrt{T}q(\tau/T)$ imply that for any $(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)\in\mathcal{W}^\tau_{n,T}(\omega_n,\mathcal{R}_{n,\tau})$ the above probability is bounded from above by

$$\mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)}\left\{\|\xi(\tau)\|_{2\to2}\geq\sqrt{T}\mathcal{R}_{n,\tau}-2\sqrt{2}(1+\epsilon)\sqrt{\omega_n}-C_\epsilon\log\left(\frac{2n}{\alpha}\right)\right\}.$$

For $\mathcal{R}_{n,\tau}$ satisfying (EC.1), this bound implies

$$\beta(\psi_{n,T},\mathcal{R}_{n,\tau})\leq\mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)}\left\{\|\xi(\tau)\|_{2\to2}>2\sqrt{2}(1+\epsilon)\sqrt{\omega_n}+C_\epsilon\log\left(\frac{2n}{\beta}\right)\right\}.$$

Applying Lemma EC.9 and using $\|\Pi\odot\Theta^0\|_{1,\infty}\leq\omega_n$, $\|\Pi\odot(\Theta^0+\Delta\Theta^\tau)\|_{1,\infty}\leq\omega_n$ we obtain the bound $\beta(\psi^\tau_{n,T})\leq\beta$. □

EC.1.2. Testing at an unknown change-point

LEMMA EC.3. *For any $\alpha\in(0,1)$ the type I error of test (13) is less than α.*

Proof. Using the union bound we can bound the type I error

$$\alpha(\psi_{n,T})=\mathrm{P}_{\mathrm{H}_0}\left\{\max_{t\in\mathcal{T}}\|Z^Y_T(t)\|_{2\to2}>H_{\alpha,n,T}\right\}$$

as

$$\alpha(\psi_{n,T})\leq\sum_{t\in\mathcal{T}}\sup_{(\Theta^0,\Pi)\in\mathcal{S}_n(\omega_n)}\mathrm{P}_{(\Theta^0,\Pi)}\left\{\|Z^Y_T(t)\|_{2\to2}>H_{\alpha,n,T}\right\},$$

where

$$\mathrm{P}_{(\Theta^0,\Pi)}\left\{\|Z^Y_T(t)\|_{2\to2}>H_{\alpha,n,T}\right\}=\mathrm{P}_{(\Theta^0,\Pi)}\left\{\|\xi(t)\|_{2\to2}>2\sqrt{2}(1+\epsilon)\sqrt{\omega_n}+C_\epsilon\log\left(\frac{4n\log_2(T)}{\alpha}\right)\right\}.$$

Using Lemma EC.9 with $\|\Pi_n\odot\Theta^0\|_{1,\infty}\leq\omega_n$, we immediately get for every $t\in\mathcal{T}$

$$\sup_{(\Theta^0,\Pi)\in\mathcal{S}_n(\omega_n)}\mathrm{P}_{(\Theta^0,\Pi)}\left\{\|\xi(t)\|_{2\to2}>2\sqrt{2}(1+\epsilon)\omega_n^{1/2}+C_\epsilon\log\left(\frac{4n\log_2(T)}{\alpha}\right)\right\}\leq\frac{\alpha}{2\log_2(T)}.$$

Using the fact that $|\mathcal{T}|\leq2\log_2(T)$, we obtain $\alpha(\psi_{n,T})\leq\alpha$. □

LEMMA EC.4. *Let $\alpha,\beta\in(0,1)$ and $H_{\alpha,n,T}$ be given by* (14). *Suppose that for some $\epsilon\in(0,1/2]$*

$$\mathcal{R}_{n,\mathcal{D}_T}\geq4\sqrt{6}(1+\epsilon)\left(\frac{\omega_n}{T}\right)^{1/2}+\frac{\sqrt{3}C_\epsilon}{\sqrt{T}}\left(\log\frac{4n\log_2(T)}{\alpha}+\log\frac{2n}{\beta}\right).$$

Then, the type II error of the test $\psi_{n,T}$ is bounded by β.

Proof. For ease of notation we denote $W^{\tau}_{n,T} := W^{\tau}_{n,T}(\omega_n, \mathcal{R}_{n,\mathcal{D}_T})$ and

$$(\mathbf{\Theta}, \Delta\mathbf{\Theta}^{\tau}) = \{(\Theta + \Delta\Theta^{\tau}, \Pi), (\Theta, \Pi)\}$$

The type II error

$$\beta(\psi_{n,T}, \mathcal{R}_{n,\mathcal{D}_T}) = \sup_{\tau\in\mathcal{D}_T} \sup_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})}\left\{\max_{t\in\mathcal{T}} \|Z^Y_T(t)\|_{2\to2} \le H_{\alpha,n,T}\right\}$$

can be bounded from above by $\inf_{t\in\mathcal{T}} \sup_{\tau\in\mathcal{D}_T} \sup_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})}\left\{\|Z^Y_T(t)\|_{2\to2} \le H_{\alpha,n,T}\right\}$. Applying the triangle inequality as it was done in Lemma EC.2, we get the bound

$$\beta(\psi_{n,T}, \mathcal{R}_{n,\mathcal{D}_T}) \le \inf_{t\in\mathcal{T}} \sup_{\substack{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}\\ \tau\in\mathcal{D}_T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})}\left\{\|\xi(t)\|_{2\to2} > \mu^{\tau}_T(t)\|\Pi\odot\Delta\Theta^{\tau}\|_{2\to2} - H_{\alpha,n,T}\right\}.$$

If $\tau \le T/2$, there exists a $t^* \in \mathcal{T}^L$ such that $\tau/2 \le t^* < \tau$. It is easy to see that

$$\mu^{\tau}_T(t^*) = \sqrt{\frac{t^*(T-\tau)}{(T-t^*)\tau}}\sqrt{\frac{\tau(T-\tau)}{T}} \ge \sqrt{\frac{\tau/2(T-\tau)}{(T-\tau/2)\tau}}\sqrt{\frac{\tau(T-\tau)}{T}} \ge \frac{1}{\sqrt{3}}\sqrt{T}q(\tau/T),$$

since $\tau < T/2$ iff $(T-\tau)/(2T-\tau) > 1/3$. If $\tau \ge T/2$, noting that $\mu^{\tau}_T(t) = \mu^{T-\tau}_T(T-t)$, we can reduce the estimation of $\mu^{\tau}_T(t)$ to the previous case: there exists $T-t' \in \mathcal{T}^R$ such that $(T-\tau)/2 < T-t' < T-\tau$ and $\mu^{\tau}_T(t') = \mu^{T-\tau}_T(T-t') \ge \frac{1}{\sqrt{3}}\sqrt{T}q(\tau/T)$. Thus, the type II error can be bounded as

$$\beta(\psi_{n,T}, \mathcal{R}_{n,\mathcal{D}_T}) \le \sup_{\substack{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}\\ \tau\in\mathcal{D}_T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})}\left\{\|\xi(t^*)\|_{2\to2} > \mu^{\tau}_T(t^*)\|\Pi\odot\Delta\Theta^{\tau}\|_{2\to2} - H_{\alpha,n,T}\right\}.$$

Since $\mu^{\tau}_T(t^*) > 1\sqrt{3}\sqrt{T}q(\tau/T)$ and $\|\Pi\odot\Delta\Theta^{\tau}\| \ge \mathcal{R}_{n,\mathcal{D}_T}$, we get

$$\beta(\psi_{n,T}, \mathcal{R}_{n,\mathcal{D}_T}) \le \sup_{\substack{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}\\ \tau\in\mathcal{D}_T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})}\left\{\|\xi(t^*)\|_{2\to2} > \frac{1}{\sqrt{3}}\sqrt{T}q(\tau/T)\mathcal{R}_{n,\mathcal{D}_T} - H_{\alpha,n,T}\right\}.$$

Using the definition of the threshold $H_{\alpha,n,T}$ and Lemma EC.9 together with $\|\Pi\odot\Theta\|_{1,\infty} \le \omega_n$ and $\|\Pi\odot(\Theta+\Delta\Theta)\|_{1,\infty} \le \omega_n$, we get

$$\beta(\psi_{n,T}, R_{n,\mathcal{D}_T}) \le \sup_{\substack{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})\in\mathcal{W}^{\tau}_{n,T}\\ \tau\in\mathcal{D}_T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^{\tau})}\left\{\|\xi(t^*)\|_{2\to2} > 2\sqrt{2}(1+\epsilon)\omega_n^{1/2} + C_{\epsilon}\log\frac{2n}{\beta}\right\} \le \beta. \quad \square$$

EC.2. Lower bound results

EC.2.1. General idea of the lower bound construction

Let $Y = (Y_1, \dots, Y_T)$ be the observations of the dynamic network following the inhomogeneous random graph model defined on a probability space $(\Omega, \mathcal{A}, \mathrm{P})$. Using (8), we can see that a lower bound on the type II error can be obtained by bounding from above the total variation distance between the measures of Y under the null and the alternantive hypotheses. The total variation distance is usually hard to bound and we can use instead the chi-squared or the Kullback–Leibler divergences as $\|\mathrm{P}_1 - \mathrm{P}_0\|_{\mathrm{TV}} \le \sqrt{\chi^2(\mathrm{P}_1, \mathrm{P}_0)}$ and $\|\mathrm{P}_1 - \mathrm{P}_0\|_{\mathrm{TV}} \le \sqrt{2\mathrm{KL}(\mathrm{P}_1, \mathrm{P}_0)}$. Thus, the problem of bounding the TV-distance is reduced to the problem of bounding one of these two divergences. This is can be done using *the second moment method* or *the fuzzy hypotheses method* as follows.

Let $\pi_{n,0}$ and $\pi_{n,1}$ be some prior distributions on the set of parameters $(\mathbf{\Theta}, \Delta\mathbf{\Theta}^\tau) = \{(\Theta + \Delta\Theta^\tau, \Pi), (\Theta, \Pi)\}$ of the network under H_0 and H_1, respectively. Define the mixture distributions $\mathrm{p}^T_{n,0}(Y) = \mathrm{E}_{\pi^T_{n,0}}\mathrm{P}(Y)$ and $\mathrm{p}^T_{n,1}(Y) = \mathrm{E}_{\pi^T_{n,1}}\mathrm{P}(Y)$, where P is the probability measure of Y. The expectations w.r.t. to the measures $\mathrm{p}^T_{n,i}$ are denoted by $\mathrm{E}^T_{n,i}$, $i = 0, 1$. For the type II error of testing at a given change-point τ, $\beta(\psi^\tau_{n,T}, \mathcal{R}_{n,\tau}) = \sup\limits_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)\in\mathcal{W}^\tau_{n,T}(\omega_n,\mathcal{R}_{n,\tau})} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)}\left\{\psi^\tau_{n,T} = 0\right\}$, the following bounds hold true (see, for example, (Ingster and Suslina 2003)):

$$
\begin{aligned}
\inf_{\psi^\tau_{n,T}\in\Psi_\alpha} \beta(\psi^\tau_{n,T}, \mathcal{R}_{n,\tau}) &\ge 1 - \frac{1}{2}\|\mathrm{p}^T_{n,1} - \mathrm{p}^T_{n,0}\|_{\mathrm{TV}} - \alpha \\
&\ge 1 - \frac{1}{2}\sqrt{\chi^2(\mathrm{p}_{n,1}, \mathrm{p}_{n,0})} - \alpha \\
&= 1 - \frac{1}{2}\left(\mathrm{E}^T_{n,0}\left[\frac{d\mathrm{p}^T_{n,1}}{d\mathrm{p}^T_{n,0}}(Y)\right]^2 - 1\right)^{1/2} - \alpha.
\end{aligned}
\tag{EC.2}
$$

Let $\alpha \in (0, 1)$ and $\beta \in (0, 1 - \alpha]$. Set $\eta = \alpha + \beta$. To establish a non-asymptotic lower bound $\inf\limits_{\psi^\tau_{n,T}\in\Psi_\alpha} \beta(\psi_{n,T}, \mathcal{R}_{n,\tau}) \ge \beta$ and the corresponding (α, β)-minimax detection rate, we need to find the conditions on $\mathcal{R}_{n,\tau}$ such that

$$
\mathrm{E}^T_{n,0}\left[\frac{d\mathrm{p}^T_{n,1}}{d\mathrm{p}^T_{n,0}}(Y)\right]^2 \le 1 + 4(1 - \alpha - \beta)^2 = 1 + 4(1 - \eta)^2.
$$

In the case of problem (P2) of unknown change-point $\tau \in \mathcal{D}_T$, bounding the type II error can be reduced to the case of a given change-point location provided in (EC.2):

$$\begin{aligned}
&\inf_{\psi_{n,T}\in\Psi_\alpha} \beta(\psi_{n,T}, \mathcal{R}_{n,\mathcal{D}_T}) \\
&= \inf_{\psi_{n,T}\in\Psi_\alpha} \sup_{\substack{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)\in\mathcal{W}^\tau_{n,T}(\omega_n,\mathcal{R}_{n,\mathcal{D}_T})\\ \tau\in\mathcal{D}_T}} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)}\left\{\psi_{n,T}=0\right\} \\
&\geq \inf_{\psi_{n,T}\in\Psi_\alpha} \sup_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)\in\mathcal{W}^{\tau^*}_{n,T}(\omega_n,\mathcal{R}_{n,\tau^*})} \mathrm{P}_{(\mathbf{\Theta},\Delta\mathbf{\Theta}^\tau)}\left\{\psi_{n,T}=0\right\},
\end{aligned}$$

where $\tau^* \in \mathcal{D}_T$ is any possible change-point from the set of alternatives. Thus, we can reduce the construction of the lower bound for the case of an unknown change-point to the case of a given change-point τ^*.

EC.2.2. Auxiliary lemma

Let $\rho_n \in (0, 1/2]$ and $q \in [-1, 1]$. Denote by p_0 and p_q the Bernoulli measures with the parameters ρ_n and $\rho_n(1+q)$ with the corresponding densities $d\mathrm{p}_0$ and $d\mathrm{p}_q$ with respect to some dominating measure λ. The following simple formulas will be useful in the proof of the lower bound.

LEMMA EC.5. *Let $\rho_n \in (0, 1/2]$, $q, q_1, q_2 \in [-1, 1]$. The following relations hold true for a Bernoulli variable $X \sim \mathrm{p}_0$:*

$$\begin{aligned}
\mathrm{E}_0\left[\frac{d\mathrm{p}_q}{d\mathrm{p}_0}\right]^2(X) &= 1 + \frac{\rho_n}{1-\rho_n}q^2, \\
\mathrm{E}_0\left[\frac{d\mathrm{p}_{q_1}}{d\mathrm{p}_0}\frac{d\mathrm{p}_{q_2}}{d\mathrm{p}_0}\right](X) &= 1 + \frac{\rho_n}{1-\rho_n}q_1 q_2.
\end{aligned}$$

EC.2.3. Lower bound for Inhomogeneous Random Graph Model

We will establish the lower bound for the case of the known change-point location τ. Let Π_n be the sampling matrix with the unit entries on the diagonal, $\mathrm{diag}(\Pi_n) = \mathbf{1}_n$ and with non-zero entries, $\min_{ij} \Pi_{ij} > 0$. In case of $\min_{ij} \Pi_{ij} = 1$ there is no missing links. Recall that $\alpha \in (0, 1)$, $\beta \in (0, 1-\alpha]$ and $\eta = \alpha + \beta$.

Proof of Theorem 2. In what follows we denote by δ_x the Dirac measure concentrated at x, where x can be a real or a matrix value. Denote by $\mathrm{P}(Y) = \prod_{t=1}^{T} \mathrm{P}(Y^t)$ the measure of the observations $Y = (Y^1, \dots, Y^T)$ from (1).

Denote by $\tilde{\Theta}^t = \Pi_n \odot \Theta^t$ the parameter of the observed adjacency matrix. We will impose the following priors on the matrix parameters $\tilde{\Theta}^t$ of the dynamic network $Y = (Y^1, \dots, Y^T)$.

Step 1. Priors on the transition matrices.

Set $\tilde{\rho}_n = (1-\varepsilon_n)\dfrac{\omega_n}{n-1}$ for some $\varepsilon_n \in (0,1)$ that will be chosen later. Assume that under the null hypothesis H_0 all the observed connections occur independently with the same probability $\tilde{\rho}_n$ for all $1 \le t \le T$. Set $V_0 = \tilde{\rho}_n(\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} - \mathrm{id}_n)$ and define the prior under H_0 on the sequence of the sampled connection probability matrices $\tilde{\Theta}^t$ $(1 \le t \le T)$:

$$\pi_{n,0}^T(\tilde{\Theta}^1, \dots, \tilde{\Theta}^T) = \prod_{t=1}^{T} \delta_{V_0}(\tilde{\Theta}^t) = \prod_{t=1}^{T} \prod_{i \neq j} \delta_{\tilde{\rho}_n}(\tilde{\Theta}^t_{ij}).$$

Here δ_{V_0} stands for the Dirac measure concentrated at V_0 and defined on the set of matrices $\Pi_n \odot \Theta$ such that $(\Theta, \Pi_n) \in \mathcal{S}_n(\omega_n)$. The prior is indeed concentrated on $S_n(\omega_n)$, since $\|V_0\|_{1,\infty} = (n-1)\tilde{\rho}_n = (1-\varepsilon)\omega_n < \omega_n$.

Let us define the prior under H_1. Let $\zeta = (\zeta_1, \dots, \zeta_n)$ be a vector of i.i.d. Rademacher random variables taking values in $\{-1, 1\}$ with probability 1/2. Assume that the sampled connection probability matrices before and after the change are defined by

$$V_{1,\zeta} = V_0 - \left(1 - \frac{\tau}{T}\right)\Lambda_{n,\zeta}, \quad V_{2,\zeta} = V_0 + \frac{\tau}{T}\Lambda_{n,\zeta},$$

where $\Lambda_{n,\zeta} = \dfrac{r_{n,\tau}}{n-1}(\zeta\zeta^{\mathrm{T}} - \mathrm{id}_n)$ is the change matrix with $r_{n,\tau} = \frac{\mathcal{R}_{n,\tau}}{q(\tau/T)}$. Note that the operator norm of $\Lambda_{n,\zeta}$ is equal to $r_{n,\tau}$ and the energy of the change-point is $q(\tau/T)\|\Lambda_{n,\zeta}\|_{2\to 2} = \mathcal{R}_{n,\tau}$.

To define the prior concentrated on $\mathcal{W}(\omega_n, \mathcal{R}_{n,\tau})$, we need to show that $\|V_{i,\zeta}\|_{1,\infty} \le \omega_n$, $i = 1, 2$ for sufficiently large n. We have that for all $n \ge 2$ and for $i = 1, 2$,

$$\begin{aligned}
\|V_{i,\zeta}\|_{1,\infty} &\le \|V_0\|_{1,\infty} + \frac{\tau \vee (T-\tau)}{T}\|\Lambda_{n,\zeta}\|_{1,\infty} \\
&= (1-\varepsilon_n)\omega_n + r_{n,\tau}\frac{\tau \vee (T-\tau)}{T} \\
&= (1-\varepsilon_n)\omega_n + q\left(\frac{\tau}{T}\right) r_{n,\tau}\sqrt{T-1} \\
&\le (1-\varepsilon_n)\omega_n + \mathcal{R}_{n,\tau}\sqrt{T}.
\end{aligned}$$

Let $\varepsilon_n = (2C_\eta)^{1/4}\omega_n^{-1/2}$, then for all $\omega_n > \sqrt{2C_\eta}$ we have $\varepsilon_n \in (0,1)$. It will be shown later in (EC.5) that $\mathcal{R}_{n,\tau}\sqrt{T} \le \varepsilon_n\omega_n$ and, consequently, $\|V_{i,\zeta}\|_{1,\infty} \le \omega_n$. Thus, the prior under H_1 is well defined and is given by

$$\pi_{n,1}^{\tau}(\tilde{\Theta}^1,\dots,\tilde{\Theta}^T) = \prod_{t=1}^{\tau}\delta_{V_{1,\zeta}}(\tilde{\Theta}^t)\prod_{t=\tau+1}^{T}\delta_{V_{2,\zeta}}(\tilde{\Theta}^t).$$

Step 2. Likelihood ratio of mixtures.

To shorten the notation, denote

$$q_{1,\tau} := -\Big(1-\frac{\tau}{T}\Big)\frac{r_{n,\tau}}{\tilde{\rho}_n(n-1)}, \quad q_{2,\tau} := \frac{\tau}{T}\frac{r_{n,\tau}}{\tilde{\rho}_n(n-1)}.$$

Let p_0 be the Bernoulli measure with the parameter $\tilde{\rho}_n$ and p_q denote the Bernoulli measure with the parameter $\tilde{\rho}_n(1+q)$, as in Lemma EC.5, Section EC.2.2. We can now calculate the mixtures under H_0 that are given by

$$\mathrm{p}_{n,0}(Y) = \mathrm{E}_{\pi_{n,0}^{T}}\mathrm{P}(Y) = \prod_{i>j}\prod_{t=1}^{T}\mathrm{p}_0(Y_{ij}^t) = \prod_{i>j}\prod_{t=1}^{T}\tilde{\rho}_n^{Y_{ij}^t}(1-\tilde{\rho}_n)^{1-Y_{ij}^t}$$

and under H_1, that are given by

$$\mathrm{p}_{n,1}^{\tau}(Y) = \mathrm{E}_{\pi_{n,1}^{T}}\mathrm{P}(Y) = \mathrm{E}_\zeta\Big[\prod_{i>j}\Big(\prod_{t=1}^{\tau}p_{q_{1,\tau}\zeta_i\zeta_j}(Y_{ij}^t)\prod_{i>j}\prod_{t=\tau+1}^{T}p_{q_{2,\tau}\zeta_i\zeta_j}(Y_{ij}^t)\Big)\Big]$$

where E_ζ stands for the expectation w.r.t. to the distribution of ζ and

$$p_{q_{k,\tau}\zeta_i\zeta_j}(Y_{ij}^t) = (\tilde{\rho}_n + \tilde{\rho}_n q_{k,\tau}\zeta_i\zeta_j)^{Y_{ij}^t}(1-\tilde{\rho}_n-\tilde{\rho}_n q_{k,\tau}\zeta_i\zeta_j)^{1-Y_{ij}^t}, \ k=1,2.$$

Denote by $\mathcal{Z} = \{-1,+1\}^n$ the set of all sequences $\zeta = (\zeta_1,\dots,\zeta_n)$ taking values in $\{-1,+1\}$. Then the likelihood ratio of mixtures is given by

$$\frac{d\mathrm{p}_{n,1}^{\tau}}{d\mathrm{p}_{n,0}}(Y) = \frac{1}{2^n}\sum_{\zeta\in\mathcal{Z}}\prod_{i>j}\Big(\prod_{t=1}^{\tau}\frac{d\mathrm{p}_{q_{1,\tau}\zeta_i\zeta_j}}{d\mathrm{p}_0}(Y_{ij}^t)\prod_{t=\tau+1}^{T}\frac{d\mathrm{p}_{q_{2,\tau}\zeta_i\zeta_j}}{d\mathrm{p}_0}(Y_{ij}^t)\Big).$$

Step 3. Second moment of the likelihood ratio.

Let $\tilde{\zeta}$ be an independent copy of the vector ζ. Then the second moment of the likelihood ratio can be written as

$$\mathrm{E}_0\Big[\frac{d\mathrm{p}_{n,1}^{\tau}}{d\mathrm{p}_{n,0}}\Big]^2(Y) = \frac{1}{2^{2n}}\sum_{\zeta,\tilde{\zeta}\in\mathcal{Z}}\mathrm{E}_0\Bigg[\prod_{i>j}\prod_{t=1}^{\tau}\frac{d\mathrm{p}_{q_{1,\tau}\zeta_i\zeta_j}}{d\mathrm{p}_0}\frac{d\mathrm{p}_{q_{1,\tau}\tilde{\zeta}_i\tilde{\zeta}_j}}{d\mathrm{p}_0}(Y_{ij}^t)\prod_{t=\tau+1}^{T}\frac{d\mathrm{p}_{q_{2,\tau}\zeta_i\zeta_j}}{d\mathrm{p}_0}\frac{d\mathrm{p}_{q_{2,\tau}\tilde{\zeta}_i\tilde{\zeta}_j}}{d\mathrm{p}_0}(Y_{ij}^t)\Bigg].$$

Using Lemma EC.5, we obtain

$$\mathrm{E}_0\Big[\frac{d\mathrm{p}_{n,1}^{\tau}}{d\mathrm{p}_{n,0}}\Big]^2(Y) = \frac{1}{2^{2n}}\sum_{\zeta,\tilde{\zeta}\in\mathcal{Z}}\prod_{i>j}\Big(1+\frac{\tilde{\rho}_n}{1-\tilde{\rho}_n}q_{1,\tau}^2\zeta_i\zeta_j\tilde{\zeta}_i\tilde{\zeta}_j\Big)^{\tau}\Big(1+\frac{\tilde{\rho}_n}{1-\tilde{\rho}_n}q_{2,\tau}^2\zeta_i\zeta_j\tilde{\zeta}_i\tilde{\zeta}_j\Big)^{T-\tau}.$$

The product in the last display can be written as

$$D_{\zeta,\tilde{\zeta}} = \exp\left(\sum_{i>j}\tau\log\Big(1+\frac{\tilde{\rho}_n}{1-\tilde{\rho}_n}q_{1,\tau}^2\zeta_i\zeta_j\tilde{\zeta}_i\tilde{\zeta}_j\Big)+(T-\tau)\log\Big(1+\frac{\tilde{\rho}_n}{1-\tilde{\rho}_n}q_{2,\tau}^2\zeta_i\zeta_j\tilde{\zeta}_i\tilde{\zeta}_j\Big)\right).$$

Note that

$$\frac{\tilde{\rho}_n}{1-\tilde{\rho}_n}\left(\tau q_{1,\tau}^2+(T-\tau)q_{2,\tau}^2\right) = \frac{Tq^2(\tau/T)r_{n,\tau}^2}{(1-\tilde{\rho}_n)\tilde{\rho}_n(n-1)^2} = \frac{T\mathcal{R}_{n,\tau}^2}{\omega_n}\frac{1}{(1-\tilde{\rho}_n)(n-1)(1-\varepsilon_n)}$$

and denote the last quantity by $\mu_n = \frac{T\mathcal{R}_{n,\tau}^2}{\omega_n}\frac{1}{(1-\tilde{\rho}_n)(n-1)(1-\varepsilon_n)}$. Applying the inequality $\log(1+x)\le x$ and using the fact that the distribution of $\sum_{i\neq j}\zeta_i\zeta_j\tilde{\zeta}_i\tilde{\zeta}_j$ is the same as the one of $\sum_{i\neq j}\zeta_i\zeta_j$, we obtain the upper bound

$$\begin{aligned}
\mathrm{E}_0\Big[\frac{d\mathrm{p}_{n,1}^{\tau}}{d\mathrm{p}_{n,0}}\Big]^2(Y) &= \frac{1}{2^{2n}}\sum_{\zeta,\tilde{\zeta}\in\mathcal{Z}} D_{\zeta,\tilde{\zeta}}\\
&\le \frac{1}{2^{2n}}\sum_{\zeta,\tilde{\zeta}\in\mathcal{Z}}\exp\Big(\frac{1}{2}\mu_n\sum_{i\neq j}\zeta_i\zeta_j\tilde{\zeta}_i\tilde{\zeta}_j\Big)\\
&= \mathrm{E}_{\zeta}\exp\Big(\frac{1}{2}\mu_n\sum_{i\neq j}\zeta_i\zeta_j\Big)\\
&= \mathrm{E}_{\zeta}\exp\Big(\frac{1}{2}\mu_n\zeta^{\mathrm{T}}(\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n-\mathrm{id}_n)\zeta\Big). \qquad \text{(EC.3)}
\end{aligned}$$

Step 4. Upper bound on the second moment.

Using Theorem 2 in (Cortinovis and Kressner 2021), we can bound the Laplace transform of the Rademacher chaos $\sum_{i\neq j}\zeta_i\zeta_j$ as follows. Let $A\in\mathbb{R}^{n\times n}$ be a symmetric matrix with zero diagonal. Then, $\forall\ 0<\mu<1/4$,

$$\log\mathrm{E}\left(e^{\mu\zeta^{\mathrm{T}}A\zeta}\right)\le\mu\|A\|_F^2\log\frac{1-2\mu}{1-4\mu}\le\frac{2\mu^2\|A\|_F^2}{1-4\mu}.$$

Using this inequality, from (EC.3) we obtain that for any $\mu_n<1/2$

$$\mathrm{E}_0\Big[\frac{d\mathrm{p}_{n,1}^{\tau}}{d\mathrm{p}_{n,0}}\Big]^2(Y)\le\mathrm{E}_{\zeta}\exp\Big\{\frac{1}{2}\mu_n\zeta^{\mathrm{T}}(\mathbf{1}_n^{\mathrm{T}}\mathbf{1}_n-\mathrm{id}_n)\zeta\Big\}\le\exp\Big(\frac{\frac{1}{2}\mu_n^2n(n-1)}{1-2\mu_n}\Big).$$

This bound implies that the second moment of the likelihood ratio is less than $1+4(1-\eta)^2$ if

$$\mu_n \leq \frac{2C_\eta}{n^2}\left(\sqrt{1+\frac{n^2}{2C_\eta}}-1\right), \tag{EC.4}$$

where $C_\eta = \log(1+4(1-\eta)^2)$. Note that this condition will imply $\mu_n < 1/2$. Now, (EC.4) is satisfied if

$$\frac{T\mathcal{R}_{n,\tau}^2}{\omega_n} \leq (1-\tilde{\rho}_n)(1-\varepsilon_n)\left(1-\frac{1}{n}\right)\sqrt{2C_\eta}\left(\sqrt{1+\frac{2C_\eta}{n^2}}-\sqrt{\frac{2C_\eta}{n^2}}\right).$$

Noting that $\sqrt{x+1}-\sqrt{x} \geq (1+2\sqrt{x})^{-1}$, we get that this inequality is satisfied if

$$\frac{T\mathcal{R}_{n,\tau}^2}{\omega_n} \leq \sqrt{2C_\eta}(1-\tilde{\rho}_n)(1-\varepsilon_n) \times\left(1-\frac{1}{n}\right)\left(1+\frac{2\sqrt{2C_\eta}}{n}\right)^{-1}. \tag{EC.5}$$

It means that all the signals with energy $\mathcal{R}_{n,\tau}$ satisfying (EC.5) are not detectable by any α-level test with the type II error smaller than β. Therefore the lower bound on the minimal detectable energy for an α-level test with type II errors bounded β is given by

$$\mathcal{R}_{n,\tau}^* \geq \left(2\log\left(1+4(1-\eta)^2\right)\right)^{1/4}\sqrt{\frac{\omega_n}{T}}$$

and the theorem follows. □

EC.3. Proof of result on the change-point localization

Proof of Proposition 1. Lemma EC.9 implies that for any $s \in [T]$, with probability at least $1-\frac{\gamma}{T}$

$$\|\xi(s)\|_{2\to 2} \leq 2\sqrt{2}(1+\epsilon)\sqrt{\omega_n}+C_\epsilon\log(2nT/\gamma). \tag{EC.6}$$

By the definition of $\widehat{\tau}_n$ we have $\|Z_T(\widehat{\tau}_n)\|_{2\to 2} \geq \|Z_T(\tau)\|_{2\to 2}$ which implies that

$$\mu_T^\tau(\tau)\Delta - \|\xi(\tau)\|_{2\to 2} \leq \mu_T^\tau(\widehat{\tau}_n)\Delta + \|\xi(\widehat{\tau}_n)\|_{2\to 2}.$$

Using (EC.6) and the union bound we get that with probability at least $1-\gamma$

$$\left(\mu_T^\tau(\tau)-\mu_T^\tau(\widehat{\tau}_n)\right)\Delta \leq 4\sqrt{2}(1+\epsilon)\sqrt{\omega_n} + C_\epsilon\log(2nT/\gamma). \tag{EC.7}$$

First, consider the case $\widehat{\tau}_n \leq \tau$. Using the definition of $\mu_T^\tau(t)$ (6), we compute

$$
\begin{aligned}
&\mu_T^\tau(\tau) - \mu_T^\tau(\widehat{\tau}_n) \\
&\quad = \sqrt{T}\left(q(x^*) - q(\widehat{x})\frac{1-x^*}{1-\widehat{x}}\right) \\
&\quad = \sqrt{T}(1-x^*)\left(\frac{q(x^*)}{1-x^*} - \frac{q(\widehat{x})}{1-\widehat{x}}\right) \\
&\quad = \sqrt{T}(1-x^*)\left(\sqrt{\frac{x^*}{1-x^*}} - \sqrt{\frac{\widehat{x}}{1-\widehat{x}}}\right) \\
&\quad = \sqrt{\frac{T(1-x^*)}{1-\widehat{x}}}\frac{x^*-\widehat{x}}{\sqrt{\widehat{x}(1-x^*)}+\sqrt{x^*(1-\widehat{x})}} \\
&\quad \geq \sqrt{T(1-x^*)}\frac{x^*-\widehat{x}}{1.5}
\end{aligned}
$$

where we use that for any $x \in (0,1)$, $x(1-x) \leq 1/4$. Plugging this calculation into (EC.7) we get

$$
(x^* - \widehat{x})\,\Delta \leq \frac{6\sqrt{2}(1+\epsilon)\sqrt{\omega_n}}{\sqrt{T(1-x^*)}} + \frac{1.5C_\epsilon \log(2nT/\gamma)}{\sqrt{T(1-x^*)}}. \tag{EC.8}
$$

Now assume that $\widehat{\tau}_n \geq \tau$. Then, using the definition of $\mu_T^\tau(t)$, (6), we compute

$$
\begin{aligned}
\mu_T^\tau(\tau) - \mu_T^\tau(\widehat{\tau}_n) &= \sqrt{T}\left(q(x^*) - q(\widehat{x})\frac{x^*}{\widehat{x}}\right) \\
&= \sqrt{T}x^*\left(\frac{q(x^*)}{x^*} - \frac{q(\widehat{x})}{\widehat{x}}\right) \\
&= \sqrt{T}x^*\left(\sqrt{\frac{1-x^*}{x^*}} - \sqrt{\frac{1-\widehat{x}}{\widehat{x}}}\right) \\
&= \sqrt{\frac{Tx^*}{\widehat{x}}}\frac{\widehat{x}-x^*}{\sqrt{\widehat{x}(1-x^*)}+\sqrt{x^*(1-\widehat{x})}} \\
&\geq \sqrt{T(1-x^*)}\frac{\widehat{x}-x^*}{1.5}
\end{aligned}
$$

which implies

$$
(\widehat{x} - x^*)\,\Delta \leq \frac{6\sqrt{2}(1+\epsilon)\sqrt{\omega_n}}{\sqrt{Tx^*}} + \frac{1.5C_\epsilon \log(2nT/\gamma)}{\sqrt{Tx^*}}. \tag{EC.9}
$$

Combining (EC.8) and (EC.9) and using $q^2(x^*) \leq x^* \wedge (1-x^*)$ we get the statement of the Proposition 1. □

EC.4. Proofs of results for the sparse graphon model

We start by summarizing the notation that we use in the proofs.

Given a matrix $\Theta \in [-1,1]^{n\times n}$, we define the empirical graphon associated with Θ as follows:

$$\tilde{f}_\Theta(x,y) = \Theta_{\lceil nx\rceil,\lceil ny\rceil}, \quad (x,y)\in[0,1]^2. \tag{EC.10}$$

In the same spirit, given a vector $v=(v_1,\dots,v_n)$, for any $x\in[0,1]$, we define the following piecewise constant function

$$\psi_v(x) = \sqrt{n}v_{\lceil nx\rceil}, \quad x\in[0,1]$$

and set

$$\mathcal{F} = \left\{\psi_v : \|v\|_{\ell_2}\le 1\right\}.$$

We have that $\|v\|_{\ell_2}\le 1$ implies

$$\|\psi_v\|_{L_2[0,1]} = \frac{1}{n}\sum_{i=1}^n n v_i^2 \le 1.$$

We will need to work with a difference of two graphons, so we extend the definition of graphon space. In what follows $\mathcal{W}$ refers to the collection of bounded symmetric measurable functions $W:[0,1]^2\to[-1,1]$.

EC.4.1. Proofs of upper bounds

Proof of Theorem 4. We have to find a threshold $H^*_{\alpha,n,T}$ such that

$$\alpha(\psi_{n,T}) = \sup_{W\in\mathcal{W}_0} \mathrm{P}_W\left\{\psi_{n,T}=1\right\} \le \alpha$$

and show that

$$\beta(\psi_{n,T},\delta_{n,T}) = \sup_{\tau\in\mathcal{D}_T}\left(\sup_{W^\tau,W^{\tau+1}\in\mathcal{W}_0(\delta_{n,T})} \mathrm{P}_{W^\tau,W^{\tau+1}}\left\{\psi_{n,T}=0\right\}\right)\le\beta.$$

Under H_0 there is no change in the graphon function W but it can be a change in the features $\boldsymbol{\varepsilon}$. Let $\Theta_1 = \left(\rho_n W(\varepsilon_i,\varepsilon_j)\right)_{(i,j)\in[n]\times[n]}$ denote the matrix of connection probabilities before the time point τ and $\Theta_2 = \left(\rho_n W(\varepsilon_i',\varepsilon_j')\right)_{(i,j)\in[n]\times[n]}$ the matrix of connection probabilities after τ. Let

$$\xi^\pi(t) = \sqrt{\frac{t(T-t)}{T}}\left(\frac{1}{t}\sum_{s=1}^t W^s - \frac{1}{T-t}\sum_{s=t+1}^T W^s\circ\pi\right) \tag{EC.11}$$

denote the centered random matrices of noise corresponding to the permutation π . We have

$$Z_T(t) = -\mu_T^\tau(t)\Delta\Theta^\tau + \xi^{\pi^*}(t), \quad t = 1, \dots, T-1,$$

where $\Delta\Theta^\tau = \Theta_1 - \Theta_2 \circ \pi^*$ and

$$\mu_T^\tau(t) = \sqrt{\frac{t(T-t)}{T}}\Big(\frac{\tau}{t}\mathbf{1}_{\{\tau+1\le t\le T\}} + \frac{T-\tau}{T-t}\mathbf{1}_{\{1\le t\le \tau\}}\Big).$$

Let π' be a permutation of $\{1, \dots, n\}$ such that

$$\pi' \in \arg\min_{\pi} \|\Theta_1 - \Theta_2 \circ \pi\|_{2\to 2}.$$

Using the definition of π^* and the triangle inequality, we have that

$$\begin{aligned}
\|Z_T^{\pi^*}(t)\|_{2\to 2} &\le \|Z_T^{\pi'}(t)\|_{2\to 2} \\
&\le \mu_T^\tau(t)\|\Theta_1 - \Theta_2 \circ \pi'\|_{2\to 2} + \|\xi^{\pi'}(t)\|_{2\to 2} \\
&\le \sqrt{T}q(t/T)\|\Theta_1 - \Theta_2 \circ \pi'\|_{2\to 2} + \|\xi^{\pi'}(t)\|_{2\to 2}.
\end{aligned}$$

Now, using Lemma EC.7, we have that with probability at least $1 - 4/n$

$$\begin{aligned}
\|\Theta_1 - \Theta_2 \circ \pi'\|_{2\to 2}^2 &\le n^2\delta_2^2(\tilde{f}_{\Theta_1}, \tilde{f}_{\Theta_2}) \\
&\le 2n^2\left(\delta_2^2(\tilde{f}_{\Theta_1}, \rho_n W) + \delta_2^2(\tilde{f}_{\Theta_2}, \rho_n W)\right) \\
&\le 32n^2\rho_n^2\sqrt{\frac{K}{n}\log(n)},
\end{aligned}$$

where $\tilde{f}_{\Theta_i}$ is the empirical graphon associated with Θ_i. Note that, for any $x > 0$, $\mathrm{P}_W(\|\xi^{\pi'}(t)\|_{2\to 2} > x) = \mathrm{E}_{\{\varepsilon_\vartheta, \varepsilon'_\vartheta\}_{\vartheta\in\mathcal{V}}}\left[\mathrm{P}\left(\|\xi^{\pi'}(t)\|_{2\to 2} > x \,\middle|\, \{\varepsilon_\vartheta, \varepsilon'_\vartheta\}_{\vartheta\in\mathcal{V}}\right)\right]$. Now, we can bound $\mathrm{P}\left(\|\xi^{\pi'}(t)\|_{2\to 2} > x \,\middle|\, \{\varepsilon_\vartheta, \varepsilon'_\vartheta\}_{\vartheta\in\mathcal{V}}\right)$ using Lemma EC.9: for every $\delta \in (0,1)$, conditionally on $\{\varepsilon_\vartheta, \varepsilon'_\vartheta\}_{\vartheta\in\mathcal{V}}$, we have

$$\|\xi^{\pi'}(t)\|_{2\to 2} \le 2\sqrt{2}(1+\epsilon)\sqrt{\frac{t(T-t)}{T}}\Big[\frac{1}{t^2}\sum_{s=1}^{t}\|\Theta^s\|_{1,\infty} + \frac{1}{(T-t)^2}\sum_{s=t+1}^{T}\|\Theta^s\|_{1,\infty}\Big]^{1/2}$$

with the probability larger than $1-\delta$, where $\Theta^t = \rho_n \left(W^t(\varepsilon_i, \varepsilon_j)\right)_{(i,j)\in[n]\times[n]}$. Note that $\|\Theta^t\|_{1,\infty} \le n\rho_n$ and we get

$$\|\xi^{\pi'}(t)\|_{2\to 2} \le 2(1+\epsilon)\sqrt{2n\rho_n} + C_\epsilon \log\left(\frac{2n}{\delta}\right)$$

with the probability larger than $1-\delta$. Taking $\delta = \alpha/(2|\mathcal{T}|)$, using the union bound and $\alpha \ge 8/n$, we obtain

$$\alpha(\psi_{n,T}) = \sup_{W\in\mathcal{W}_0} \mathrm{P}_W\left\{\max_{t\in\mathcal{T}} \|Z_T^{\pi^*}(t)\|_{2\to 2} \ge H^*_{\alpha,n,T}\right\} \le \alpha.$$

Let us turn to the type II error. The proof is close to the proof of Lemma EC.4 but, compared to the case of inhomogeneous random graph model considered in Lemma EC.4, we need to account for possible change in the latent variables ε. The key point is to link the change in the operator norm of the matrix of parameters to the operator norm of the corresponding empirical graphon. Following the proof of Lemma EC.4 we can show that, if $\tau \le T/2$, there exists $t^* \in \mathcal{T}^L$ such that $\tau/2 \le t^* < \tau$ and

$$\beta(\psi_{n,T}, \delta_{n,T}) \le \sup_{W^\tau, W^{\tau+1}\in\mathcal{W}_0(\delta_{n,T})} \mathrm{P}_{W^\tau,W^{\tau+1}}\left\{\|\xi^{\pi^*}(t^*)\|_{2\to 2} > \frac{\sqrt{T}q(\tau/T)}{\sqrt{3}}\|\Theta_1 - \Theta_2\circ\pi^*\|_{2\to 2} - H^*_{\alpha,n,t^*}\right\}.$$

By the triangle inequality,

$$\|\xi^{\pi^*}(t^*)\|_{2\to 2} \ge \frac{\sqrt{T}q(\tau/T)}{\sqrt{3}}\|\Theta_1 - \Theta_2\circ\pi^*\|_{2\to 2} - H^*_{\alpha,n,t^*}.$$

Applying Lemma EC.6 to the right-hand side of the last display we obtain the bound

$$\|\xi^{\pi^*}(t^*)\|_{2\to 2} \ge \frac{n\sqrt{T}q(\tau/T)}{\sqrt{3}}\|\tilde{f}_{\Theta_1-\Theta_2\circ\pi^*}\|_{2\to 2} - H^*_{\alpha,n,t^*} = \frac{n\sqrt{T}q(\tau/T)}{\sqrt{3}}\|\tilde{f}_{\Theta_1} - \tilde{f}_{\Theta_2\circ\pi^*}\|_{2\to 2} - H^*_{\alpha,n,t^*}.$$

Finally, using (EC.10) we get the bound

$$\|\xi^{\pi^*}(t^*)\|_{2\to 2} \ge \frac{n\sqrt{T}q(\tau/T)}{\sqrt{3}}\delta(\tilde{f}_{\Theta_1}, \tilde{f}_{\Theta_2}) - H^*_{\alpha,n,t^*}.$$

Now, using twice the triangle inequality, we get

$$\|\xi^{\pi^*}(t^*)\|_{2\to 2} \ge \frac{n\sqrt{T}q(\tau/T)}{\sqrt{3}}\left(\delta(\rho_n W_1, \rho_n W_2) - \delta(\tilde{f}_{\Theta_1}, \rho_n W_1) - \delta(\rho_n W_2, \tilde{f}_{\Theta_2})\right) - H_{\alpha,n,t^*}.$$

Note that Lemma EC.7 and $\delta(\tilde{f}_{\Theta_i}, \rho_n W_i) \le \delta_2(\tilde{f}_{\Theta_i}, \rho_n W_i)$ imply

$$\delta(\tilde{f}_{\Theta_i}, \rho_n W_i) \le 4\rho_n\left(\frac{K_i \log n}{n}\right)^{1/4}$$

with probability at least $1 - 2/n$. Take $\beta \geq 6/n$. Using Lemma EC.10 with $\delta = \beta/3$ we obtain

$$\|\xi^{\pi^*}(t^*)\|_{2\to 2} \leq \sup_{\pi} \|\xi^{\pi}(t^*)\|_{2\to 2} \leq 2(1+\epsilon)\sqrt{2n\rho_n} + C_\epsilon \log\left(\frac{12n}{\beta}\right).$$

with probability $1 - \beta/3$. This implies that $\beta(\psi_{n,T}, \delta_{n,T}) \leq \beta$ if

$$\frac{n\sqrt{T}q(\tau/T)}{\sqrt{3}}\left(\rho_n\delta(W_1, W_2) - 4\rho_n\left(\frac{(K_1+K_2)\log n}{n}\right)^{1/4}\right) \geq H^*_{\alpha,n,t^*} + 2(1+\epsilon)\sqrt{2n\rho_n} + C_\epsilon \log\left(\frac{12n}{\beta}\right).$$

Combining this condition with the threshold $H^*_{\alpha,n,t}$ defined in (19) and the facts that $q(t^*/T) \leq q(\tau/T)$ for $t^* < \tau \leq T/2$ and $K_1, K_2 \leq K$, we obtain the detection condition (20). The case of $\tau > T/2$ is analogous. □

Proof of Theorem 6. Theorem 6 follows from combining the bounds obtained in Lemma EC.6, Lemma EC.8 and Theorem 3. □

LEMMA EC.6. *Let $\Theta = (\Theta_{ij}) \in [-1, 1]^{n\times n}$ be symmetric matrix. Then*

$$\|\Theta\|_{2\to 2} \geq n\|\tilde{f}_\Theta\|_{2\to 2}.$$

Proof. By definition of the operator norm,

$$\|\tilde{f}_\Theta\|_{2\to 2} = \sup_{\psi\in L_2[0,1], \|\psi\|_{L_2}\leq 1} \left|\iint_{[0,1]^2} \tilde{f}_\Theta(x, y)\psi(x)\psi(y)\, dx\, dy\right|.$$

Then we have

$$\|\tilde{f}_\Theta\|_{2\to 2} = \sup_{\psi\in L_2[0,1], \|\psi\|_{L_2}\leq 1} \left|\sum_{ij} \Theta_{ij} \int_{i/n}^{(i+1)/n} \psi(x)\, dx \int_{j/n}^{(j+1)/n} \psi(y)\, dy\right|$$

that can be written as

$$\|\tilde{f}_\Theta\|_{2\to 2} = \sup_{\psi\in L_2[0,1], \|\psi\|_{L_2}\leq 1} \left|\frac{1}{n}\sum_{ij} \Theta_{ij} v_i^\psi v_j^\psi\right|,$$

where $v_i^\psi = \sqrt{n}\int_{i/n}^{(i+1)/n} \psi(x)\, dx$ and $v^\psi = (v_i^\psi)_{i=1}^n$. Note that the Cauchy-Schwartz inequality implies

$$\left(\int_{i/n}^{(i+1)/n} \psi(x)\, dx\right)^2 \leq \frac{1}{n}\int_{i/n}^{(i+1)/n} \psi^2(x)\, dx$$

and we get that $\|v^\psi\|_{\ell_2} \leq 1$ for ψ such that $\|\psi\|_{L_2} \leq 1$. Now we can write

$$n\|\tilde{f}_\Theta\|_{2\to 2} \leq \sup_{\|v\|_{\ell_2}\leq 1} \left|\sum_{ij} \Theta_{ij} v_i v_j\right| = \|\Theta\|_{2\to 2}$$

and the statement of Lemma EC.6 follows. □

LEMMA EC.7. *For any $K \le \frac{n}{\log(n)}$ assume that $W \in \mathcal{W}_K$. Let $\Theta = (\Theta_{ij}) \in [0,1]^{n\times n}$ be symmetric matrix with entries $\Theta_{ij} = W(\xi_i, \xi_j)$ for $i < j$, where ξ_i are i.i.d. uniform random variables on $[0,1]$. We have that, with probability large than $1-2/n$,*

$$\delta(\tilde{f}_\Theta, W) \le 4\Big(\frac{K}{n}\log n\Big)^{1/4}.$$

Proof. Following the proof of Proposition 3.2 in (Klopp et al. 2017), we get

$$\delta_2(\tilde{f}_{\Delta\Theta^\tau}, \Delta W^\tau) \le \frac{1}{n} + \sum_{a=1}^{K} |\lambda_a - \widehat{\lambda}_a|,$$

with

$$\widehat{\lambda}_a = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{\varepsilon_i \in \phi^{-1}(a)\}}$$

and $\lambda_a = \lambda(\phi^{-1}(a))$, where λ stands for the Lebesgue measure. Since $\varepsilon_1, \dots \varepsilon_n$ are i.i.d. uniform random variables, $n\widehat{\lambda}_a$ has a binomial distribution with parameters (n, λ_a). We have $n\widehat{\lambda}_a - n\lambda_a = \sum_{i=1}^{n}(Y_i - \lambda_a)$, where $Y_i \sim \text{Bernoulli}(\lambda_a)$. Applying the Bernstein inequality we obtain that for any $t > 0$

$$|n\widehat{\lambda}_a - n\lambda_a| \le \left(2t\sum_{a=1}^{K}\lambda_a(1-\lambda_a)\right)^{1/2} + 2t/3$$

with probability $1 - 2e^{-t}$. Taking $t = \log(nK)$ implies that with probability $1 - 2/(nK)$

$$|n\widehat{\lambda}_a - n\lambda_a| \le (2n\lambda_a \log(nK))^{1/2} + \frac{2}{3}\log(nK).$$

Using $K \le n$ and the union bound we obtain that, with probability $1-2/n$,

$$\delta^2(\tilde{f}_\theta, W) \le \frac{1}{n} + \frac{2}{n}\sum_{a=1}^{K}(n\lambda_a \log n)^{1/2} + \frac{4K\log(n)}{3n} \le \frac{1}{n} + 2\left(\frac{K\log n}{n}\right)^{1/2} + \frac{4K\log n}{3n}$$

where we use $\sum_{a=1}^{K}\lambda_a = 1$ and the Cauchy–Schwarz inequality. Using $\frac{K\log n}{n} \le 1$ we complete the proof of Lemma EC.7. □

LEMMA EC.8. *Assume that $W \in \Sigma(\gamma, L)$. Let $\Theta = (\Theta_{ij}) \in [0,1]^{n\times n}$ be symmetric matrix with entries $\Theta_{ij} = W(\varepsilon_i, \varepsilon_j)$ for $i < j$, where ε_i are i.i.d. uniform random variables on $[0,1]$. We have that, with probability at least $1-2/n$,*

$$\delta(\tilde{f}_\Theta, W) \le 2\left(\frac{\log n}{n}\right)^{\frac{\gamma\wedge 1}{2}}$$

Proof. Following the proof of Proposition 3.6 in (Klopp et al. 2017), we get

$$\delta^2(\tilde{f}_\Theta, W) \leq \frac{2}{n} + \frac{1}{n}\sum_{m=1}^{n}\left|\frac{m}{n+1} - \varepsilon_{(m)}\right|^{2\gamma'}$$

where $\gamma' = \gamma \wedge 1$ and $\varepsilon_{(m)}$ stands for the m-th largest element of the set $\{\varepsilon_1, \dots, \varepsilon_n\}$. Note that, the random variable $\varepsilon_{(m)}$ follows β-distribution with parameters $(m, n+1-m)$, $\varepsilon_{(m)} \sim \mathrm{Beta}(m, n+1-m)$. The β-distribution is sub-Gaussian and the proxy variance σ^2 for $\mathrm{Beta}(m, n+1-m)$ is bounded by $\frac{1}{4(n+2)}$ (see, for example, (Marchal and Arbel 2017)). By the exponential Markov inequality (see, for example, (Vershynin 2018), Lemma 5.5) we get

$$\mathrm{P}\left\{\left|\varepsilon_{(m)} - \frac{m}{n+1}\right| > t\right\} \leq 2e^{-t^2/(4\sigma^2)}.$$

Taking $t = (\log n/(n+2))^{1/2}$ implies that, with probability at least $1 - 2/n^2$,

$$\left|\varepsilon_{(m)} - \frac{m}{n+1}\right| < \left(\frac{\log n}{n+2}\right)^{1/2}.$$

Now, applying the union bound we obtain

$$\delta^2(\tilde{f}_\Theta, W) \leq \frac{2}{n} + \left(\frac{\log n}{n+2}\right)^{\gamma'}$$

and Lemma EC.8 follows. □

EC.4.2. Proof of the lower bound for K-step graphons

We will start with the definition of a class of K-step graphons used throughout the proof. Let $u = (u_1, \dots, u_K) \in (-\frac{1}{K}, \frac{1}{K})^K$ be a given vector satisfying $\sum_{k=1}^{K} u_k = 0$. Define the partition $\Pi = \bigcup_{1\leq k,l\leq K} \Pi_{kl}$ of the set $[0,1]^2$ into K^2 blocks:

$$\Pi_{kl}(u) = \Big[\frac{k-1}{K} + \sum_{i=1}^{k-1} u_i, \frac{k}{K} + \sum_{i=1}^{k} u_i\Big) \times \Big[\frac{l-1}{K} + \sum_{i=1}^{l-1} u_i, \frac{l}{K} + \sum_{i=1}^{l} u_i\Big) \quad \forall 1 \leq k, l \leq K.$$

Let $Q = (Q_{kl})_{1\leq k,l\leq K} \in [0,1]^{K\times K}$ be a matrix of connection probabilities. The K-step graphon W_u is a blockwise constant function defined by

$$W_u(x, y) = \sum_{k,l\in[K]^2} Q_{kl}\mathbf{1}\{(x, y) \in \Pi_{kl}(u)\}. \qquad \text{(EC.12)}$$

Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n) \in [0,1]^n$ be the vector of i.i.d. features uniformly distributed over $[0,1]$. Note that the community assignment of each vertex ϑ is defined by the corresponding variable ε_ϑ:

$$\mathrm{P}\Big\{\vartheta \text{ belongs to the block } k\Big\} = \mathrm{P}\Big\{\varepsilon_\vartheta \in \Big[\frac{k-1}{K} + \sum_{i=1}^{k-1} u_i, \frac{k}{K} + \sum_{i=1}^{k} u_i\Big)\Big\} = \frac{1}{K} + u_k.$$

We can introduce a new random variable $\xi_\vartheta = \xi_\vartheta(u)$, $\vartheta \in [n]$ of the block assignment following multinomial distribution $\mathcal{M}(K, p_1, \dots, p_K)$ with parameters $p_k = 1/K + u_k$. Denote the corresponding vector of i.i.d. multinomial variables by $\xi_u = (\xi_1(u), \dots, \xi_n(u)) \in [K]^n$. Given ξ_u, the connection probabilities are given by

$$\Theta_{ij}(\xi_u) = \begin{cases} \rho_n Q_{\xi_i(u),\xi_j(u)}, & i \neq j \\ 0, & i = j. \end{cases}$$

Denote by $\mathrm{P}_{W_u,\xi_u}(A^t)$ the conditional distribution of the dynamic network at time t given the node assignment ξ_u:

$$\mathrm{P}_{W_u,\xi_u}(A^t) = \prod_{i<j} \Theta_{ij}(\xi_u)^{A^t_{ij}} (1 - \Theta_{ij}(\xi_u))^{1-A^t_{ij}}.$$

Note that it can be written as

$$\mathrm{P}_{W_u,\xi_u}(A^t) = \prod_{i<j} \sum_{k,l \in [K]^2} (\rho_n Q_{kl})^{A^t_{ij}} (1 - \rho_n Q_{kl})^{1-A^t_{ij}} \mathbf{1}_{\{(\xi_i(u),\xi_j(u))=(k,l)\}}.$$

In what follows we denote by $A = (A^1, \dots, A^T)$ the full set of observations and by $A^{\le\tau} = (A^1, \dots, A^\tau)$ and $A^{>\tau} = (A^{\tau+1}, \dots, A^T)$ the realizations before and after the time τ. We denote by $\mathrm{P}^{\otimes\tau}(A^{\le\tau})$ and $\mathrm{P}^{\otimes(T-\tau)}(A^{>\tau})$ the corresponding product measures.

A K-step graphon depends on two main ingredients: the partition Π_K of $[0,1]^2$ and the connection probability matrix Q. We will see that choosing different prior distributions on Q and Π will lead to two different lower bounds. The first lower bound, that we call *agnostic error lower bound*, will be derived from the uncertainty of sampling vector of features ε. The second one, that we call *network sampling lower bound*, comes from the uncertainty of random realizations of the network.

Without loss of generality, we will prove the result for $K = 2$. Indeed, for any $K > 2$,

$$\inf_{\psi \in \Psi_\alpha} \sup_{W^\tau, W^{\tau+1} \in \mathcal{W}_K(\delta_{n,T})} \mathrm{P}_{W^\tau, W^{\tau+1}}\{\psi = 0\} \ge \inf_{\psi \in \Psi_\alpha} \sup_{W^\tau, W^{\tau+1} \in \mathcal{W}_2(\delta_{n,T})} \mathrm{P}_{W^\tau, W^{\tau+1}}\{\psi = 0\},$$

and the boundary for the case of two blocks will imply the one for K blocks.

Proof of Theorem 5.

I. Agnostic error lower bound. The first lower bound is related to the error coming from the sampling of $\boldsymbol{\varepsilon}$. We start by choosing a prior distribution on the graphons and the assignment vectors. Based on the prior, we will bound the Kullback-Leibler divergence between measures under the null and the alternative hypotheses. Note that it follows from (EC.2) and the inequality $\frac{1}{2}\|\mathrm{P}_1 - \mathrm{P}_0\|_{\mathrm{TV}} \le \sqrt{\frac{1}{2}\mathrm{KL}(\mathrm{P}_0, \mathrm{P}_1)}$ that the type II error is bounded from below by β if $\mathrm{KL}(\mathrm{P}_0, \mathrm{P}_1) \le 2(1 - \alpha - \beta)^2 = 2(1-\eta)^2$. Thus, we need to provide an upper bound on the Kullvack–Leibler divergence that will imply the corresponding lower bound on the minimax detectable distance between graphons.

Step 1. Choice of priors. We will use W_u graphons defined in (EC.12). We suppose that the connection probability matrix Q is the same under H_0 and under H_1 and is defined as $Q = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$.

1. *Prior under* H_0: we choose vector $u = 0$ and get a partition Π with blocks of a constant size $1/2$. The corresponding graphon is denoted by W_0. Let ξ_0 and ξ_0' be two independent block assignment vectors following the multinomial distribution with class probabilities $p_k = 1/2$, $k = 1, 2$. Since matrix Q does not change, the corresponding conditional distributions P_{W_0,ξ_0} and $\mathrm{P}_{W_0,\xi_0'}$ coincide. Then, the measure under H_0 be given by

$$\mathrm{P}_0(A) = \Big(\sum_{a\in\{1,2\}^n} \mathrm{P}(\xi_0 = a)\mathrm{P}^{\otimes\tau}_{W_0,\xi_0=a}(A^{\le\tau})\Big)\Big(\sum_{b\in\{1,2\}^n} \mathrm{P}(\xi_0' = b)\mathrm{P}^{\otimes(T-\tau)}_{W_0,\xi_0'=b}(A^{>\tau})\Big).$$

2. *Prior under* H_1*:* fix some $0 < \lambda < 1/2$. Let $u = (\lambda, -\lambda)$, $v = (-\lambda, \lambda)$ and W_u, W_v be the corresponding graphons with the probabilities of classes $1/2 + \lambda$ and $1/2 - \lambda$. The only difference between these two graphons is a slight disequilibrium around $(x, y) \in [1/2 - \lambda, 1/2 + \lambda]^2$. It is not difficult to see that $\delta^2(W_u, W_v) \ge 2\lambda$.

Let $\xi_u \in \{1, 2\}^n$ and $\xi_v \in \{1, 2\}^n$ be two independent class assignment vectors before and after the change in the graphon such that $\mathrm{P}(\xi_i(u) = k) = \frac{1}{2} + u_k$, $\mathrm{P}(\xi_i(v) = k) = \frac{1}{2} + v_k$ $\forall i = 1, \dots, n$, $k = 1, 2$. Note that the prior with two different assignment vectors and the same the connection probability matrix Q takes into account the case of a possible label mismatch before and after the change. The measure under H_1 is defined as

$$\mathrm{P}_1(A) = \Big(\sum_{a\in\{1,2\}^n} \mathrm{P}(\xi_u = a)\mathrm{P}^{\otimes\tau}_{W_u,\xi_u=a}(A^{\le\tau})\Big)\Big(\sum_{b\in\{1,2\}^n} \mathrm{P}(\xi_v = b)\mathrm{P}^{\otimes(T-\tau)}_{W_v,\xi_v=b}(A^{>\tau})\Big).$$

Step 2. Bounding the divergence. Denote for brevity $\mathrm{P}(\xi = a)$ by $\mathrm{P}_\xi(a)$. Since the matrix Q is the same for all graphons, the conditional probabilities generating the networks under H_0 and under H_1 are the same. Denote

$$\mathrm{P}_Q(A^{\le\tau}|a) := \mathrm{P}^{\otimes\tau}_{W_0,\xi_0=a}(A^{\le\tau}) = \mathrm{P}^{\otimes\tau}_{W_u,\xi_u=a}(A^{\le\tau})$$

and

$$\mathrm{P}_Q(A^{>\tau}|b) := \mathrm{P}^{\otimes(T-\tau)}_{W_o,\xi_0'=b}(A^{>\tau}) = \mathrm{P}^{\otimes(T-\tau)}_{W_v,\xi_v=b}(A^{>\tau}).$$

Then, we have

$$\mathrm{KL}(\mathrm{P}_0, \mathrm{P}_1) = \mathrm{KL}^{\le\tau}(\mathrm{P}_0, \mathrm{P}_1) + \mathrm{KL}^{>\tau}(\mathrm{P}_0, \mathrm{P}_1)$$

where

$$\mathrm{KL}^{\le\tau}(\mathrm{P}_0, \mathrm{P}_1) = \sum_{A^{\le\tau}} \sum_{a\in\{1,2\}^n} \mathrm{P}_{\xi_0}(a)\mathrm{P}_Q(A^{\le\tau}|a) \log\left(\frac{\sum\limits_{a\in\{1,2\}^n} \mathrm{P}_{\xi_0}(a)\mathrm{P}_Q(A^{\le\tau}|a)}{\sum\limits_{a\in\{1,2\}^n} \mathrm{P}_{\xi_u}(a)\mathrm{P}_Q(A^{\le\tau}|a)}\right)$$

and

$$\mathrm{KL}^{>\tau}(\mathrm{P}_0,\mathrm{P}_1)=\sum_{A^{>\tau}}\sum_{b\in\{1,2\}^n}\mathrm{P}_{\xi_0'}(b)\mathrm{P}_Q(A^{>\tau}|b)\log\left(\frac{\sum\limits_{b\in\{1,2\}^n}\mathrm{P}_{\xi_0'}(b)\mathrm{P}_Q(A^{>\tau}|b)}{\sum\limits_{b\in\{1,2\}^n}\mathrm{P}_{\xi_v}(b)\mathrm{P}_Q(A^{>\tau}|b)}\right).$$

Taking into account that the function $f(x,y)=x\log(x/y)$ is convex, we can apply the Jensen's inequality and obtain that

$$\mathrm{KL}(\mathrm{P}_0,\mathrm{P}_1)\le\sum_{a\in\{1,2\}^n}\mathrm{P}_{\xi_0}(a)\log\frac{\mathrm{P}_{\xi_0}(a)}{\mathrm{P}_{\xi_u}(a)}+\sum_{b\in\{1,2\}^n}\mathrm{P}_{\xi_0'}(b)\log\frac{\mathrm{P}_{\xi_0}(b)}{\mathrm{P}_{\xi_v}(b)}=n\Big(\mathrm{KL}(\mathrm{P}_{\xi_0},\mathrm{P}_{\xi_u})+\mathrm{KL}(\mathrm{P}_{\xi_0'},\mathrm{P}_{\xi_v})\Big).$$

The last equality follows from the fact that $\mathrm{P}_\xi(a)$ are product probabilities. Thus we have to bound the Kullback–Leibler divergence between two binomial distributions. Using the inequality $\log(1+x)\ge x/(1+x)\ \forall x>-1$, we obtain

$$\mathrm{KL}(\mathrm{P}_{\xi_0},\mathrm{P}_{\xi_u})=\frac{1}{2}\log\Big(\frac{1/4}{1/4-\lambda^2}\Big)\le\frac{2\lambda^2}{1-4\lambda^2}$$

which implies $\mathrm{KL}(\mathrm{P}_0,\mathrm{P}_1)\le 4n\lambda^2/(1-4\lambda^2)$. Recall that $\delta^2(W_u,W_v)\ge 2\lambda$. Consequently, if $\delta=\delta(W_u,W_v)$, we can write

$$\mathrm{KL}(\mathrm{P}_0,\mathrm{P}_1)\le\frac{n}{2}\frac{\delta^4}{1-\delta^4}\le 2(1-\eta)^2$$

if

$$\delta^4\le\frac{4n^{-1}(1-\eta)^2}{1+4n^{-1}(1-\eta)^2}.$$

The last inequality is true if $\delta^4\le\frac{8}{3}n^{-1}(1-\eta)^2$ for all $n\ge 8$. It implies the lower bound condition on the distance between graphons:

$$\delta(W_u,W_v)\le\Big(\frac{8}{3}\Big)^{1/4}(1-\eta)^{1/2}n^{-1/4}.$$

II. Network Sampling lower bound. In this part we will suppose that the transition matrix Q changes but the partition Π does not change. In order to bound the type II error by β from below, we need to show that the chi-squared divergence between the mixtures under H_0 and H_1 is smaller that $4(1-\eta)^2$.

Step 1. Choice of priors. It will be sufficient to show the result for the case of $K=2$, since as we will see the lower bound on the separation rate is independent of K. We will work with 2-step graphons with fixed partition Π into 4 equal blocks $\Pi_{kl}=[k-1/2,k/2)\times[l-1/2,l/2)$, $1\le k,l\le 2$.

Prior under H_0. We suppose that under H_0 the connection probabilities are all equal to $1/2$, that is $Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ and $\Theta_{ij} = \rho_n/2$, $\forall i \neq j$. The corresponding graphon is denoted by W_0. Denote $p_0 = \rho_n/2$. Then, independently of the feature vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$,

$$\mathrm{P}_0(A) = \prod_{i<j} \prod_{i=1}^{T} p_0^{\sum_{t=1}^T A_{ij}^t} (1-p_0)^{T-\sum_{t=1}^T A_{ij}^t}$$

Prior under H_1. Denote by Q_1 and Q_2 the connection probability matrices before and after the change-point. Let $\lambda > 0$. We assume that

$$Q_1 = \rho_n^{-1} \begin{pmatrix} p_1 & p_2 \\ p_2 & p_1 \end{pmatrix} \quad \text{and} \quad Q_2 = \rho_n^{-1} \begin{pmatrix} p_3 & p_4 \\ p_4 & p_3 \end{pmatrix}$$

where

$$p_1 = \rho_n\Big(\frac{1}{2} + \Big(1 - \frac{\tau}{T}\Big)\lambda\Big), \quad p_2 = \rho_n\Big(\frac{1}{2} - \Big(1 - \frac{\tau}{T}\Big)\lambda\Big), \quad p_3 = \rho_n\Big(\frac{1}{2} - \frac{\tau}{T}\lambda\Big), \quad p_4 = \rho_n\Big(\frac{1}{2} + \frac{\tau}{T}\lambda\Big).$$

Denote the corresponding graphons by W_1 and W_2 and the corresponding matrices of connection probabilities by Θ_i, $i = 1, 2$. Let $\xi = (\xi_1, \ldots, \xi_n)$ be the class assignment vector of i.i.d. variables taking values $\{1, 2\}$ with probability 1/2. Then

$$\mathrm{P}_1(A) = \mathrm{P}_{W_1}^{\otimes \tau}(A^{\leq \tau}) \mathrm{P}_{W_2}^{\otimes (T-\tau)}(A^{>\tau})$$

where

$$P_{W_1}(A^t) = \prod_{i<j} \Big(p_1^{A_{ij}^t} (1-p_1)^{1-A_{ij}^t} \mathbf{1}\{\xi_i = \xi_j\} + p_2^{A_{ij}^t} (1-p_2)^{1-A_{ij}^1} \mathbf{1}\{\xi_i \neq \xi_j\} \Big)$$

and

$$P_{W_2}(A^t) = \prod_{i<j} \Big(p_3^{A_{ij}^t} (1-p_3)^{1-A_{ij}^t} \mathbf{1}\{\xi_i = \xi_j\} + p_4^{A_{ij}^t} (1-p_4)^{1-A_{ij}^1} \mathbf{1}\{\xi_i \neq \xi_j\} \Big)$$

Step 2. Bounding the χ^2-divergence. To prove the lower bound, we will need an upper bound on the chi-squared divergence

$$\chi^2(\mathrm{P}_0, \mathrm{P}_1) = \mathrm{E}_{\mathrm{P}_0} \Big(\frac{d\mathrm{P}_1}{d\mathrm{P}_0} \Big)^2 - 1 \leq 4(1 - \alpha - \beta)^2 = 4(1-\eta)^2.$$

This bound will follow from the upper bound on the second moment of the likelihood ratio:

$$\mathrm{E}_{\mathrm{P}_0} L^2(A) \leq 1 + 4(1-\eta)^2,$$

where $L(A) = \frac{d\mathrm{P}_1}{d\mathrm{P}_0}(A)$. Define the set $S = \Big\{\{a,b\} \in [n]^2 :\ a < b,\ \xi_a = \xi_b\Big\}$ and its compliment S^c. Denote by $N = n(n-1)/2$ the cardinality of $S \cup S^c$ and by μ the distribution of S. Denote for brevity $q_i = (1-p_i)/(1-p_0)$, $i = 1,2,3,4$. Then, $L(A) = \int L_S(A) d\mu(S)$, where $L_S(A) = L_S(A) L_{S^c}(A)$ with

$$L_S(A) = q_1^{\tau|S|} q_3^{(T-\tau)|S|} \prod_{\{i,j\}\in S} \left(\frac{p_1}{p_0 q_1}\right)^{\sum_{t=1}^{\tau} A_{ij}^t} \left(\frac{p_3}{p_0 q_3}\right)^{\sum_{t=\tau+1}^{T} A_{ij}^t}$$

and

$$L_{S^c}(A) = q_2^{\tau|S^c|} q_4^{(T-\tau)|S^c|} \prod_{\{i,j\}\in S^c} \left(\frac{p_2}{p_0 q_2}\right)^{\sum_{t=1}^{\tau} A_{ij}^t} \left(\frac{p_4}{p_0 q_4}\right)^{\sum_{t=\tau+1}^{T} A_{ij}^t}.$$

We need to find the second moment of $L(A)$. Let S_1 and S_2 be two independent copies of S, then $\mathrm{E}_{\mathrm{P}_0}[L^2(A)] = \int \mathrm{E}_{\mathrm{P}_0}[L_{S_1}(A) L_{S_2}(A)] d\mu(S_1) d\mu(S_2)$. We can write

$$\mathrm{E}_{\mathrm{P}_0}[L_{S_1}(A) L_{S_2}(A)] = \mathrm{E}_{\mathrm{P}_0}[L_{S_1\cap S_2}(A) L_{S_1^c\cap S_2^c}(A) L_{S_1\triangle S_2}(A)]$$

where

$$L_{S_1\cap S_2}(A) = q_1^{\tau|S_1\cap S_2|} q_3^{(T-\tau)|S_1\cap S_2|} \prod_{\{i,j\}\in S_1\cap S_2} \left(\frac{p_1^2}{p_0^2 q_1^2}\right)^{\sum_{t=1}^{\tau} A_{ij}^t} \left(\frac{p_3^2}{p_0^2 q_3^2}\right)^{\sum_{t=\tau+1}^{T} A_{ij}^t},$$

$$L_{S_1^c\cap S_2^c}(A) = q_2^{\tau(|S_1^c\cap S_2^c|} q_4^{(T-\tau)(|S_1^c\cap S_2^c|)} \prod_{\{i,j\}\in S_1^c\cap S_2^c} \left(\frac{p_2^2}{p_0^2 q_2^2}\right)^{\sum_{t=1}^{\tau} A_{ij}^t} \left(\frac{p_4^2}{p_0^2 q_4^2}\right)^{\sum_{t=\tau+1}^{T} A_{ij}^t}$$

and

$$L_{S_1^c\triangle S_2^c}(A) = (q_1 q_2)^{\tau|S_1\triangle S_2|} (q_3 q_4)^{(T-\tau)|S_1\triangle S_2|} \prod_{\{i,j\}\in S_1\triangle S_2} \left(\frac{p_1 p_2}{p_0^2 q_1 q_2}\right)^{\sum_{t=1}^{\tau} A_{ij}^t} \left(\frac{p_3 p_4}{p_0^2 q_3 q_4}\right)^{\sum_{t=\tau+1}^{T} A_{ij}^t}$$

Taking into account the relations $p_1 - p_0 = (1-\tau/T)\rho_n\lambda$, $p_2 - p_0 = -(1-\tau/T)\rho_n\lambda$, $p_3 - p_0 = -(\tau/T)\rho_n\lambda$, $p_4 - p_0 = (\tau/T)\rho_n\lambda$ and Lemma EC.2.2, we obtain

$$\mathrm{E}_{\mathrm{P}_0}[L_{S_1}(A) L_{S_2}(A)] = f(\tau, p_0)^{|S_1\cap S_2| + |S_1^c\cap S_2^c|} g(\tau, p_0)^{|S_1\triangle S_2|}$$

where

$$f(\tau, p_0) = \left(1 + \left(1 - \frac{\tau}{T}\right)^2 \frac{\lambda^2\rho_n^2}{p_0(1-p_0)}\right)^{\tau} \left(1 + \left(\frac{\tau}{T}\right)^2 \frac{\lambda^2\rho_n^2}{p_0(1-p_0)}\right)^{T-\tau}$$

and

$$g(\tau, p_0) = \left(1 - \left(1 - \frac{\tau}{T}\right)^2 \frac{\lambda^2\rho_n^2}{p_0(1-p_0)}\right)^{\tau} \left(1 - \left(\frac{\tau}{T}\right)^2 \frac{\lambda^2\rho_n^2}{p_0(1-p_0)}\right)^{T-\tau}.$$

We have

$$\mathrm{E}_{\mathrm{P}_0}[L_{S_1}(A)L_{S_2}(A)] = \exp\{|S_1 \cap S_2| + |S_1^c \cap S_2^c| \log f(\tau, p_0) + |S_1 \triangle S_2| \log g(\tau, p_0)\}.$$

Using the bounds

$$\log f(\tau, p_0) \le \left(\tau\left(1-\frac{\tau}{T}\right)^2 + (T-\tau)\left(\frac{\tau}{T}\right)^2\right)\frac{\lambda^2\rho_n^2}{p_0(1-p_0)}$$

and

$$\log g(\tau, p_0) \le -\left(\tau\left(1-\frac{\tau}{T}\right)^2 + (T-\tau)\left(\frac{\tau}{T}\right)^2\right)\frac{\lambda^2\rho_n^2}{p_0(1-p_0)},$$

and the fact that $|S_1 \triangle S_2| + |S_1 \cap S_2| + |S_1^c \cap S_2^c| = N$, we obtain

$$\mathrm{E}_{\mathrm{P}_0}[L_{S_1}(A)L_{S_2}(A)] \le \exp\left[q^2\left(\frac{\tau}{T}\right)\frac{T\lambda^2\rho_n^2}{p_0(1-p_0)}(2|S_1 \cap S_2| + 2|S_1^c \cap S_2^c| - N)\right].$$

Note that $p_0 = \rho_n/2 < 1/2$. Thus, in order to bound the second moment likelihood ratio, we need to control the exponential moment

$$\mathrm{E}_{\mathrm{P}_0}[L^2(A)] \le \mathrm{E}_{S_1,S_2}\left\{\exp\left[\left(2|S_1 \cap S_2| + 2|S_1^c \cap S_2^c| - N\right)4Tq^2\left(\frac{\tau}{T}\right)\lambda^2\rho_n\right]\right\}.$$

We need to control the exponential moment of the random variable $U = |S_1 \cap S_2| + |S_1^c \cap S_2^c|$, where S_1 and S_2 are independent random variables distributed according to μ. Following the last lines of Lemma 4.9 in (Klopp et al. 2017), denote by $\xi^{(1)} = (\xi_1, \ldots, \xi_n)$ and $\xi^{(2)} = (\xi_1, \ldots, \xi_n)$ the assignment vectors corresponding to the variables S_1 and S_2, respectively. For any $(i, j) \in \{1, 2\}^2$ introduce the random variable that counts the number of nodes in the classes i and j according to the first and the second assignement:

$$N_{ij} = \left|\left\{a \in [n] : \ \xi_a^{(1)} = i, \ \xi_a^{(2)} = j\right\}\right|, \quad (i, j) \in \{1, 2\}^2.$$

Then $2|S_1 \cap S_2| + n = N_{11}^2 + N_{12}^2 + N_{21}^2 + N_{22}^2$ and $2|S_1^c \cap S_2^c| = 2N_{11}N_{22} + 2N_{12}N_{21}$. Hence, $2U + n = (N_{11} + N_{22})^2 + (N_{12} + N_{21})^2$. Note that $N_{11} + N_{22} + N_{12} + N_{21} = n$. Let $Z := N_{11} + N_{22} - n/2$. It is a centered binomial random variable with parameters (n, 1/2) and

$$2U - N = (n/2 + Z)^2 + (n/2 - Z)^2 - n - N = 2Z^2 - n/2.$$

Consequently, we need to control the exponential moment of Z^2:

$$\mathrm{E}_{\mathrm{P}_0}[L^2(A)] \le \mathrm{E}\exp\left[8Tq^2\left(\frac{\tau}{T}\right)\lambda^2\rho_n Z^2\right].$$

Using Hoeffding's inequality, we can show that $P(Z^2 > t) \leq 2e^{-2t/n}$, thus Z^2 is subexponential with the moments $E[Z^{2k}] \leq n^k k!$. Consequently, for any γ_n such that $0 < n\gamma_n < 1$ we have

$$Ee^{\gamma_n Z^2} \leq 1 + \sum_{k=1}^{+\infty} \frac{\gamma_n^k E[Z^{2k}]}{k!} \leq \sum_{k=0}^{+\infty} (n\gamma_n)^k = \frac{1}{1 - n\gamma_n}.$$

Set $\gamma_n = 8Tq^2\left(\frac{\tau}{T}\right)\lambda^2\rho_n$. We can see that if $n\gamma_n \leq 4(1-\eta)^2(1+4(1-\eta)^2)^{-1} < 1$, then $E_{P_0}[L^2(A)] \leq 1 + 4(1-\eta)^2$. Since $\delta^2(W_1, W_2) = 4\lambda^2$, we have $\gamma_n = 2Tn\rho_n q^2\left(\frac{\tau}{T}\right)\delta^2$ and the above condition on $n\gamma_n$ implies the lower bound

$$q\left(\frac{\tau}{T}\right)\delta(W_1, W_2) \leq \frac{\sqrt{2}(1-\eta)}{(1+4(1-\eta)^2)^{1/2}} \frac{1}{\sqrt{n\rho_n T}}$$

and the second part of the theorem follows. □

EC.5. Auxiliary results

EC.5.1. Concentration inequalities for matrix processes

The first result is the concentration inequality for the operator norm of a random matrix with independent entries (see Bandeira and van Handel (2016), Corollary 3.12 and Remark 3.13):

Proposition EC.1 (**Bandeira and Van Handel, 2016**). *Let W be an $m \times m$ symmetric matrix whose entries W_{ij} are independent centered random variables bounded (in absolute value) by some $\sigma_* > 0$. Then, for any $0 < \epsilon \leq 1/2$ there exists a universal constant c_ϵ such that, for every $x \geq 0$*

$$P\left\{\|W\|_{2\to 2} \geq 2\sqrt{2}(1+\epsilon)\sigma + x\right\} \leq m \exp\left(-\frac{x^2}{c_\epsilon \sigma_*^2}\right),$$

where $\sigma = \max_i \left[\sum_j \mathrm{var}(W_{ij})\right]^{1/2}$.

Next we apply this concentration inequality to the Matrix CUSUM statistics.

Let $X^t \in [-1,1]^{n\times n}$ $(1 \leq t \leq T)$ be a sequence of matrices with independent entries X^t_{ij} for any $1 \leq i, j \leq n$ and for any $t = 1, \ldots T$. Assume that X^t_{ij} are centered Bernoulli random variables taking values in $\{1 - B^t_{ij}, -B^t_{ij}\}$ with success probability B^t_{ij}. Let π denote a permutation of $\{1, \ldots, n\}$. Consider the following centered matrix processes defined for $1 \leq t \leq T-1$ in (7) and (EC.11):

$$\xi(t) = \sqrt{\frac{t(T-t)}{T}}\left(\frac{1}{t}\sum_{s=1}^{t} X^s - \frac{1}{T-t}\sum_{s=t+1}^{T} X^s\right),$$

$$\xi^\pi(t) = \sqrt{\frac{t(T-t)}{T}}\left(\frac{1}{t}\sum_{s=1}^{t} X^s - \frac{1}{T-t}\sum_{s=t+1}^{T} X^s \circ \pi\right).$$

LEMMA EC.9. *For any $\epsilon \in (0, 1/2]$ there exists an absolute constant C_ϵ such that, for every $\delta \in (0,1)$ we have*

$$\|\xi(t)\|_{2\to 2} \le 2\sqrt{2}(1+\epsilon)\sqrt{\frac{t(T-t)}{T}}\left[\frac{1}{t^2}\sum_{s=1}^{t}\|B^s\|_{1,\infty} + \frac{1}{(T-t)^2}\sum_{s=t+1}^{T}\|B^s\|_{1,\infty}\right]^{1/2} + C_\epsilon \log\frac{2n}{\delta}$$

with the probability larger than $1-\delta$.

Proof. The result follows from the direct application of Proposition EC.1. Since X^s are independent, we can easily estimate $\sigma^2 = \max_i \sum_j \mathrm{Var}[\xi_{ij}(t)]$ from above. Indeed,

$$\mathrm{Var}[\xi_{ij}(t)] = \frac{t(T-t)}{T}\left(\frac{1}{t^2}\sum_{s=1}^{t} B^s_{ij}(1-B^s_{ij}) + \frac{1}{(T-t)^2}\sum_{s=t+1}^{T} B^s_{ij}(1-B^s_{ij})\right).$$

Therefore, since $0 \le B_{ij} \le 1$,

$$\sigma^2 \le \frac{t(T-t)}{T}\left(\frac{1}{t^2}\sum_{s=1}^{t}\|B^s\|_{1,\infty} + \frac{1}{(T-t)^2}\sum_{s=t+1}^{T}\|B^s\|_{1,\infty}\right).$$

We will show now that the norm $\|\xi(t)\|_\infty$ is bounded by some σ^* with high probability. Consider the entries of the matrix $\xi(t)$ defined by

$$\xi_{ij}(t) = \sqrt{\frac{t(T-t)}{T}}\left(\frac{1}{t}\sum_{s=1}^{t} X^s_{ij} - \frac{1}{T-t}\sum_{s=t+1}^{T} X^s_{ij}\right) = \sum_{s=1}^{T} V^s_{ij}$$

Since $-B^s_{ij} \le X^s_{ij} \le 1 - B^s_{ij}$ we have $a_s \le V^s_{ij} \le b_s$ with

$$b_s - a_s \le \sqrt{\frac{t(T-t)}{T}}\left(\frac{1}{t}\mathbf{1}_{\{1\le s\le t\}} + \frac{1}{T-t}\mathbf{1}_{\{t<s\le T\}}\right).$$

Since X^s_{ij} are independent, applying the Hoeffding inequality, we obtain for any $x > 0$ that for any $1 \le i, j \le n$,

$$\mathrm{P}\left\{|\xi_{ij}(t)| > x\right\} \le 2\exp\left\{-\frac{2x^2}{\sum_s (b_s - a_s)^2}\right\} = 2e^{-2x^2}.$$

Using the union bound, we get that

$$\mathrm{P}\left\{\|\xi(t)\|_\infty > x\right\} \le n^2\mathrm{P}\left\{|\xi_{ij}(t)| > x\right\} = 2n^2 e^{-2x^2}.$$

Consequently, for any $\delta \in (0,1)$ we have $\|\xi(t)\|_\infty \le \log^{1/2}(2n/\sqrt{\delta}) \le \log^{1/2}(2n/\delta)$ with probability larger than $1-\delta/2$. Applying Proposition EC.1 given the event $\left\{\|\xi(t)\|_\infty \le \sigma_*\right\}$ with $\sigma_* = \log^{1/2}(2n/\delta)$ we get

$$\mathrm{P}\left\{\|\xi(t)\|_{2\to 2} \ge 2\sqrt{2}(1+\epsilon)\sigma + x\right\} \le n\exp\left\{-\frac{x^2}{c_\epsilon \log(2n/\delta)}\right\} + \frac{\delta}{2}.$$

Choosing $x = \sqrt{c_\epsilon}\log(2n/\delta)$ and $C_\epsilon = c_\epsilon^{1/2}$ we obtain the statement of the lemma. □

Lemma EC.10. *For any $\epsilon \in (0, 1/2]$ there exists an absolute constant C_ϵ such that, for every $\delta \in (0,1)$ we have*

$$\sup_\pi \|\xi^\pi(t)\|_{2\to 2} \le 2\sqrt{2}(1+\epsilon)\sqrt{\frac{t(T-t)}{T}}\Big(\frac{1}{t}\Big[\sum_{s=1}^{t}\|B^s\|_{1,\infty}\Big]^{1/2} + \frac{1}{(T-t)}\Big[\sum_{s=t+1}^{T}\|B^s\|_{1,\infty}\Big]^{1/2}\Big) + C_\epsilon \log\frac{4n}{\delta}$$

with probability larger than $1-\delta$.

Proof. Using the triangle inequality, for any fixed permutation π, we get

$$\|\xi^\pi(t)\|_{2\to 2} \le \sqrt{\frac{t(T-t)}{T}}\Big(\Big\|\frac{1}{t}\sum_{s=1}^{t} X^s\Big\|_{2\to 2} + \Big\|\frac{1}{T-t}\sum_{s=t+1}^{T} X^s \circ \pi\Big\|_{2\to 2}\Big).$$

The invariance of the operator norm under permutations implies

$$\|\xi^\pi(t)\|_{2\to 2} \le \sqrt{\frac{t(T-t)}{T}}\Big(\Big\|\frac{1}{t}\sum_{s=1}^{t} X^s\Big\|_{2\to 2} + \Big\|\frac{1}{T-t}\sum_{s=t+1}^{T} X^s\Big\|_{2\to 2}\Big)$$

Applying Proposition EC.1 to each term of the right-hand side of the above inequality as it is done in the proof of Lemma EC.9 we get the result.

EC.5.2. Result on the Hadamard product of two matrices

Lemma EC.11. *Let $A = (A_{ij}) \in [0,\infty)^{n\times n}$ and $B = (B_{ij}) \in \mathbb{R}^{n\times n}$. Assume that* $\operatorname{diag}(B) = 0$. *Then,*

$$\|A \odot B\|_{2\to 2} \ge \frac{\min\limits_{(ij):\, i\ne j} A_{ij}}{\sqrt{r \vee 1}}\|B\|_{2\to 2}, \tag{EC.13}$$

where $r = \operatorname{rank}(A\odot B)$. Moreover, if A and B are symmetric, we have that

$$\|A \odot B\|_{2\to 2} \ge \frac{\min_{ij} A_{ij}}{2\sqrt{r^* \vee 1}}\|B\|_{2\to 2}, \tag{EC.14}$$

where

$$r^* = \min_{M=(M_{ij}):\, M_{ij}=(A\odot B)_{ij}\, for\, i\ne j} \operatorname{rank}(M).$$

Proof. If $\min\limits_{(ij):\, i\ne j} A_{ij} = 0$, then the statement of the Lemma is trivially true. Now assume that $\min\limits_{(ij):\, i\ne j} A_{ij} > 0$. We have that

$$\begin{aligned}
\|B\|^2_{2\to 2} &\le \|B\|_F^2 \\
&= \sum_{i\ne j} A_{ij}^2 (A_{ij}^{-1})^2 B_{ij}^2 \\
&\le \max_{i\ne j}(A_{ij}^{-1})^2 \|A\odot B\|_F^2 \\
&\le r \max_{i\ne j}(A_{ij}^{-1})^2 \|A\odot B\|^2_{2\to 2}
\end{aligned}$$

which implies (EC.13). On the other hand, let M be a solution to

$$M \in \underset{M=(M_{ij}):\, M_{ij}=(A\odot B)_{ij} \text{ for } i\neq j}{\arg\min} \operatorname{rank}(M).$$

Let $\operatorname{rank}(M) = r^* < r$. We have that

$$\begin{aligned}
\|B\|_{2\to 2}^2 &\le \|B\|_F^2 \\
&= \sum_{i\neq j} A_{ij}^2 (A_{ij}^{-1})^2 B_{ij}^2 \\
&\le \max_{i\neq j} (A_{ij}^{-1})^2 \left\{ \sum_{i\neq j} \left(A_{ij} B\right)_{ij}^2 + \sum_i M_{ii}^2 \right\} \\
&= \max_{i\neq j} (A_{ij}^{-1})^2 \|M\|_F^2 \\
&\le r^* \max_{i\neq j} (A_{ij}^{-1})^2 \|M\|_{2\to 2}^2
\end{aligned}$$

which implies

$$\begin{aligned}
\|B\|_{2\to 2} &\le \sqrt{r^*} \max_{i\neq j} (A_{ij}^{-1}) \|M\|_{2\to 2} \\
&\le 2\sqrt{r^*} \max_{i\neq j} (A_{ij}^{-1}) \|A\odot B\|_{2\to 2}
\end{aligned} \tag{EC.15}$$

where in the last inequality we use that $\|M\|_{2\to 2} \le 2\|A\odot B\|_{2\to 2}$. To prove it, using the triangle inequality, it is enough to prove that $\|M - A\odot B\|_{2\to 2} \le \|A\odot B\|_{2\to 2}$. Let denote by $\lambda_1(X) \le \lambda_2(X) \le \cdots \le \lambda_n(X)$ the eigenvalues of a symmetric matrix X. Then, using Weyl's inequality, we have that

$$\begin{aligned}
\lambda_{j+k-n}(A\odot B) &\le \lambda_j(A\odot B - M) + \lambda_k(M) \\
&\le \lambda_{j+k-1}(A\odot B).
\end{aligned} \tag{EC.16}$$

Note that $r^* < r$ implies that there exist a k such that $\lambda_k(M) = 0$ but $\lambda_k(A\odot B) \neq 0$. Assume first that $\|M - A\odot B\|_{2\to 2} = -\lambda_1(M - A\odot B)$. Then, taking in (EC.16) $j = 1$ we get $-\lambda_1(M - A\odot B) \le \lambda_k(A\odot B) \le \|A\odot B\|_{2\to 2}$. Now, if $\|M - A\odot B\|_{2\to 2} = \lambda_n(M - A\odot B)$, taking $j = n$ we also get $\lambda_n(M - A\odot B) \le -\lambda_k(A\odot B) \le \|A\odot B\|_{2\to 2}$.

To conclude the proof, note that (EC.15) implies (EC.14). □

EC.6. Numerical experiments

In this section, we provide the study of numerical performance of our method. For each setting we applied three tests: the test $\psi^{\tau}_{n,T}$ at the given change-point τ defined in (9), the test $\psi_{n,T}$ over the dyadic grid $\mathcal{T}^d$ defined in (13) (we add the point $\lfloor T/2 \rfloor$ to the grid in our simulations) and the test $\psi^{full}_{n,T}(Y)$ based on the maximum over the whole set $\mathcal{D}_T = \{1, \dots, T-1\}$. Each test is calibrated to the significance level $\alpha = 0.05$.

The test defined in (9) is based on the threshold (10) depending on some given value $\epsilon \in (0, 1/2]$ and an unknown universal constant C_ϵ. To overcome this difficulty, we use a slightly different threshold obtained from the matrix Bernstein inequality (see, for example, Theorem 1.4 in (Tropp 2011)). The threshold for the test $\psi^{\tau}_{n,T}$ is given by

$$\tilde{H}^{\tau}_{\alpha,n,T} = \frac{1}{3}\frac{\log(n/\alpha)}{\sqrt{T}q(\tau/T)} + \Big(\frac{1}{9}\frac{\log^2(n/\alpha)}{Tq^2(\tau/T)} + 2\kappa_n \log(n/\alpha)\Big)^{1/2}.$$

For test over the dyadic grid $\psi_{n,T}(Y)$ and the test $\psi^{full}_{n,T}(Y)$ we use the same threshold

$$\tilde{H}_{\alpha,n,T}(t) = \frac{1}{3}\frac{\log(n|\mathcal{T}|/\alpha)}{\sqrt{T}q(t/T)} + \Big(\frac{1}{9}\frac{\log^2(n|\mathcal{T}|/\alpha)}{Tq^2(t/T)} + 2\kappa_n \log(n/\alpha)\Big)^{1/2}, \quad t \in \mathcal{T}.$$

In order to compare the performance of the tests under different regimes (n, T, τ), we introduce “energy-to-noise ratio" defined by

$$\mathrm{ENR} := \mathrm{ENR}_{n,T}(\tau/T, \Delta\Theta^{\tau}) = \frac{q(\tau/T)\|\Delta\Theta^{\tau}\|_{2\to 2}}{\sqrt{\kappa_n/T}}.$$

This ratio provides a numerical upper bound on the minimax testing constant (see Theorem 1). We denote by ENR^{τ}, ENR^{d}, ENR^{f} the minimal detectable ENR for the tests $\psi^{\tau}_{n,T}$, $\psi_{n,T}$, and $\psi^{full}_{n,T}$, respectively. Here “detectable" means that the average power of the corresponding test is equal to 1 over 100 simulations. Note that the lower bound constant for any test of level α with $\beta = 0$ is equal to $c^* = \log^{1/4}(1 + (1-\alpha)^2)/(4\sqrt{2})$ (see Theorem 2).

EC.6.1. Results for Stochastic Block Models

In this section, we apply our method to four different scenarios of Stochastic Block Models (SBM). Recall that for an SBM model with K communities and connection probability matrix Q between the communities the matrix of connection probabilities Θ is defined as $\Theta = Z^T Q Z$, where $Z \in \{0, 1\}^{K\times n}$ is the membership matrix. Each row i of the matrix Z contains only zeros except one entry Z_{ij} that is equal to 1 if the node i belongs to the community j. We suppose in these simulations that the membership matrix Z does not change.

Scenario 1: SBM with 2 communities, change in connection probability between communities. We suppose that the network follows the Stochastic Block Model with two balanced communities (block sizes are $\lfloor n/2 \rfloor$ and $n - \lfloor n/2 \rfloor$.) The probabilities of connection between the communities change at some point and are given by the following matrices Q_1 (before the change) and Q_2 (after the change point):

$$Q_1 = \rho_n \begin{pmatrix} 0.6 & 1 \\ 1 & 0.6 \end{pmatrix}, \quad Q_2 = \rho_n \begin{pmatrix} 0.6 & \delta \\ \delta & 0.6 \end{pmatrix}, \quad \delta \in [0, 1].$$

Scenario 2: SBM with 2 communities, change in connection probability within one community. As in the previous scenario, we assume that the network follow the stochastic block model with two balanced communities. In this scenario, the matrices Q_1 (before the change) and Q_2 (after the change point) are defined by

$$Q_1 = \rho_n \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.6 \end{pmatrix}, \quad Q_2 = \rho_n \begin{pmatrix} \delta & 0.5 \\ 0.5 & 0.6 \end{pmatrix}, \quad \delta \in [0, 1].$$

Scenario 3: SBM with 2 communities, change in connection probability within two communities. Same setting as before, but now connection probabilities inside of both communities change:

$$Q_1 = \rho_n \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}, \quad Q_2 = \rho_n \begin{pmatrix} \delta & 0.2 \\ 0.2 & \delta \end{pmatrix}, \quad \delta \in [0, 1].$$

Scenario 4: SBM with 3 communities and change in connection probability between communities. We suppose that the network follow the stochastic block model with three balanced communities (the block sizes are $k_1 = k_2 = \lfloor n/3 \rfloor$ and $k_3 = n - k_1 - k_2$). The probabilities of connection between the communities change at some point and are given by the following matrices Q_1 (before the change) and Q_2 (after the change point)

$$Q_1 = \rho_n \begin{pmatrix} 0.6 & 1 & 0.6 \\ 1 & 0.6 & 0.5 \\ 0.6 & 0.5 & 0.6 \end{pmatrix}, \quad Q_2 = \rho_n \begin{pmatrix} 0.6 & 1-\delta & 0.6 \\ 1-\delta & 0.6 & 0.5+\delta \\ 0.6 & 0.5+\delta & 0.6 \end{pmatrix}, \quad \delta \in [0, 0.5].$$

This scenario was considered in (Yu et al. 2021) with the constant parameter $\delta = 0.5$. In our study we vary δ and report the test power in terms of the change in the energy.

Scenario 5: SBM with 2 communities, change in connection probability within communities. We suppose that the network follows the Stochastic Block Model with two balanced communities (block sizes are $\lfloor n/2 \rfloor$ and $n - \lfloor n/2 \rfloor$.) The probabilities of connection within the communities change at some point and are given by the following matrices Q_1 (before the change) and Q_2 (after the change point):

$$Q_1 = \rho_n \begin{pmatrix} 0.6 & 1 \\ 1 & 0.6 \end{pmatrix}, \quad Q_2 = \rho_n \begin{pmatrix} \delta & 1 \\ 1 & \delta \end{pmatrix}, \quad \delta \in [0, 1].$$

The sparsity ρ_n is set to $n^{-1/2}$. The sparsity parameter κ_n is set to $\kappa_n = \frac{1}{2} n\rho_n(\|Q_1\|_{1,\infty} \vee \|Q_2\|_{1,\infty})$ for Scenarios 1–3, 5 and to $\kappa_n = \frac{1}{3} n\rho_n(\|Q_1\|_{1,\infty} \vee \|Q_2\|_{1,\infty})$ for Scenario 4. In all the scenarios we have $\kappa_n \leq \sqrt{n}$.

EC.6.1.1. Varying *n* and *T*. In this part we study the dependency of the energy-to-noise ratio ENR on n and T. We report the results of simulations for five scenarios in Table EC.1. We see that globally the ENR decreases when the number of observations T increases. Some changes cannot be detected by our tests. For example, for Scenarios 1, 2 and 4 the change-point is undetectable for $T = 20$ and $n = 100$ by any test. It can be explained by the small number of observations T implying the threshold that is systematically greater than the value of the test statistic. Scenario 4 seems to be more difficult than the other ones, it might be, in particular, due to the fact that the allowed changes are within the interval $[0, 0.5]$ that is smaller than in Scenarios 1–3.

The ENR of the test $\psi^{\tau}_{n,T}$ is always smaller than the ENR of two other tests. Concerning the tests over the dyadic grid and over the whole set of observations, the test $\psi_{n,T}$ outperforms the test $\psi^{full}_{n,T}$ in the majority of parameter settings and scenarios.

EC.6.1.2. Estimating the sparsity. In this section, we study the performance of our tests with the thresholds based on the estimated sparsity parameter $\widehat{\kappa}_n$. Taking the maximum of $1,\infty$-norms $\max_t \max_j A^t_{\cdot j}$ as an estimator will systematically overestimate κ_n. Indeed, since $\mathrm{E}\max_j \xi_j = \sqrt{2\log n}(1 + o(1))$, for $\xi_j \sim \mathcal{N}(0, 1)$ i.i.d., using the Gaussian approximation for binomial variables we can show that $\mathrm{E}(\max_j A^t_{\cdot j}) \asymp \kappa_n + \sqrt{2\kappa_n \log n}/n$. To estimate κ_n, we first calculate $A^t_{\cdot j} = \sum_{i=1}^n A^t_{ij}$ for each $t = 1, \dots, T$. Then, we obtain a robust estimator of the sparsity taking the 0.9-level empirical quantile of $A^t_{\cdot j}$. The final estimator of κ_n maximizes the obtained estimated sparsities $\widehat{\kappa}^t_n$ over t: $\widehat{\kappa}_n = \max_t Q\left(\left\{\sum_{i=1}^n A^t_{ij},\ j = 1, \dots, n\right\}, 0.9\right)$. Here $Q(Z, \alpha)$ denotes the α-level empirical quantile of the sample Z. The relative risk of estimation of κ_n is shown for different values of n and sparsity

Table EC.1 The results of simulations for four proposed scenarios for $n = 100$ and 150.

	n	T	ENR$^\tau$	ENR^d	ENR^f		n	T	ENR$^\tau$	ENR^d	ENR^f
Scenario 1	100	20	NA	NA	NA	Scenario 2	100	20	NA	NA	NA
$\tau/T = 0.5$	100	50	2.1546	2.4397	2.6298	$\tau/T = 0.5$	100	50	2.1550	2.4766	2.7018
	100	100	2.0612	2.4197	2.6885		100	100	2.1834	2.5018	2.7292
	100	250	2.0546	2.4089	2.6923		100	250	2.0857	2.5172	2.8049
	150	20	2.1928	NA	NA		150	20	2.2389	NA	NA
	150	50	2.1363	2.4515	2.6266		150	50	2.1812	2.5387	2.7175
	150	100	2.0801	2.4763	2.6745		50	100	2.1239	2.5284	2.7812
	150	250	2.0360	2.4276	2.7408		150	250	2.1588	2.5586	2.7984
Scenario 3	100	20	2.2098	NA	NA	Scenario 4	100	50	NA	NA	NA
$\tau/T = 0.5$	100	50	2.1253	2.4495	2.6296	$\tau/T = 0.5$	100	100	NA	NA	NA
	100	100	2.0377	2.4452	2.6999		100	150	2.0235	NA	NA
	100	250	2.0137	2.4146	2.7386		100	250	2.0483	2.3748	2.7013
	150	20	2.2528	NA	NA		150	50	NA	NA	NA
	150	50	2.1212	2.4814	2.6815		150	100	NA	NA	NA
	150	100	2.1508	2.4904	2.7168		150	150	2.0664	2.4236	NA
	150	250	2.0583	2.4163	2.7743		150	250	2.0420	2.3713	2.2007

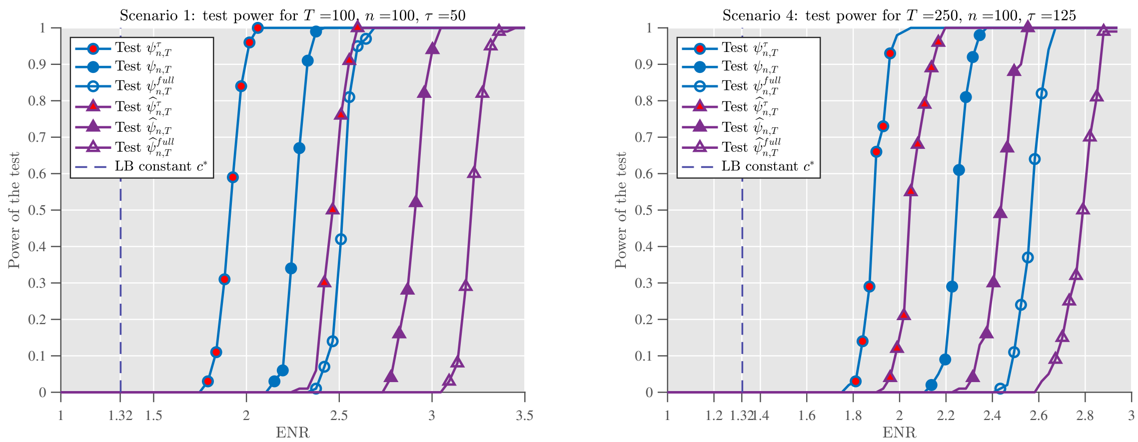
ρ_n in Table EC.2. We can see that the choice of the quantile $Q = 0.9$ guarantees $\widehat{\kappa}_n \geq \kappa_n$ which is important to maintain the test significance at level smaller than α and it also has a good estimation precision for a large range of graph sparsity regimes.

Table EC.2 Approximation of the relative risk of estimation $R(\widehat{\kappa}_n) = \mathrm{E}(\widehat{\kappa}_n - \kappa_n)/\kappa_n$ over 100 simulations for $T = 10$, $\tau = T/2$, $n \in \{100, 200, 500, 1000, 1200\}$, three different levels of sparsity ρ_n and the quantiles $Q \in \{0.7, 0.8, 0.9\}$.

Sparsity	Q	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 1200$
	0.7	-0.0334	-0.0352	-0.0407	-0.0294	-0.0212
$\rho_n = n^{-0.1}$	0.8	-0.0018	0.0070	-0.0082	-0.0200	-0.0067
	0.9	0.0586	0.0545	0.0230	0.0029	0.0154
	0.6	0.0239	0.0090	-0.0008	-0.0066	-0.0114
$\rho_n = n^{-0.25}$	0.8	0.0350	0.0823	0.0460	0.0310	0.0269
	0.9	0.1342	0.1631	0.1029	0.0746	0.0677
	0.7	0.1299	0.1412	0.0837	0.0806	0.0850
$\rho_n = n^{-0.5}$	0.8	0.2712	0.2189	0.1800	0.1489	0.1486
	0.9	0.4892	0.3719	0.3396	0.2623	0.2577

In Fig. EC.1 we compare the performance of the test adaptive to the unknown sparsity level with the test where we use the true value κ_n. We consider Scenario 1 with $n = 100$, $T = 100$ and Scenario 4 with $n = 100$, $T = 250$. For Scenario 1, $\kappa_n \approx 0.8n\rho_n$ and, for Scenario 4, $\kappa_n \approx 0.77n\rho_n$. In our simulations the change-point is located in the middle. For Scenario 1 with $T = 100$, our estimator $\widehat{\kappa}_n$ slightly overestimates the true value $\kappa_n = 7.94$ with the average value $\widehat{\kappa}_n = 12.9129$ calculated over 100 simulations and over all values of the parameter δ. For Scenario 4 and $T = 250$ we obtain

Figure EC.1 The test powers with known (tests ψ) and estimated sparsity parameters (tests $\widehat{\psi}$) for $n = 100$, $\tau/T = 0.5$ and $T = 100$ for Scenario 1 (on the left) and $T = 250$ for Scenario 4 (on the right).



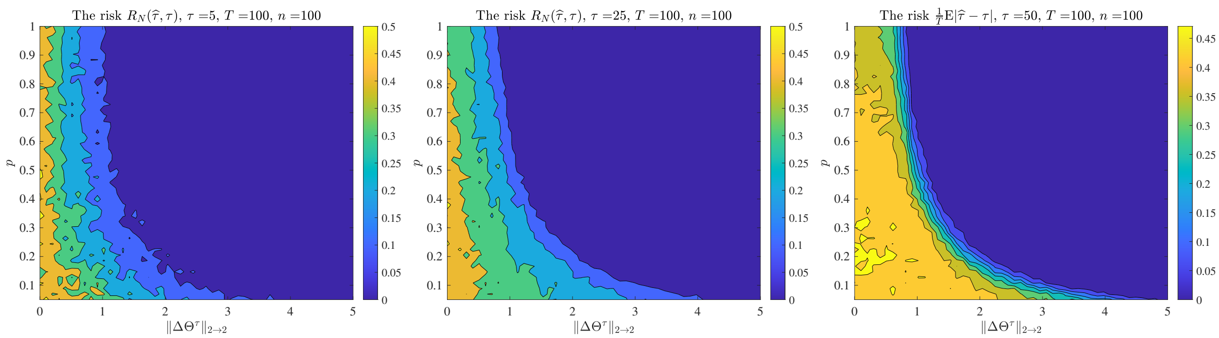
a better estimation that is equal to $\widehat{\kappa}_n = 11.6858$ while the true value $\kappa_n = 7.26$. For Scenario 1, the tests adaptive to the unknown sparsity level behave quite well with a reasonable power and with the ENR that is about 1.25 times greater than the ENR of the corresponding test with known κ_n. For Scenario 4 the proportion between the ENRs is about 1.07 times. Better performances of the adaptive test for Scenario 4 is expected as in this case the change in the sparsity level of the network is less important than in the Scenario 1. Our test construction is based on the upper bound for the sparsity level for all t and naturally gives better results in the case of the change which is more homogeneous in terms of sparsity.

EC.6.2. Coping with missing links

EC.6.2.1. Change-point localization in the case of uniform sampling. We have simulated the networks of size $n = 100$ from Scenario 1 with $T = 100$ and a change-point located at $\tau \in \{5, 25, 50\}$. The networks have missing links generated according to the uniform sampling matrix $\Pi = p_n(\mathbf{1}_n\mathbf{1}_n^t - \mathrm{id}_n)$ with the sampling rate $p_n \in (0, 1]$. We compute the average absolute error of our estimator $\widehat{\tau}$ defined in (16) over $N = 100$ simulations normalized by the number of observations T: $R_N(\widehat{\tau}, \tau) = (NT)^{-1} \sum_{i=1}^{N} |\widehat{\tau}_i - \tau|$. We present the dependence of this risk on the sampling rate p_n and on the norm of the jump $\Delta\Theta^\tau$.

In Fig. EC.2 we observe the dependence of the risk on the location τ: the closer τ is to the middle of the interval, the easier the estimation is. The dependence of the rate of convergence of $\widehat{\tau}$ on the norm $\|\Pi \odot \Delta\Theta^\tau\|_{2\to 2} = p_n\|\Delta\Theta^\tau\|_{2\to 2}$ is represented by the level curve $p_n\|\Delta\Theta^t\|_{2\to 2} \approx \mathrm{const}$ separating the black area corresponding to a low change-point localization error from the light one with the higher error.

Figure EC.2 The risk of the change-point estimator under Scenario 1 for $T = 100$, $n = 100$ and the change-points $\tau \in \{5, 25, 50\}$ (left to right). The links are observed at the constant sampling rate $p_n \in (0, 1]$. The estimation is easier when the change-point is located in the middle (the graph to the right).



EC.6.2.2. Change-point localization for non-uniform sampling patterns. We have simulated the networks of size $n = 100$ with the change point in the middle of $T = 100$ observations following three different sampling patterns:

Setting A. Change-point in missing communication.

The networks before and after the change follow Scenario 1. The sampling matrix $\Pi = Z^T\tilde{\Pi}Z$ has the same community structure as the networks before and after the change and follows “missing in communication" pattern: $\tilde{\Pi} = \begin{pmatrix} 1 & p \\ p & 1 \end{pmatrix}$. This example was considered in Section 3.4.

Setting B. Change-point in communication, within groups missing values.

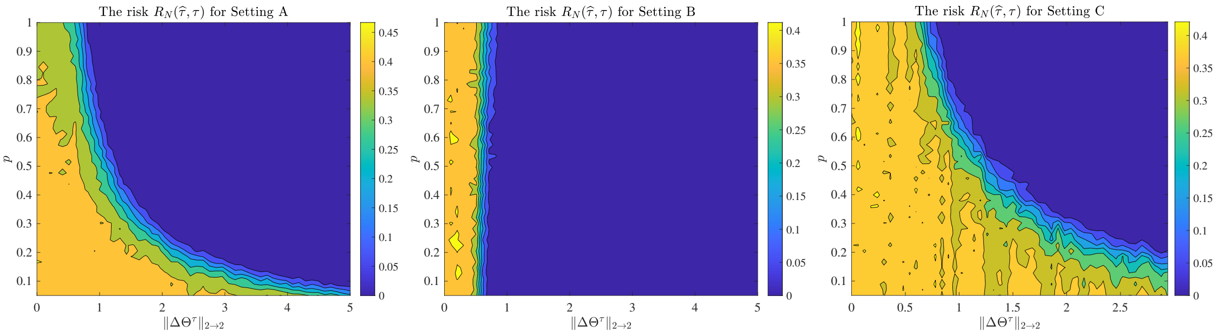
The networks before and after the change follow Scenario 1. The sampling matrix $\Pi = Z^T\tilde{\Pi}Z$ has the same community structure as the networks before and after the change and follows “missing within groups" pattern: $\tilde{\Pi} = \begin{pmatrix} p & 1 \\ 1 & p \end{pmatrix}$.

Setting C. Change-point and missing links withing communities.

The networks before and after the change follow Scenario 5. The sampling matrix $\Pi = Z^T\tilde{\Pi}Z$ has the same community structure as the networks before and after the change and follows “missing in communication" pattern: $\tilde{\Pi} = \begin{pmatrix} p & 1 \\ 1 & 1-p \end{pmatrix}$.

In Fig. EC.3 we present the results of the simulations that show the dependence of the risk of estimation on the sampling probability p and the jump norm $\|\Delta\Theta\|_{2\to 2}$. Under Setting A, the distortion parameter is $\delta_n(\Pi, \Theta) = p/\sqrt{2}$ and we see that the level curve is given by $p\|\Delta\Theta^{\tau}\|_{2\to 2} \approx$ const since the links between the communities with a change-point are sampled at the uniform rate p. Under Setting B, the missing links do not affect the change-point estimation since $\|\Pi \odot \Delta\Theta\|_{2\to 2} = \|\Delta\Theta\|_{2\to 2}$, the level curve is constant and the change-point estimation risk is independent of the link

Figure EC.3 The risk of the change-point estimator under three different patterns of missing links for $T = 100$, $n = 100$ and $\tau = 50$ (Settings A–C, from left to right).



sampling. Finally, under Setting C, the distortion parameter is equal to $\delta_n(\Pi, \Theta) = (p \wedge 1-p)/\sqrt{2}$ and we see a different level curve $(p \wedge 1-p)\|\Delta\Theta^\tau\|_{2\to 2} \approx$ const with the change in connection probabilities within blocks and links observed with different sampling probabilities p and $1-p$.

EC.6.3. Change-point detection under temporal dependence

Following the reviewer's suggestion, we study the robustness of our testing procedure to the temporal dependency in the observations. We observe a realization of the network $A = \{A^t,\ 1 \le t \le T\}$ without missing links and with temporal dependence for given nodes i, j. We suppose that the process A^t_{ij} is a Markov chain with values in $\{0, 1\}$ and the stationary connection probabilities Θ^0_{ij} for $1 \le t \le \tau$ and Θ^1_{ij} for $\tau < t \le T$. The Markov model for Bernoulli trials that we will use was proposed and studied by Klotz (1973). Let $\lambda \in (0, 1)$ be some given value and

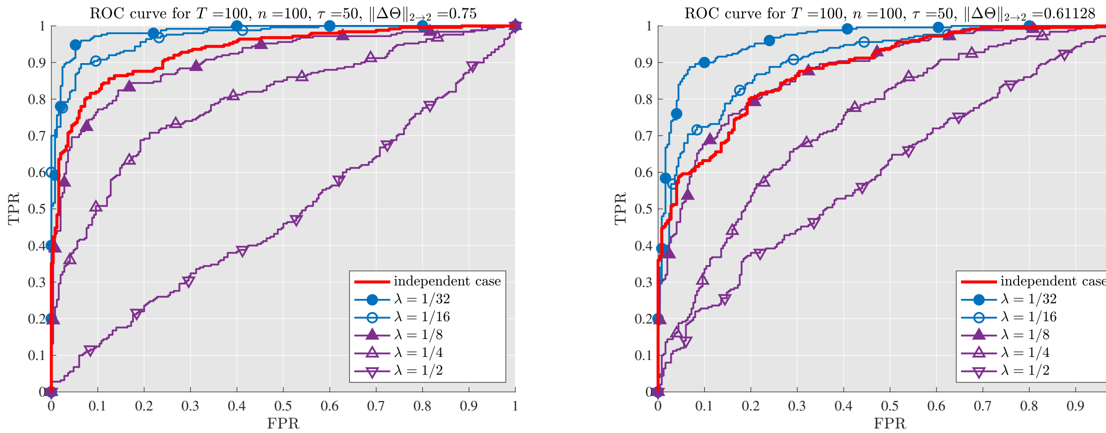
$$\Pi_{ij} = \begin{pmatrix} 1-(1-\lambda)\frac{\Theta_{ij}}{1-\Theta_{ij}} & (1-\lambda)\frac{\Theta_{ij}}{1-\Theta_{ij}} \\ 1-\lambda & \lambda \end{pmatrix}$$

be the transition matrix of the Markov chain (A^t_{ij}) for given i,j with the connection probability $\Theta_{ij} = \Theta^0_{ij}\mathbf{1}_{\{1\le t\le\tau\}} + \Theta^1_{ij}\mathbf{1}_{\{\tau<t\le T\}}$. Note that for $\lambda = \Theta_{ij}$ we have $\Pi_{ij} = \Theta_{ij}$ and the observations A^t_{ij} are independent. We have $\mathrm{P}\left\{A^{t+1}_{ij} = 1 | A^t_{ij} = 1\right\} = \lambda$ and

$$\mathrm{P}\left\{A^{t+1}_{ij} = 1 | A^t_{ij} = 0\right\} = (1-\lambda)\frac{\Theta_{ij}}{1-\Theta_{ij}}.$$

It can be easily seen that if $\lambda > \Theta_{ij}$ the probability to observe a link between i and j given the presence of a link in the past is greater than Θ_{ij} and the probability of absence of link given the absent link in the past is greater than $1-\Theta_{ij}$. Thus, in this case we will have the observations $\{A^t_{ij}, 1 \le t \le T\}$ of zeros and ones that form clusters. Vice versa, if $\lambda < \Theta_{ij}$, it will be less probable that the observations

Figure EC.4 The test power for the Markov dependent networks with $n = 100$ vertexes and $\lambda = 0.6$ following Scenario 1 (on the left) for $T = 100$, and Scenario 4 (on the right) for $T = 250$.

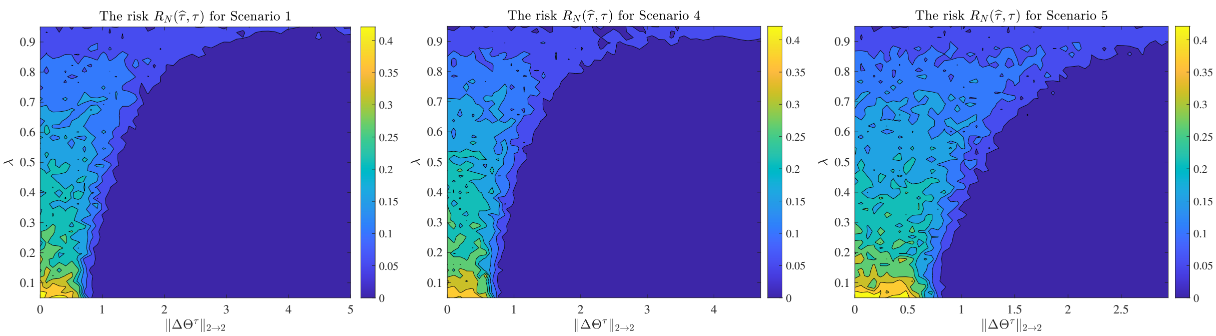


form clusters. The correlations between X_t and X_{t+h} are given by $\rho_{ij}(X_t, X_{t+h}) = \left(\frac{\lambda - \Theta_{ij}}{1-\Theta_{ij}}\right)^{|h|}$, $h \in \mathbb{Z}$ and decrease exponentially with h. It suggests that for small values of λ, when the dependence becomes weak, the testing procedure would give better results rather than for the values of λ that are close to one.

We will use the Matrix CUSUM test statistic for detection of a change-point at a given location for different degrees of dependency. In order to check the test behavior under the dependency, we will trace the ROC curves for $\lambda \in \{2^{-k},\ 1 \le k \le 5\}$ and compare them to the ROC curve obtained for the independent case. We have performed 250 simulations of networks with and without a change following Scenarios 1 and 4 for $n = 100$, $T = 100$, and with the change-point located in the middle. Then, according to the values of the test statistic, we calculated the True Positive Rate (TPR), corresponding to the correctly detected alternative hypothesis of presence of a change and the False Positive Rate (FPR) that corresponds to incorrectly rejected null hypothesis. In Fig. EC.4 the ROC curves are presented for the SBM networks following Scenario 1 and 4. It turns out that for $\lambda = 1/2$, when the dependency is high, the test behaves almost as a random guess. On the other hand, for small values of $\lambda = 1/32, 1/16$ (the blue curves), the test statistic works better on the dependent data than for the independent case. It is known that the testing procedure can benefit from the presence of dependency, see, for example, (Enikeeva et al. 2020) where the influence of dependency on the change-point detection was studied in the case of Gaussian time series. Finally, we obseve that for $\lambda = 1/8$ the test statistic have the same performance as in the independent case.

We have also tested the influence of the dependency on the estimation procedure. We performed 100 simulations of networks following Scenarios 1, 4 and 5 for $T = 100$, $n = 100$ and the change-

Figure EC.5 The risk of the change-point estimator for $T = 100$, $n = 100$ and the change-point in the middle, $\tau = 50$ for Scenarios 1, 4, and 5.



point in the middle. In Fig. EC.5 we present the empirical risk of the change-point estimator $\widehat{\tau}$ depending on the norm of the jump $\|\Delta\Theta\|_{2\to2}$ and on $\lambda \in (0, 1)$. We see that the results for all scenarios show the same dependence of the risk on λ and $\|\Delta\Theta\|_{2\to2}$. In the detectability zone (deep blue and blue zones), we see that the performance of the estimator increases if λ decreases, as in the case of testing. If we compare the result for Scenario 1 to the one for the independent case presented in Fig. EC.3 in the middle (there is no influence of the missing links here), we see that beyond the detectability zone when $\|\Delta\Theta\|_{2\to2} < 0.5$, the change-point estimation benefits from the dependency if λ increases. Overall, the Matrix CUSUM statistic shows a pretty good performance for this model of dependency both in terms of estimation and testing.

EC.6.4. Results for graphon model

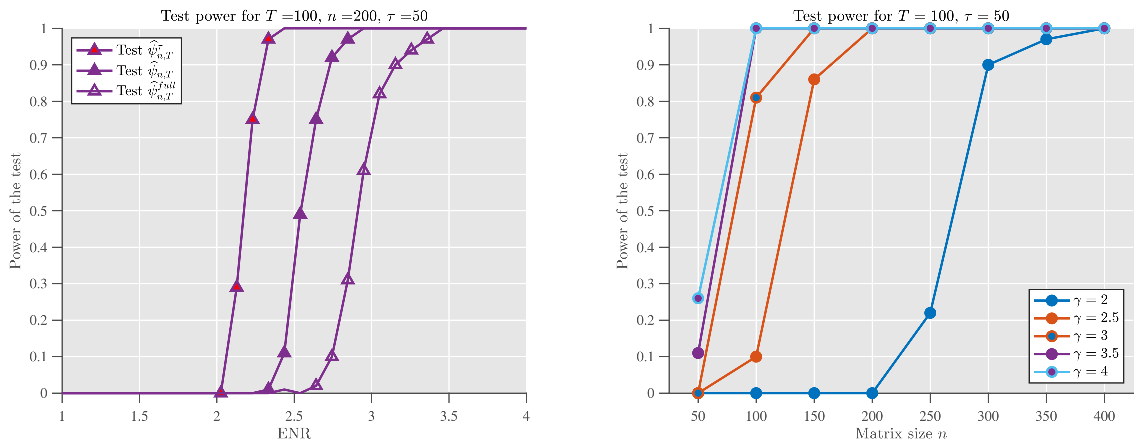
In this section, we simulate a dynamic network from graphon model with the graphon function in Hölder classes with

$$W_1(x, y) = xy \quad \text{before the change and} \quad W_2^\gamma(x, y) = (xy)^\gamma \quad \text{after the change.}$$

Here $\gamma \geq 1$ is the smoothness parameter that defines the impact of the change. We suppose that the assignment vector ε does not change. The sparsity parameter is set to $\rho_n = 1/\sqrt{n}$. The smoothness parameter γ varies from 1 to 5, the change-point is at the middle of the interval, that is $\tau = T/2$, and the number of observations is $T = 100$. The matrix size varies from 50 to 400.

On the left hand side of Fig. EC.6 we see the dependence of the power of three tests $\psi^\tau_{n,T}$, $\psi_{n,T}$, $\psi^{full}_{n,T}$ on the smoothness parameter γ. The results are similar to those obtained for the SBMs, the test over the dyadic grid outperforms the test over the whole grid and both are less powerful that testing at a given change-point τ. The change in graphons with smoothness $\gamma > 3$ can be detected with power close to 1. The graph on the right hand side of Fig. EC.6 shows the power of the

Figure EC.6 Power of testing the change in the Hölder class graphons at $\tau = T/2$, $T = 100$. The graph to the left displays the power for three different test for $n = 200$. The graph to the right shows the power of the test $\psi^{\tau}_{n,T}$ for different values of γ depending on the matrix size n.



test $\psi^{\tau}_{n,T}$ for different sizes of networks $n \in \{50, 100, 200, 300, 400\}$ and for different values of the smoothness γ. We can see that the detection power grows with n that confirms the detection rate $1/\sqrt{n\rho_n T}$. On the other hand, the smaller is the smoothness γ, the harder the detection will be. For example, if $\gamma = 4$, the detection power is 1 starting from $n = 100$ and for $\gamma = 2$, the detection power becomes close to 1 only for $n = 400$.