Change-point detection in dynamic networks with missing links

Farida Enikeeva
joint work with **Olga Klopp**

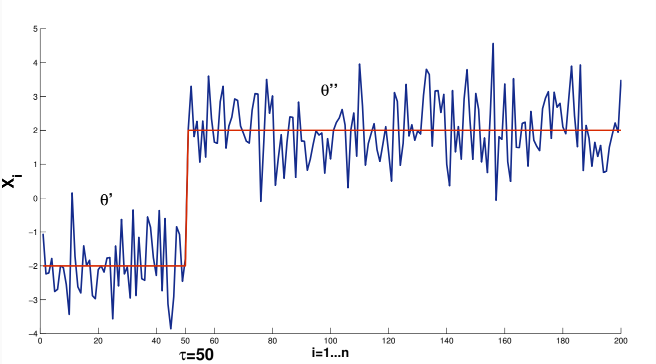
28 Juin 2022

Laboratoire de Mathématiques et Applications, Université de Poitiers
and CREST, ESSEC

Classical change-point problem

Gaussian data with **change-point in mean**:

$$X_i = \begin{cases} \theta' + \varepsilon\xi_i, & i \le \tau \\ \theta'' + \varepsilon\xi_i, & i > \tau \end{cases}, \quad \xi_i \sim \mathcal{N}(0, I_d),\ i = 1, \ldots, n$$



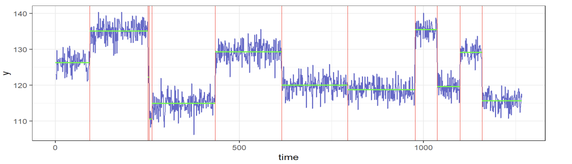
Goal

Propose a statistic that will detect the presence of a change-point and estimate it

Extensions of classical model

Classical example: Gaussian independent data with **changes in mean**

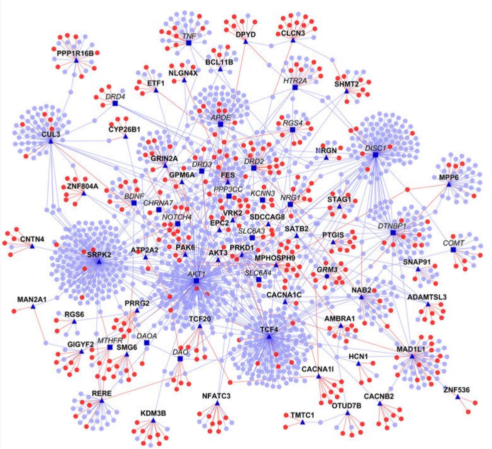
$$X_t = \sum_{k=1}^{K} \theta_k \mathbf{1}_{\tau_k \le t < \tau_{k+1}} + \xi_t, \quad t = 1, \ldots, T$$



Extensions:

- **Dependency in the data**: the noise ξ_t is a time series (ARIMA, GARCH...); changes in mean
- **Changes in mean and variance**
- **Changes in the parameter** of $X_1, \ldots, X_T$ (variance, parameter of ARMA), mean is constant
- **Non-Gaussian processes**: Poisson process, random graph models

Network analysis is an important research field driven by applications in social sciences, computer science, biology, and other fields.



Protein-Protein Interaction Network of Schizophrenia

Ganapathiraju et al, '16, *NPJ Schizophr.*

Real-life networks

- The observed network is modeled by a random graph
- The network can evolve over time $\Longrightarrow$ dynamic network
- The real-life network is usually sparse : the graph contains few links
- Many of the real-life networks are only partially observed $\Longrightarrow$ missing links
- The parameters of a networks might abruptly change

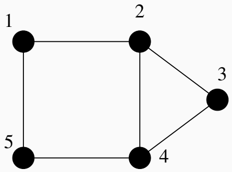
Goals

Detect and localize abrupt changes in the parameter matrix of a sparse dynamic network: **change-point detection and estimation**

Graph Notations

A (simple, undirected graph) $G = (E, V)$ consists of

- a set of vertices $V = \{1, \ldots n\}$
- a set of edges $E \subset \{\{i, j\} : i, j \in V \text{ and } i \neq j\}$



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

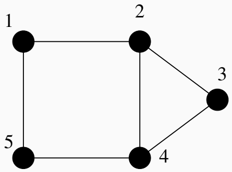
The corresponding adjacency matrix is denoted $A = (A_{ij}) \in \{0, 1\}^{n \times n}$, where

$$A_{i,} = 1 \iff (i, j) \in E$$

Graph Notations

A (simple, undirected graph) $G = (E, V)$ consists of

- a set of vertices $V = \{1, \dots n\}$
- a set of edges $E \subset \{\{i, j\} : i, j \in V \text{ and } i \neq j\}$



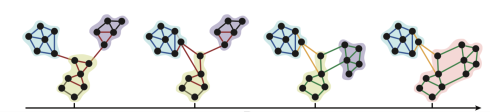
$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The corresponding adjacency matrix is denoted $A = (A_{ij}) \in \{0, 1\}^{n \times n}$, where

$$A_{i,} = 1 \iff (i, j) \in E$$

Inhomogeneous random graph model

Observations: a sequence of random graphs represented by their adjacency matrices $A = \left\{A^t,\ 1 \leq t \leq T\right\}$



- Simple undirected graphs G^t with n vertices
- Adjacency matrix $A^t = (A^t_{ij}) \in \{0,1\}^{n\times n}$:
    - The graph G^t contains the edge (i,j) iff $A^t_{ij} = 1$
    - A^t_{ij} are **independent Bernoulli** with the connection probability

$$\Theta^t_{ij} = \mathbb{P}(A^t_{ij} = 1), \quad 1 \leq j < i \leq n.$$

    - Connection probability matrix $\Theta^t = (\Theta^t_{ij}) \in [0,1]^{n\times n}$ is symmetric with zero diagonal
- Sparsity parameter: $\rho_n = \max_t \|\Theta^t\|_\infty$

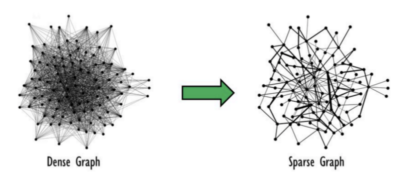
Modeling sparsity

Data: sequence of adjacency matrices $A = \{A^t,\ 1 \leq t \leq T\}$ with i.i.d. $A^t_{ij} \sim \mathrm{Bernoulli}(\Theta^t_{ij})$

Sparsity

- A sparse graph has a few links, connection probability tends to zero: $\Theta^t_{ij} \to 0$ as $n \to \infty$ for some (or all) (i, j)
- Sparsity parameter: $\rho_n = \max_t \|\Theta^t\|_\infty$ with the sparsity assumption

$$\rho_n \to 0, \quad n \to \infty$$



$\rho_n \gtrsim \frac{1}{\sqrt{n}}$ $\qquad$ $\rho_n \lesssim \frac{1}{\sqrt{n}}$

Change-point problem

“Signal + Noise” model:

$$A^t = \Theta^t + W^t, \quad 1 \leq t \leq T,$$

where W^t is a matrix of centered independent Bernoulli variables

The connection probability matrix Θ^t might abruptly change:

$$\Theta^t = \begin{cases} \Theta^0, & 1 \leq t \leq \tau \\ \Theta^0 + \Delta\Theta^\tau, & \tau + 1 \leq t \leq T \end{cases}$$

- The change-point τ is unknown
- Θ^0, $\Theta^0 + \Delta\Theta^\tau$ are the connection probability matrices before and after the change
- $\Delta\Theta^\tau$ is a symmetric jump matrix
- The amount of change is measured by $\|\Delta\Theta^\tau\|$

Goals:

Test $\mathrm{H}_0: \ \Delta\Theta^\tau = 0 \ \forall\tau$ vs $\mathrm{H}_1: \ \exists\tau : \|\Delta\Theta^\tau\| \geq r_{n,T} > 0$ and estimate τ

References

- **Change-point localization**
  - Bhattacharjee, Banerjee & Michailidis '20
    Single CP estimation in SBM
  - Wang, Yu & Rinaldo '21
    Multiple CP estimation, minimax rates
- **Change-point detection**
  - Chen, Zhou & Lin '20
    test statistic based on operator norm of the jump matrix estimator; asymptotic results
  - **Wang, Yu & Rinaldo '21**
    results on detection of zero CP
- **Two-sample testing**
  - Tang, Athreya, Sussman, Lyzinski & Priebe '17
    test statistic based on spectral embedding, bootstrap
  - **Ghoshdastidar, Gutzeit, Carpentier & von Luxburg '20**
    two tests, minimax separation rates

Problem Statement

We observe the sequence of adjacency matrices $\mathbf{A} = (A^1, \dots, A^T)$ such that $A^t = \Theta^t + W^T$, where

$$\Theta^t = \begin{cases} \Theta^0, & 1 \leq t \leq \tau \\ \Theta^0 + \Delta\Theta^\tau, & \tau + 1 \leq t \leq T \end{cases}$$

Define the set of parameters of a ρ_n-sparse network:

$$\mathcal{M}_n(n\rho_n) = \left\{ \Theta \in [0,1]^{n \times n} : \ \Theta = \Theta^T, \ \mathrm{diag}(\Theta) = 0, \ \|\Theta\|_{1,\infty} \leq n\rho_n \right\}$$

Testing problem

$$\mathrm{H}_0 : \ \Theta^0 \in \mathcal{M}_n(\rho_n), \ \Delta\Theta^\tau = 0 \quad \text{for all } 1 \leq \tau < T$$

against

$$\mathrm{H}_1 : \exists 1 \leq \tau < T : \ \Theta^0, \Theta^0 + \Delta\Theta^\tau \in \mathcal{M}_n(\rho_n) \text{ and}$$

$$q\left(\frac{\tau}{T}\right) \|\Delta\Theta^\tau\|_{2\to 2} \geq R_{n,T}$$

Change-point energy

- **The change-point energy** quantifies the detection difficulty:

$$\mathcal{E}(\tau) = q\left(\frac{\tau}{T}\right)\|\Delta\Theta^{\tau}\|_{2\to 2}.$$

- $q(t) = \sqrt{t(1-t)}$, $t \in [0,1] \implies$ impact of the CP location $\implies$ it is easier to detect/estimate a CP in the middle rather that at the border
- CP detection: testing whether the energy is zero or at least $R_{n,T} > 0$:

$$\mathrm{H}_0 : \ \mathcal{E}(\tau) = 0 \text{ for all } \tau \quad \text{against} \quad \mathrm{H}_1 : \ \mathcal{E}(\tau) \geq R_{n,T} \text{ for some } \tau$$

Goals

- Find a minimal detectable energy rate: $R_{n,T} \asymp ?$
- Propose an optimal test/estimator that will detects a CP having minimal detectable energy

References: testing problem

Partial answers

- **Optimal change point detection and localization in sparse dynamic networks** [Wang, Yu & Rinaldo '21]
  - Change-point localization and detection
  - Detection procedure requires two independent samples
  - Separation rate in terms of the Frobenius norm $\|\Delta\Theta\|_F$
  - Up to a $\log^{1+\epsilon} T$ factor in the localization rate
- **Two-sample hypothesis testing for inhomogeneous random graphs** [Ghoshdastidar, Gutzeit, Carpentier & von Luxburg '20]
  - Two-sample testing problem: τ is known
  - Separation rate for spectral and Frobenius norms and corresponding tests
  - Additional $\log n$ factor in the minimax detection rate

CUSUM statistic

Classical Gaussian case

- **Observations:** $X_i = \theta + \Delta\theta \mathbf{1}\{i > \tau\} + \xi_i$, $i = 1, \ldots, n$
- $\xi \sim \mathcal{N}_d(0, I_d)$ are i.i.d. d-dimensional Gaussian r.v.
- Means θ and $\theta + \Delta\theta$ before and after the change-point τ

CUSUM statistic

Compare means before and after each point $1 \leq t \leq T - 1$:

$$Z_T(t) = \sqrt{\frac{t(T-t)}{T}} \Big(\frac{1}{t} \sum_{i=1}^{t} X_i - \frac{1}{T-t} \sum_{i=t+1}^{T} X_i \Big)$$

- Change-point estimator $\widehat{\tau} = \arg \max_{1 \leq t \leq T-1} \|Z_T(t)\|_2$
- Test of level α with critical set $\Big\{ \|Z_T(t)\|_2 > q_{1-\alpha, T, d} \Big\}$

Test statistic

We observe $\mathbf{A} = (A^1, \ldots, A^T)$ such that $A^t = \Theta^t + W^t$, where at some unknown time $1 \leq \tau < T$ there is a change in Θ^t:

$$\Theta^t = \begin{cases} \Theta^0, & 1 \leq t \leq \tau \\ \Theta^0 + \Delta\Theta^\tau, & \tau + 1 \leq t \leq T \end{cases}$$

Matrix CUSUM statistic

$$Z_T(t) = \sqrt{\frac{t(T-t)}{T}} \left(\frac{1}{t} \sum_{s=1}^{t} A^s - \frac{1}{T-t} \sum_{s=t+1}^{T} A^s \right), \quad t = 1, \ldots, T-1.$$

- $Z_T(t)$: difference between the average number of connections before and after time t
- Change in Θ^t at time $\tau \implies$ the value of $Z_T(t)$ is maximal in the neighborhood of τ
- We detect a change in Θ^t at τ if $\|Z_T(\tau)\|$ is sufficiently large

Model for the Matrix CUSUM statistic

"Signal +noise" model for Z_T

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

- The centered noise is

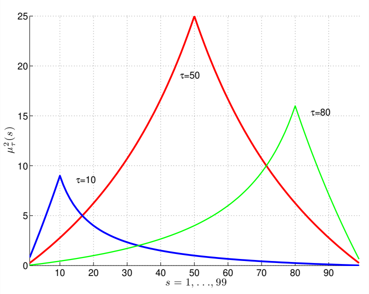
$$\xi(t) = \sqrt{\frac{t(T-t)}{T}} \left(\frac{1}{t} \sum_{s=1}^{t} W^s - \frac{1}{T-t} \sum_{s=t+1}^{T} W^s \right)$$

The true CP τ maximizes $\mu_T(t)$:

$$\max_{t=1,\dots,T-1} \mu_T(t) = \mu_T(\tau)$$

$$= \sqrt{\frac{\tau(T-\tau)}{T}} = \sqrt{T} q\left(\frac{\tau}{T}\right)$$

$\Longrightarrow$ CP energy: $q(\tau/T)\|\Delta\Theta^\tau\|$



What norm to choose?

The test is based on the norm of the Matrix CUSUM statistic:

$$\psi_{n,T} = \mathbf{1}\left\{\max_{t\in\mathcal{T}} \|Z_T(t)\| > H_{\alpha,n,T}\right\} \implies \text{what norm to choose?}$$

Intuitive idea

We have $Z_T(t) = -\mu_T(t)\Delta\Theta^\tau + \xi(t), \quad t = 1,\dots,T-1$

- $\|\Delta\Theta^\tau\|$ should be larger than $\|\xi(t)\|$
- What about the Frobenius norm? Assume that $\Theta^t_{ij} \approx \rho_n$. Then $\mathbb{E}[\xi^2_{ij}(t)] \approx \rho_n$ and

$$\frac{\mathbb{E}\|\xi(t)\|_F^2}{n^2} \approx \rho_n \gg \frac{\|\Delta\Theta^\tau\|_F^2}{n^2} \approx \rho_n^2$$

  $\implies$ the Frobenius norm is not a suitable one
- The operator norm of $\xi(t)$ can be controlled by the Matrix Bernstein inequality

  $\implies$ **choose $\|Z_T(t)\|_{\mathrm{op}}$ as a test statistic**

Testing for a change at a given point τ

- Calculate the Matrix CUSUM statistic:

$$Z_T(\tau) = \sqrt{\frac{\tau(T-\tau)}{T}} \left(\frac{1}{t} \sum_{s=1}^{\tau} A^s - \frac{1}{T-t} \sum_{s=\tau+1}^{T} A^s \right), \quad t = 1, \ldots, T-1.$$

- Calculate the operator norm of $Z_T(\tau)$: $\|Z_T(\tau)\|_{\mathrm{op}} := \sigma_{\mathsf{max}}(Z_T(\tau))$

Testing for a change at a given point τ

- The test of significance level α :

$$\boxed{\text{detect the change if } \; \|Z_T(\tau)\|_{\mathrm{op}} > H_{\alpha,n}}$$

- Approximate quantile $H_{\alpha,n}$ is found from the matrix Bernstein inequality:

$$H_{\alpha,n} = c_* \sqrt{n\rho_n \log(n/\alpha)}$$

where c_* is an absolute constant

Testing for a change at an unknown point

- Caculate the test statistics $\|Z_T(t)\|_{\mathrm{op}}$ for $t = 1, \ldots, T-1$.
- Take a grid $\mathcal{T}$ that approximates the set $1, \ldots, T$

Testing at an unknown point

- The test of significance level α :

$$\boxed{\text{detect the change if } \max_{t \in \mathcal{T}} \|Z_T(t)\|_{\mathrm{op}} > H_{\alpha,n,T}}$$

- $H_{\alpha,n} = c_* \sqrt{n\rho_n \log(n \log |\mathcal{T}|/\alpha)}$

Choice of the grid $\mathcal{T}$:

- If $\mathcal{T} = \{1, \ldots, T-1\} \implies \log T$ factor in the minimal detectable signal
- Dyadic grid (Liu, Gao & Samworth '21): take $\mathcal{T} = \mathcal{T}^L \cup \mathcal{T}^R$, where

$$\mathcal{T}^L = \left\{2^k,\ k = 0, \ldots, \lfloor \log_2\left(\frac{T}{2}\right) \rfloor\right\},\ \mathcal{T}^R = \left\{T - 2^k,\ k = 0, \ldots, \lfloor \log_2\left(\frac{T}{2}\right) \rfloor\right\}.$$

$\implies$ **only** $\log\log T$ **factor** in the minimal detectable signal

Theoretical results

Let $\alpha \in (0, 1)$ be a given significance level and $n \geq 1/\alpha$.

Minimal detectable energy

$$q(\tau/T)\|\Delta\Theta\|_{\text{op}} \asymp \sqrt{\frac{n\rho_n}{T}}$$

Testing at a given change-point τ:

- Assume that $q(\tau/T)\|\Delta\Theta\|_{\text{op}} \geq C_1\sqrt{\frac{n\rho_n}{T}\log(n/\alpha)}$.

  Then our test detects a change with type I and II errors $\leq \alpha$

Testing at an unknown change -point:

- Assume that $q(\tau/T)\|\Delta\Theta\|_{\text{op}} \geq C_2\sqrt{\frac{n\rho_n}{T}\log((\log T)n/\alpha)}$.

  Then the test over the dyadic grid detects a change with type I and II errors $\leq \alpha$

Change-point estimation

Change-point estimator

The estimator of τ maximizes the operator norm of the Matrix CUSUM test statistic:

$$\widehat{\tau}_n = \arg \max_{1 \le t \le T-1} \|Z_T(t)\|_{\mathrm{op}}$$

Theoretical guarantees (Enikeeva and Klopp '21):

- Let $\gamma \in (0,1)$. The estimated change-point $\widehat{\tau}$ satisfies

$$\mathbb{P}\left\{ \frac{1}{T}|\widehat{\tau} - \tau| \le 3\, c_* \frac{\left(\frac{n\rho_n}{T} \log(nT/\gamma)\right)^{1/2}}{q(\tau/T)\|\Delta\Theta\|_{\mathrm{op}}} \right\} \ge 1 - \gamma.$$

- **Wang et al (2021)**: if

$$q(\tau/T)\|\Delta\Theta\|_F \lesssim \sqrt{\frac{n\rho_n \log T}{T}}$$

then no consistent change-point estimator can exist.

Numerical experiments

We applied three tests to the observations $A^1, \dots, A^T$:

- test $\psi^{\tau}_{n,T}$ at the given change-point τ with the threshold
- test $\psi_{n,T}$ over the dyadic grid $\mathcal{T}$
- test $\psi^{full}_{n,T}(Y)$ based on the maximum over $\{1, \dots, T-1\}$

Parameters

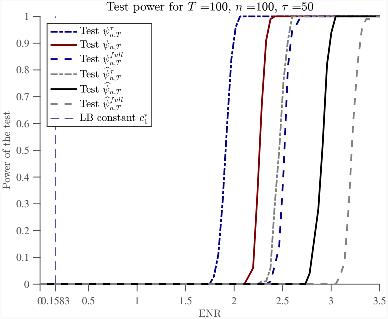
- Test are calibrated at the level $\alpha = 0.05$, the threshold is

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

- **Sparsity:** ρ_n is set to $n^{-1/2}$, $n\rho_n = n^{1/2}$
- **Estimated sparsity:** use the average link number $n\widehat{\rho}_n = \max_t Q\Big(\Big\{\sum_{i=1}^n A^t_{ij},\ j = 1, \dots, n\Big\}, 0.9\Big)$, where $Q(Z, \alpha)$ is the α-level empirical quantile of the sample Z.
- **"Energy-to-noise ratio"**: $\mathrm{ENR} := \frac{\mathcal{E}(\tau)}{\sqrt{n\rho_n/T}}$.

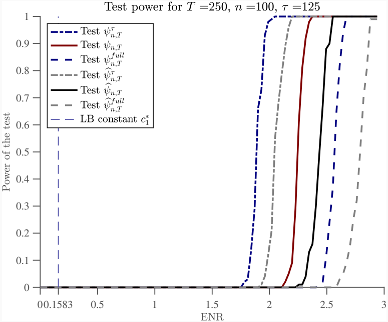
Adaptation to the unknown sparsity level

Test powers depending on the ENR for $n = 100$, $\tau/T = 0.5$ for known $n\rho_n$ and estimated sparsity



SBM with 2 communities and change in connection probability between communities, $T = 100$.

Sparsity: $n\rho_n = 7.94$, $n\widehat{\rho}_n = 12.91$



SBM with 3 communities and change in connection probability between communities, $T = 250$.

Sparsity: $n\rho_n = 7.26$, $n\widehat{\rho}_n = 11.68$

Transport for London (TfL) Open Data

London Bicycle Sharing Network data[1]

- Data collected since 2012
- ID of each bicycle
- ID and name of the origin and the destination trip stations
- journey (rental) starting and ending time and date
- ID and the duration of each trip

We focus on

- the period from June 24, 2012 to August 31, 2012
- London Olympics: July 27, 2012 – August 12, 2012

[1] https://api.tfl.gov.uk

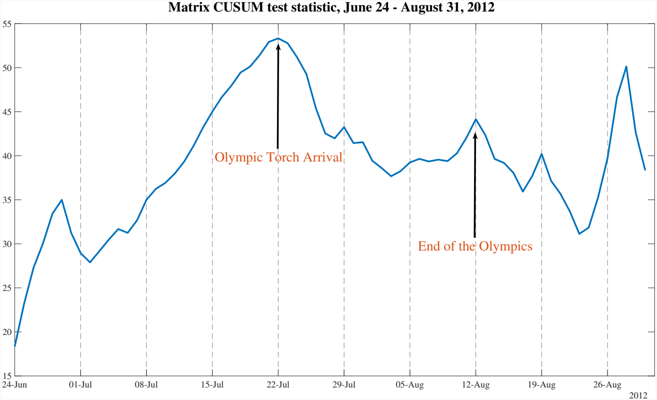
Transport for London (TfL) Open Data

Transport for London (TfL) Open Data

- Dynamic network: a sequence of $T = 69$ daily observations
- Each observation: a graph with $n = 595$ vertices corresponding to the bike rental stations
- Two vertices are connected if
  - the minimal trip duration is not less than 3 minutes
  - the number of trips is greater than a predefined threshold
    - 0.9975-level empirical quantile of the distribution of the total number of trips between every couple of stations
- Average number of links $n\rho_n = 43.2319$ (over $T = 69$ observations)
- Sparsity $\rho_n = 0.0727 \asymp n^{-0.4}$

Transport for London (TfL) Open Data

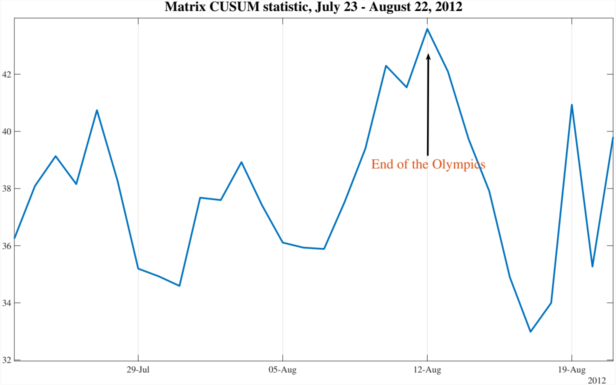
The values of the matrix CUSUM statistic during the whole period



The test detects the change at $\alpha = 0.05$; the estimated CP corresponds to the day of the arrival of the Olympic Torch (July, 22 2012).

Transport for London (TfL) Open Data

Zoom to the period of 31 day from July 23 to August, 22



The test detects the presence of a change. The CP estimator gives the End of the Olympics day. $\Longrightarrow$ Use multiple testing methods for multiple change-point localization

Summary

- Minimax separation rate for the change-point energy

$$q(\tau/T)\|\Delta\Theta^{\tau}\|_{2\to 2} \asymp \sqrt{\frac{n\rho_n}{T}}$$

  - Sparsity: $\rho_n = \max_t \|\Theta^t\|_\infty$
  - Change points that are away from the end points may be detected at lower size of jump in the parameter matrix

- Test based on the spectral norm of the Matrix CUSUM statistics
  - minimax optimal in T up to $\log\log T$; up to $\log n$ in n
  - robust to missing links
  - works for networks with changing size: generalization to graphon

- Localization of the change point