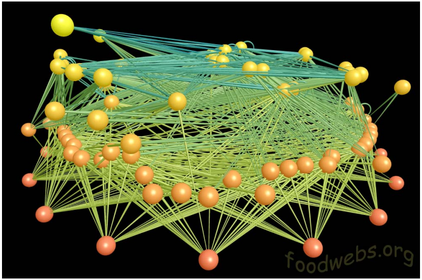
Detecting change points in dynamic networks

Olga Klopp

joint work with F. Enikeeva

Network model

Network analysis has become an important research field driven by applications in social sciences, computer sciences, statistical physics, biology,. . .



East-river trophic network [Yoon et al.(04)]

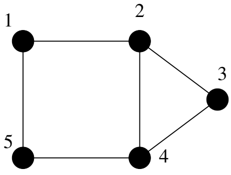
Approach

- The modeling of real networks as random graphs.
- Model-based statistical analysis.

Graph Notations

A (simple, undirected graph) $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ consists of

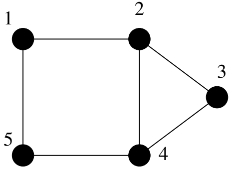
- a set of vertices $V = \{1, \ldots n\}$
- a set of edges $E \subset \{\{i,j\} : i,j \in V$ and $i \neq j\}$



Graph Notations

A (simple, undirected graph) $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ consists of

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



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

The corresponding adjacency matrix is denoted $\mathbf{A} = (\mathbf{A}_{i,j}) \in \{0,1\}^{n \times n}$, where $\mathbf{A}_{i,j} = 1 \Leftrightarrow (i,j) \in E$

Network Model

- Our observations: the adjacency $\mathbf{A} = (\mathbf{A}_{ij})$
- $\mathbf{A}_{ij}$ independent Bernoulli random variables with **probability of connection**

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

- $\mathbf{\Theta}$ $n \times n$ symmetric matrix with the coefficients $\mathbf{\Theta}_{ij}$ for $1 \leq j < i \leq n$ and zeros on the diagonal.

- **Inhomogeneous random graph model**

Dynamic networks

- Most of the real-life networks evolve over the time
- We observe a sequence of sparse graphs:

$$\mathbf{A}^1, \mathbf{A}^2, \ldots$$

- Each $\mathbf{A}^t$ = independent realization from an **unknown inhomogeneous random graph model**
- The underlying distribution of this sequence of graphs may change at some unknown time moment
- **Problem of detection of possible changes in a time sequence of networks**
  - intrusion detection
  - health care monitoring
  - fraud detection

Change Point Detection and Localization

Modeling dynamic networks

- "Signal + Noise" model of the observed data $\mathbf{A}^t$:

$$\mathbf{A}^t = \mathbf{\Theta}^t + \mathbf{W}^t \quad (1 \leq t \leq T),$$

- The connection probability matrix $\mathbf{\Theta}^t$ might change at an unknown time $1 \leq \tau \leq T-1$:

$$\mathbf{\Theta}^t = \mathbf{\Theta}^0 \mathbf{1}_{\{1 \leq t \leq \tau\}} + (\mathbf{\Theta}^0 + \Delta\mathbf{\Theta}^\tau)\mathbf{1}_{\{\tau+1 \leq t \leq T\}} \quad (t = 1, \ldots, T).$$

  - $\mathbf{\Theta}^0$ the connection probability matrix before the change
  - $\Delta\mathbf{\Theta}^\tau$ symmetric jump matrix of a change that occurs at time τ.

Change point Detection/Estimation in Dynamic Networks

- Detection:
  - **Chen et al (2019)**
  - **Wang et al (2021)**
  - ...
- Localization:
  - **Bhattacharjee et al (2018)**
  - **Wang et al (2021)**
  - ...
- Two-Sample Testing:
  - **Tang et al (2015)**
  - **Ghoshdastidar et al (2020)**
  - ...

Test problem

- We would like to test whether there is no change in $\mathbf{\Theta}^t$:

$$H_0 : \ \Delta\mathbf{\Theta}^\tau = 0 \quad \text{for all } \tau$$

against the alternative hypothesis of a change in $\mathbf{\Theta}^t$ at some point τ

$$H_1 : \|\Delta\mathbf{\Theta}^\tau\| \geq \rho_n \text{ for some } \tau.$$

- $\rho_n > 0$ is the minimal value of the jump that guarantees the change-point detection

Partial answer

Problem of the minimax distinguishability of hypotheses

- **Ghoshdastidar et al (2020)** (two sample test)
  - Spectral and Frobenius norm
  - Known τ
  - Additional $\log(n)$ factor
- **Wang et al (2021)** (localization)
  - Frobenius norm
  - up to a $\log$ factor
  - upper bound for a two-sample procedure

Sparsity

Main integral characteristics

- number of vertices n
- number of edges $|E|$

Sparsity

Main integral characteristics

- number of vertices n
- number of edges $|E|$

Maximal number of edges

$$\frac{n(n-1)}{2}$$

Sparsity

Main integral characteristics

- number of vertices n
- number of edges $|E|$

Maximal number of edges

$$\frac{n(n-1)}{2}$$

- Dense graph $|E| \sim n^2$
- Real world networks are sparse : $|E| = o(n^2)$
  - more difficult to handle

Modeling sparsity

- $\mathbf{\Theta}_{ij}^t \to 0$ as $n \to \infty$ for some (or all) (i,j)
- Usual assumption: $r_n = \max_t \|\mathbf{\Theta}^t\|_\infty$ and $r_n \to 0$ as $n \to \infty$
- **Sparsity:** $\kappa_n = \max_t \|\mathbf{\Theta}^t\|_{1,\infty} = \max_{t,j} \sum_i \mathbf{\Theta}_{ij}$:
  - $\kappa_n / n \to 0$ as $n \to \infty$
  - $\kappa_n \leq r_n n$
  - more flexible
  - can be easily estimated!

Energy of the change point

- Difficulty of assessing the existence of a change-point: *energy*

$$\mathcal{E}(\tau) = \mathbf{q}(\tau/\mathbf{T}) \|\boldsymbol{\Delta\Theta}^{\tau}\|_{\mathbf{2\to 2}}$$

- $q(t) = \sqrt{t(1-t)}$: quantifies the impact of change-point location on the difficulty of detecting the change
- Problem of testing whether the energy is zero or is at least $\rho_n > 0$:

$$H_0 : \ \mathcal{E}(\tau) = 0 \text{ for all } \tau$$

against the alternative hypothesis of a change in Θ^t:

$$H_1 : \ \mathcal{E}(\tau) \geq \rho_n > 0 \text{ for some } \tau$$

Minimax detection rate

- *Energy*: $\mathcal{E}(\tau) = q(\tau/T)\|\Delta\Theta^{\tau}\|_{2\to 2}$
- $H_0: \ \mathcal{E}(\tau) = 0$ for all τ against the alternative hypothesis

  $H_1: \ \mathcal{E}(\tau) \geq \rho_n > 0$ for some τ

Goal

Find the minimal amount of energy that guarantees the change-point detection

Minimax detection rate

- *Energy*: $\mathcal{E}(\tau) = q(\tau/T)\|\Delta\Theta^{\tau}\|_{2\to 2}$
- $H_0 : \ \mathcal{E}(\tau) = 0$ for all τ vs $H_1 : \ \mathcal{E}(\tau) \geq \rho_n > 0$ for some τ

Detection threshold (up to log factors)

$$\left(\frac{\kappa_n}{T}\right)^{1/2}$$

- $\kappa_n = \max_t \|\mathbf{\Theta}^t\|_{1,\infty}$: sparsity
- T time points
- Lower and upper bound:
  - if the energy of the change point is smaller $\Rightarrow$ we can not distinguish between these two hypothesis.
  - a test which satisfies the upper detection condition

Matrix CUSUM statistic

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

- Measures of the difference between the average number of connections before and after the point t
- A change of the parameter matrix Θ^t at time $\tau \Rightarrow$ the value of the process Z_T will be maximal in the neighborhood of τ
- If $\|Z_T(t)\|$ is sufficiently large at some point $t \Rightarrow$ there is a change in the connection matrix $\mathbf{\Theta}^t$ of the network
- **In what norm?**

Matrix CUSUM statistic

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

- "signal +noise" model

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

  - $\mu_T(t) = \sqrt{\frac{t(T-t)}{T}} \left((\tau/t)\mathbf{1}_{\{\tau+1 \le t \le T\}} + (T-\tau)/(T-t)\mathbf{1}_{\{1 \le t \le \tau\}} \right)$ (max at the true change-point $t = \tau$)
  - centered noise

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

In what norm?

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

- $\|\Delta\boldsymbol{\Theta}^\tau\|$ should be larger than $\|\xi(t)\|$
- We need to control $\|\xi(t)\|$
- **Matrix Hoeffding inequality:**

$$\mathbb{E}\|\xi(t)\|_{2\to 2} \leq 3\sqrt{2}\left(\sqrt{1-t/T} + \sqrt{t/T}\right)(\kappa_n)^{1/2}$$

- Can not work with Frobenius norm as in sparse regime: assuming $\Theta^t_{ij} \approx r_n \to 0$

$$\frac{\mathbb{E}\|\xi(t)\|_2^2}{n^2} \approx r_n \gg \frac{\|\Delta\Theta^\tau\|_2^2}{n^2} \approx r_n^2$$

Decision rule

- Matrix CUSUM statistics:

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

- Decision rule:

$$\psi_{n,T}(Y) = \mathbb{1}_{\left\{\max\limits_{t\in\mathcal{T}} \|Z_T(t)\|_{2\to 2} > H_{\alpha,n,T}\right\}},$$

- $\mathcal{T} = \mathcal{T}^L \cup \mathcal{T}^R$: dyadic grid of $\{1, \ldots, T-1\}$ where

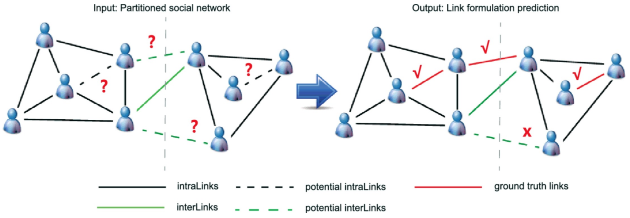
$$\mathcal{T}^L = \left\{2^k,\ k = 0, \ldots, \lfloor \log_2(T/2) \rfloor\right\}$$

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

Missing Links

Real-life networks are only partially observed

- Exhaustive exploration of all interactions in a network is expensive
- Survey data: non-response or drop-out of participants
- Online social network data: sub-sample of the network



A balanced modularity maximization link prediction model in social networks [Wu et al.(2017)]

Missing links

- Imputation methods require observations from a homogeneous distribution
- Change-point detection and estimation methods are designed for the case of complete observations
- Can we detect the change point from partial observations of our networks?
- How will missing links affect the detection rate?

Change-point localization

- An estimator of τ, $\widehat{\tau}_n$, such that $|\widehat{\tau}_n - \tau| \leq \epsilon$ with high probability
- $\epsilon/T =$ *localization rate*
- an estimator is consistent if, as $T \to \infty$, its localization rate vanishes
- Using CUSUM statistics: $\widehat{\tau}_n \in \arg\max_{1 \leq t \leq T-1} \|Z_T(t)\|_{2\to 2}$

Theorem (Enikeeva and K., 2021)

Let $\gamma \in (0,1)$ *and* $x^* = \tau/T$. *Then, the estimated change-point* $\widehat{x} = \widehat{\tau}/T$ *satisfies*

$$|\widehat{x} - x^*| \leq \frac{3\, c_* \sqrt{\omega_n \log(nT/\gamma)}}{\mathcal{E}(\tau)\sqrt{T}}$$

with probability larger than $1-\gamma$.

Wang et al (2021): if $\mathcal{E}(\tau) \leq \dfrac{\sqrt{\kappa_n}}{\sqrt{33T}}$, then no consistent change-point estimator can exist.

Numerical experiments

- We applied three tests:
  - test $\psi^{\tau}_{n,T}$ at the given change-point τ
  - test $\psi_{n,T}$ over the dyadic grid $\mathcal{T}^d$
  - test $\psi^{full}_{n,T}(Y)$ based on the maximum over the whole set $\{1,\dots,T-1\}$
- Each test is calibrated at the significance level $\alpha = 0.05$
- The sparsity ρ_n is set to $n^{-1/2}$
- "Energy-to-noise ratio"

$$\mathrm{ENR} := \frac{\mathcal{E}(\tau)}{\sqrt{\kappa_n/T}}.$$

Adaptation to the unknown sparsity level

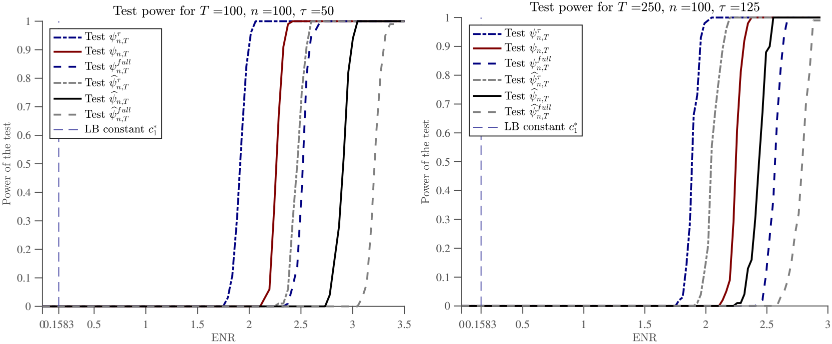


Figure: The test powers with known and estimated sparsity parameters for $n = 100$, $\tau/T = 0.5$. Left: SBM with two communities and change in connection probability between communities and $T = 100$. Right: SBM with three communities and change in connection probability between communities and $T = 250$.

Coping with missing links

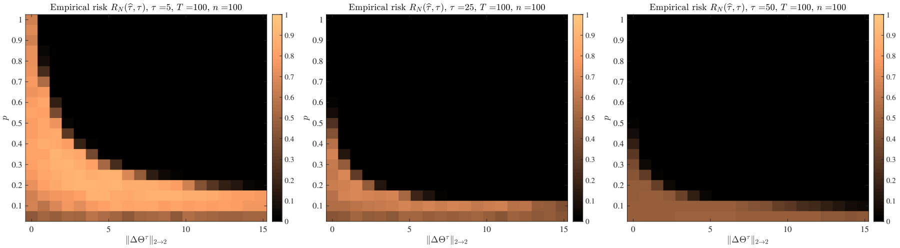


Figure: The risk of the change-point estimator for $T = 100$, $n = 100$ and $\tau = 5, 25, 50$ (left to right).

- We sample the links at the uniform rate p_n
- Average absolute error of $\hat{\tau}$ over $N = 100$ simulations: $R_N(\hat{\tau},\tau) = (NT)^{-1}\sum_{i=1}^{N}|\hat{\tau}_i - \tau|$
- dependence of the error on the sampling rate p_n and on the norm of the jump $\Delta\Theta^{\tau}$.

Transport for London (TfL) Open Data

- https://api.tfl.gov.uk
- Information about London Bicycle Sharing Network 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
- Two-month period from June 24, 2012 to August 31, 2012
- Games of the XXX Olympiad

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:
  - 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 sparsity $\bar{\kappa}_n = 43.2319$ (over $T = 69$ observations)
- The corresponding value of $\rho_n = \kappa_n / n = 0.0727 \asymp n^{-0.4}$

Transport for London (TfL) Open Data

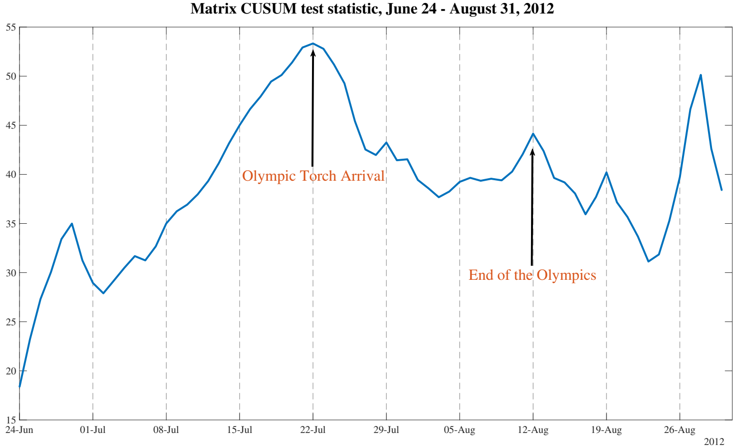


Figure: The value of the matrix CUSUM statistic calculated during the whole period of observations. The vertical grid lines correspond to Sundays.

Transport for London (TfL) Open Data

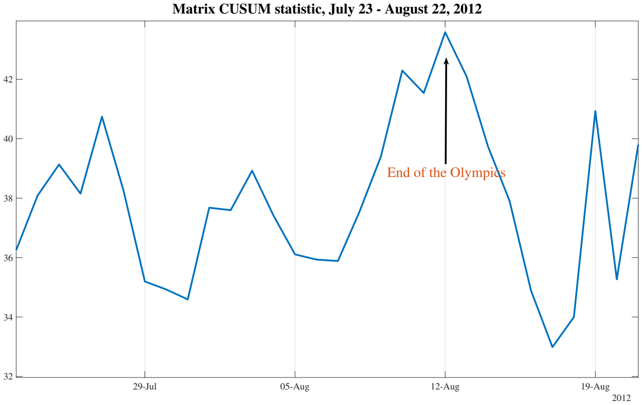


Figure: The value of the matrix CUSUM statistic calculated during 31 day from July 23 to August, 22.

Segmentation methods for multiple change-point localization

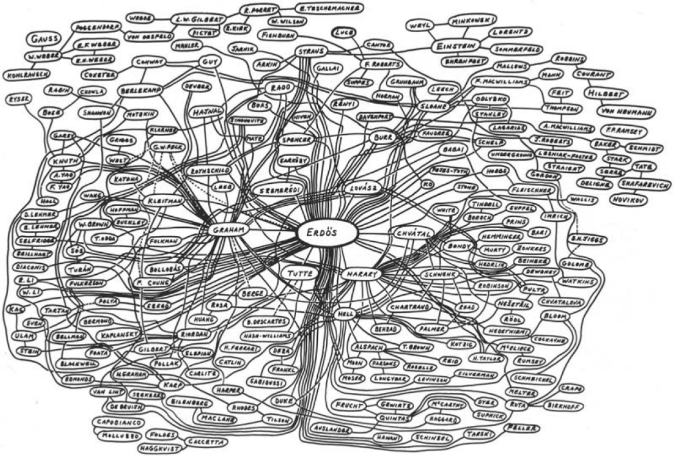
Summary

Problem of change point detection in dynamic network:

- **Minimax separation rate for the energy of the change point** $\mathcal{E}(\tau) = q(\tau/T)\|\Delta\Theta^{\tau}\|_{2\to 2}$

$$\sqrt{\kappa_n/T}$$

  - Sparsity: $\kappa_n = \max_t \|\boldsymbol{\Theta}^t\|_{1,\infty}$
  - Change points that are away from the end point 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
  - robust to missing links
  - works for networks with changing size
- Localisation of the change point(s)


ERDÖS
GRAHAM
HARARY
TUTTE
CHVÁTAL
LOVÁSZ
SZEMERÉDI
EINSTEIN
GAUSS
HILBERT
VON NEUMANN
MINKOWSKI
LORENTZ
SOMMERFELD
WEYL
STRAUS
KNUTH
CONWAY
GUY
BERLEKAMP
SHANNON
KLEITMAN
F. CHUNG
BERGE
ROSA
HAJNAL
RADO
SPENCER
BURR
SLOANE
BABAI
KAC
ULAM
STEIN
TURÁN
SÓS
KAPLANSKY
RIORDAN
GILBERT
POLLAK
KARP
DUKE
HELL
CHARTRAND
PALMER
ROTA
BIRKHOFF
FELLER
TARSKI
SCHINZEL
HANANI
QUINTAS
FRUCHT
MOSER
MOON
BLOOM
WATKINS
GORDON
DELIGNE
SHAFAREVICH
NOVIKOV
SERRE
TATE
SCHMIDT
F.P. RAMSEY
COURANT
ROBBINS
MANN
DAVENPORT
RÉNYI
GALLAI
CANTOR
F. ROBERTS
MAHLER
JARNIK
FIENBURG
W. WILSON
R. PORRET
E. TESCHEMACHER
L.W. GILBERT
WEBER
POGGENDORF
W. WEBER
KOHLRAUSCH
COXETER
RYSER
RABIN
CHOWLA
BOSE
HALL
D. LEHMER
SELFRIDGE
DIACONIS
R. LI
W. LI
EVEN
BELLMAN
BLACKWELL
EDMONDS
VAN LINT
DE BRUIJN
CAPOBIANCO
MOLLUZZO
FOLDES
CACCETTA
HAGGKVIST
MAC LANE
RHODES
HARPER
TILSON
FRANKL
B. DESCARTES
NASH-WILLIAMS
HUANG
KATONA
WANG
HOFFMAN
W. BROWN
FOLKMAN
BOLLOBÁS
ROTHSCHILD
SÁRKÖZY
SCHWENK
ROBINSON
NEŠETŘIL
RÖDL
ALSPACH
T. BROWN
REID
LEVINSON
GEWIRTZ
McCARTHY
HAGGARD
SUPNICK
DYER
MELTER
SCHMEICHEL
WHITE
BOESCH
PRINS
HEMMINGER
MURTY
BONDY
ZONKERS
BEINEKE
HEDETNIEMI
FLEISCHNER
STONE
TINDELL
WALLIS
GOLOMB
CHVATALOVA

Thank You !