CRIM

CENTRE DE
RECHERCHE INFORMATIQUE
DE MONTRÉAL

Utilisation du package rEMM pour la reconnaissance de motifs spatio-temporels

PRÉSENTÉ À : R à Quebec

PAR : Mohamed Dahmane

DATE : 15/05/2019

crim.ca

PRINCIPAL PARTENAIRE FINANCIER

Québec

Markov Chain vs Extented Markov Models (EMM)

Motivation

- **Markov Chains** are suitable to model complex systems
- **Spatio-temporal** event prediction: time may be difficult to modelize an update with MC structures
- **Clustering** alone neglects the **temporal proximity** of spatial data
- The EMM adds the temporal component by **superimposing a dynamically** adapting Markov Chain.
- The Extended Markov Models (EMM) are particularly well suited to model spatiotemporal data such as: Network traffic, environmental data, weather data etc ….

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KK} \end{pmatrix}$$

Extended Markov Models

EMM structure

- Is an evolving MC => EMM(t) = MC

Manipulations

- A data point can be **added** to an existing cluster,
- **delete/create** a new cluster,
- **merge/split** clusters,
- **fading** and **pruning** the cluster structure: Fading is achieved by reducing the weight of old observations

$$w_t = 2^{-\lambda t}$$

Example EMM

```
> EMMTraffic
   Loc_1 Loc_2 Loc_3 Loc_4 Loc_5 Loc_6 Loc_7
1     20    50   100    30    25     4    10
2     20    80    50    20    10    10    10
3     40    30    75    20    30    20    25
4     15    60    30    30    10    10    15
5     40    15    25    10    35    40     9
6      5     5    40    35    10     5     4
7      0    35    55     2     1     3     5
8     20    60    30    11    20    15    10
9     45    40    15    18    20    20    15
10    15    20    40    40    10    10    14
11     5    45    55    10    10    15     0
12    10    30    10     4    15    15    10
```

A → `> emm <- EMM(threshold=0.2, measure="eJaccard")`

B → `> build(emm, EMMTraffic)`

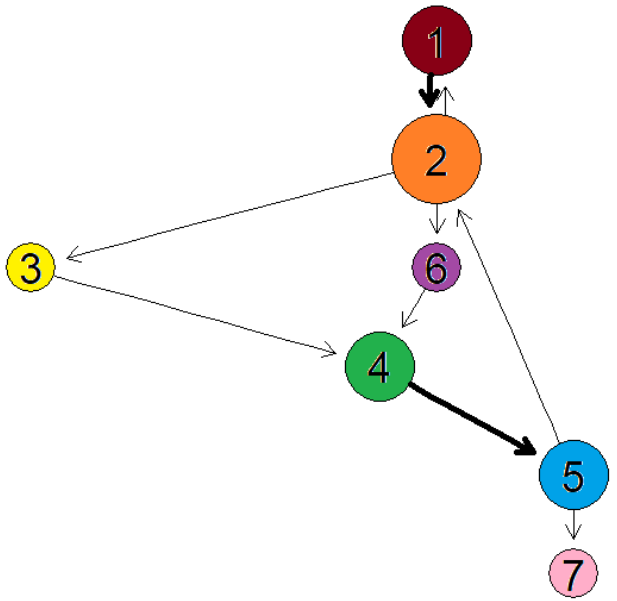
$$eJaccard = \frac{x * y}{x^2 + y^2 - x * y}$$

```
> size(emm)
[1] 7
> find_clusters(emm, tabPlus)
 [1] "1" "2" "1" "2" "3" "4" "5" "2" "6" "4" "5" "7"
> cluster_centers(emm)
  Loc_1 Loc_2 Loc_3 Loc_4 Loc_5 Loc_6 Loc_7
1    20    50   100    30    25     4    10
2    20    80    50    20    10    10    10
3    40    15    25    10    35    40     9
4     5     5    40    35    10     5     4
5     0    35    55     2     1     3     5
6    45    40    15    18    20    20    15
7    10    30    10     4    15    15    10
> |
```

EMM structure

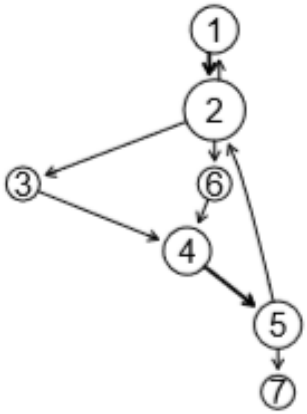
```
> EMMTraffic
   Loc_1 Loc_2 Loc_3 Loc_4 Loc_5 Loc_6 Loc_7
1     20    50   100    30    25     4    10
2     20    80    50    20    10    10    10
3     40    30    75    20    30    20    25
4     15    60    30    30    10    10    15
5     40    15    25    10    35    40     9
6      5     5    40    35    10     5     4
7      0    35    55     2     1     3     5
8     20    60    30    11    20    15    10
9     45    40    15    18    20    20    15
10    15    20    40    40    10    10    14
11     5    45    55    10    10    15     0
12    10    30    10     4    15    15    10
```

```
> cluster_centers(emm)
  Loc_1 Loc_2 Loc_3 Loc_4 Loc_5 Loc_6 Loc_7
1    20    50   100    30    25     4    10
2    20    80    50    20    10    10    10
3    40    15    25    10    35    40     9
4     5     5    40    35    10     5     4
5     0    35    55     2     1     3     5
6    45    40    15    18    20    20    15
7    10    30    10     4    15    15    10
```

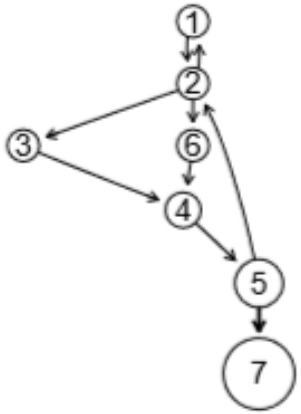


EMM manipulations

original



(a)



(b)

Fading

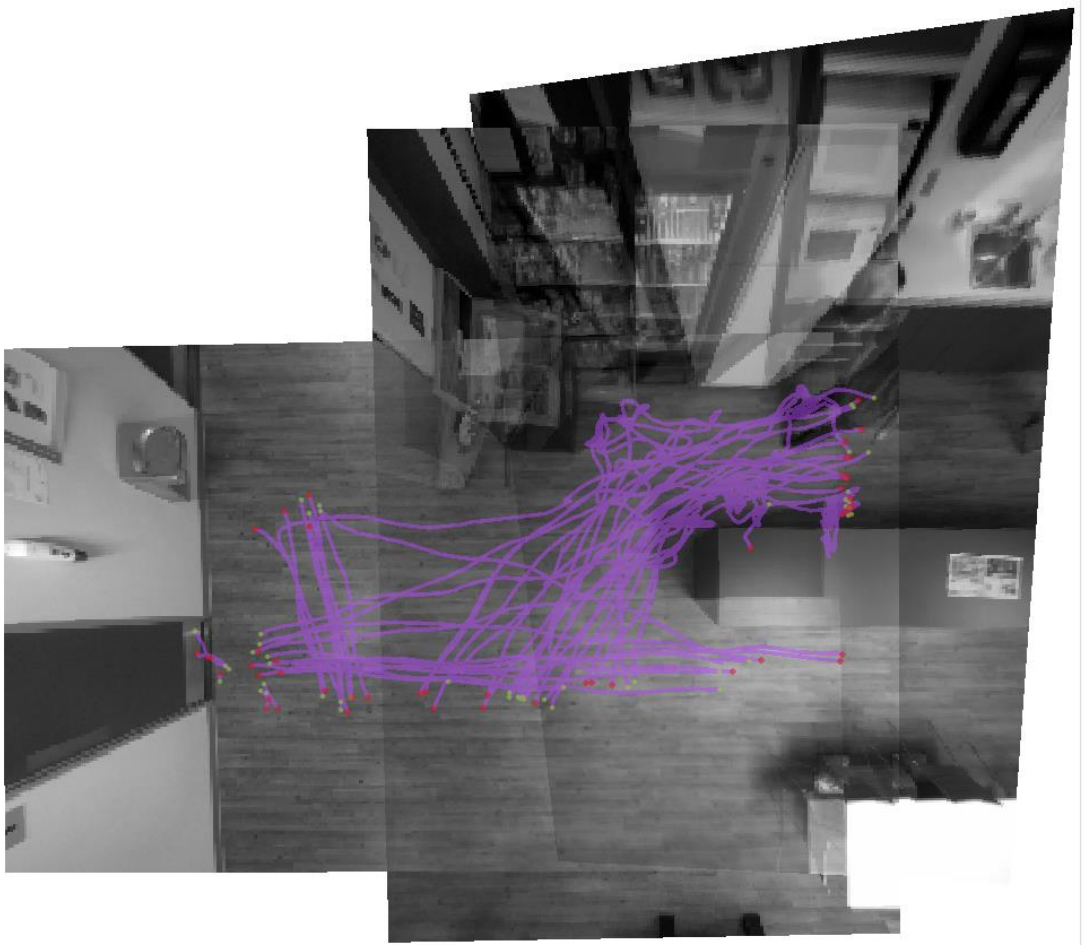
Pruning

(c)

Analyze of a stream of tracking data

Data :

- (x,y,t)



Analyze of a stream of tracking data

Data

- (x,y,t)

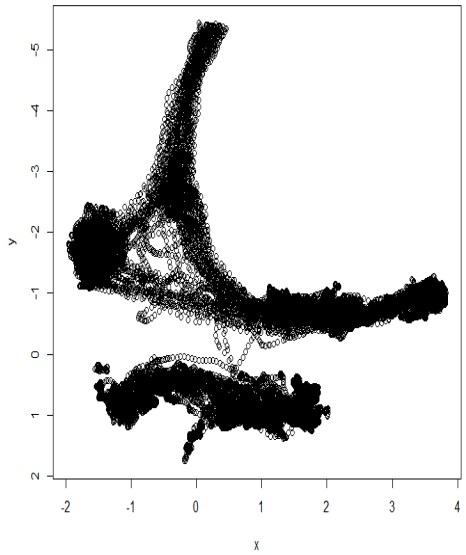
Filter

	id	x	y	date_time	row_name	h
1	0	1.875499	0.8730250	2017-03-23 14:23:41	1458	1.051
2	0	1.875360	0.8843182	2017-03-23 14:23:42	1464	1.050
3	0	1.875360	0.8843182	2017-03-23 14:23:42	1470	1.050

85651	81	1.30824840	-0.7371868	2017-03-23 14:59:20	87080	1.774
85652	81	1.31512890	-0.7487430	2017-03-23 14:59:20	87084	1.764
85653	81	1.31768570	-0.7473654	2017-03-23 14:59:20	87088	1.769

Analyze of a stream of tracking data

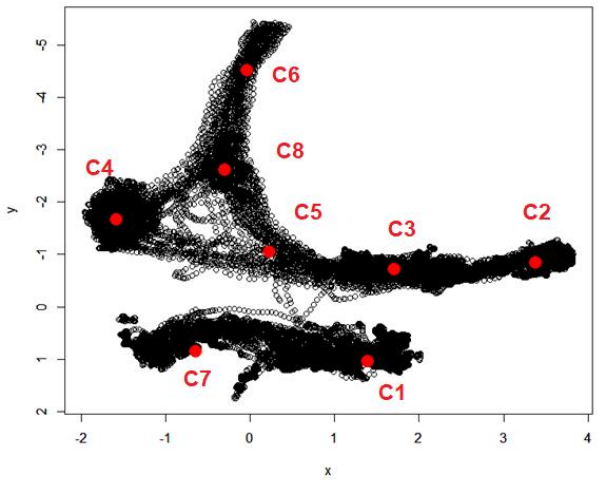
Summary of (x, y) data stream

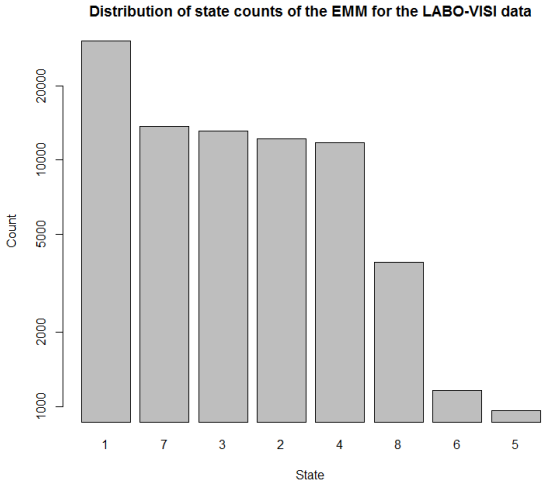


Stat.	x	y
Min.	-1.9630	-5.4294
1st Qu.	-0.3604	-0.9316
Median	1.1794	0.3171
Mean	0.8857	-0.1562
3rd Qu.	1.7244	1.0583
Max.	3.8231	1.7534

Analyze of a stream of tracking data : Clustering

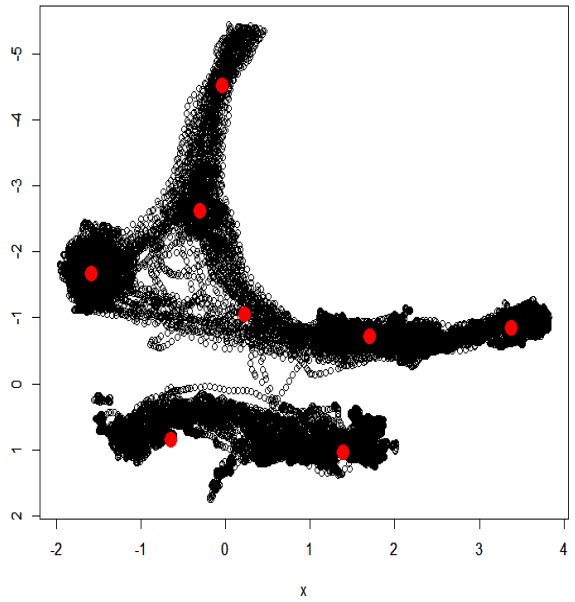
- EMM Clustering gives :
- 8 clusters
- Resulting states :
- State 6 and state 5 have integrate a few observations (transit states)
- State 8 can be considered as a waiting sate.
- Clusters 1 to 4 and 7 represent most observations (engagement state)

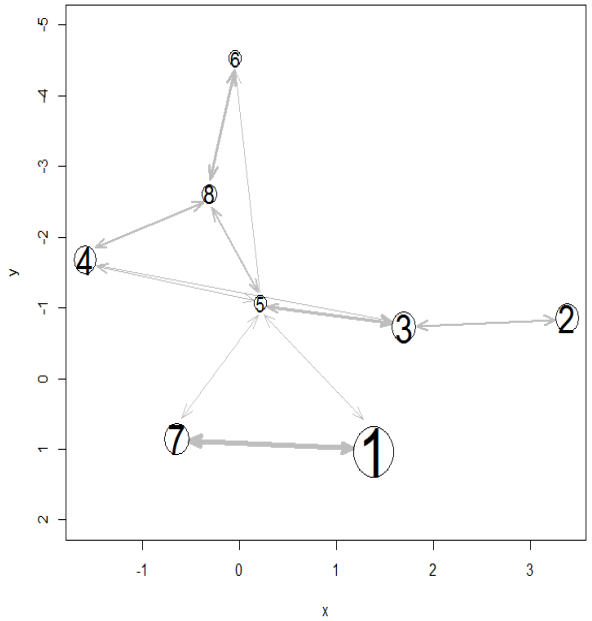




Analyze of a stream of tracking data : State graph

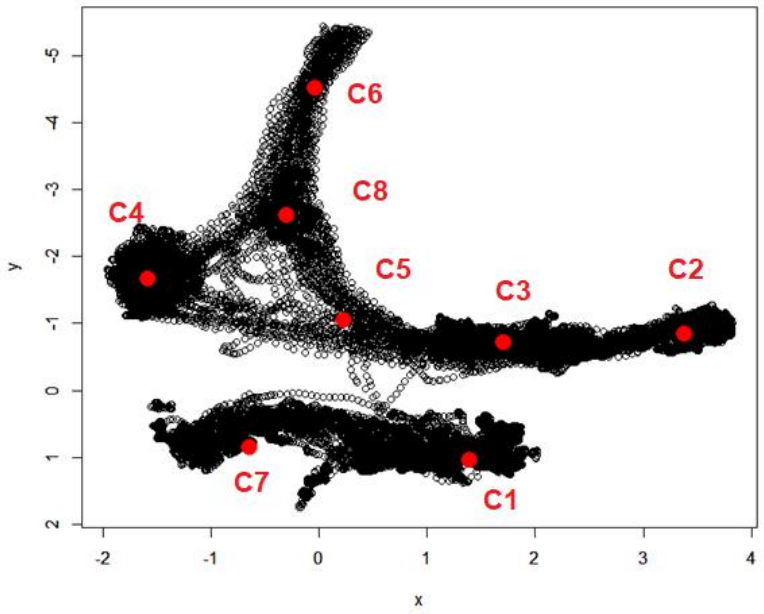
- State graph:
  - each state radius is proportional to the sate counts

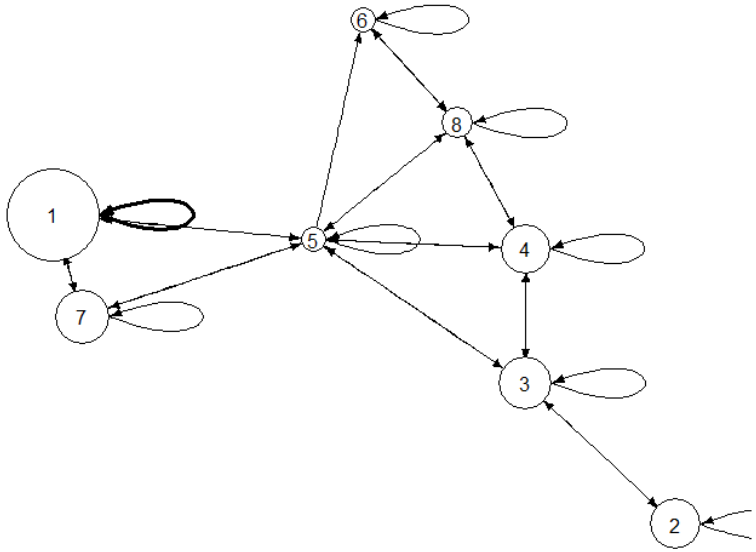




Analyze of a stream of tracking data : Transition graph

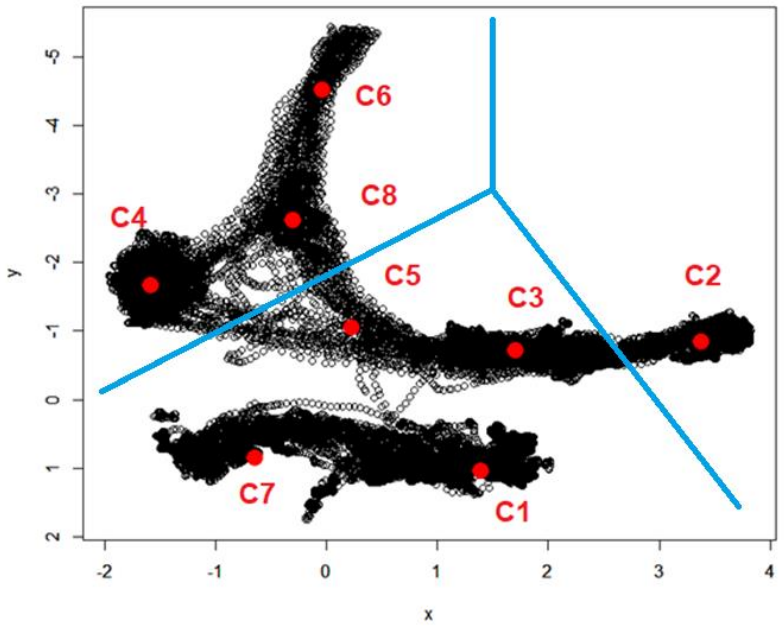
- Transition graph :
  - each transition is proportional to the transition probability

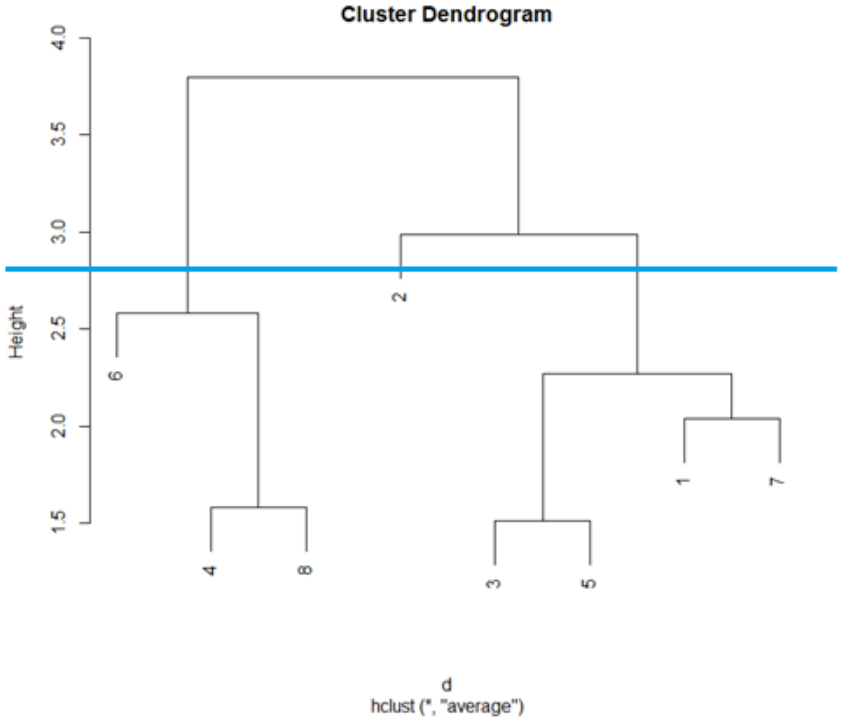




Analyze of a stream of tracking data : Reclustering

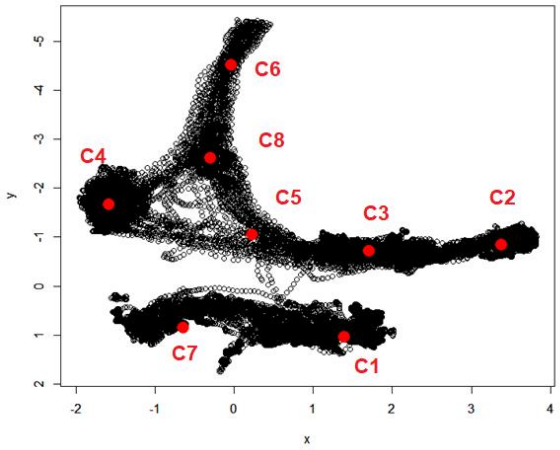
- Reclustering
  - We can also re-cluster the states

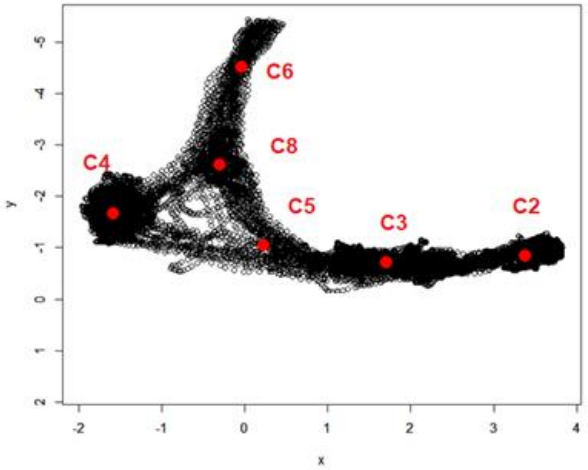


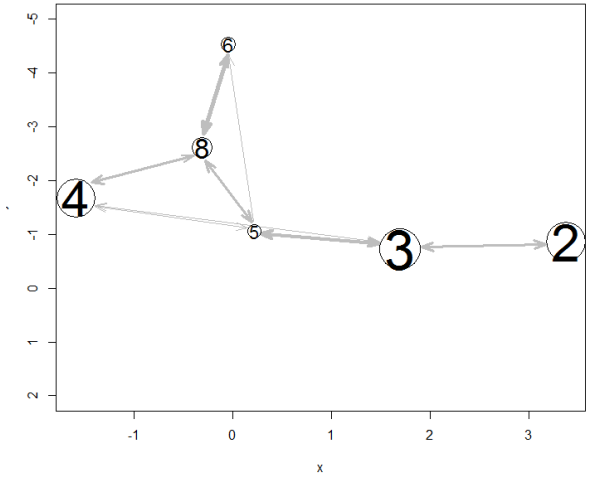


Analyze of a stream of tracking data : Graph pruning

- **Graph pruning** :
  - Removing states 1 and 7



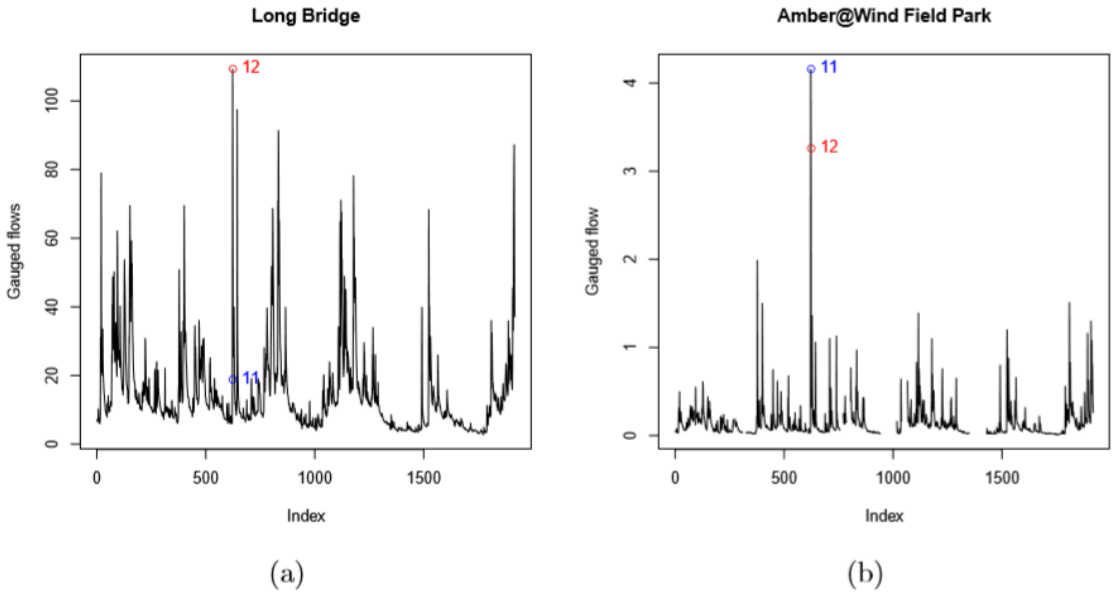




EMM applications

Application

- 6.1. Analyzing river flow data
- 6.2. Genetic sequence analysis



Reference

- https://cran.r-project.org/web/packages/rEMM/vignettes/rEMM.pdf
- https://cran.r-project.org/web/packages/rEMM/rEMM.pdf

Mohamed Dahmane, Ph. D.
Chercheur

dahmanmo@crim.ca
www.crim.ca

101 - 405, avenue Ogilvy, Montréal (Québec) H3N 1M3 514 840-1234 / 1 877 840-2746