A Tri-Clustering Method for Temporal Interaction Analysis

Fabrice Rossi, Romain Guigourès et Marc Boullé

SAMM (Université Paris 1) et Orange Labs (Lannion)

September 18, 2014

Temporal Interaction Data

Time stamped interactions between actors

- X calls Y at time t
- X sends an email to Y at time t
- X likes/answers to Y's post at time t
- and also: citations (patents, articles), web links, tweets, moving objects, etc.

Temporal Interaction Data

- a set of sources S (emitters)
- a set of destinations D (receivers)
- a temporal interaction data set $E = (s_n, d_n, t_n)_{1 \leq n \leq m}$ with $s_n \in S$, $d_n \in D$ and $t_n \in \mathbb{R}$ (time stamps)

Time-Varying Graph

Graph point of view

- interactions as edges in a directed graph
- vertices $V = S \cup D$, edges $\simeq E$
- presence function ρ from $V^2 \times \mathbb{R}$ to $\{0, 1\}$: $\rho(s, d, t) = 1$ if and only if $(s, d, t) \in E$

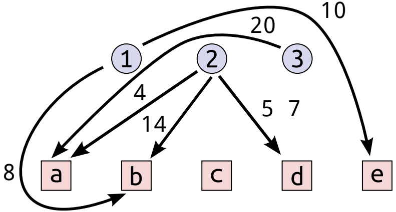
Complex time-varying graphs

- directed graph (possibly bipartite)
- multiple edges: s can send several messages to d (at different times)
- no "snapshot" assumption: time stamps are continuous

Example

$$S = \{1, 2, 3\} \quad D = \{a, b, c, d, e\}$$

source	dest.	time
2	*a*	4
2	*d*	5
2	*d*	7
1	*b*	8
1	*e*	10
2	*b*	14
3	*a*	20



Static Graph Analysis

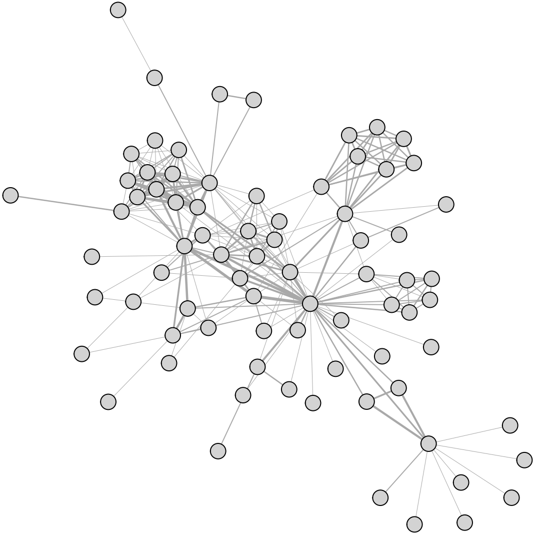
Role based analysis

- Groups of “equivalent” actors
- Structure based equivalence: interacting in the same way with other (groups of) actors
- Strongly related to graph clustering

Notable patterns

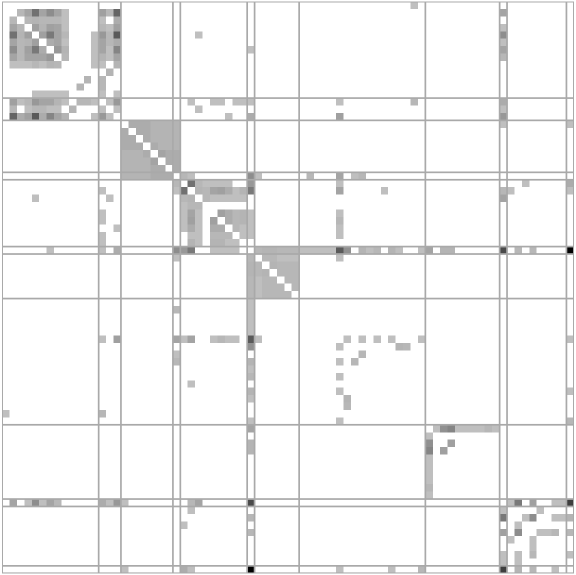
- *community*: internal connections and no external ones
- *bipartite*: external connections and no internal ones
- *hub*: very high degree vertex

Example

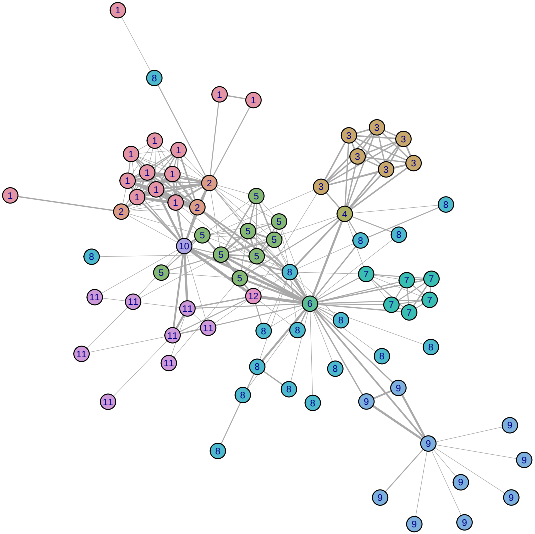


Example

Example

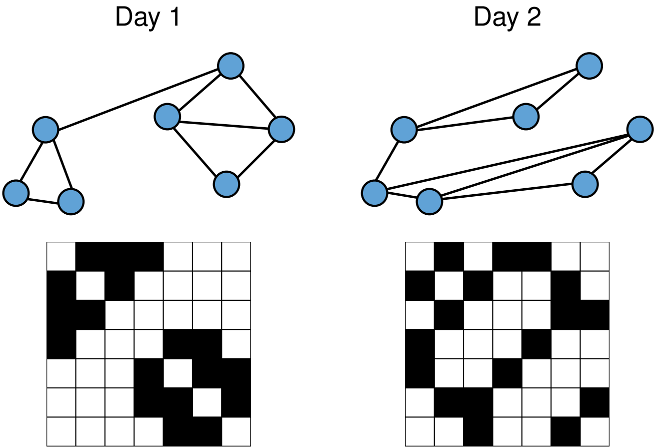


Example



What is Evolving?

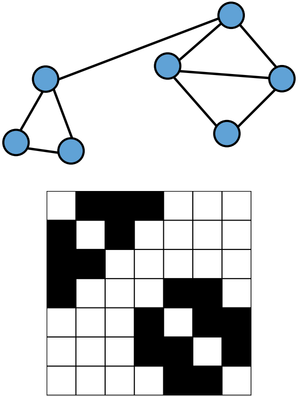
Evolving clusters, fixed patterns



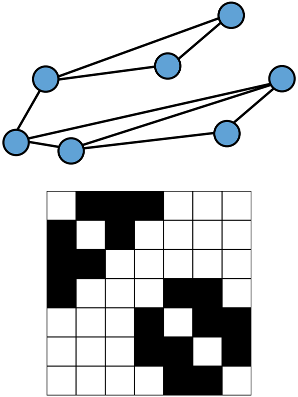
What is Evolving?

Evolving clusters, fixed patterns

Day 1

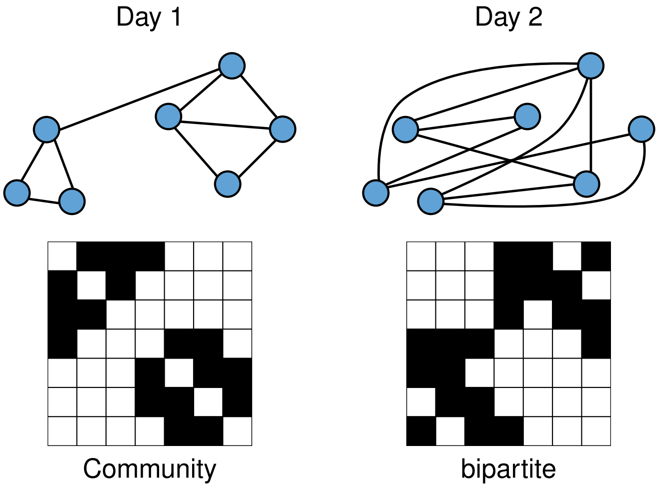


Day 2



What is Evolving?

Fixed clustering, evolving patterns



Our Point of View

Temporal Block Models

- stable partitions of sources and destinations
- a series of time intervals
- one block model per time interval
- no assumption on the number of interactions in each 3D block

Example

	D_1	D_2
S_1	18	15
S_2	2	16
S_3	1	0

$[t_1, t_2]$

	D_1	D_2
S_1	16	10
S_2	2	12
S_3	2	6

$]t_2, t_3]$

	D_1	D_2
S_1	14	5
S_2	2	6
S_3	10	10

$]t_3, t_4]$

A Generative Model for Temporal Interaction Data

Time structure

- time stamps provide here only an order
- no collision assumption (could be lifted)
- rank based representation

Parameters

- three partitions $\mathbf{C}^S$, $\mathbf{C}^D$ and $\mathbf{C}^T$
- an edge/interaction count 3D table $\boldsymbol{\mu}$: μ_{ijl} is the number of interactions between sources in c_i^S and destinations in c_j^D that take place during c_l^T
- out-degrees $\boldsymbol{\delta}^S$ of sources and in-degrees $\boldsymbol{\delta}^D$ of destinations

An example

- $S = \{1, \dots, 6\}$, $D = \{a, b, \dots, h\}$.
- $\mathbf{C}^S = \{\{1, 2, 3\}, \{4, 5\}, \{6\}\}$, $\mathbf{C}^D = \{\{a, b, c, d, e\}, \{f, g, h\}\}$
- $\mathbf{C}^T = \{\{1, \dots, 12\}, \{13, \dots, 33\}, \{34, \dots, 50\}\}$
- μ

c_1^T:

	c_1^D	c_2^D
c_1^S	5	1
c_2^S	2	0
c_3^S	4	0

c_2^T:

	c_1^D	c_2^D
c_1^S	2	2
c_2^S	2	5
c_3^S	5	5

c_3^T:

	c_1^D	c_2^D
c_1^S	0	0
c_2^S	1	0
c_3^S	1	15

- degrees

s	1	2	3	4	5	6
δ_s^S	3	6	1	2	8	30

d	a	b	c	d	e	f	g	h
δ_d^D	3	6	2	6	5	13	8	7

Generative Model

Parameters

- three partitions $\mathbf{C}^S$, $\mathbf{C}^D$ and $\mathbf{C}^T$
- an edge/interaction count 3D table $\boldsymbol{\mu}$: μ_{ijl} is the number of interactions between sources in c_i^S and destinations in c_j^D that take place during c_l^T
- out-degrees $\boldsymbol{\delta}^S$ of sources and in-degrees $\boldsymbol{\delta}^D$ of destinations

Consistency constraints

- degree balance equations

$$\sum_{1 \leq j \leq k_D, 1 \leq l \leq k_T} \mu_{ijl} = \sum_{s \in c_i^S} \delta_s^S \text{ and } \sum_{1 \leq i \leq k_S, 1 \leq l \leq k_T} \mu_{ijl} = \sum_{d \in c_j^D} \delta_d^D$$

- for a given $\boldsymbol{\mu}$, there is only one $\mathbf{C}^T$ because of the ordering constraint

Generation process

Principles

- hierarchical model
- independence inside each level
- uniform distribution for each independent part

The distribution

Generating $E = (s_n, d_n, t_n)_{1 \leq n \leq \nu}$ from a parameter list (with $\nu = \sum_{ijl} \mu_{ijl}$)

1. assign each (s_n, d_n, t_n) to a tri-cluster $c_i^S \times c_j^S \times c_l^S$ while fulfilling $\boldsymbol{\mu}$ constraints
2. independently on each variable (S, D and T), assign s_n, d_n and t_n based on the tri-cluster constraints, on δ^D and on δ^S

An example

- $S = \{1, \ldots, 6\}$, $D = \{a, b, \ldots, h\}$.
- $\mathbf{C}^S = \{\{1,2,3\},\{4,5\},\{6\}\}$, $\mathbf{C}^D = \{\{a,b,c,d,e\},\{f,g,h\}\}$
- $\mathbf{C}^T = \{\{1,\ldots,12\},\{13,\ldots,33\},\{34,\ldots,50\}\}$
- μ

c_1^T:

	c_1^D	c_2^D
c_1^S	5	1
c_2^S	2	0
c_3^S	4	0

c_2^T:

	c_1^D	c_2^D
c_1^S	2	2
c_2^S	2	5
c_3^S	5	5

c_3^T:

	c_1^D	c_2^D
c_1^S	0	0
c_2^S	1	0
c_3^S	1	15

- degrees

s	1	2	3	4	5	6
δ_s^S	3	6	1	2	8	30

d	a	b	c	d	e	f	g	h
δ_d^D	3	6	2	6	5	13	8	7

An example (continued)

- here $\nu = 50$
- a possible edge ids assignment:

	c_1^D	c_2^D
c_1^S	$\{1,\ldots,5\}$	$\{8\}$
c_2^S	$\{11,12\}$	$\emptyset$
c_3^S	$\{21,\ldots,24\}$	$\emptyset$

c_1^T

	c_1^D	c_2^D
c_1^S	$\{6,7\}$	$\{9,10\}$
c_2^S	$\{13,14\}$	$\{16,\ldots,20\}$
c_3^S	$\{25,\ldots,29\}$	$\{31,\ldots,35\}$

c_2^T

	c_1^D	c_2^D
c_1^S	$\emptyset$	$\emptyset$
c_2^S	$\{15\}$	$\emptyset$
c_3^S	$\{30\}$	$\{36,\ldots,50\}$

c_3^T

- then the sources in c_1^S are sources of the following edges

$$\{1,\ldots,5\} \cup \{8\} \cup \{6,7\} \cup \{9,10\} = \{1,\ldots,10\}.$$

- a δ^S compatible assignment is

interaction	1	2	3	4	5	6	7	8	9	10
source	2	2	1	2	1	3	2	1	2	2

An example (continued)

- Similarly, entities in c_1^D are the destination entity for the following edges

$$\{1,\ldots,5\}\cup\{6,7\}\cup\{11,12\}\cup\{13,14\}\cup\{15\}\cup\{21,\ldots,24\}\cup\{25,\ldots,29\}\cup\{30\},$$

which can be obtained using the following assignment

interaction	1	2	3	4	5	6	7	11	12	13	14	15
destination	d	d	e	a	b	a	b	e	d	d	b	b

interaction	21	22	23	24	25	26	27	28	29	30
destination	b	d	a	e	c	d	e	e	b	c

- for time stamp ranks, a possible assignment for c_1^T is

interaction	1	2	3	4	5	8	11	12	21	22	23	24
time stamp rank	5	7	10	4	8	2	9	6	1	3	12	11

An example (continued)

Final data set

interaction	source	destination	time stamp rank
1	2	*d*	5
2	2	*d*	7
3	1	*e*	10
4	2	*a*	4
5	1	*b*	8
6	3	*a*	20
7	2	*b*	14
⋮	⋮	⋮	⋮
50	6	*f*	43

Likelihood function

Compatibility

Consider $E = (s_n, d_n, t_n)_{1 \leq n \leq m}$ and $\mathcal{M} = (\mathbf{C}^S, \mathbf{C}^D, \boldsymbol{\mu}, \boldsymbol{\delta}^S, \boldsymbol{\delta}^D)$, then $\mathcal{L}(\mathcal{M}|E) \neq 0$ if and only if

1. $m = \sum_{ijl} \boldsymbol{\mu}_{ijl}$;
2. for all $s \in S$, $\delta^S_s = |\{n \in \{1, \dots, m\} | s_n = s\}|$;
3. for all $d \in D$, $\delta^D_d = |\{n \in \{1, \dots, m\} | d_n = d\}|$;
4. for all $i \in \{1, \dots, k_S\}$, $j \in \{1, \dots, k_D\}$ and $l \in \{1, \dots, k_T\}$,

$$\boldsymbol{\mu}_{ijl} = \left| \left\{ \{n \in \{1, \dots, m\} | s_n \in c^S_i, d_n \in c^D_j, t_n \in c^T_l \right\} \right| .$$

E and $\mathcal{M}$ are said to be compatible.

Likelihood function

Formula

If $\mathcal{M}$ and E are compatible

$$\mathcal{L}(\mathcal{M}|E) = \frac{\left(\prod_{i=1}^{k_S}\prod_{j=1}^{k_D}\prod_{l=1}^{k_T}\mu_{ijl}!\right)\left(\prod_{s\in S}\delta_s^S!\right)\left(\prod_{d\in D}\delta_d^D!\right)}{\nu!\left(\prod_{i=1}^{k_S}\mu_{i..}!\right)\left(\prod_{j=1}^{k_D}\mu_{.j.}!\right)\left(\prod_{l=1}^{k_T}\mu_{..l}!\right)}.$$

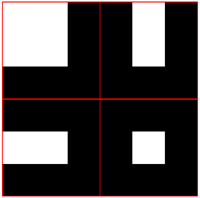
Can be rewritten to depend only on $\mathbf{C}^S$, $\mathbf{C}^D$, $\mathbf{C}^T$ and E.

Interpretation

- the likelihood increases with the number of empty tri-clusters ($\mu_{ijl} = 0$)
- the likelihood decreases when clusters are imbalanced (edge wise)

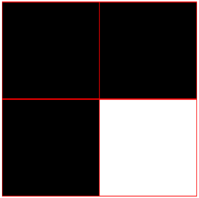
Favored tri-clusterings

Empty blocks



	D_1	D_2
S_1	5	7
S_2	7	8

88 597 190 167 200



	D_1	D_2
S_1	9	9
S_2	9	0

227 873 431 500

- identical parameters: 2×2 clusters, 27 edges
- two different partitions
- more mapping possibilities on the left, each one is then less likely

Fitting the parameters

Difficulties

- number of classes?
- combinatorial optimization

Maximum A Posteriori

- $P(\mathcal{M}|E) = \frac{P(E|\mathcal{M})P(\mathcal{M})}{P(E)}$
- we use a MAP (maximum a posteriori) approach

$$\mathcal{M}^* = \arg\max_{\mathcal{M}} P(E|\mathcal{M})P(\mathcal{M})$$

- rather than choosing directly the parameters, we choose a prior distribution on them $P(\mathcal{M})$
- strongly related to regularization approaches

Prior distribution on the parameters

Principles

- uniform distributions everywhere (combinatorial approach)
- hierarchical model
- independence at each level of the hierarchy (conditionally on the upper layers)

Prior on parameters

- $k_S^{\max} \sim \mathcal{U}(\{1,\ldots,|S|\})$, $k_D^{\max} \sim \mathcal{U}(\{1,\ldots,|D|\})$, and $k_T \sim \mathcal{U}(\{1,\ldots,m\})$
- $P_S \sim \mathcal{U}(\mathcal{P}_{k_S^{\max}}(\{1,\ldots,|S|\}))$ and $P_D \sim \mathcal{U}(\mathcal{P}_{k_D^{\max}}(\{1,\ldots,|D|\}))$
- similar uniform distribution for μ, δ^S and δ^D

The MAP Criterion

$$
\begin{aligned}
-\log P(E|\mathcal{M})P(\mathcal{M}) = \log|S| + \log|D| + \log m + \underbrace{\log \mathcal{B}(|S|, k_S) + \log \mathcal{B}(|D|, k_D)}_{\text{partitions}} \\
+ \underbrace{\log \binom{m + k_S k_D k_T - 1}{k_S k_D k_T - 1}}_{\text{number of edges}} + \sum_{i=1}^{k_S} \underbrace{\log \binom{\mu_{i..} + |c_i^S| - 1}{|c_i^S| - 1}}_{\text{degree in } c_i^S} \\
+ \sum_{j=1}^{k_D} \underbrace{\log \binom{\mu_{.j.} + |c_j^D| - 1}{|c_j^D| - 1}}_{\text{degree in } c_j^D} + \underbrace{\log(m!) - \sum_{i,j,l} \log(\mu_{ijl}!)}_{\text{edges}} \\
+ \underbrace{\sum_{i=1}^{k_S} \log \mu_{i..}! - \sum_{s \in S} \log \delta_s^S!}_{\text{edges in } c_i^S} \\
+ \underbrace{\sum_{j=1}^{k_D} \log \mu_{.j.}! - \sum_{d \in D} \log \delta_d^D!}_{\text{edges in } c_j^D} + \underbrace{\sum_{l=1}^{k_T} \log \mu_{..l}!}_{\text{time}}
\end{aligned}
$$

The MAP Criterion

$$-\log P(E|\mathcal{M})P(\mathcal{M}) = \log|S| + \log|D| + \log m + \underbrace{\log\mathcal{B}(|S|,k_S) + \log\mathcal{B}(|D|,k_D)}_{\text{partitions}}$$

$$+ \underbrace{\log\binom{m + k_S k_D k_T - 1}{k_S k_D k_T - 1}}_{\text{number of edges}} + \sum_{i=1}^{k_S}\underbrace{\log\binom{\mu_{i..} + |c_i^S| - 1}{|c_i^S| - 1}}_{\text{degree in } c_i^S}$$

$$+ \sum_{j=1}^{k_D}\underbrace{\log\binom{\mu_{.j.} + |c_j^D| - 1}{|c_j^D| - 1}}_{\text{degree in } c_j^D} + \underbrace{\log(m!) - \sum_{i,j,l}\log(\mu_{ijl}!)}_{\text{edges}}$$

$$+ \underbrace{\sum_{i=1}^{k_S}\log\mu_{i..}! - \sum_{s\in S}\log\delta_s^S!}_{\text{edges in } c_i^S}$$

$$+ \underbrace{\sum_{j=1}^{k_D}\log\mu_{.j.}! - \sum_{d\in D}\log\delta_d^D!}_{\text{edges in } c_j^D} + \underbrace{\sum_{l=1}^{k_T}\log\mu_{..l}!}_{\text{time}}$$

The MAP Criterion

$$
\begin{aligned}
-\log P(E|\mathcal{M})P(\mathcal{M}) = {} & \log|S| + \log|D| + \log m + \underbrace{\log \mathcal{B}(|S|, k_S) + \log \mathcal{B}(|D|, k_D)}_{\text{partitions}} \\
& + \underbrace{\log \binom{m + k_S k_D k_T - 1}{k_S k_D k_T - 1}}_{\text{number of edges}} + \sum_{i=1}^{k_S} \underbrace{\log \binom{\mu_{i..} + |c_i^S| - 1}{|c_i^S| - 1}}_{\text{degree in } c_i^S} \\
& + \sum_{j=1}^{k_D} \underbrace{\log \binom{\mu_{.j.} + |c_j^D| - 1}{|c_j^D| - 1}}_{\text{degree in } c_j^D} + \underbrace{\log(m!) - \sum_{i,j,l} \log(\mu_{ijl}!)}_{\text{edges}} \\
& + \underbrace{\sum_{i=1}^{k_S} \log \mu_{i..}! - \sum_{s \in S} \log \delta_s^S!}_{\text{edges in } c_i^S} \\
& + \underbrace{\sum_{j=1}^{k_D} \log \mu_{.j.}! - \sum_{d \in D} \log \delta_d^D!}_{\text{edges in } c_j^D} + \underbrace{\sum_{l=1}^{k_T} \log \mu_{..l}!}_{\text{time}}
\end{aligned}
$$

Optimization

Difficult Combinatorial Problem

- large parameter space
- discrete and complex criterion

Simple Heuristic

- greedy block merging
  - starts with the most refined triclustering
  - choose the best merge at each step
- specific data structures: $O(m)$ operations for evaluating a parameter list and $O(m\sqrt{m}\log m)$ for the full merging operation

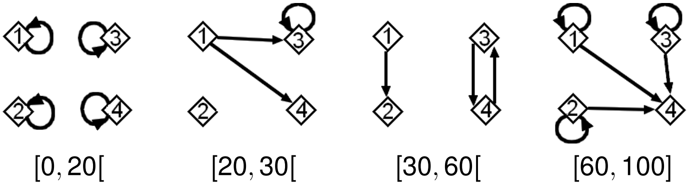
Extensions

- local improvements (vertex swapping for instance)
- greedy merging starting from semi-random partitions

Experiments

Synthetic Data

- block structure



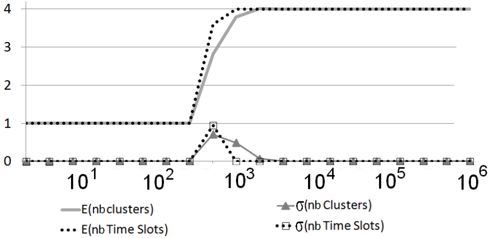
- cluster sizes

cluster	1	2	3	4
size	5	5	10	20

- edges are built according to this model, with 30 % of random rewiring
- results as a function of m, the number of edges

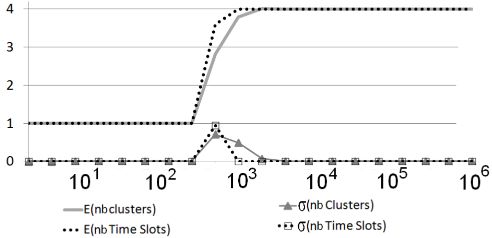
Results

1. With the data just described

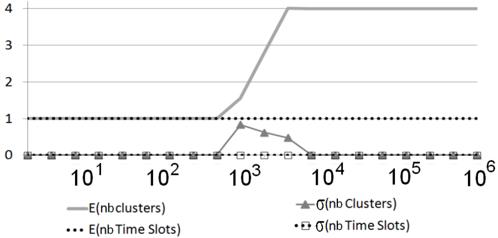


Results

1. With the data just described



2. When the temporal structured is removed



Real Data

Phone Calls in Ivory Coast

- Cellular phone calls to Ivory Coast from other countries
- Emitters: countries ($\sim$ 190)
- Receivers: cellular antenna (1216 antennas)
- minute level timestamps
- two months of communication: roughly 13 millions of incoming calls

Raw results

- very fine clustering: 286 clusters of antennas, 33 clusters of countries and 10 temporal intervals
- greedy simplification: 12 clusters of antennas, 11 clusters of countries and 6 temporal intervals

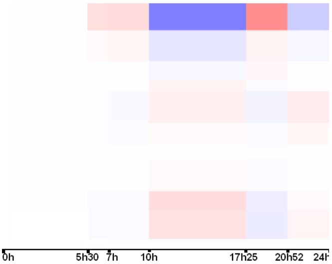
Burkina Faso

Burkina Faso

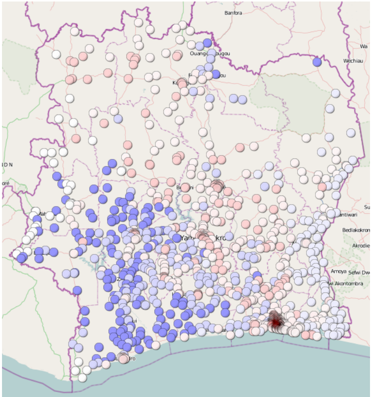
- neighbor of Ivory Coast
- provider of the first group of non Ivorian inhabitants of the Ivory Coast (roughly 15 % of the population)
- largest emitter of phone calls to Ivory Coast
- found isolated in a cluster of countries (even after simplification)

A typical result

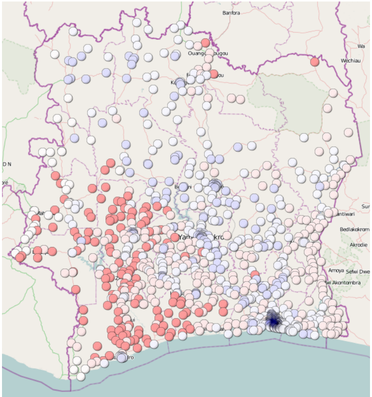
Mutual information between antenna clusters and time interval in the Burkina's cluster



Geographical view



$[10h; 17h25]$



$[17h25; 20h52[$

Real Data

Bike sharing in London

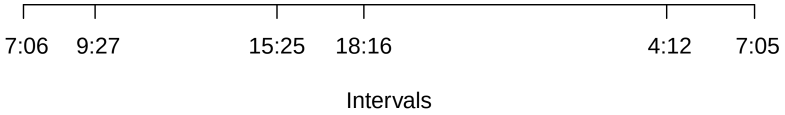
- classical bike share system
- 488 stations
- 4.8 millions of journey from 7 months

Analysis

- stationary point of view: ride hour (minute resolution)
- departure time
- on a standard PC, 50 minutes of calculation leads to:
  - 296 source clusters, 281 destination clusters
  - 5 time intervals

Analysis

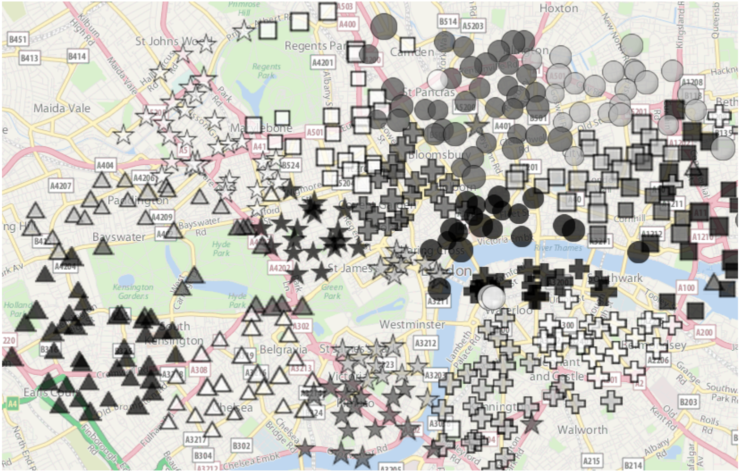
Time intervals



Too many clusters

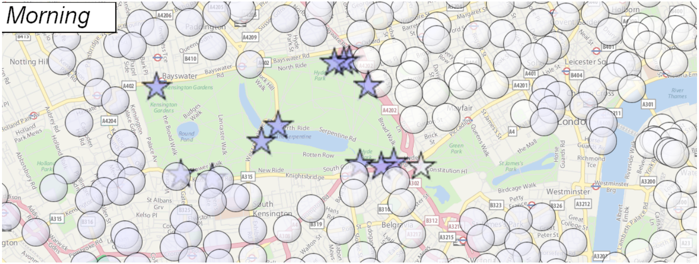
- density estimation, not clustering
- bid data $\Rightarrow$ fine patterns
- greedy simplification by cluster merging
  - uses the same algorithm
  - automatic balance between merges

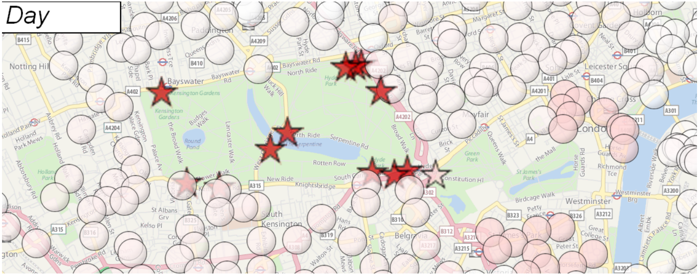
Simplified triclustering



Only 20 clusters of stations but still 5 time intervals

Comparisons





Conclusion

Summary

- MODL based temporal graph block modeling
  - complex structure detection
  - adapted to large volumes of data (in term of the number of interaction)
- automatic time segmentation
- no shown here: a full set of associated exploratory tools

Perspectives

- extensive comparisons with other techniques (already done for static graphs)
- how to handle weighted graphs?
- in general, the obtained models are too fine grained. Can we do better than greedy coarsening?

Publications

Romain Guigourès, Marc Boullé, and Fabrice Rossi.
Segmentation géographique par étude d'un journal d'appels téléphoniques.
In *2ème Journée thématique : Fouille de grands graphes*, Grenoble (France), octobre 2011.

Romain Guigourès, Marc Boullé, and Fabrice Rossi.
A triclustering approach for time evolving graphs.
In *Co-clustering and Applications, IEEE 12th International Conference on Data Mining Workshops (ICDMW 2012)*, pages 115–122, Brussels, Belgium, décembre 2012.

Romain Guigourès, Marc Boullé, and Fabrice Rossi.
Triclustering pour la détection de structures temporelles dans les graphes.
In *3ème conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques (MARAMI 2012)*, Villetaneuse, France, octobre 2012.

Romain Guigourès, Marc Boullé, and Fabrice Rossi.
étude des corrélations spatio-temporelles des appels mobiles en france.
In Christel Vrain, André Péninou, and Florence Sedes, editors, *Actes de 13ème Conférence Internationale Francophone sur l'Extraction et gestion des connaissances (EGC'2013)*, volume RNTI-E-24, pages 437–448, Toulouse, France, février 2013. Hermann-Éditions.