Temporal Graph Clustering

Fabrice Rossi, Romain Guigourès et Marc Boullé

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

October 20, 2015

Temporal Graphs

A variable notion...

- a time series of graphs? (e.g., one per day)
- transient nodes with permanent connections
- edges with duration
- etc.

Temporal Graphs

A variable notion...

- a time series of graphs? (e.g., one per day)
- transient nodes with permanent connections
- edges with duration
- etc.

with a unifying model (Casteigts et al. [2012])

- a set of vertices V and a set of edges E
- a time domain $\mathcal{T}$
- a presence function ρ from $E \times \mathcal{T}$ to $\{0, 1\}$
- a latency function ζ from $E \times \mathcal{T}$ to $\mathbb{R}^+$

Temporal Interaction Data

Time stamped interactions between actors

- X sends a SMS to 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 $G = (V, E')$
- vertices $V = S \cup D$, edges $E' \simeq E$

$$E' = \{(s, d) \in V^2 \mid \exists t\ (s, d, t) \in E\}$$

- presence function ρ from $V^2 \times \mathbb{R}$ to $\{0, 1\}$: $\rho(s, d, t) = 1$ if and only if $(s, d, t) \in E$

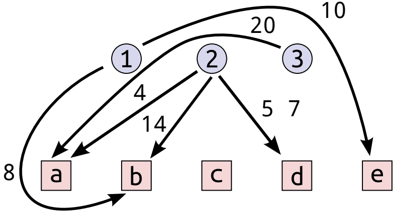
Complex time-varying graphs

- 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

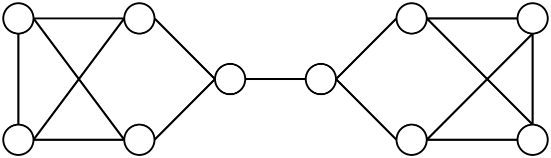


Outline

Static Graph Analysis

Role based analysis

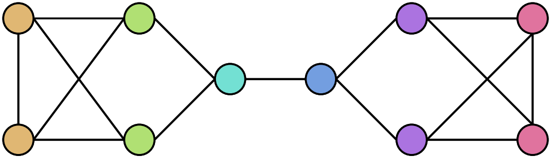
- Groups of “equivalent” actors (roles)
- Structure based equivalence: interacting in the same way with other (groups of) actors
- Strongly related to graph clustering



Static Graph Analysis

Role based analysis

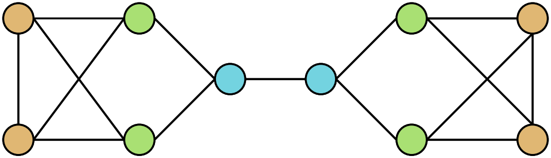
- Groups of “equivalent” actors (roles)
- Structure based equivalence: interacting in the same way with other (groups of) actors
- Strongly related to graph clustering



Static Graph Analysis

Role based analysis

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



Static Graph Analysis

Role based analysis

- Groups of “equivalent” actors (roles)
- 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

Block Models

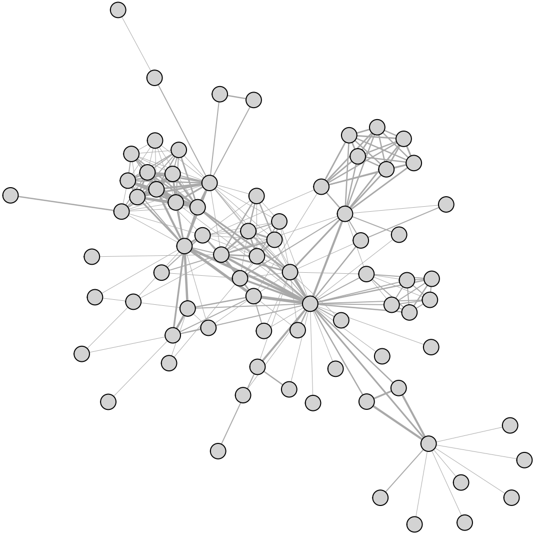
Principles

- Each actor (vertex) has a *hidden* role chosen among a finite set of possibilities (classes)
- The connectivity is explained *only* by the hidden roles

Stochastic Block Model

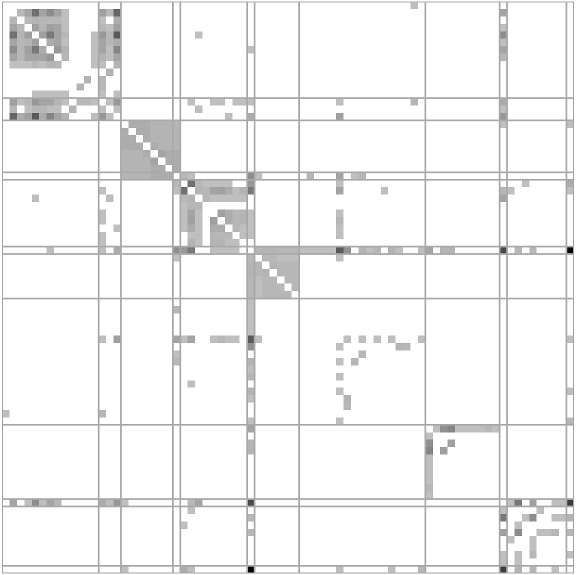
- K classes (roles)
- $Z_i \in \{1, \dots, K\}$ role of vertex/actor i
- conditional independence of connections $\mathbb{P}(X|Z) = \prod_{i \neq j} \mathbb{P}(X_{ij}|Z_i, Z_j)$ where $X_{ij} = 1$ when i and j are connected
- $\mathbb{P}(X_{ij} = 1|Z_i = k, Z_j = l) = \gamma_{kl}$ connection probability between roles k and l
- given X, we infer Z (clustering) and γ

Example

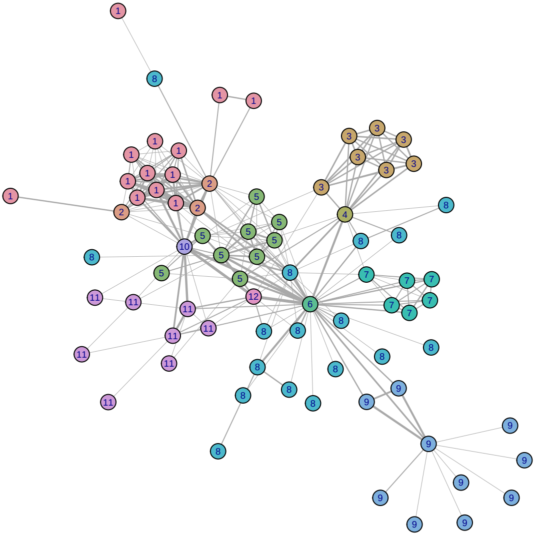


Example

Example



Example



Temporal Models

Snapshot Assumption

- Time series of static graphs: $G_1, G_2, \ldots, G_T$
- Each graph covers a time interval
- Nothing happens (on a temporal point of view) during a time interval

A Naive Analysis...

- Analyze each graph G_k independently
- Hope for the results to show some consistency

Temporal Models

Snapshot Assumption

- Time series of static graphs: $G_1, G_2, \ldots, G_T$
- Each graph covers a time interval
- Nothing happens (on a temporal point of view) during a time interval

A Naive Analysis...

- Analyze each graph G_k independently
- Hope for the results to show some consistency

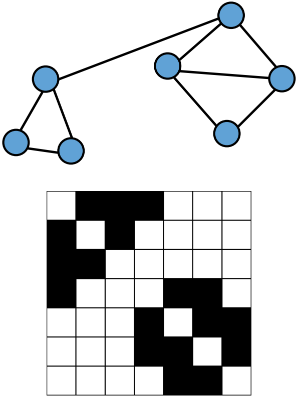
Fails

1. Fitting a model is a complex combinatorial optimization problem: results are unstable
2. Intrinsic redundancy: what is evolving?

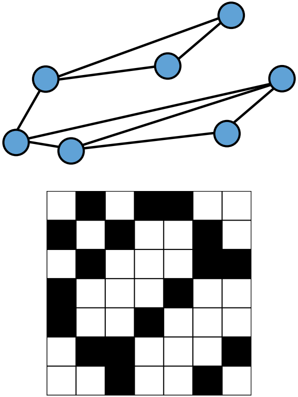
What is Evolving?

Evolving clusters, fixed patterns

Day 1



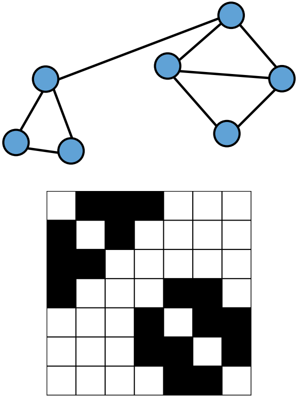
Day 2



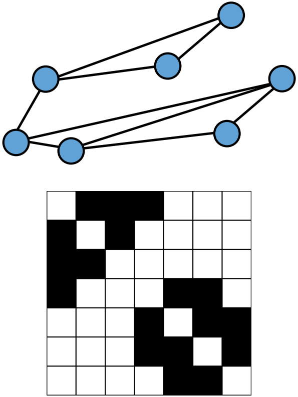
What is Evolving?

Evolving clusters, fixed patterns

Day 1

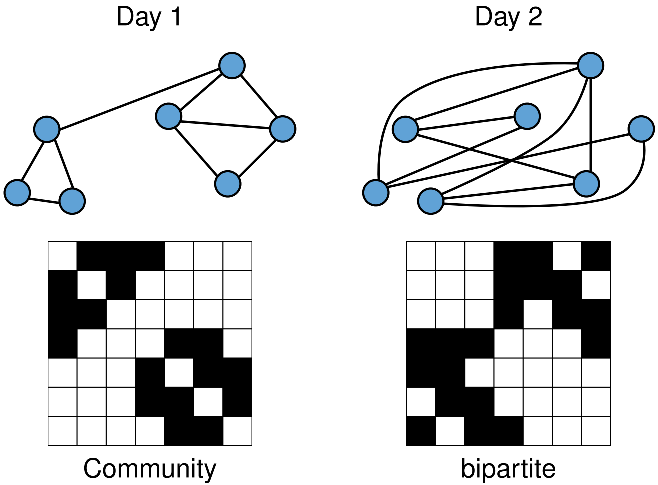


Day 2



What is Evolving?

Fixed clustering, evolving patterns



Possible solutions

Soft Constraints

- Clusters (roles) at time $t+1$ are influenced by clusters at time t: Markov chain models for instance
- Constrained evolution of connection probabilities (e.g. friendship increases with the number of encounters)

Hard Constraints

- Fixed patterns: modularity
- Fixed clustering

Possible solutions

Soft Constraints

- Clusters (roles) at time $t+1$ are influenced by clusters at time t: Markov chain models for instance
- Constrained evolution of connection probabilities (e.g. friendship increases with the number of encounters)

Hard Constraints

- Fixed patterns: modularity
- Fixed clustering

Lifting the Snapshot Constraint

- Continuous time models
- Change detection point of view: find intervals on which the connectivity pattern is stable

Temporal Block Models

Main principle

- S: source vertices, D: destination vertices
- k_S source roles, k_D destination roles and k_T time intervals
- μ_{ijl} is the number of interactions between sources with role i and destinations with role j that take place during the time interval l
- given the roles and the time intervals, the μ_{ijl} are independent

Non parametric approach

- we do not use a parametric distribution for μ_{ijl}
- μ_{ijl} becomes a parameter in (discrete) generative model
- implies a rank based representation of the time stamps

A Generative Model for Temporal Interaction Data

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

Over parametrized

- allows switching from a clustering point of view to a numerical one
- ease the design of the generative model
- ease the design of a prior distribution

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

c_1^T

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

c_2^T

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

c_3^T

- 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

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

A MAP approach

Generative model 101

- chose probability distribution over set of objects, with a parameter "vector" $\mathcal{M}$
- quality measure for $\mathcal{M}$ given an object E, the likelihood $\mathcal{L}(\mathcal{M}) = P(E|\mathcal{M})$

A MAP approach

Generative model 101

- chose probability distribution over set of objects, with a parameter "vector" $\mathcal{M}$
- quality measure for $\mathcal{M}$ given an object E, the likelihood $\mathcal{L}(\mathcal{M}) = P(E|\mathcal{M})$

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

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

- $\mathcal{M}$ can include what would be meta-parameters in other approaches (the number of clusters, for instance)
- strongly related to regularization approaches

MAP implementation

Difficult Combinatorial Optimization 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

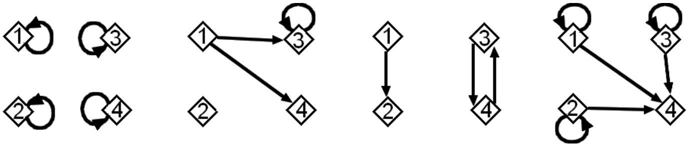
Extensions

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

Experiments

Synthetic Data

- block structure



$[0, 20[$ $[20, 30[$ $[30, 60[$ $[60, 100]$

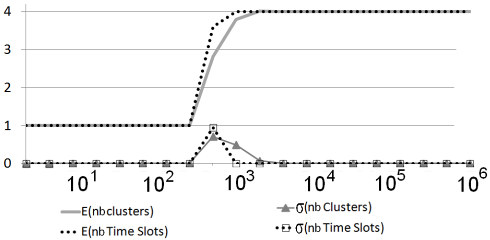
- cluster sizes

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

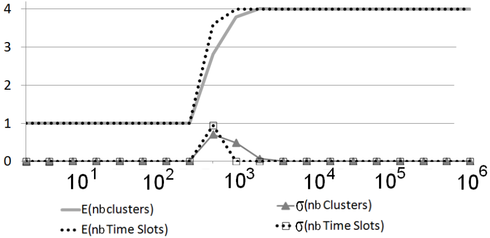
Results

1. With the data just described

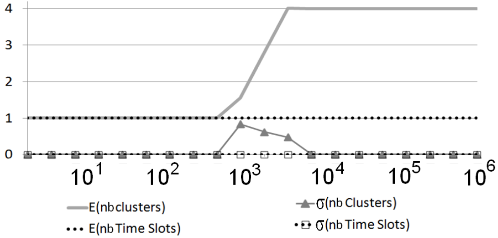


Results

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

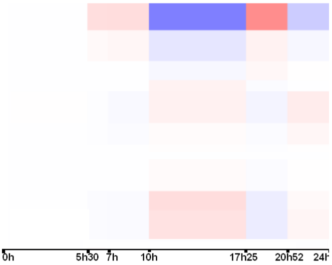
Burkina Faso

Burkina Faso

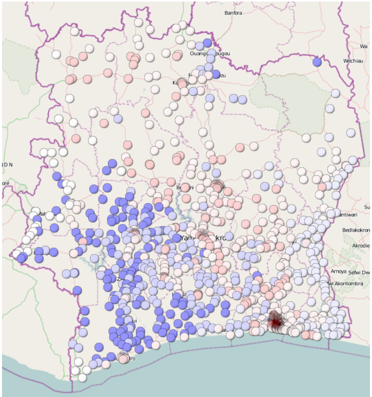
- neighbor of Ivory Coast
- provider of the first group of non Ivorian inhabitants of the Ivory Coast (roughly 15 % of the population)
- largest emitter of phone calls to Ivory Coast
- found isolated in a cluster of countries (even after simplification)

A typical result

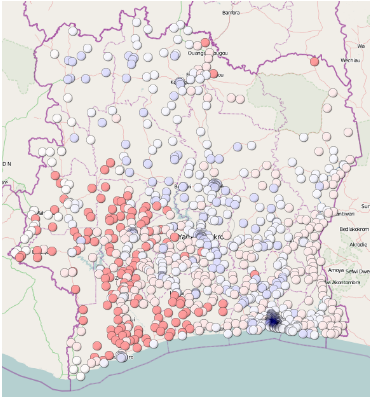
Mutual information between antenna clusters and time interval in the Burkina's cluster



Geographical view



$[10h; 17h25]$



$[17h25; 20h52[$

Real Data

Bike sharing in London

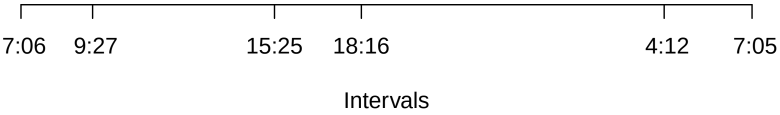
- classical bike share system
- 488 stations
- 4.8 millions of journey from 7 months

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

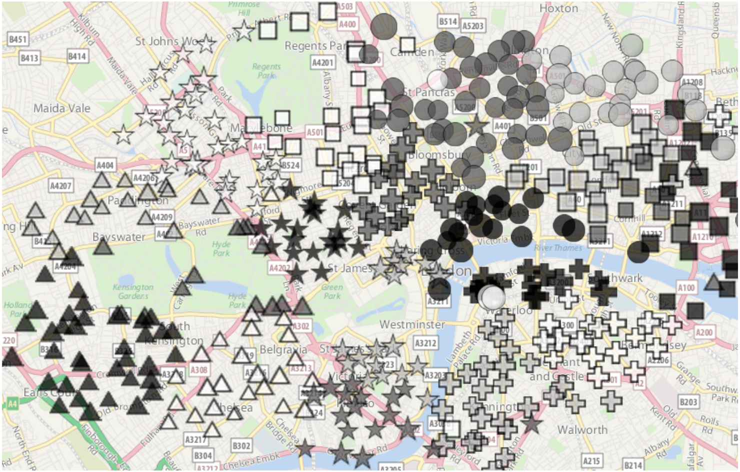
Time intervals



Too many clusters

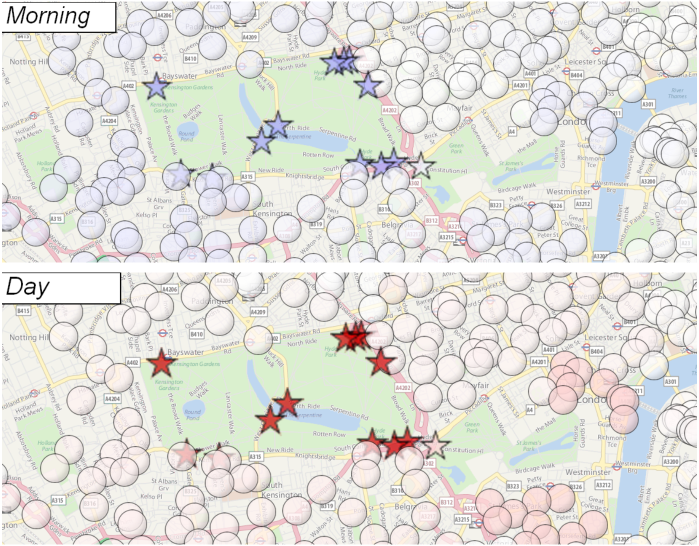
- density estimation, not clustering
- bid data $\Rightarrow$ fine patterns
- greedy simplification by cluster merging
  - uses the same algorithm
  - automatic balance between merges

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?

References

A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro. Time-varying graphs and dynamic networks. *International Journal of Parallel, Emergent and Distributed Systems*, 27(5):387–408, 2012. doi: 10.1080/17445760.2012.668546.

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

R. Guigourès, M. Boullé, and F. 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 2012a. ISBN 978-1-4673-5164-5. doi: 10.1109/ICDMW.2012.61.

R. Guigourès, M. Boullé, and F. 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 2012b.

R. Guigourès, M. Boullé, and F. Rossi. étude des corrélations spatio-temporelles des appels mobiles en france. In C. Vrain, A. Péninou, and F. 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.

R. Guigourès, M. Boullé, and F. Rossi. Discovering patterns in time-varying graphs: a triclustering approach. *Advances in Data Analysis and Classification*, pages 1–28, 2015. ISSN 1862-5347. doi: 10.1007/s11634-015-0218-6. URL http://dx.doi.org/10.1007/s11634-015-0218-6.

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

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

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

c_1^T

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

c_2^T

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

c_3^T

- 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

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

  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
$\vdots$	$\vdots$	$\vdots$	$\vdots$
50	6	f	43

Likelihood function

Compatibility

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

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

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

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)

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

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

The MAP Criterion

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