

Signaux-graphes pour l'étude d'interactions sociales

Partie 2 - Temporalité dans les interactions sociales

Pierre Borgnat

CR1 CNRS – Laboratoire de Physique, ENS de Lyon, Université de Lyon

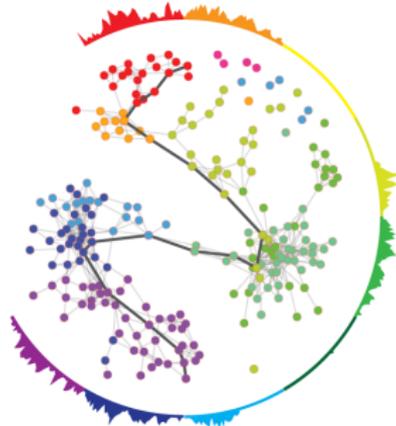
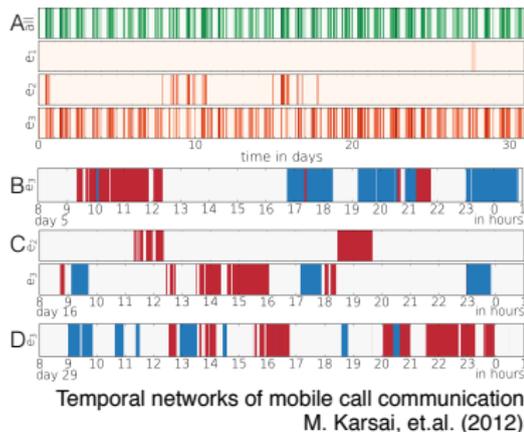
Équipe SiSYPHE : Signaux, Systèmes et Physique

06/2014



Généralités sur les dynamiques sociales

- Quelques exemples de réseaux dynamiques



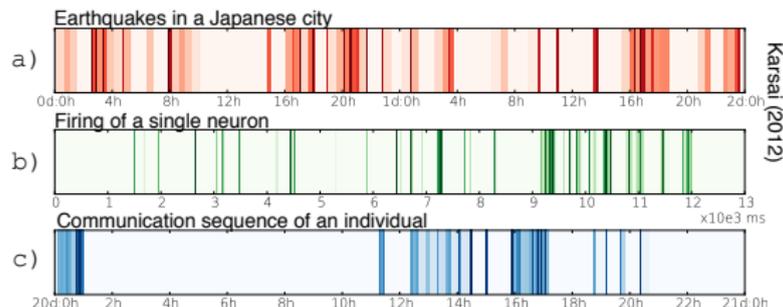
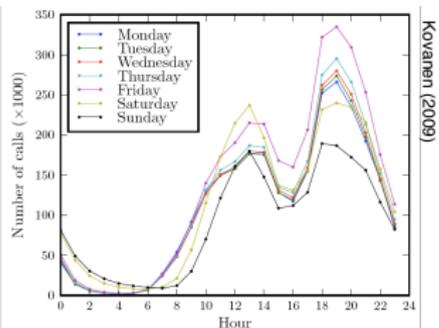
Téléphones

Sociopatterns: One day in a museum
[Dublin, 2011]

- Article de review :
[Holme & Saramäki, “Temporal networks”, 2012]

Généralités sur les dynamiques sociales

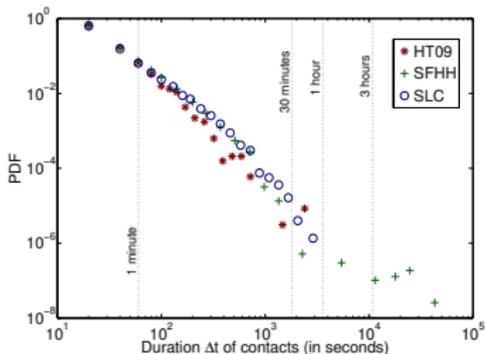
- Dynamique non stationnaire : rythmes cycliques, “burstiness”,...



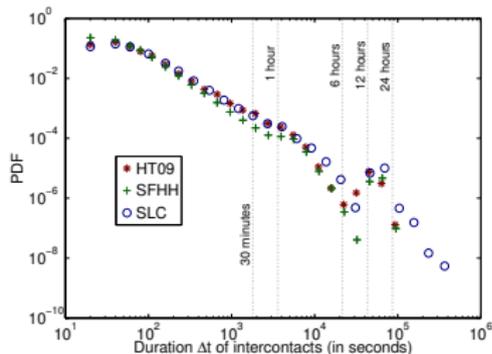
Généralités sur les dynamiques sociales

- Les durées contact et inter-contact suivent une loi de puissance
- Exemple sur le conférence à SLC

Durée de contact

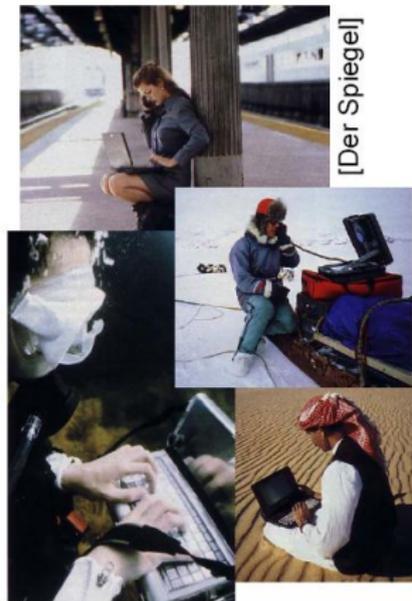


Durées d'inter-contact



Réseaux dynamiques

- Complexification progressive :
topologie fixe ; réseaux en croissance ;
processus dynamiques sur topologie
fixe ; processus dynamique en
co-évolution avec réseaux
dynamiques ; réseaux temporels
- Contextes : épidémies, diffusion
d'information (de rumeur), réseaux de
communication ad-hoc, sans fil,
micro-blogs, réseaux d'évolution en
biologie, etc.
- => **Réseaux temporels**
- Quels modèles, quels outils ?



[Scherrer et al. "Description and simulation of dynamic mobility network", *Computer Networks* (2008)]

- Methodology: empirical analyses from data
- Available datasets at that time:
 - IMOTE: 41 nodes, 3 days, sampling 2' [Chaintreau et al., 2006]
 - MIT: 100 nodes, 9 months, sampling 5' [Eagle et al., 2007]
- Power-laws strike back:
 - Contact and inter-contact duration PDF

[Chaintreau et al., 2006]

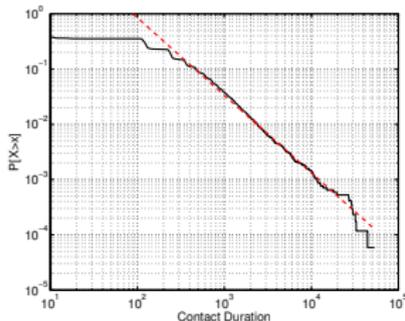
-Scale-free networks, Small worlds networks,...

[Barabasi, 1999]

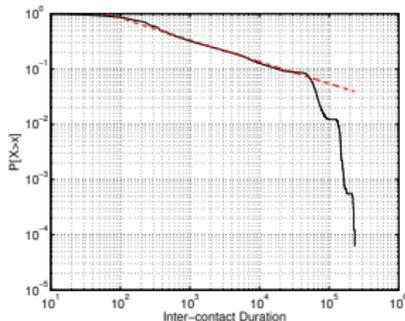
Analysis of Dynamical Network

Scale invariance in dynamical networks

[IMOTE dataset]



Contacts



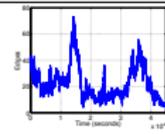
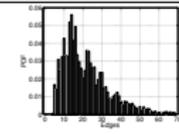
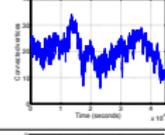
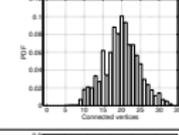
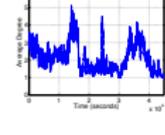
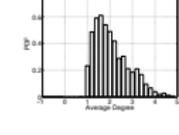
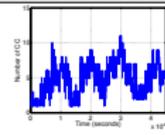
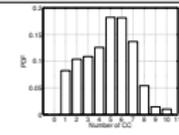
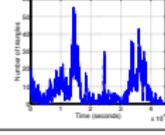
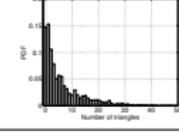
Inter-contacts

$$\text{Model: } P(\tau > T) \sim cT^{-\alpha}.$$

Data	Type	α	c	Mean (s)
IMOTE	Inter-contact	0.60	156	3680
IMOTE	Contact	1.66	98	140

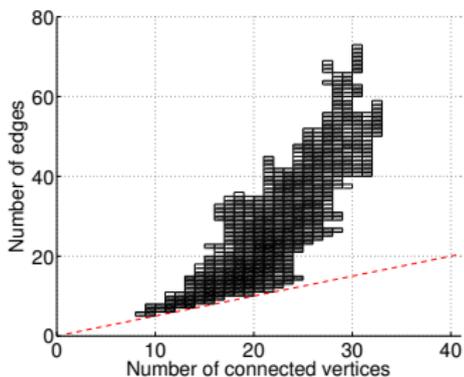
Analysis of Dynamical Network

Basic Statistics [IMOTE, 2006]

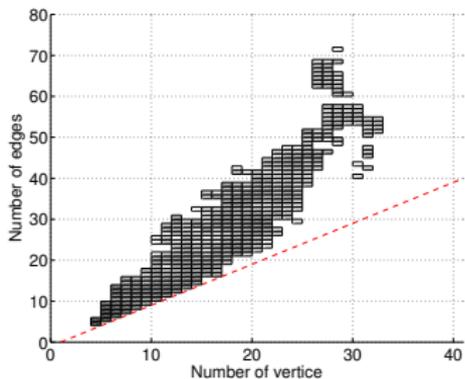
Prop.	Data	PDF	Mean	Std.	Corr. T (s)
# Liens			21.9	12.4	3930
# Nœuds			19.9	4.73	5540
Degré Moyen			2.08	0.77	3600
# CC			4.75	2.12	3380
# Triangles			6.93	8.30	4440

Analysis of Dynamical Network

Joint Distribution



Connected Nodes/Links



Nonnected Nodes/Links in CC

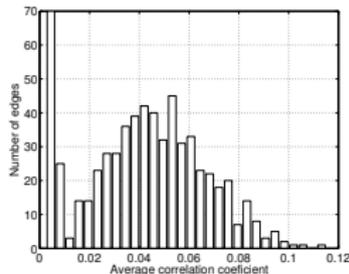
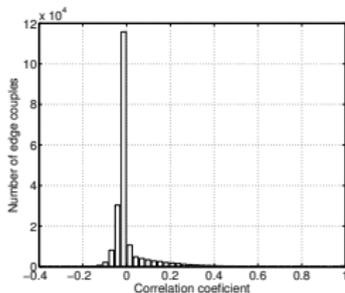
- **Large Variability**

For a given number of connected nodes, the # of connected links may be widely spread.

Dynamical properties: link creation/deletion process

Prop.	Data	PDF	Mean	Std. Dev.	Corr. T (s)
$E_+(t)$			0.15	0.55	900
$E_-(t)$			0.15	0.55	680

- Link creation/deletion is not much correlated along time



- Not many correlation between evolution of different nodes.

Exemple d'un modèle de réseau dynamique

- Travail de [Scherrer et al. "Description and simulation of dynamic mobility network", *Computer Networks* (2008)]
- Principe de base pour proposer un modèle de réseau dynamique :
modèle dynamique d'activation / inactivation des liens
- Objectif : respecter les contraintes mesurées = statistiques globales
- Nombre de nœuds / liens connectés
- Degré des nœuds
- Nombre de triangles, etc.

Exemple d'un modèle de réseau dynamique

- Simulation du processus d'ajout/suppression de lien, pour chaque lien et chaque pas de temps indépendamment
- Contraintes : on garde les distrib. temporelles des temps de contacts / inter-contacts
- On force la création de triangles

For i=0..Simulation-Time

For each link e

Pr = Uniform(0,1);

Ptr(e,t) = TransitionProbability(e,t);

if (Pr \leq Ptr(e,t))

ChangeState(e);

Prise en compte des contacts / inter-contacts

- Objectif ? Reproduire la distribution stationnaire des temps de contact $P_{ON}(t)$ / inter-contact $P_{OFF}(t)$.
- Pourquoi ? Ces distributions ont une queue lourde.
- Comment ? Calculer $P_+(t)$ et $P_-(t)$ en fonction de $P_{ON}(t)$ et $P_{OFF}(t)$.

$$P_{ON}(t) = P_-(t) \times \prod_{i=1}^{t-1} (1 - P_-(i))$$

On peut inverser cette relation pour obtenir $P_-(t)$ récursivement:

$$P_-(t) = \frac{P_{ON}(t)}{\prod_{i=1}^{t-1} (1 - P_-(i))}, \quad t \geq 2, \quad P_-(1) = P_{ON}(1)$$

$$P_+(t) = \frac{P_{OFF}(t)}{\prod_{i=1}^{t-1} (1 - P_+(i))}, \quad t \geq 2, \quad P_+(1) = P_{OFF}(1)$$

Prise en compte des créations de triangles

- Objectif ? Favoriser les ajouts de liens qui créent un nouveau triangle.
- Pourquoi ? Reproduire la proportion de ces évènements observée dans les données.
- Comment ? Pondération de la probabilité de transition $P_+(t)$.

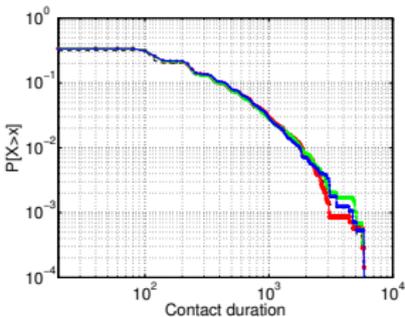
$$P_+(t) \frac{P_{+/tri=}}{f_{+/tri=}}$$

Pour les créations de liens qui mène à une création de triangle

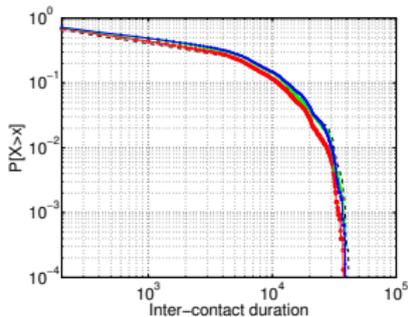
$$P_+(t) \frac{P_{+/tri=}}{f_{+/tri=}}$$

Pour les créations de liens qui ne mène pas à une création de triangle

Informations injectées dans le modèle



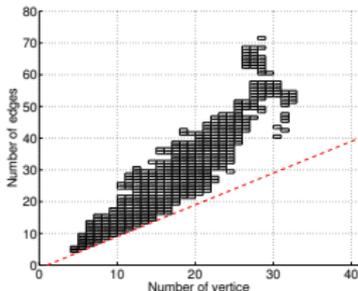
Contacts



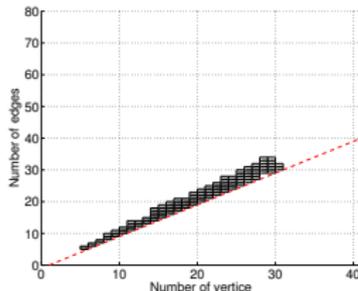
Inter-contacts

Données	$P_{+/tri+}$	$P_{+/tri=}$
IMOTE	44 %	56 %
CI	5 %	95 %
TRI	60 %	40 %

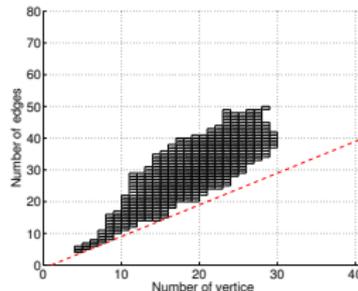
Distributions conjointes obtenues



IMOTE



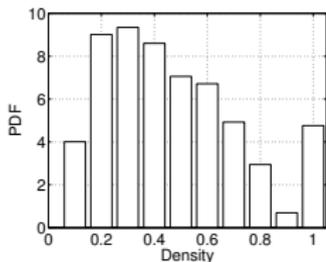
CI



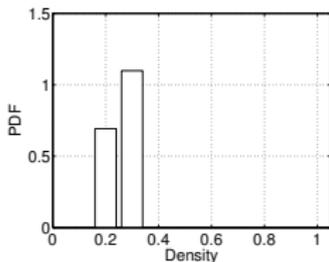
TRI

- La variabilité des configurations n'est reproduite que par le modèle favorisant les triangles

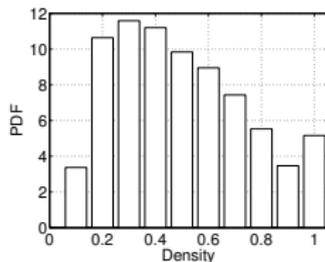
Densités des communautés



IMOTE



CI



TRI

- La densité des communautés n'est reproduite que par le modèle favorisant les triangles

Autres questions ou modèles pour les réseaux dynamiques

- Modèles écrits au niveau des groupes
cf. [Stehlé, Barrat et al., 2010, 2011]
- Flèche du temps dans les interactions sociales ?
(poser la question à Nicolas Tremblay...)
- Trouver des modèles de graphes temporels $\mathcal{G}^t = (\mathcal{V}^t, \mathcal{E}^t)$
ou bien de flots de liens (A_i, B_i, t_i)
- Adapter la notion de communautés aux réseaux dynamiques

Décomposition en groupes dépendant du temps

- Sur cette question, il existe plusieurs solutions et pas encore de consensus
- Présentation ici de quelques travaux dans ce sens
- Autre piste qui ne sera pas présenté dans le cours : dans le cadre de la thèse de Ronan Hamon (transformation de graphes en signaux et NMF)

Décomposition en groupes dépendant du temps - 1

- Définition d'un groupe dynamique de manière formelle :
 - sous-graphe maximal ayant au moins σ liens pendant plus que τ pas de temps
 - critère de densité : $2|E||V|(|V| - 1) \geq \delta$
- Recherche exhaustive (en data mining) par *D-miner* (utilise la monotocité de la définition)

Data

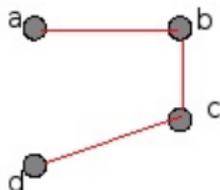
	t1	t2	t3	t4	t5	t6
a-b	1	1	1	1	0	0
a-c	0	0	0	0	0	0
a-d	0	0	0	1	1	1
b-c	1	1	1	1	0	0
b-d	0	0	0	1	1	1
c-d	1	1	1	1	1	1



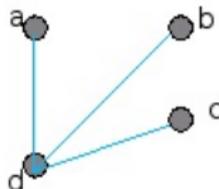
Two formal concepts

	t1	t2	t3	t4	t5	t6
a-b	1	1	1	1	0	0
a-c	0	0	0	0	0	0
a-d	0	0	0	1	1	1
b-c	1	1	1	1	0	0
b-d	0	0	0	1	1	1
c-d	1	1	1	1	1	1

That represents two maximal subgraphs



From time step 1 to time step 4

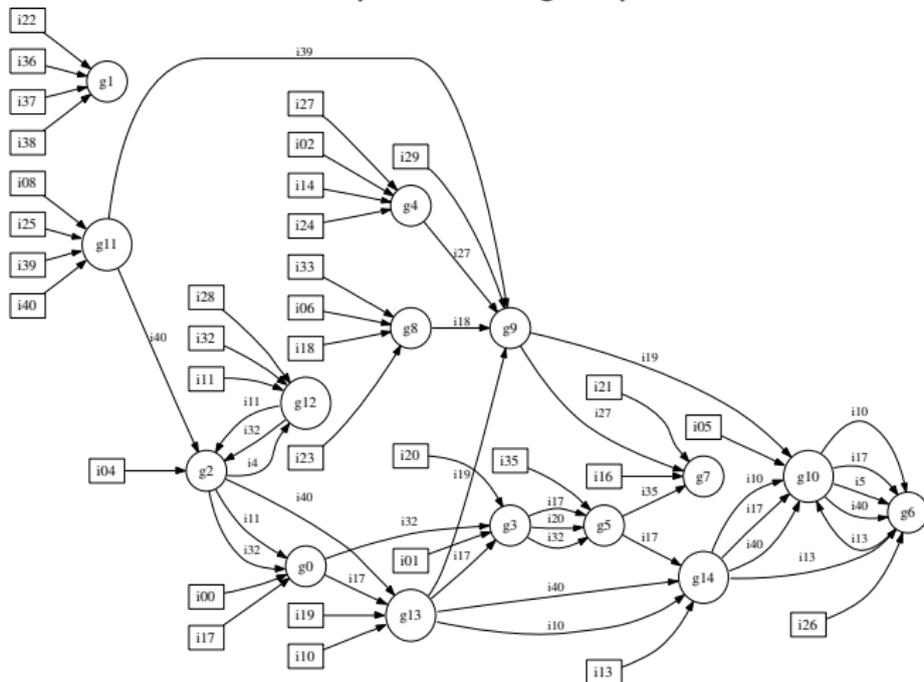


From time step 4 to time step 6

Décomposition en groupes dépendant du temps -1

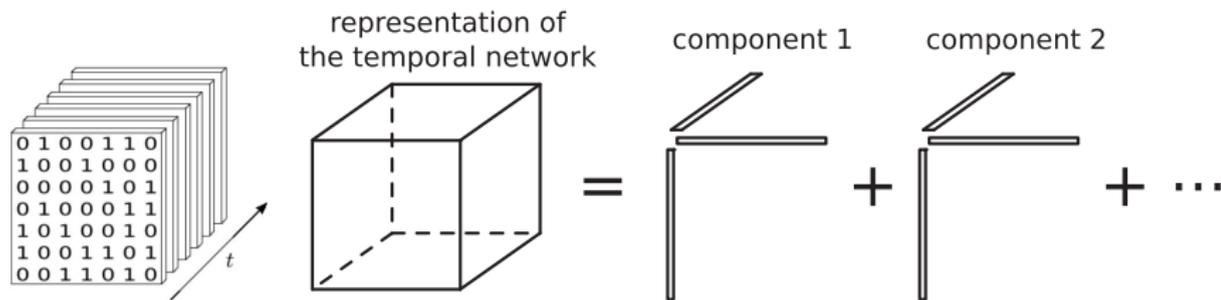
- Résultat sur les données IMOTE [Scherrer et al., 2008]
- Cercles = groupes ; carrés = individus

Trajectoires des individus parmi les groupes.



Décomposition en groupes dépendant du temps -2

- Travail de [Gauvin et al. "Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach", *PLOS One* (2014)]
- Codage Tensoriel du cube de données des $\mathcal{G}^t = (\mathcal{V}, \mathcal{E}^t)$



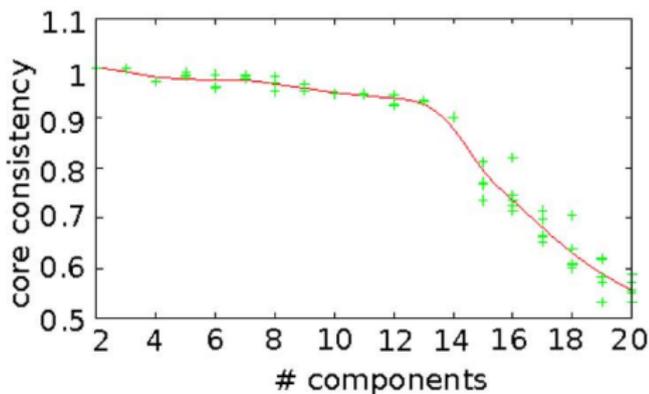
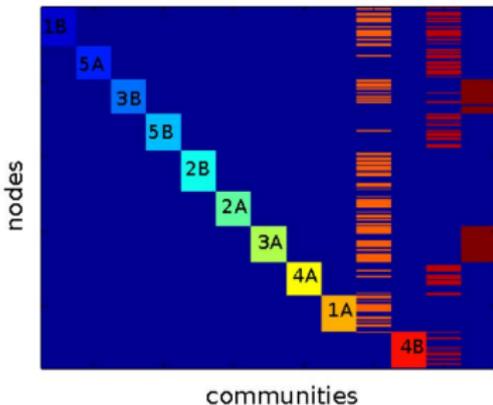
- Décomposition tensorielle de type PARAFAC

$$\mathcal{T} = \sum_r^{R_T} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

Décomposition en groupes dépendant du temps -2

- Applications : école primaire [Sociopatterns]
- Groupes obtenus

Cohérence des groupes



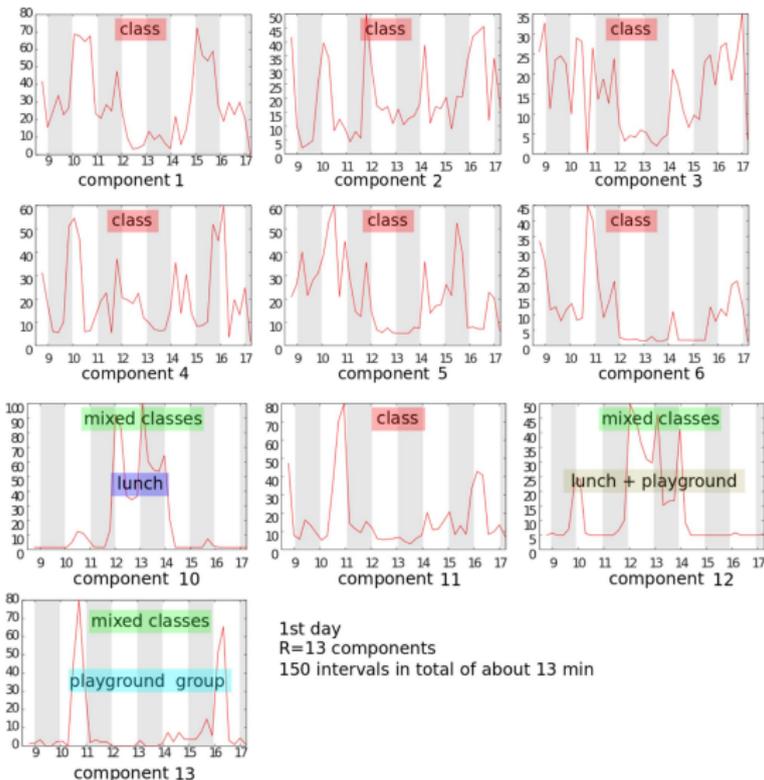
Décomposition en groupes dépendant du temps -2

- Interprétation des groupes : classes vs. groupes

$$\begin{pmatrix} 23 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 23 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 22 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 21 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 23 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 22 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 24 & 0 \end{pmatrix} \quad \begin{pmatrix} 23 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 13 & 0 \\ 0 & 25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 15 & 14 \\ 0 & 0 & 22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 26 & 0 & 0 & 0 & 0 & 0 & 0 & 23 & 16 & 0 \\ 0 & 0 & 0 & 0 & 23 & 0 & 0 & 0 & 0 & 0 & 22 & 17 & 0 \\ 0 & 0 & 0 & 0 & 0 & 22 & 0 & 0 & 0 & 0 & 0 & 9 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 21 & 0 & 0 & 0 & 0 & 3 & 12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 23 & 0 & 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 21 & 0 & 0 & 1 & 12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 24 & 0 & 1 & 13 \end{pmatrix}$$

Décomposition en groupes dépendant du temps -2

- Activités au cours du temps des groupes



Décomposition en groupes dépendant du temps -3

- Travail de [Xu, Klinger, Hero "Tracking Communities in Dynamic Social Networks", (2011, 2013)]
- Introduire de la continuité temporelle dans les méthodes de recherches de groupes ou clusters
- Idée principale : *Adaptive Forgetting Factor for Evolutionary Clustering and Tracking (AFFECT)*
- Modèle supposé :

$$W^t = \Psi^t + N^t$$

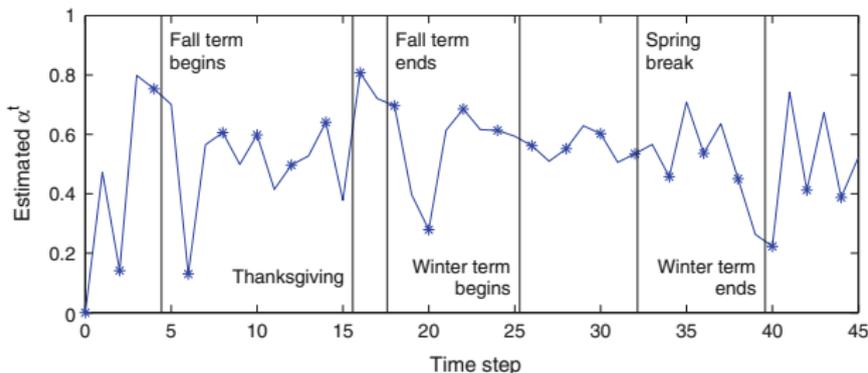
- Suivi à la Kalman
- Au premier ordre : du lissage exponentiel de la matrice d'adjacence

$$\bar{W}^t = \alpha^t \bar{W}^{t-1} + (1 - \alpha^t) W^t$$

Décomposition en groupes dépendant du temps -3

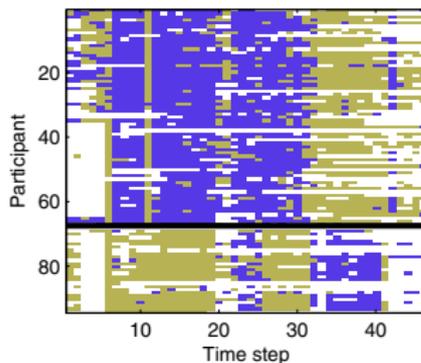
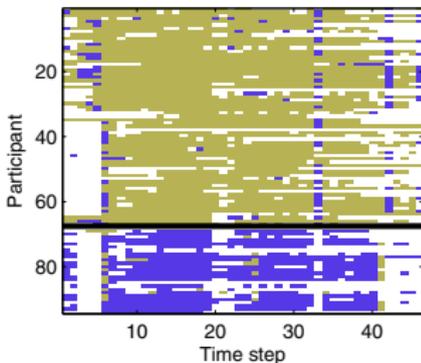
- Estimation de α^t : minimisation du MSE

$$(\alpha^t)^* = \frac{\sum_{i=1}^n \sum_{j=1}^n \text{var}(w_{ij}^t)}{\sum_{i=1}^n \sum_{j=1}^n \left[(\bar{w}_{ij}^{t-1} - \psi_{ij}^t)^2 + \text{var}(w_{ij}^t) \right]}$$



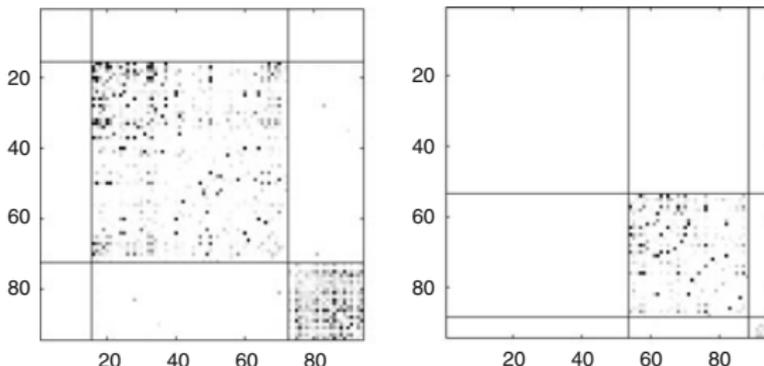
Décomposition en groupes dépendant du temps -3

- Comparaison entre le clustering ainsi obtenue et celui sans continuité temporelle



Décomposition en groupes dépendant du temps -3

- Changement des groupes instantanés obtenus juste avant et après les vacances d'hiver



- Développements à reprendre ou faire :
nombre de groupes ? réseaux multiplexés ?

Conclusion

- Beaucoup de ces sujets sont très actifs
 - Analyse des données sociales numériques
 - Étude des réseaux complexes, avec l'apport du traitement du signal
 - Réseaux dynamiques
- Beaucoup d'enjeux, d'opportunités, de sujets
- Des données et des questions originales par rapport au TSI habituel

<http://perso.ens-lyon.fr/pierre.borgnat>

Remerciements: merci à Nicolas Tremblay et à Marton Karsai (LIP, ENSL) à qui j'ai emprunté beaucoup de figures ou diapo.