

L'analyse des parcours de vie : *approches typologiques*

Journée "Trajectoires'11"

14 octobre 2011

Nicolas Robette

Laboratoire Printemps (CNRS-UVSQ)



Émergence de l'analyse des parcours de vie

- Elder (1974): « *Children of the Great Depression* »
- Fin 70's-80's: perspective de recherche importante
- Parallèlement au développement :
 - des **données longitudinales micro-individuelles** (panels, enquêtes rétrospectives, biographiques)
 - des **méthodes d'analyse statistique**

Une approche « atomiste »

Unité d'analyse	événement
Objectif	modélisation de probabilités de transitions/durée
Genèse du parcours	processus stochastique (paramétrique)
Démarche	explicative, causale
Méthodes	modèles de Markov, analyse de survie, modèle de Cox, multilevel-multiprocess...

Une approche « holiste »

Unité d'analyse	parcours
Objectif	recherche des différences/régularités
Genèse du parcours	inconnue (non paramétrique)
Démarche	descriptive, exploratoire, culture "algorithmique"
Méthodes	mesures de similarité (analyse factorielle, de séquences...), classifications (CAH, arbres de décision...)

Démarche typologique

Similarité entre l'ensemble des parcours

→ Matrice de distance

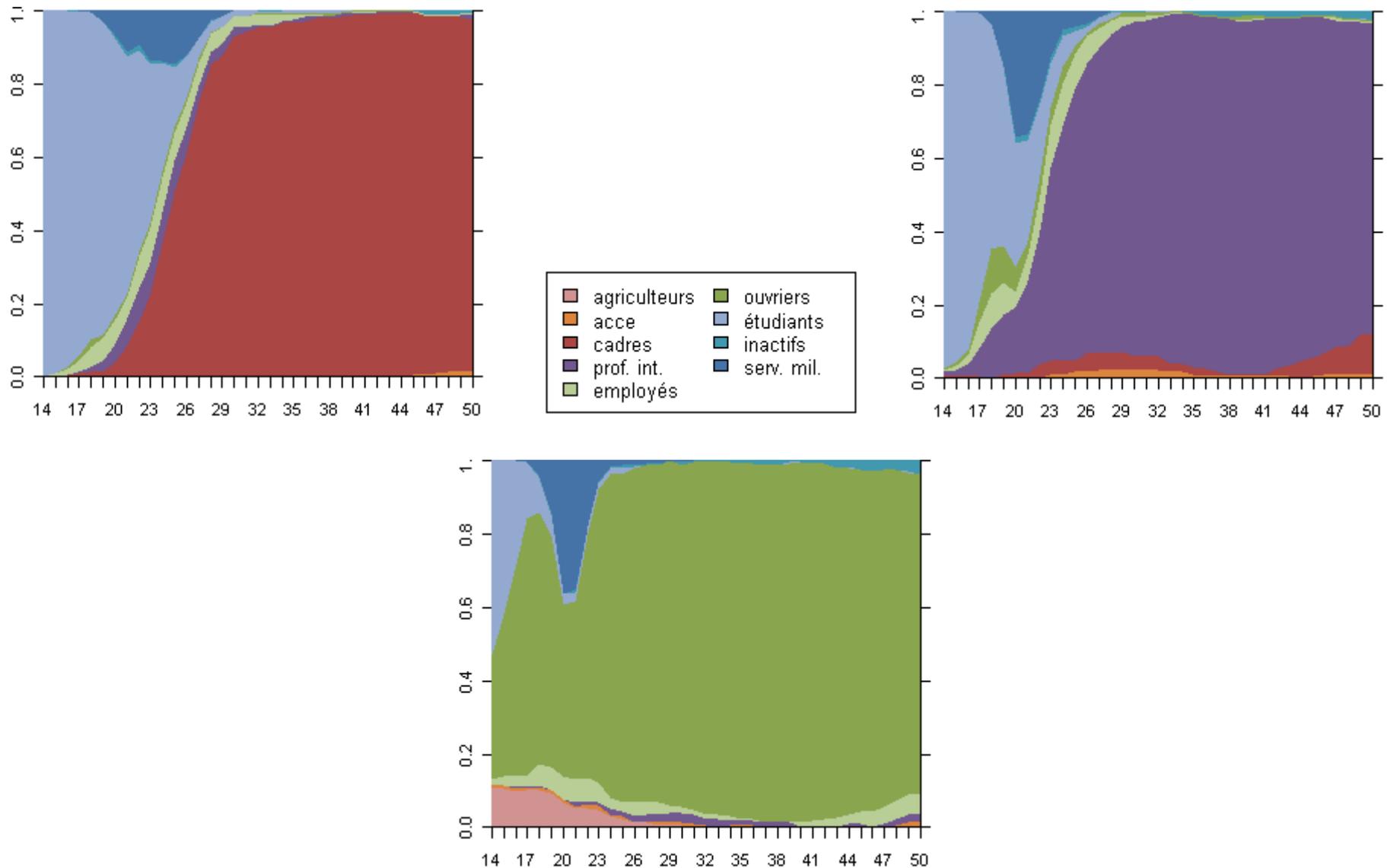
→ Classification (CAH, ...)

→ **Typologie** de parcours

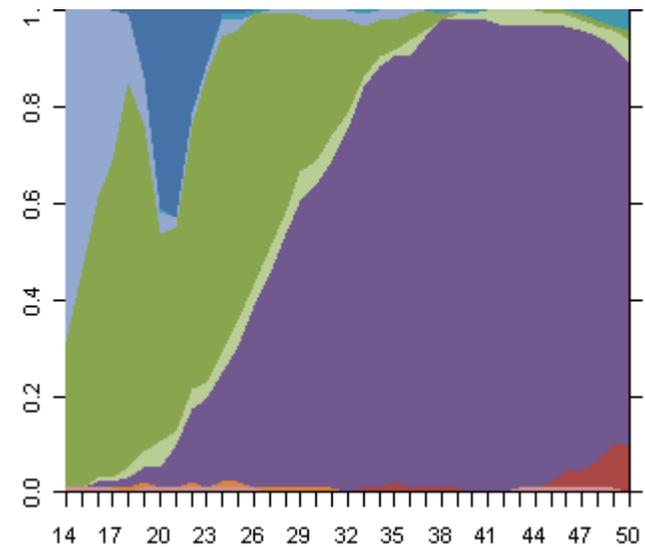
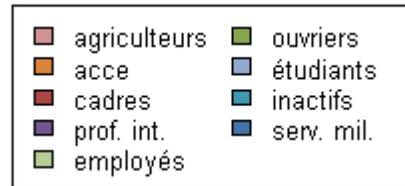
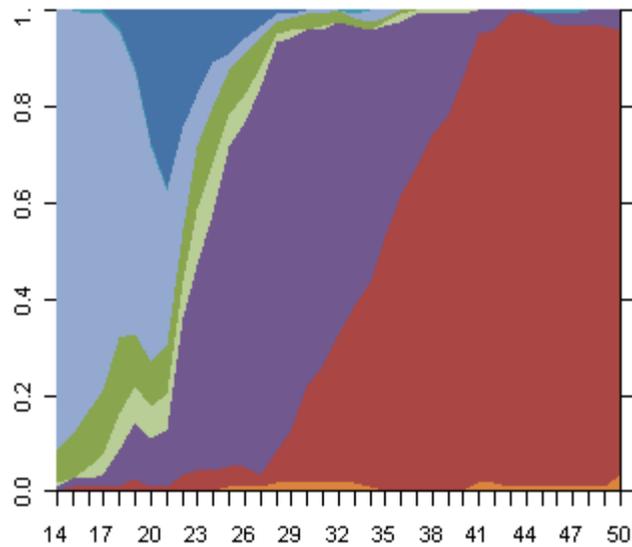
Exemple de typologie (1)

- Enquête *Biographies et entourage* (INED, 2001)
- **Carrières professionnelles** de 1421 hommes
- 37 années, de 14 à 50 ans
- 9 états:
 - 6 catégories socioprofessionnelles,
 - étudiant,
 - service militaire,
 - autre inactivité

Exemple de typologie (2)



Exemple de typologie (3)



Exemple de trajectoire

Trajectoire d'insertion professionnelle:

E = étudiant

C = chômeur

S = salarié

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

Les méthodes factorielles (1)

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

- **Codage disjonctif**

18E	18C	18S	...	25E	25C	25S
1	0	0	...	0	0	1

avec ACP -> distance euclidienne

avec AFC -> distance du χ^2

→ prise en compte de la **durée** et du **moment**

Les méthodes factorielles (2)

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

Calendrier simplifié (Analyse Harmonique Qualitative)

18-20 E	18-20 C	18-20 S	21-25 E	21-25 C	21-25 S
1	0	0	0	0,2	0,8

avec AFC -> distance du χ^2

→ prise en compte de la **durée** et du **moment**

(moment moins précis, mais moins sensible au « décalage »)

→ permet de « pondérer » des sous-périodes

Les méthodes factorielles (3)

18	19	20	21	22	23	24	25
E	E	E	C	S	S	S	S

- **Indicateurs:**

- Durées

E	C	S
3	1	4

- Transitions

EE	EC	ES	CE	CC	CS	SE	SC	SS
2	1	0	0	0	1	0	0	3

- Nombre d'épisodes, temps d'accès à un état...

Voir (Grelet, 2002)

Les méthodes séquentielles

- Les parcours individuels sont construits comme des **séquences d'états**
- Mesure de la **similarité** entre paires de séquences
- **Algorithmes** = **optimal matching analysis**, ...

Optimal Matching Analysis (1)

- Méthode utilisée en bio-informatique (ADN)
- Introduction dans les sciences sociales par Andrew **Abbott** (années 80)
- **Principe:** mesurer la dissimilarité entre paires de séquences en évaluant le **coût** représenté par la **transformation** de l'une des séquences en l'autre

Voir par exemple (Macindoe & Abbott, 2004)

Optimal Matching Analysis (2)

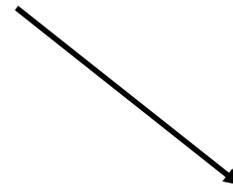
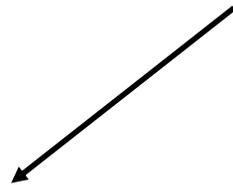
- 3 opérations élémentaires:
 - insertion
 - suppression
 - substitution

Optimal Matching Analysis (2)

Exemple :

x: B B A B A B

y: B A B A B B



x: B B A B A B

y: ~~B~~ **B** ~~A~~ **B** ~~B~~

→ 4 substitutions

x: B B A B A B

y: **B** ~~B~~ ~~A~~ ~~B~~ ~~A~~ ~~B~~ ~~B~~

→ 1 insertion, 1 suppression

Optimal Matching Analysis (2)

- 3 opérations élémentaires:
 - insertion
 - suppression
 - substitution
- à chaque opération est associé un **coût**
- la **distance** entre deux séquences est équivalente au **coût minimal** de transformation de l'une des séquences en l'autre

Le choix des coûts (1)

Question centrale de l'OMA:

- **Substitution:**

conserve la structure temporelle (**moment**)
mais altère les événements (ordre)

- **Insertion/suppression:**

déforme le temps
mais conserve l'**ordre** des événements

Le choix des coûts (2)

- Matrice des **coûts de substitution**:
 - selon **hypothèses théoriques**: hiérarchie des états...
 - à partir des **données**: en fonction des probabilités de transition...
- Coûts d'**insertion/suppression** (*indel*) :
 - **ordre** privilégié → *indel* faible /substitution
 - **moment** privilégié → *indel* élevé /substitution

Les métriques d'Elzinga (2003;2008)

- **Critique** : OMA prend mal en compte l'ordre (substituer A à B ou B à A est équivalent)
- **Plusieurs propositions** :
 - Plus long préfixe commun
 - Plus longue sous-séquence commune
 - Nombre de sous-séquences communes
 - ...

Dynamic Hamming (*Lesnard*)

- **Critique:** Les probabilités de transition dépendent du temps
- **Principe:**
 - Pas d'insertion/suppression
 - Coûts de substitution calculés à chaque période
- Applications à des emplois du temps

Quelles différences ? (1)

- Un grand nombre de méthodes, chacune appréhendant de manière spécifique:
moment + durée + ordre
- Nécessité de comparaisons systématiques
- Premiers résultats → **robustesse**

Quelles différences ? (2)

- 3 sous-ensembles =
OMA *vs* AF *vs* Elzinga
- Principales **différences** :
 - Transition *vs* durée (i.e. mobilité *vs* stabilité)
 - AF : parfois petites classes avec *états rares*
 - Parcours très *chaotiques* et/ou totalement distincts

Choix du nombre de classes

- **Selon des indicateurs statistiques:**
 - Inertie, variance inter/intra, critère de Calinski...
- **Empiriquement :**
 - arbitrage entre nb de classes interprétable et prise en compte de la diversité
 - Jusqu'à observer un processus donné, jusqu'à obtenir classes homogènes...
 - Éventuellement, observer à plusieurs niveaux de la classification

Typologie en 5 classes

cadres	32,6%
PI	27,0%
ouvriers	26,2%
employés	8,7%
--> acce	5,4%

Comment caractériser les classes ?

- **Indicateurs:**
 - ✓ durées, transitions, nb d'épisodes, temps d'accès...
 - ✓ homogénéité (distance intra, entropie...)
- **Parangon** : trajectoire « idéal-typique »
- **Représentations graphiques:**
 - ✓ chronogrammes
 - ✓ tapis (= *index plots*)

Indicateurs

		cadres	PI	ouvriers	empl	acce
Effectif		437	362	352	117	73
%		32,6	27,0	26,2	8,7	5,4
Entropie		0,310	0,357	0,351	0,434	0,471
Dintra		20,8	23,2	21,9	28,0	31,8
nb états		3,5	3,7	3,0	3,6	4,2
nb transitions		3,0	3,5	2,9	3,5	3,9
durée moyenne	agri	0,0	0,1	0,8	0,8	0,1
	acce	0,2	0,2	1,7	0,4	19,9
	cadre	22,1	0,7	0,0	0,4	2,2
	pi	3,8	23,8	1,8	0,9	4,3
	empl	1,1	2,9	1,0	23,5	2,1
	ouvr	0,7	3,0	28,3	5,5	2,7
	sm	8,0	4,7	1,8	3,2	4,1
	inact	0,1	0,3	0,3	1,2	0,4
etu	1,1	1,3	1,2	1,1	1,2	
nombre d'épisodes	agri	0,0	0,0	0,1	0,1	0,0
	acce	0,1	0,1	0,2	0,1	1,2
	cadre	1,2	0,2	0,0	0,1	0,4
	pi	0,6	1,3	0,2	0,3	0,7
	empl	0,2	0,5	0,2	1,4	0,4
	ouvr	0,2	0,7	1,8	1,0	0,7
	sm	1,1	0,9	0,6	0,7	0,9
	inact	0,1	0,1	0,1	0,2	0,1
etu	0,6	0,7	0,6	0,6	0,7	

Parangons

CADRES: 26 ans → professeur de maths

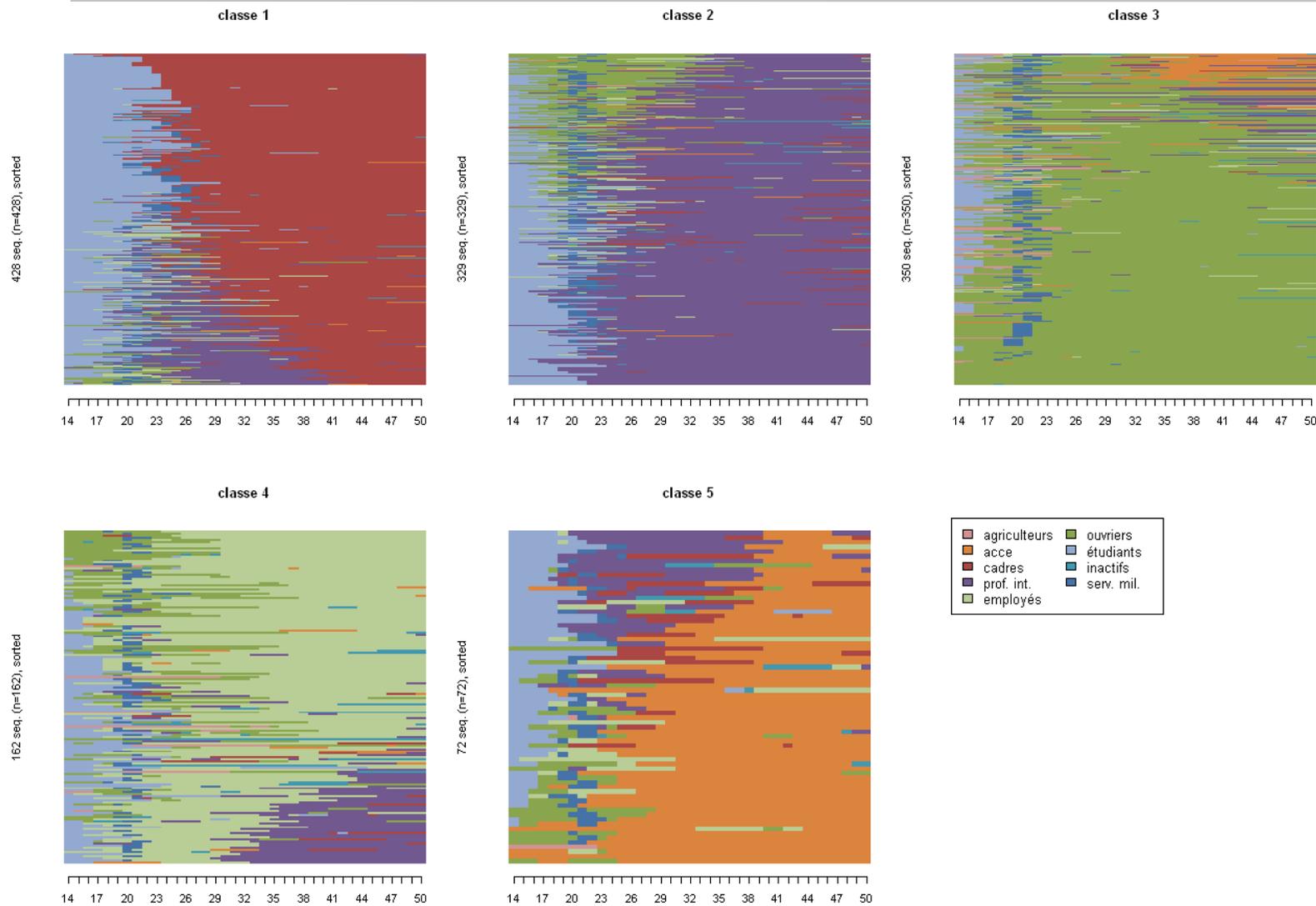
PI: 19 ans → ouvrier (ouvr)
23 ans → chef d'atelier
27 ans → professeur d'atelier

OUVRIERS: 16 ans → manutentionnaire, monteur de cloisons,
chauffeur magasinier

EMPLOYES: 17 ans → maroquinier (ouvr)
24 ans → employé de presse

ACCE: 21 ans → représentant dans l'entreprise familiale (PI)
28 ans → directeur

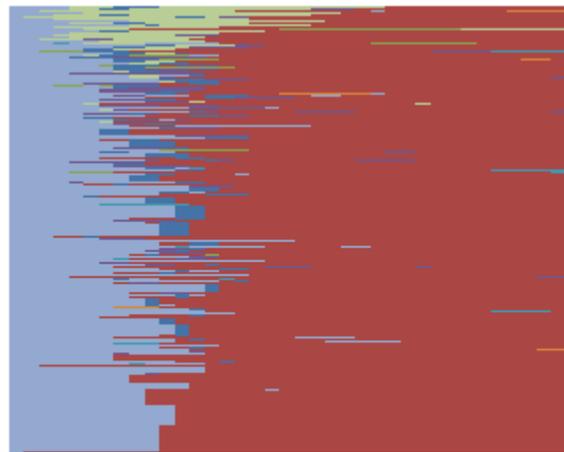
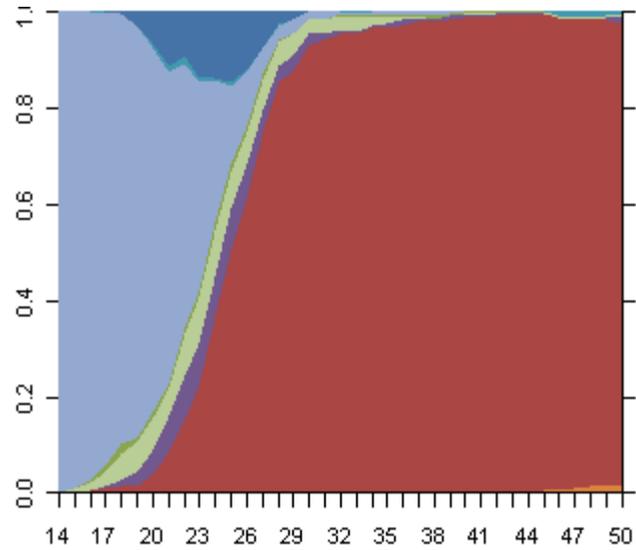
Tapis (ou index plots)



Typologies en 5 et 10 classes

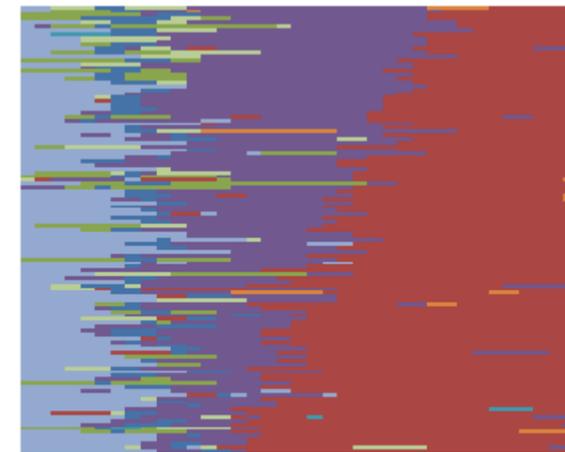
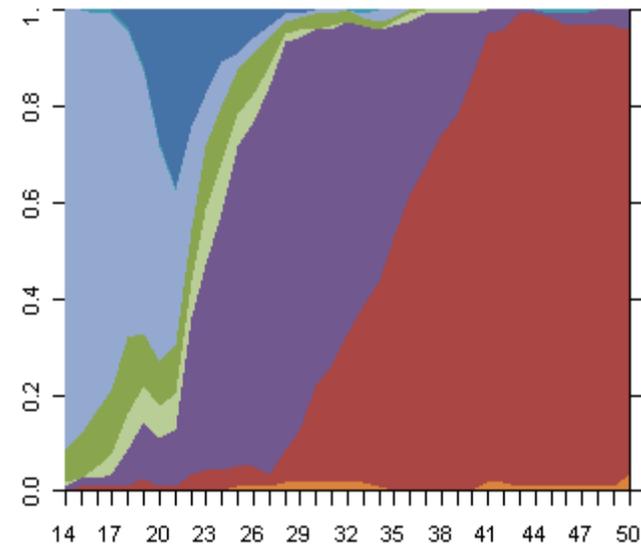
Typologie en 5 classes		Typologie en 10 classes	
cadres	32,6%	cadres	26,2%
		PI --> cadres	6,4%
PI	27,0%	PI	15,5%
		ouvriers --> PI	6,3%
		employés --> PI	5,2%
ouvriers	26,2%	ouvriers	19,2%
		ouvriers --> PI/acce	7,0%
employés	8,7%	employés	6,5%
		ouvriers --> employés	2,2%
* --> acce	5,4%	* --> acce	5,4%

Cadres



14 17 20 23 26 29 32 35 38 41 44 47 50

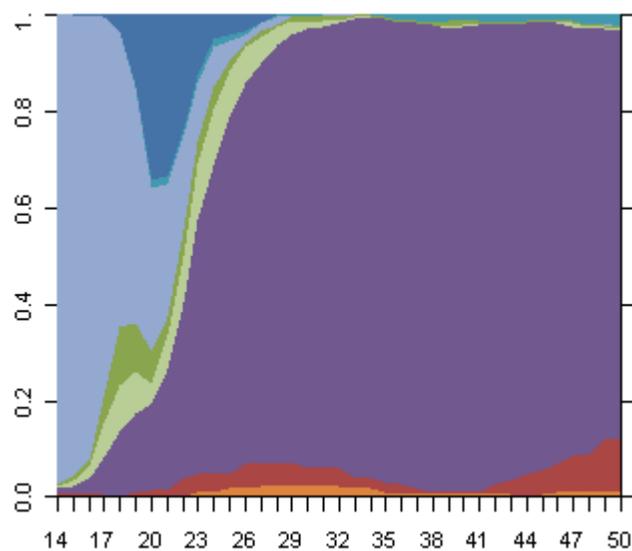
PI→Cadres



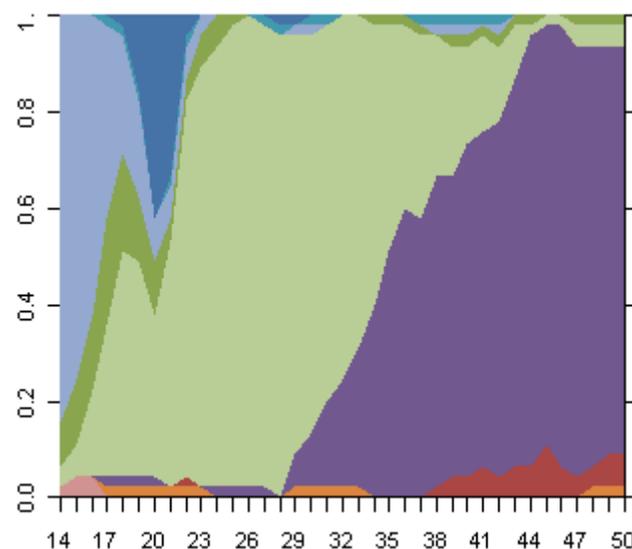
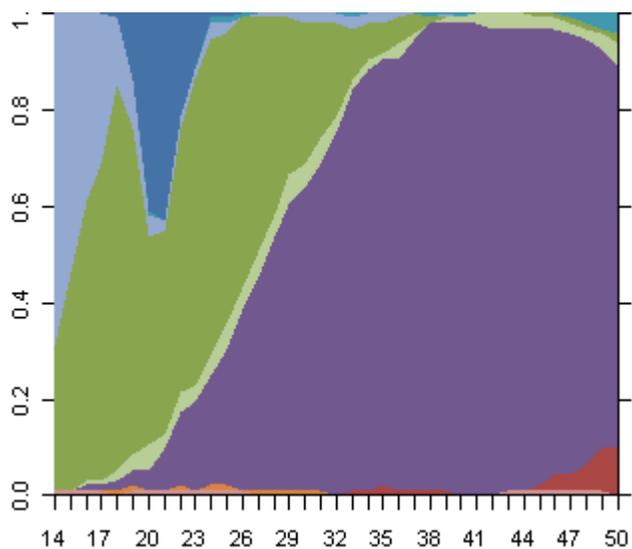
14 17 20 23 26 29 32 35 38 41 44 47 50

Professions intermédiaires

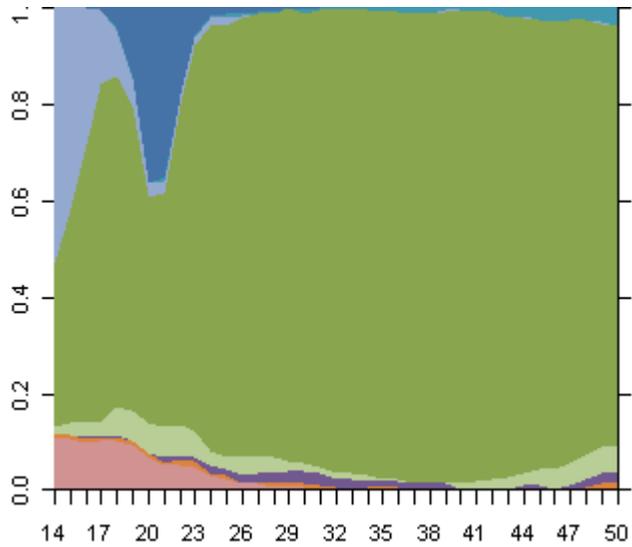
Ouvr → PI



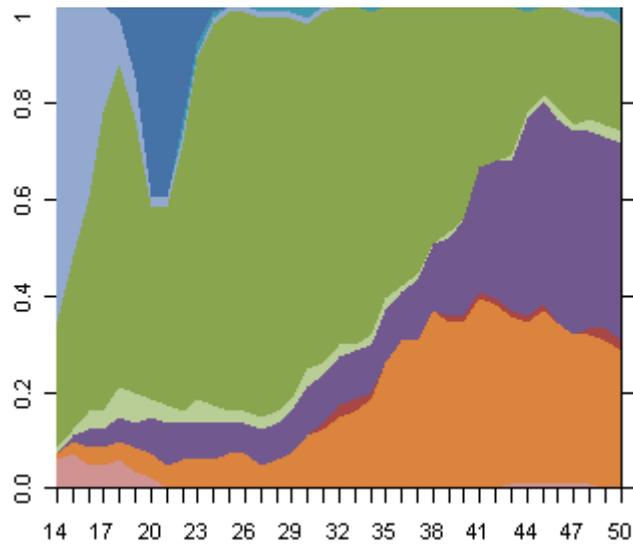
Empl → PI



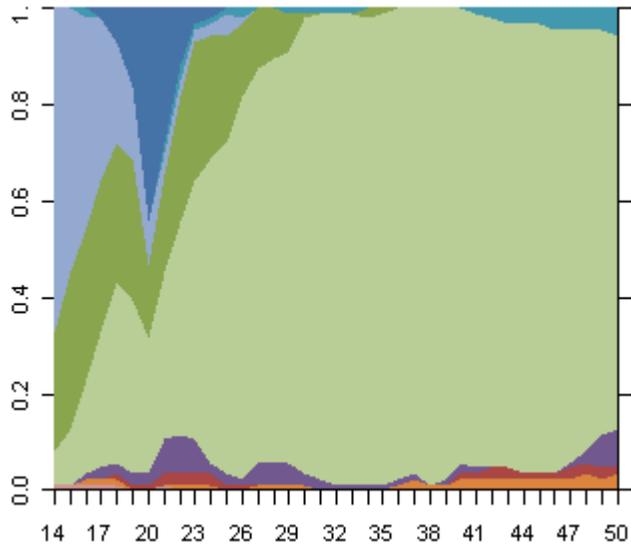
Ouvriers



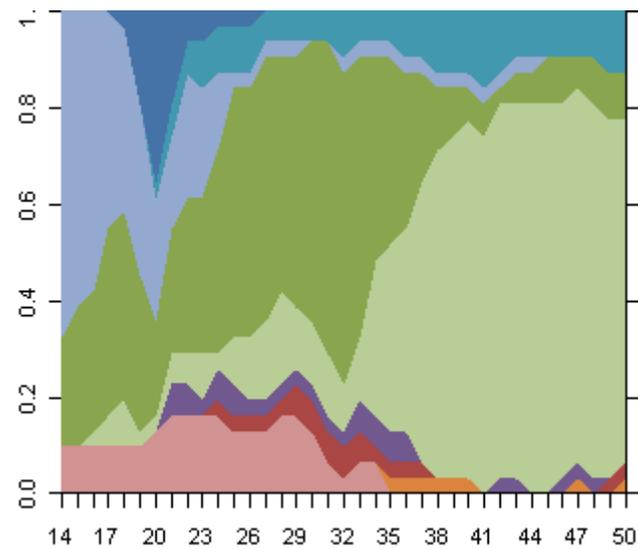
Ouvr \rightarrow PI/acce



Employés



Ouvr \rightarrow empl

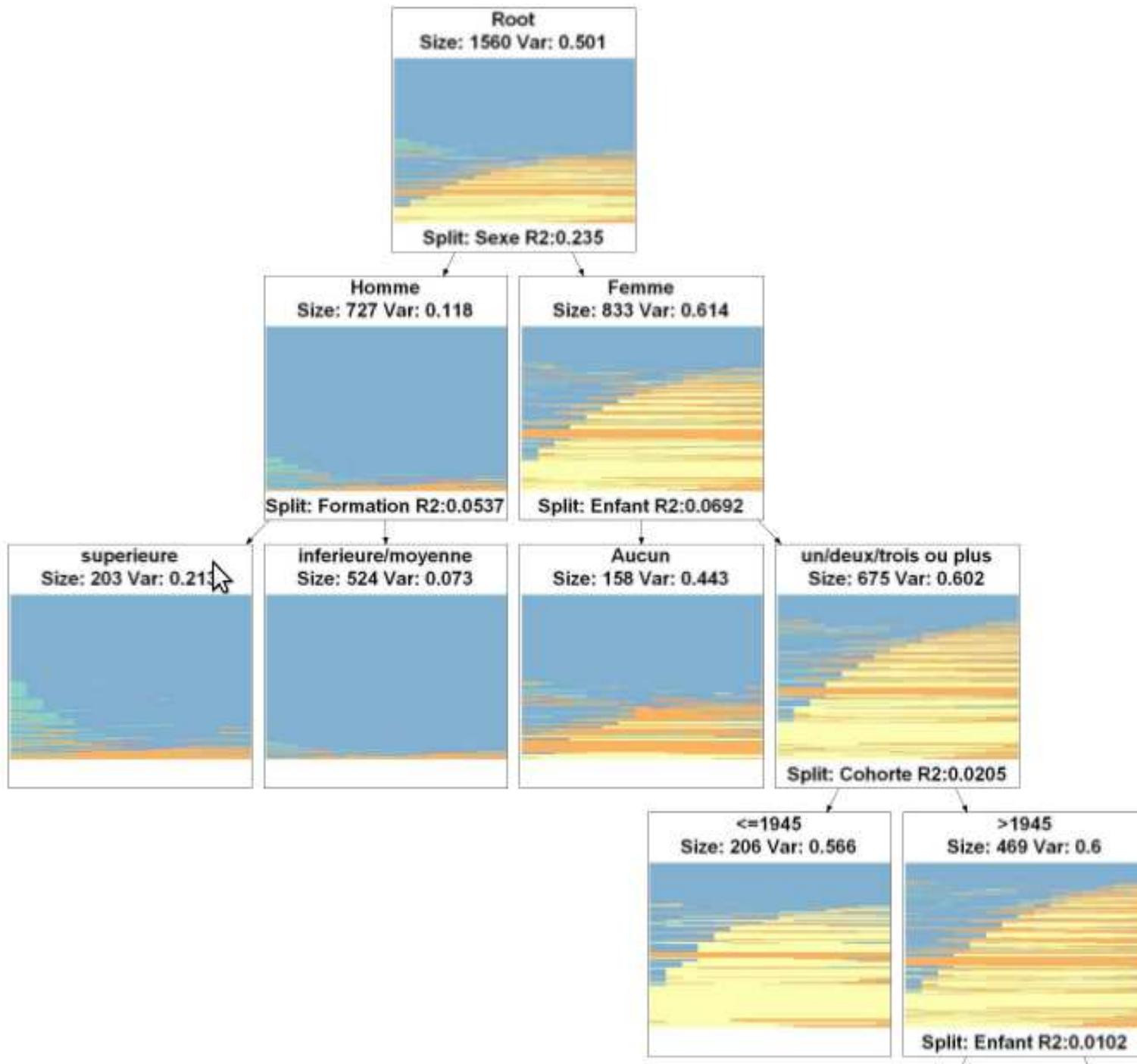


Poursuivre l'exploration

- **Caractérisation** statistique et graphique, à différents niveaux de partition (nb de classes)
- Quels sont les **déterminants** de l'appartenance à un type de trajectoire ? *Génération, origine sociale...*
 - classe = variable à expliquer
- La trajectoire a-t-elle des **conséquences** sur le parcours ultérieur ? *Passage à la retraite...*
 - classe = variable explicative

Démarches « non-typologiques »

- Projection sur **plan factoriel** (MDS)
- Distance à un **parcours de référence**, modal, « normal »...
- Similarité moyenne par **groupes** d'individus
- pseudo-ANOVA et arbres d'induction



Quelles applications ?

- Carrières professionnelles...
 - ... ou autres trajectoires individuelles...
 - ... ou tous types de séquences
- Différents éléments des parcours de vie (Elder):
 - dans le temps et l'espace
 - parcours « multidimensionnels »
 - parcours « liés »

Logiciels pour l'OMA

- package R (*TraMiner*)
- TDA (téléchargeable gratuitement)
- Module STATA (*SQ*)
- Logiciels spécifiques: Optimize, Distance...

Conclusion

- Démarche **exploratoire** 
 - souple
 - robuste
 - puissante
- **Complémentarité** avec analyses explicatives:
 - à développer !

Bibliographie

- **ma page perso:** <http://nicolas.robette.free.fr/Publis.htm>
- Robette N., 2011, *Explorer et décrire les parcours de vie : les typologies de trajectoires*, Paris : Ceped (série « les clefs pour »), 86 p.
- Robette N., 2010, « The diversity of pathways to adulthood in France: evidence from a holistic approach », *Advances in Life Course Research*
- Robette N., Thibault N., 2008, « L'analyse exploratoire de trajectoires professionnelles: analyse harmonique qualitative ou appariement optimal? », *Population-F*, 64(3), p.621-646.

- Billari F., 2005, « Life course analysis: two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood », in R. Levy, P. Ghisletta, J.-M. Le Goff, D. Spini and E. Widmer (Eds.), *Towards an Interdisciplinary Perspective on the Life Course*, Oxford: Elsevier, pp. 261–282.
- Elzinga C., 2008, « Sequence analysis: Metric representations of categorical time series », *Sociological Methods and Research*, in revision.
- Grelet Y., 2002, « Des typologies de parcours. Méthodes et usages », *Document Génération* 92, (20), 47 p.
- Lesnard L., Saint-Pol T. (de), 2004, « Introduction aux méthodes d'appariement optimal (Optimal Matching Analysis) », *Document de travail du Crest*, (15), 30 p.
- Macindoe H., Abbott A., 2004, « Sequence analysis and optimal matching techniques for social science data », in Hardy Melissa, Bryman Alan, *Handbook of Data Analysis*, London, Sage, p. 387-406.