

Réseaux sociaux : l'exemple des vélos partagés

Patrick Flandrin

CNRS & École Normale Supérieure de Lyon
Laboratoire de Physique & IXXI



INSTITUT DE FRANCE
Académie des sciences

29 mars 2011

un concept original

- **Vélo'v** = système de vélos partagés à *grande échelle* (343 stations, 4000 vélos)
- *grand succès* depuis sa mise en place à Lyon en mai 2005 (partenariat JCDecaux et Grand Lyon)
- *exemple suivi* dans de nombreuses autres villes de France (Vélib' à Paris, Vélhop à Strasbourg, ...), d'Europe (Bicing à Barcelone, ...), d'Amérique (Bixi à Montréal, ...), d'Asie (Hangzhou, Shanghai), ...

[http://fr.wikipedia.org/wiki/Vélos_en_libre_service]

vélo'v

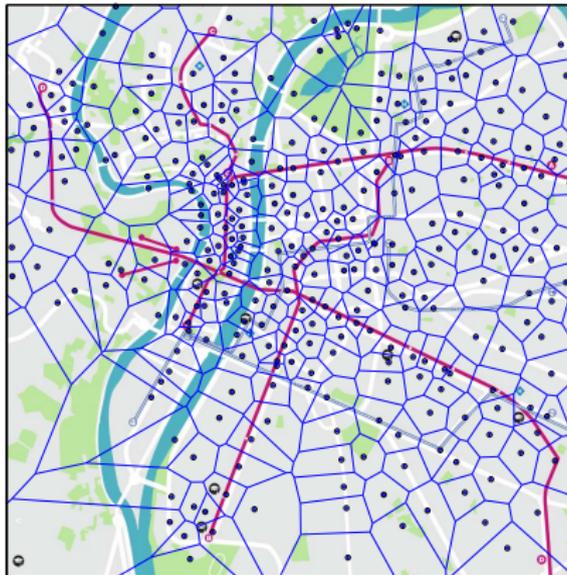


GRANDLYON

communauté urbaine

un réseau de stations et de flux

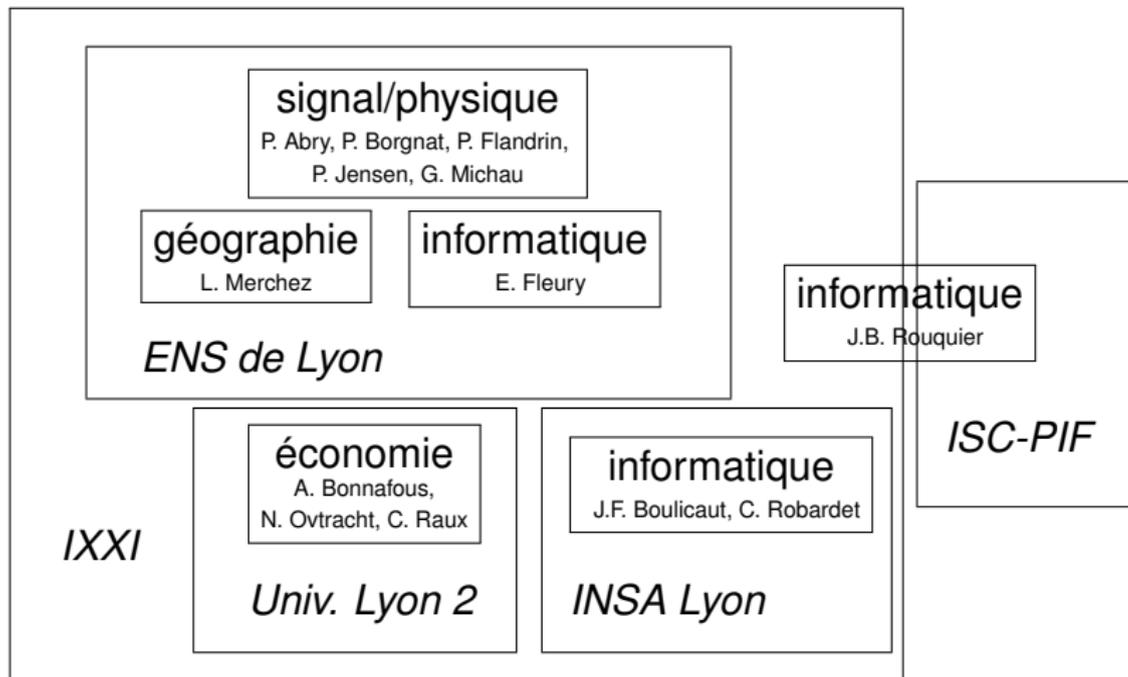
Map of Lyon with Velo'v stations: Voronoi



un objet d'étude remarquable

- un exemple de **réseau dynamique défini par ses données d'usage**
- accès à un **ensemble d'informations très riche** :
enregistrement des données de location relatives aux 13 millions de trajets effectués du 25/05/05 au 12/12/07
 - *données temporelles* : heures de début et de fin de chaque trajet
 - *données spatiales* : stations de départ et d'arrivée
 - *données annexes* : internes au système (maintenance, redéploiement) et externes (localisation GPS des stations, météo, vacances, grèves + variables socio-économiques)

une recherche pluri-disciplinaire

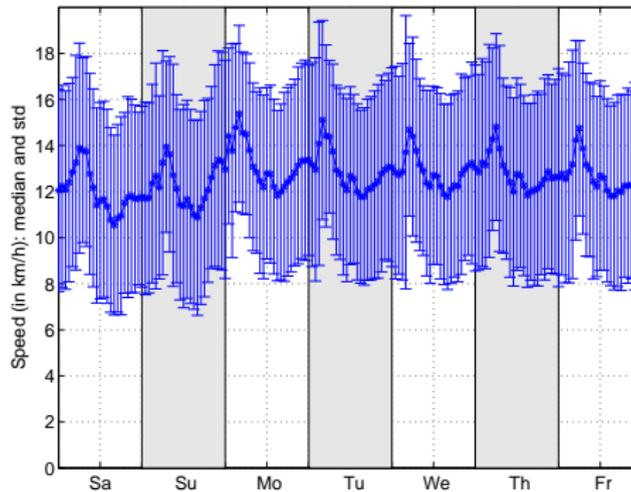


obtenir une offre adaptée

de bonnes raisons d'**utiliser** un tel système

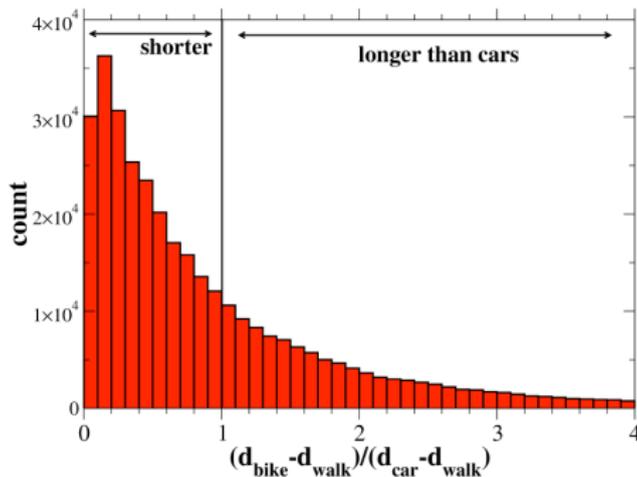
- *réaliser un **gain** par rapport à d'autres modes de déplacements*
- *avoir toujours une station "**à proximité**"*
- *trouver une station **non vide** au départ et **non pleine** à l'arrivée*

vitesse



[P. Borgnat *et al.*, *Adv. Complex Syst.*, 2011]

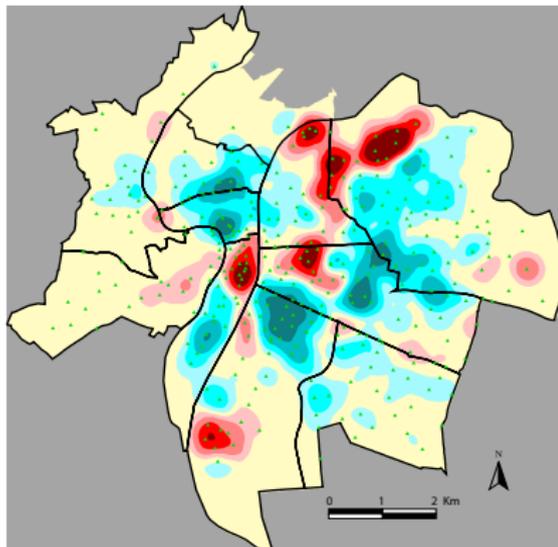
distance



[P. Jensen *et al.*, *Transp. Res. Part D*, 2010]

disponibilité

Variation de densité de Velo'v entre 7h et 8h*

période :
mai-juin (jours ouvrés)* densité relative calculée
dans un rayon de 500m,
à partir des effectifs moyens
de Velo'v présents dans les
stations

- stations Velo'v
- limites d'arrondissements

Variation de densité
(en velo'v / km²)

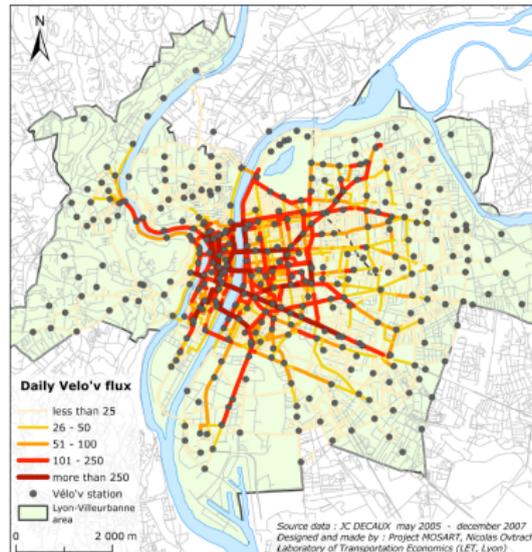
[L. Merchez, 2010]

assurer une qualité de service

de bonnes raisons de **déployer** de tels systèmes

- **garantir** le bon fonctionnement du système actuel
 - *prévoir le trafic*
 - *minimiser les redéploiements*
- **concevoir** les systèmes futurs
 - *"lire" la ville à travers le trafic*
 - *dimensionner à partir de données socio-économiques, géographiques, etc.*

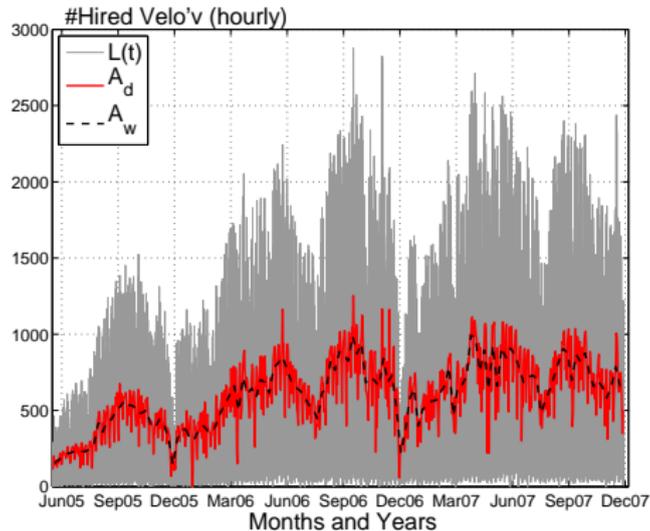
flux journaliers moyens



[P. Jensen et al., *Transp. Res. Part D*, 2010]

Vélo'v comme série temporelle

nombre global de locations par heure



[P. Borgnat *et al.*, *Adv. Complex Syst.*, 2011]

modélisation

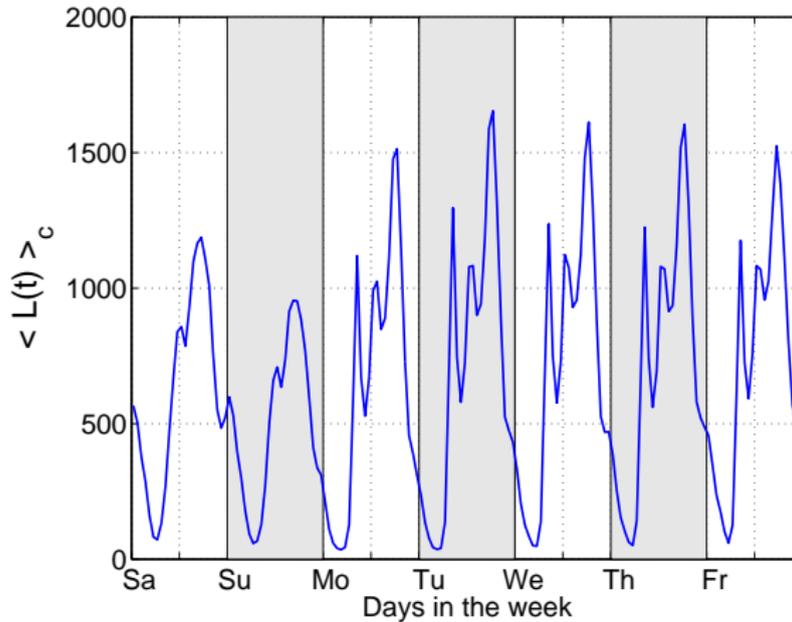
- comportements **multi-échelles**

- *cycliques* (jour, semaine, année) \Rightarrow gabarit comme moyenne cyclique horaire sur une semaine $L_c(t); t = 0, \dots, 24 \times 7 - 1 = 167$ heures
- *non stationnaires* (montée en puissance du système, saisons, semaine/week-end) \Rightarrow gabarit renormalisé pour chaque jour de la semaine $\langle L_c(t) \rangle = L_c(t) / \sum_{t \in d_7} L_c(t)$

modèle horaire

$$L(t) = \underbrace{A_d(d)}_{\text{moyenne par jour}} \times \langle L_c(t) \rangle + \underbrace{F(t)}_{\text{fluctuations}}$$

modélisation



prédiction

1. amplitude moyenne journalière

$$L(t) = \underbrace{A_d(d)} \times \langle L_c(t) \rangle + F(t)$$

- estimée par **régression linéaire multivariée** :

$\widehat{A_d(d)} = \sum_k \alpha_k Y_k$, où les Y_k sont des variables explicatives

- de conditions météorologiques (température, pluie)
- de taille du système (nombre de vélos) et de popularité (nombre d'abonnés)
- de conditions spécifiques (vacances scolaires, grèves)

prédiction

1. amplitude moyenne journalière

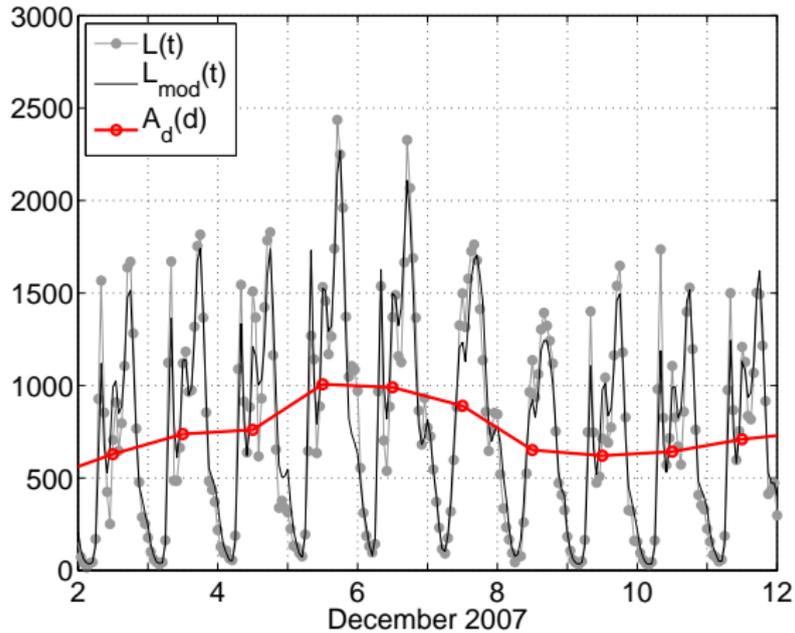
$$L(t) = \underbrace{A_d(d)} \times \langle L_c(t) \rangle + F(t)$$

- estimée par **régression linéaire multivariée** :

$\widehat{A_d(d)} = \sum_k \alpha_k Y_k$, où les Y_k sont des variables explicatives

- de conditions météorologiques (**température**, pluie)
- de taille du système (nombre de vélos) et de popularité (**nombre d'abonnés**)
- de conditions spécifiques (**vacances scolaires**, grèves)

prédiction



prédiction

2. fluctuations horaires

$$L(t) = A_d(d) \times \langle L_c(t) \rangle + \underbrace{F(t)}$$

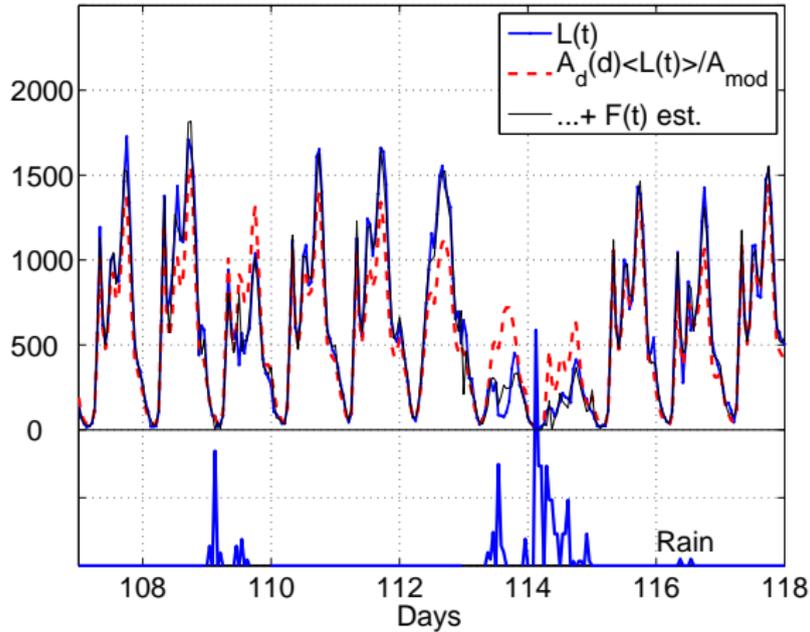
- modélisées par un processus ARX(1)

$$F(t) = \alpha F(t-1) + \beta \underbrace{R(t)}_{\text{pluie}} + \underbrace{\varepsilon(t)}_{\text{bruit blanc}}$$

- estimées par

$$\widehat{F}(t) = \hat{\alpha} \left[L(t-1) - \widehat{A_d(d)} \times \langle L_c(t-1) \rangle \right] + \hat{\beta} \widehat{R}(t)$$

prédiction



Vélo'v comme réseau dynamique

graphe dynamique

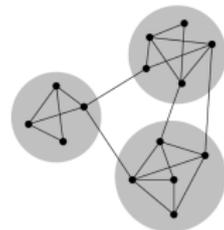
- **réseau** caractérisé par ses
 - *sommets* = stations
 - *arêtes* = couplages entre stations, mesurés par $T[n, m](t)$, nombre de trajets de la station n à la station m à l'instant t , agrégés sur une durée Δ (flux dirigés de trafic)
- analyse d'un **graphe dynamique** orienté
 - à *grande échelle* ($\Delta = 1$ semaine) pour la recherche de communautés
 - à *petite échelle* ($\Delta = 1$ heure) pour l'analyse des flux

[P. Borgnat *et al.*, Proc. ECCS, 2009]

recherche de communautés

objectif

trouver des **groupes de stations**
échangeant entre elles des vélos de façon
privilegiée et différenciée



- principe = maximiser la **modularité de Newman**

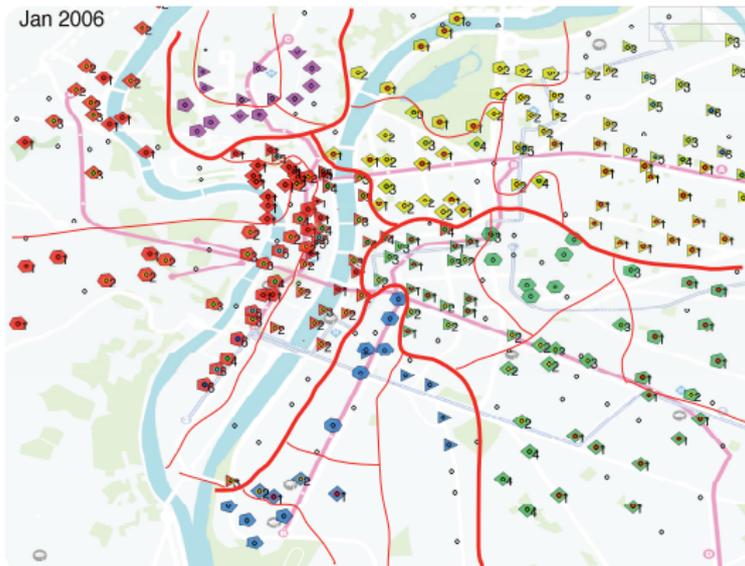
$$Q = \frac{1}{2W} \sum_{n,m} \left\{ T[n, m] - \frac{\sum_{j \neq m} T[n, j] \times \sum_{k \neq n} T[k, m]}{2W} \right\} \delta_{C_n, C_m}$$

[M.E.J. Newman, *EPJ B*, 2004]

- méthode = algorithme (hiérarchique) dit “**de Louvain**”

[V.D. Blondel *et al.*, *J. Stat. Mech.*, 2008]

communautés identifiées



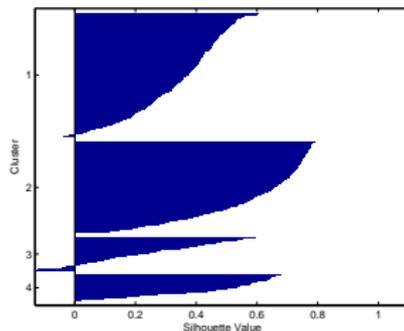
analyse des flux

● réduction de dimension

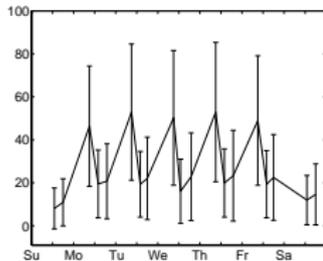
- en gardant les 1046 *flux les plus actifs* ($\sim 2\%$ des 343×343 connexions possibles)
- en se focalisant sur les 19 *pics horaires d'activité* identifiés par l'analyse temporelle globale

catégorisation

*mesure de distance entre flux
par corrélation et regroupement
par un algorithme "K-means" \Rightarrow
4 groupes significatifs*

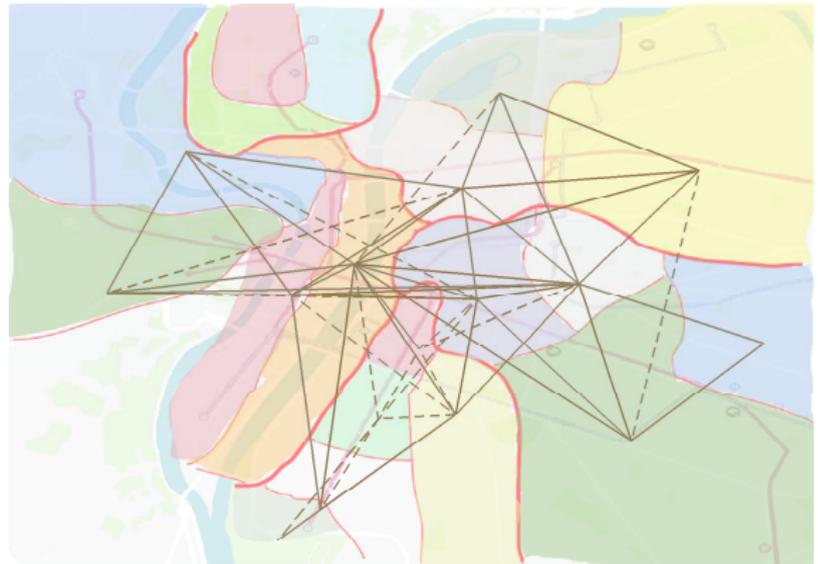
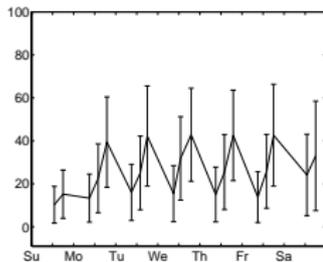


groupe 1



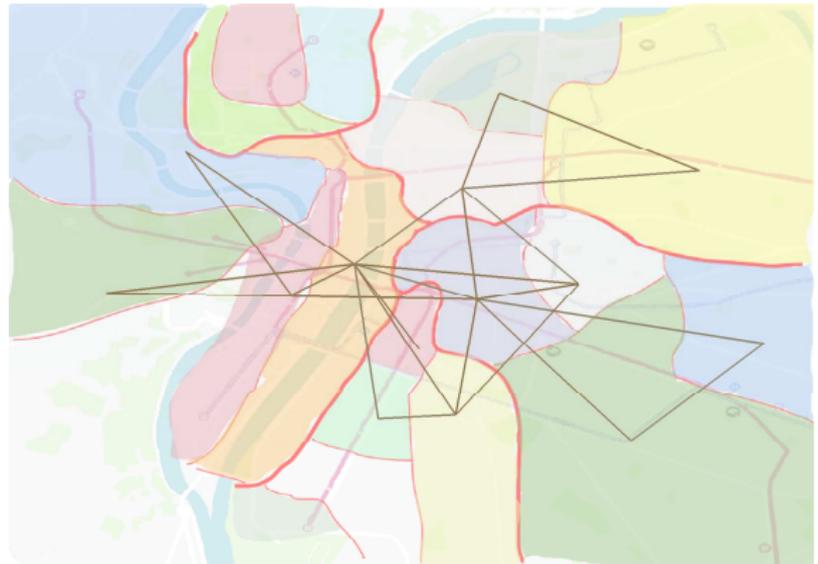
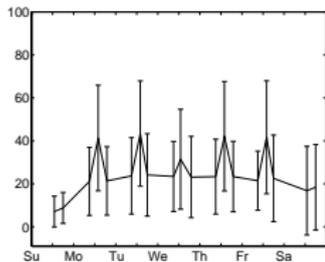
semaine, matin

groupe 3



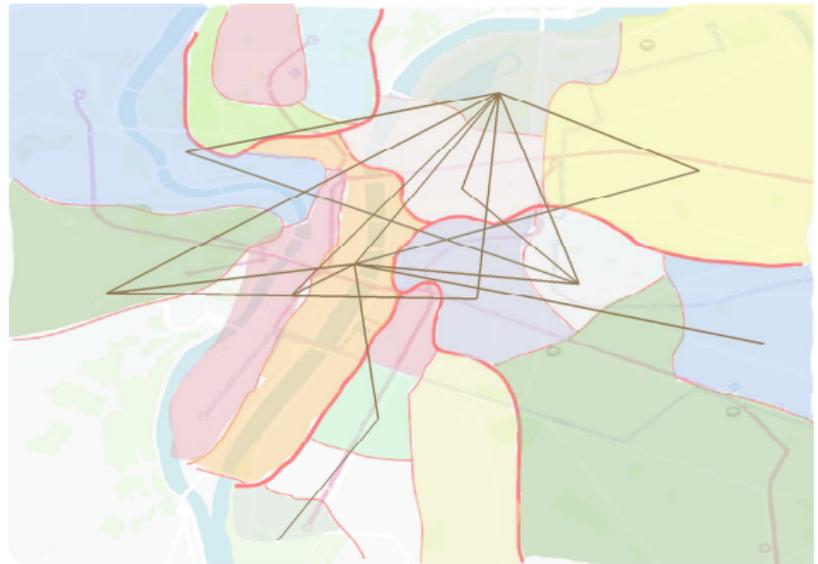
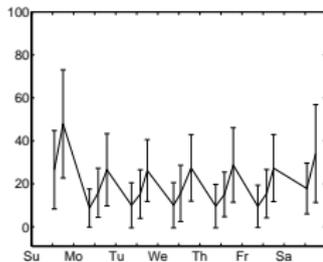
semaine, soir

groupe 2



semaine, midi

groupe 4



dimanche, après-midi

Vélo'v comme réseau social

usages

- objectif
 - relier les usages de Vélo'v à des **données socio-économiques** (*INSEE*) : population active, tranches d'âge, ménages sans voiture, revenu, emploi, etc.
 - en s'appuyant sur des variables à l'échelle des **IRIS** ("Ilots Regroupés pour l'Inférence Statistique")
- difficultés
 - affecter les 343 stations aux 230 IRIS
 - s'affranchir de la forte corrélation des 227 variables *INSEE*

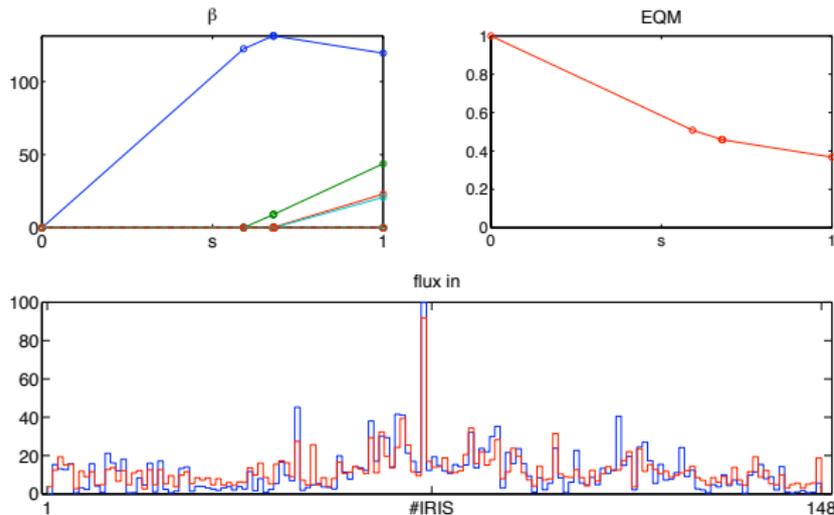
couplage entre flux et données IRIS

- modèle **linéaire (instantané)** $F^{in/out} = \mathbf{X}\beta$, avec
 - $F^{in/out} \in \mathbb{R}_+^n$; $n = 148$: flux (entrant/sortant)
 - $\{\mathbf{X} \in \mathbb{R}^{n \times p}; p < n\}$: variables INSEE par IRIS (échantillon "expert" $\rightarrow p = 24$)
 - $\beta^{in/out} \in \mathbb{R}_+^p$: vecteur de régression
- régression **parcimonieuse** (LASSO/LARS)
 - 1 $\hat{\beta}^{in/out}(s) = \arg \min_{\beta \geq 0} \|F^{in/out} - \mathbf{X}\beta\|_2$
 - 2 sous la contrainte $\|\beta\|_1 \leq s$
 - 3 en imposant également $\beta \geq 0$

[R. Tibshirani, *J. Roy. Stat. Soc. Ser. B*, 1996]

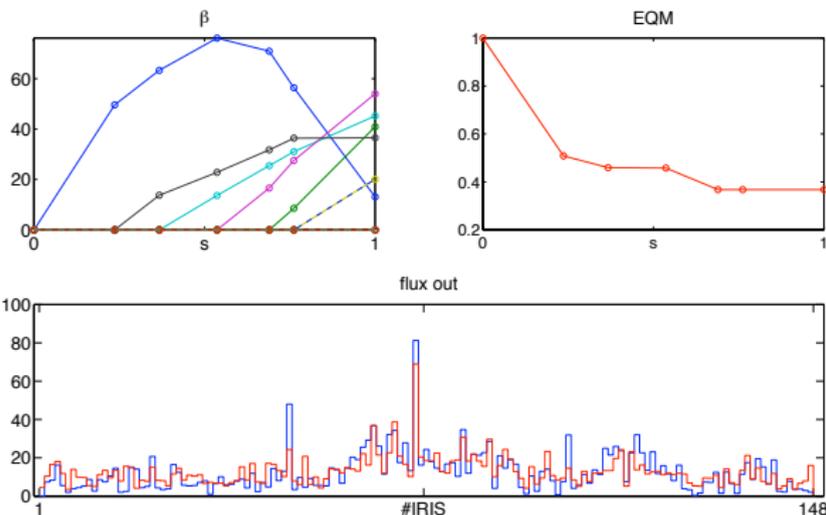
[B. Efron *et al.*, *Ann. Stat.*, 2004]

exemple : flux entrant à 8 heures en semaine



1. emplois totaux
2. population active au lieu de travail
3. population poursuivant des études
4. nombre de commerces

exemple : flux sortant à 8 heures en semaine



1. emplois totaux
2. population active déclarant marcher à pied
3. nombre de commerces
4. nombre de salariés dans le secteur tertiaire
5. population active au lieu de travail
6. population entre 20 et 24 ans

analyse/synthèse

- améliorer la **modélisation** du système actuel
 - prise en compte des contraintes de *capacité* des stations
 - modèles “*vraiment*” dynamiques
 - dynamiques *de/sur* graphes
- aider à la **conception** de systèmes futurs
 - prise en compte de données *a priori* (géographiques, socio-économiques, etc.)
 - couplage à d'*autres modalités* de transport (métro, tram, bus)
 - simulation à base d'*agents*
 - comparaison avec d'*autres systèmes* équivalents

- des **enjeux sociétaux** importants
 - nouvelles pratiques de mobilité urbaine quotidienne
 - modes de vie, modalités de transport et “sociétés innovantes”
- de **nouveaux défis** pour l’analyse et la modélisation
 - grandes bases de données hétérogènes (réduction de données sous contraintes)
 - graphes dynamiques (transposables à d’autres réseaux d’interaction)
- des **approches transversales**
 - point de vue “systèmes complexes”
 - complémentarité sciences “dures” / sciences humaines (mesures / enquêtes, SIG, etc.)

Vélo'v

Problèmes et enjeux
Analyse des données
Perspectives

problèmes ouverts
conclusion

montée en puissance du système

