# Comprendre la propagation spatiale de l'épidémie de Covid-19 avec des observations de mobilité

**Olga Mula**

**(U. Paris Dauphine PSL & Inria Paris)**

**GdT Covid – 18/06/2020**

Collaboration avec:

- PSL – Initiative Face Au Virus
  (J. Atif, O. Cappé, A. Kazakci, Y. Leo, L. Massoulié)
- Sorbonne Univesité (A. Bakhta, T. Boiveau, Y. Maday)

# Contexte

Accords d'exploitation de données à des fins de recherche entre PSL et différents acteurs: **Facebook**, Roofstreet, RATP...

# Nos objectifs

**But ultime:** Développer des outils d'analyse et de prédiction pour aider à la prise de décisions.

**Méthodologie:** Croiser données de mobilité avec données de santé et utilisation de modèles statistiques et/ou épidémiologiques.

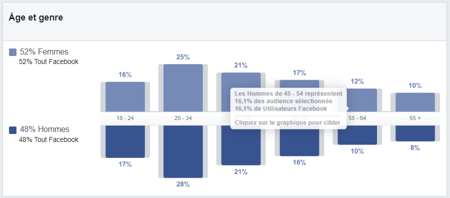
# Plan de l'exposé

1. Nature des données de Facebook.
2. Ce que les données nous apprennent sur les déplacements.
   Rapport en ligne.
3. Nos développements pour aider à mieux comprendre la propagation spatiale de l'épidémie.

# Nature des données

**Côté Facebook:**

- **Données brutes:** Données GPS d'utilisateurs qui acceptent géolocalisation. Environ 4M personnes. Certains biais d'âge.

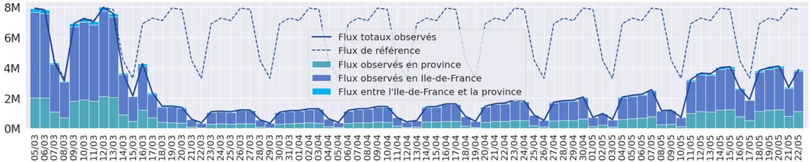



- **Traitement d'anonymisation:** Identification des comportements individuels devient impossible.

**Côté PSL:** Étude de mobilité avec données anonymisées.

**Remarques:** Facebook n'est pas le seul opérateur capable de fournir ce type de données. Nous récupérons actuellement des informations complémentaires auprès d'autres acteurs (ex: trajets de train, avion...).

**Nombre de déplacements entre départements enregistrés quotidiennement.**

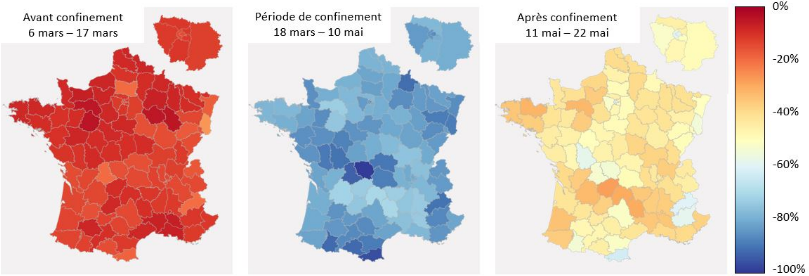



**Chiffres clés:**

-80% déplacements pendant le confinement

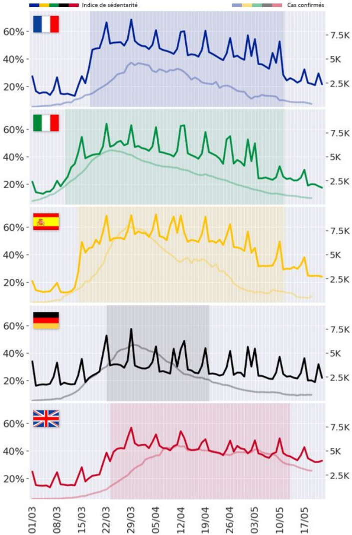
50% des déplacements habituels depuis la reprise

**Réduction des déplacements entre départements par rapport au niveau de référence de chaque département.**




**Chiffres clés:** Réduction des déplacements habituels homogène pour chaque département.

## Évolution de l'indice de sédentarité



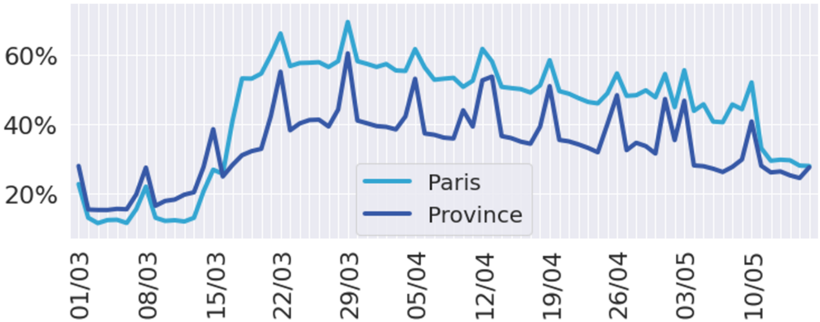

**Indice de sédentarité:**
Nb. pers. ayant passé leur journée dans un carré de 600 m par 600 m.

**FR, IT, ESP:**

- 15% avant confinement
- 50% pendant confinement
- 30% après

**UK, DE:** Effets plus modérés

# Sédentarité à Paris et en Province




**Chiffres clés:**

- Différence de 15 points pendant confinement
- Différence disparaît rapidement après confinement

## Conclusions de cette partie

- Impact fort des mesures de confinement sur la mobilité
- Timide reprise dans les premiers jours de déconfinement

## Nos travaux en cours

- Prédictions nouveaux cas par départements à l'horizon de la semaine
- Étudier l'influence sur la propagation spatiale de facteurs comme:
    - la densité de population
    - les trajets avec beaucoup de contacts (ex. gares)

# Modèle SIR Mulit-Départemental avec mobilité

Setting:

- $K$ régions
- $N_i(t)$ personnes au temps $t$ dans la région $i$.
- Population totale supposée constante:

$$N = \sum_{i=1}^{K} N_i(t), \quad \forall t \geq 0.$$

Evolution population par région:

- $\lambda_{i \to j}(t) \in [0, 1]$ probabilité qu'une personne en $i$ voyage en $j$ au temps $t$.

$$\frac{\mathrm{d}N_i}{\mathrm{d}t}(t) = \sum_{j \neq i} \lambda_{j \to i}(t) N_j(t) - \sum_{j \neq i} \lambda_{i \to j}(t) N_i(t)$$
$$= \text{Nombre arrivées - Nombre de départs.}$$

- Avec les données Facebook, nous avons des estimations des $\lambda_{i \to j}(t)$ par tranches de 8 heures.

On répartit la population $N_i(t)$ de chaque département en trois compartiments

$$N_i(t) = S_i(t) + I_i(t) + R_i(t)$$

avec:

- $S_i(t)$ nombre de Susceptibles dans la région $i$ au temps $t$
- $I_i(t)$ nombre d'Infectieux
- $R_i(t)$ nombre de Rétablis (=guéris+morts)

# Modèle SIR Multi-Départemental avec mobilité

Un modèle simple porte sur les densités

$$\mathcal{S}_i(t) = \frac{S_i(t)}{N_i(t)}, \quad \mathcal{I}_i(t) = \frac{I_i(t)}{N_i(t)}, \quad \mathcal{R}_i(t) = \frac{R_i(t)}{N_i(t)},$$

et on postule une évolution suivant

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{S}_i}{\mathrm{dt}}(t) &= -\beta_i(t)\mathcal{I}_i(t)\mathcal{S}_i(t) \\
\frac{\mathrm{d}\mathcal{I}_i}{\mathrm{dt}}(t) &= -\frac{\mathrm{d}\mathcal{S}_i}{\mathrm{dt}}(t) - \gamma_i(t)\mathcal{I}_i(t) \\
\frac{\mathrm{d}\mathcal{R}_i}{\mathrm{dt}} &= \gamma_i(t)\mathcal{I}_i(t),
\end{aligned}$$

avec:

- $\beta_i(t)$ taux de transmission dans la région $i$
- $\gamma_i(t)$ taux de guérison dans la région $i$

**Ces coefficients ne sont pas connus a priori**
**$\rightarrow$ besoin de les fitter avec données de santé.**

# Modèle SIR Multi-Départemental avec mobilité

**Stratégie de modélisation de $t = 0$ à $t =$ aujourd'hui:**

1. Résoudre évolution pour les $N_i(t)$ suivant

$$\frac{\mathrm{d}N_i}{\mathrm{dt}}(t) = \sum_{j \neq i} \lambda_{j \to i}(t) N_j(t) - \sum_{j \neq i} \lambda_{i \to j}(t) N_i(t).$$

2. Caler paramètres $\beta_i(t)$ et $\gamma_i(t)$
3. Résoudre évolution pour les $\mathcal{S}_i(t)$, $\mathcal{I}_i(t)$, $\mathcal{R}_i(t)$ avec modèle SIR.
4. Déduire $S_i(t) = \mathcal{S}_i(t) N_i(t)$ et idem pour $I_i(t)$, $R_i(t)$.

**Stratégie de prédiction pour $t \geq$ demain**: en cours de discussion

# Calage des paramètres $\beta_i(t)$ et $\gamma_i(t)$

En principe, si on observait $I_i^{(\mathrm{obs})}(t)$ et $R_i^{(\mathrm{obs})}(t)$, on pourrait trouver $\beta_i(t)$ et $\gamma_i(t)$ par:

- Différences finies:

$$\hat{\gamma}_i(t) = \mathrm{DF}\left(\frac{1}{\mathcal{I}_i(t)}\frac{\mathrm{d}\mathcal{R}_i}{\mathrm{dt}}(t)\right)$$
$$\hat{\beta}_i(t) = \mathrm{DF}\left(\frac{1}{\mathcal{I}_i(t)\mathcal{S}_i(t)}\left(\frac{\mathrm{d}\mathcal{I}_i}{\mathrm{dt}}(t) + \hat{\gamma}_i(t)\mathcal{I}_i(t)\right)\right)$$

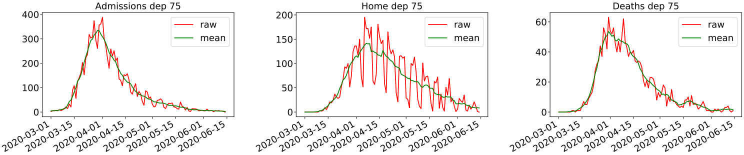
- Résolution d'un problème de contrôle optimal:

$$\min_{\substack{\beta_i(t),\gamma_i(t)\\ \in L^\infty([0,T],\mathbb{R})}} \sum_{i=1}^{K} \int_0^T \left(\|\mathcal{I}_i^{(\mathrm{obs})}(t) - \mathcal{I}_i(t)\|^2 + \|\mathcal{R}_i^{(\mathrm{obs})}(t) - \mathcal{R}_i(t)\|^2\right) \mathrm{dt}$$

tel que $\mathcal{I}_i(t)$, $\mathcal{R}_i(t) \sim$ SIR Multirégional.

**Base SIVIC: données de santé par département et par jour:**

- **Nb de guéris, décès** $\rightarrow$ estimation de $R_i(t) \rightarrow$ estimation $\gamma_i(t)$
- **Nb d'admissions hospitalières** $H_i(t)$

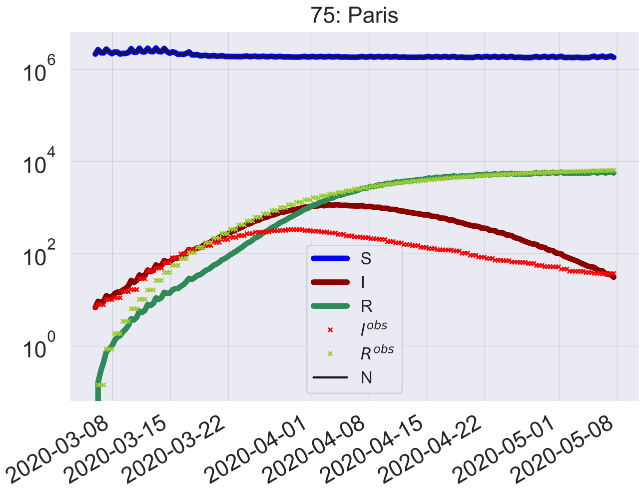



**Problème:** Pour espérer estimer $\beta_i(t)$, besoin d'estimations sur $I_i(t)$ mais on ne voit que les $H_i(t)$.

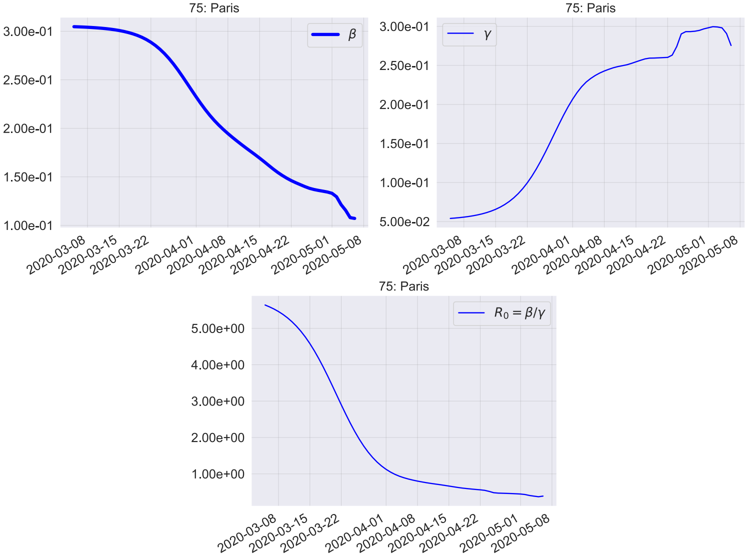
**Besoin de faire un choix:**

- **Raffiner modèle:** au prix de plus de paramètres difficiles à estimer
- **Redresser données:** Supposer que $I_i(t) = \alpha H_i(t)$ avec $\alpha \geq 1$.
- **SIR sur les hospitalisations** ($\alpha = 1$): interprétation inhabituelle du modèle.

# Quelques résultats préliminaires de calage d'un modèle SIR sur les hospitalisations



# Quelques résultats préliminaires de calage d'un modèle SIR sur les hospitalisations



# Conclusions de la partie SIR Multirégional

- Nous arrivons à reproduire l'évolution de l'épidémie par départements avec une approche de type contrôle optimal,
- Efforts actuels portent sur le développement de méthodes efficaces de prédiction.

## Merci pour votre attention!