

Avancées en apprentissage statistique pour la détection d'anomalies et le transfert

Soutenance de HDR

Rémi Emonet — 2024-02-05

UJM FST / Lab. Hubert Curien

Rapporteurs	Marianne CLAUSEL	Professeure	Université de Lorraine
	Nicolas COURTY	Professeur	Université de Bretagne Sud
	Paulo GONCALVES	DR Inria	ENS Lyon
Examinatrices	Élisa FROMONT	Professeure	Université de Rennes
	Christine SOLNON	Professeure	INSA de Lyon
Tuteur	Marc SEBBAN	Professeur	Université Jean Monnet



Inria



**Laboratoire
Hubert Curien**

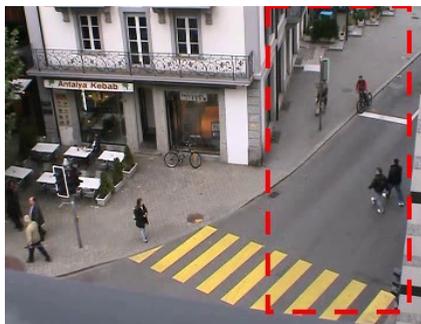
UMR • CNRS • 5516 • Saint-Étienne



**Université
Jean Monnet**
Saint-Étienne



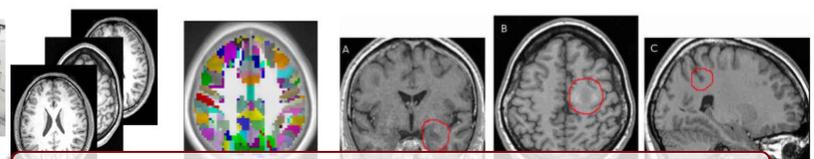
Comment gérer la rareté de certaines données d'intérêt, les anomalies ?



Projet VANAHEIM, GTT, RATP



CIFRE Bluecime, Projet MIVAO



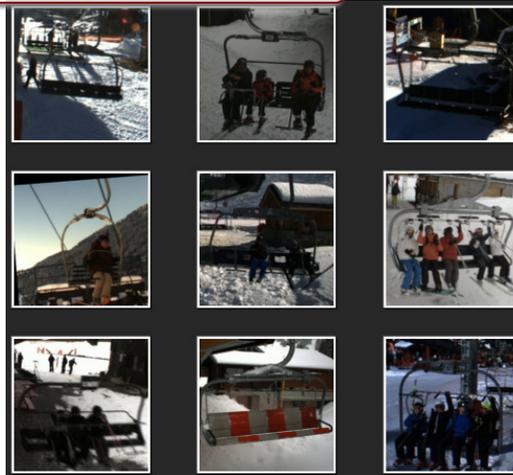
Projet région TADALoT, collab. CREATIS



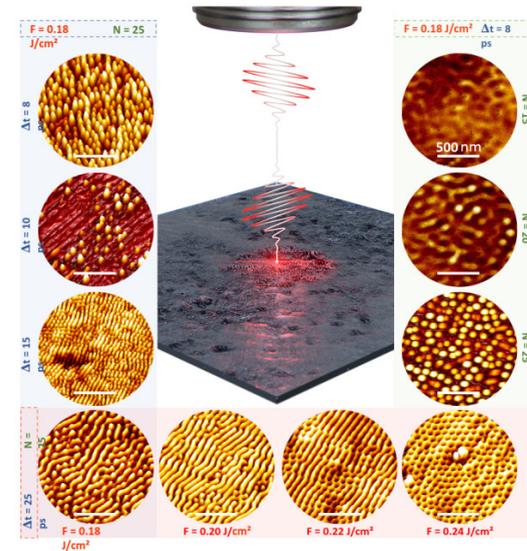
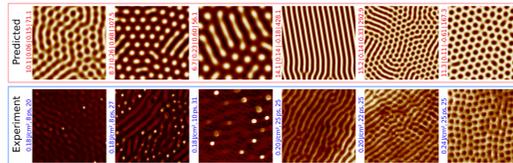
Blucime, MIVAO



Ségmentation sémantique



Collab LabHC, interaction laser-matière



1. [Trajectoire](#)
2. Focus 1 : approches pour la détection d'anomalies supervisée
3. Focus 2 : transfert hétérogène par *optimal tensor transport*
4. Projet de recherche

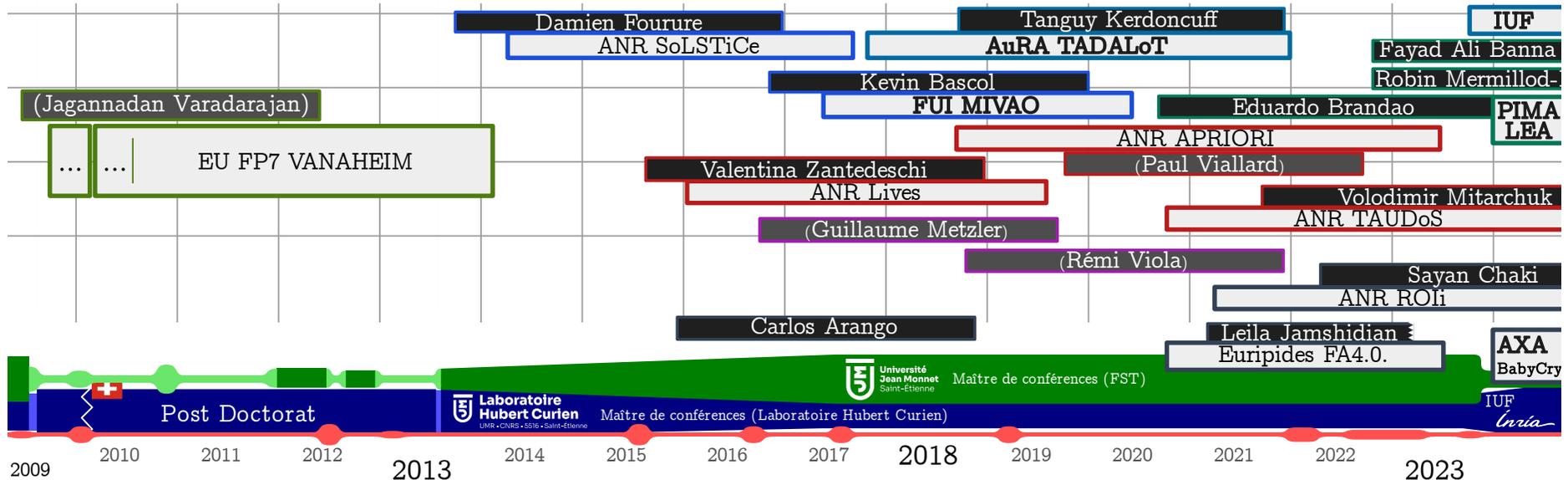
Trajectoire



Projets et encadrements

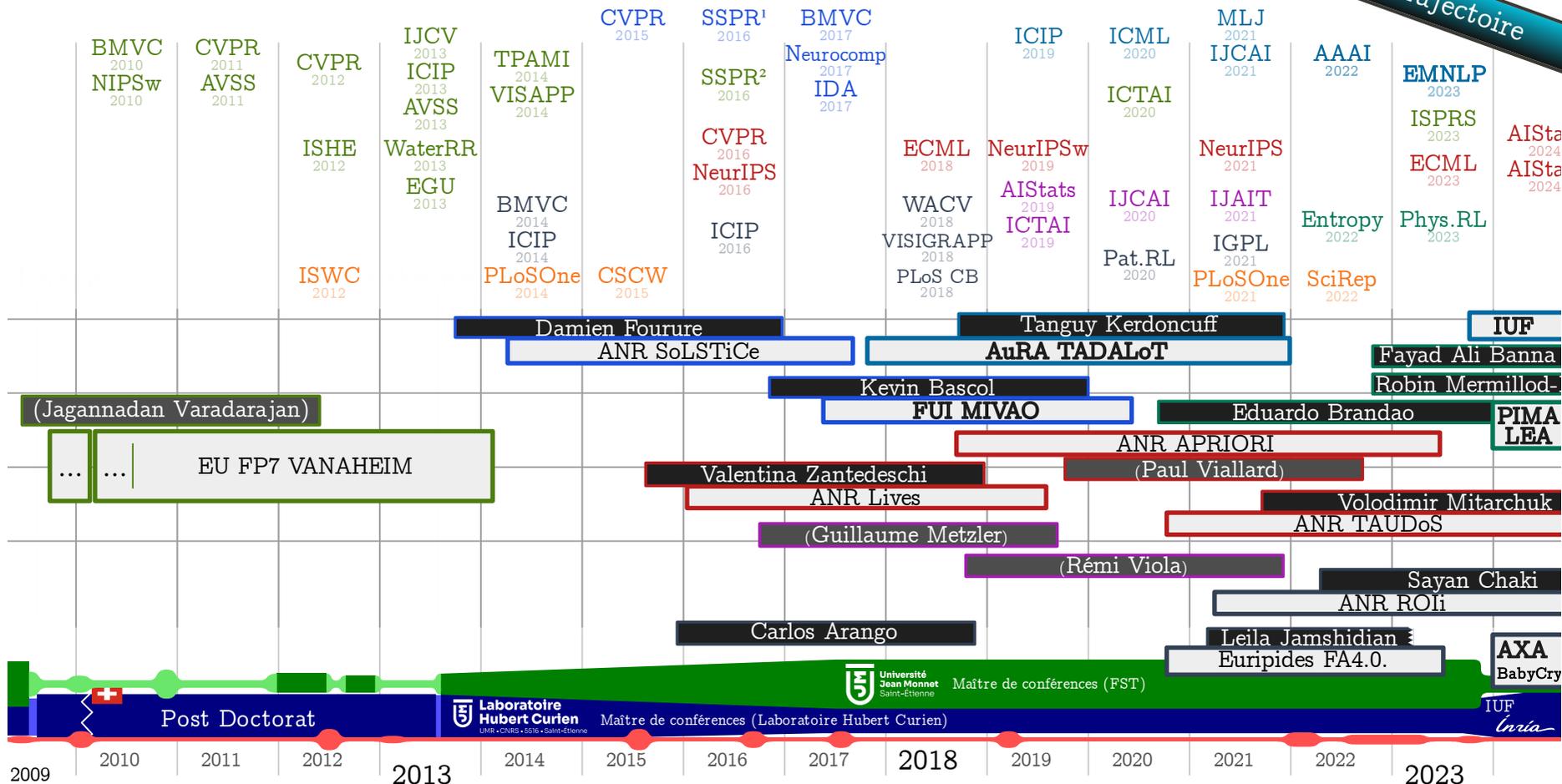
trajectoire

...



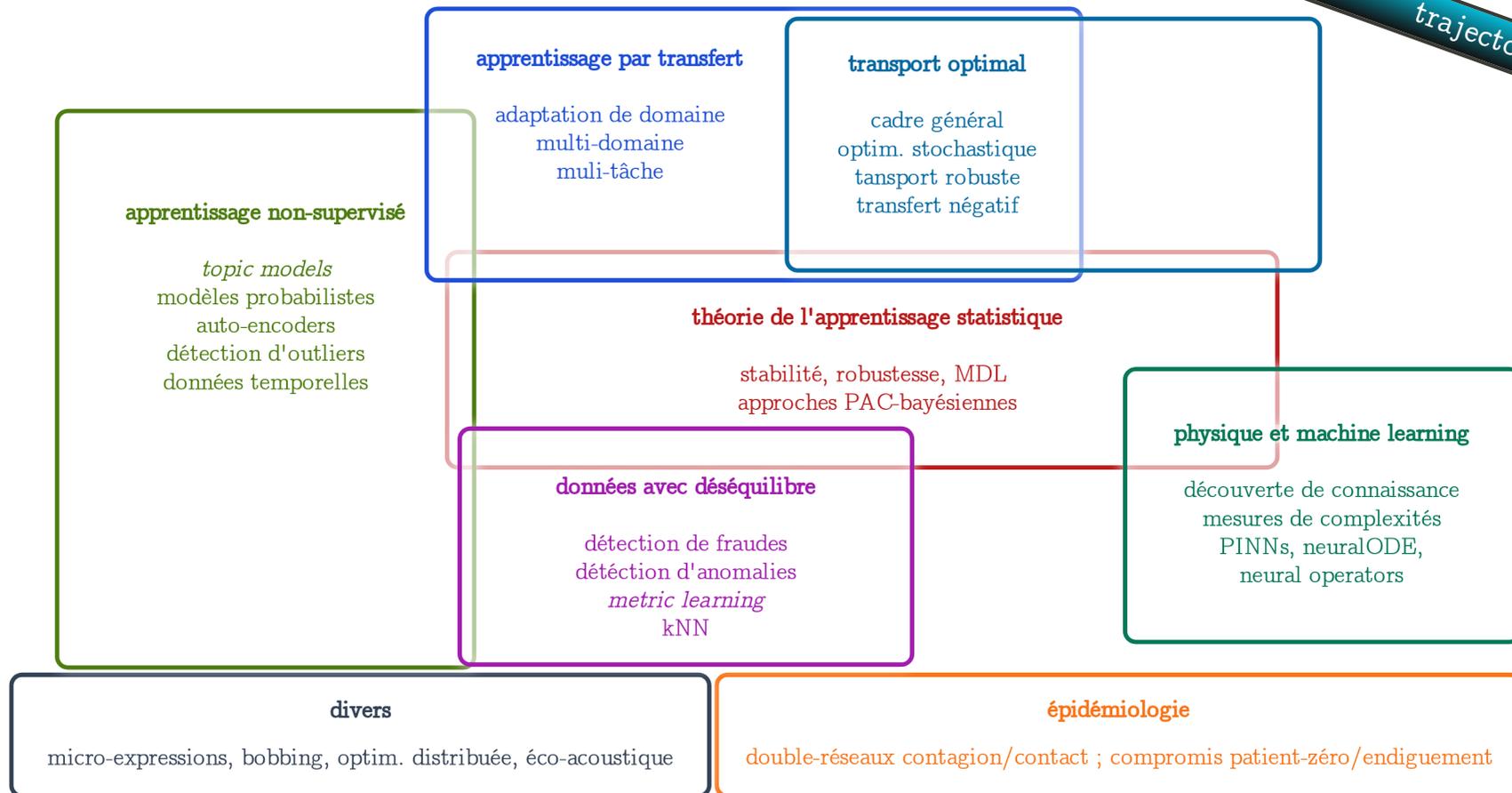
Publications (conférences et revues)

trajectoire

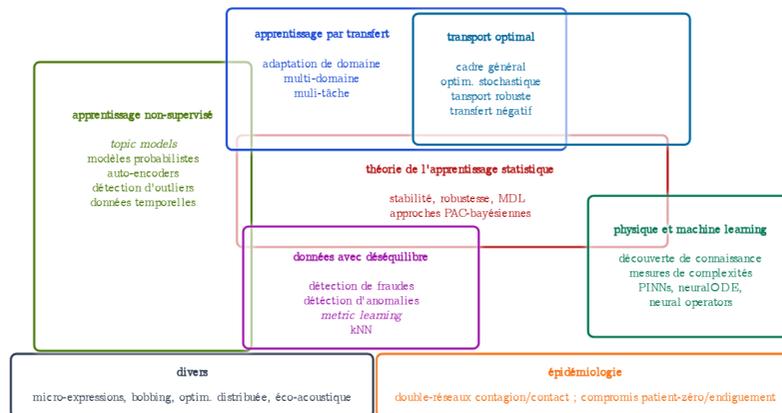


Thèmes de recherche

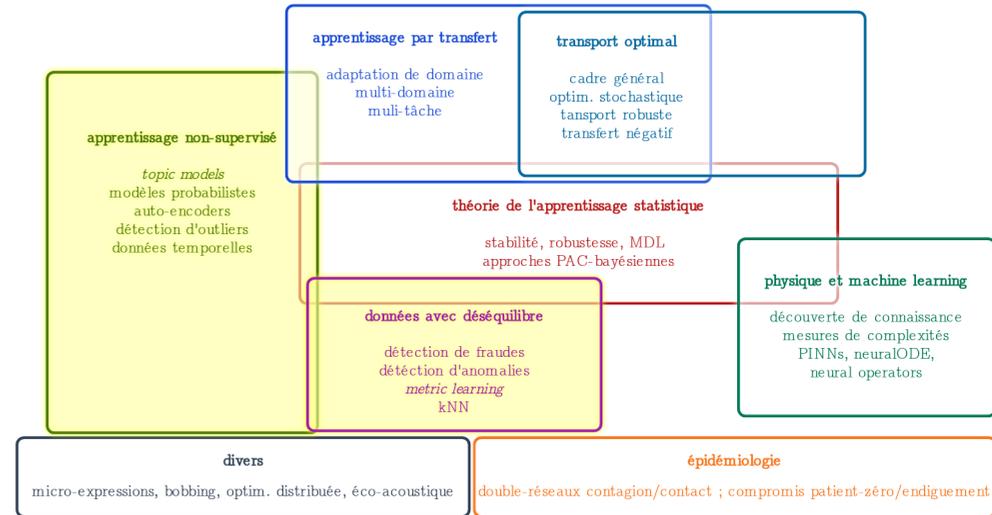
trajectoire

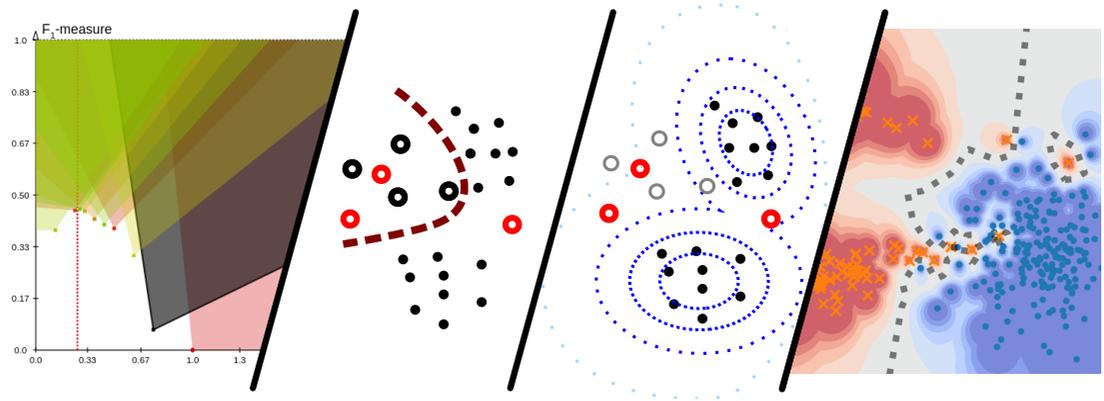
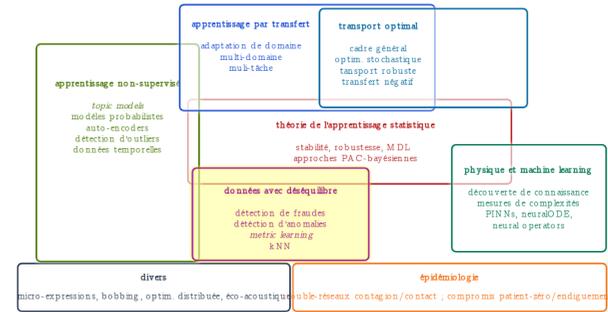


1. Trajectoire
2. Focus 1 : approches pour la détection d'anomalies supervisée
3. Focus 2 : transfert hétérogène par *optimal tensor transport*
4. Projet de recherche



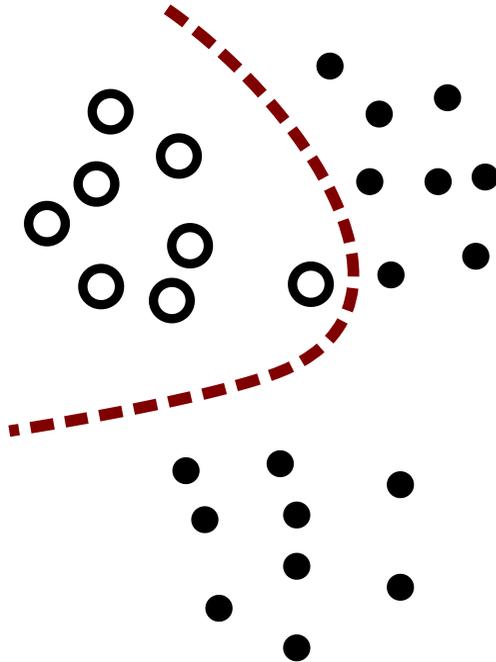
Focus 1 : approches pour la détection d'anomalies supervisée



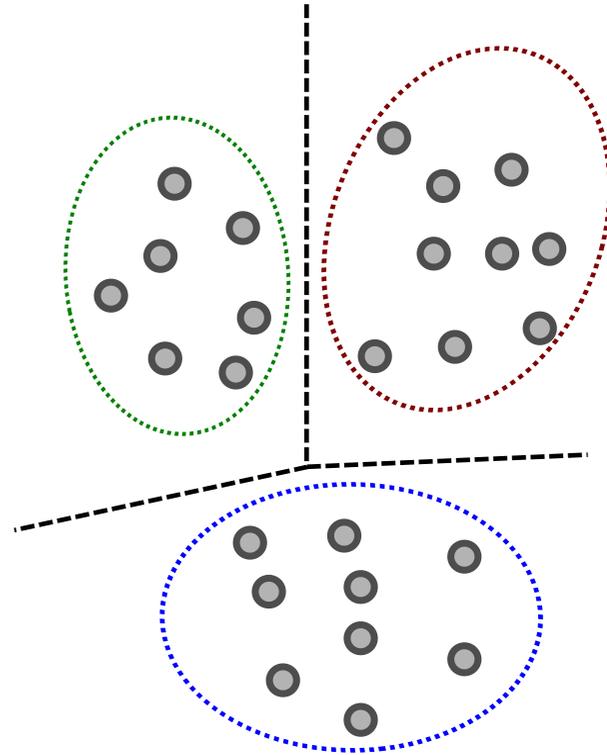


Détection d'anomalies comme classification déséquilibrée

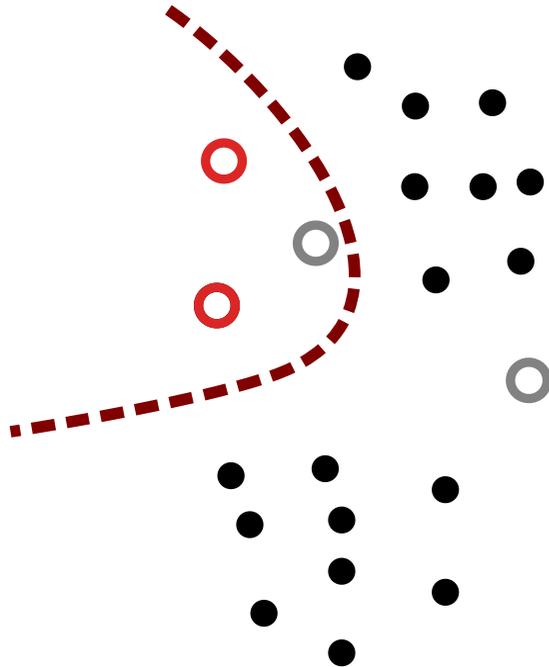
Supervisé
(avec étiquettes)



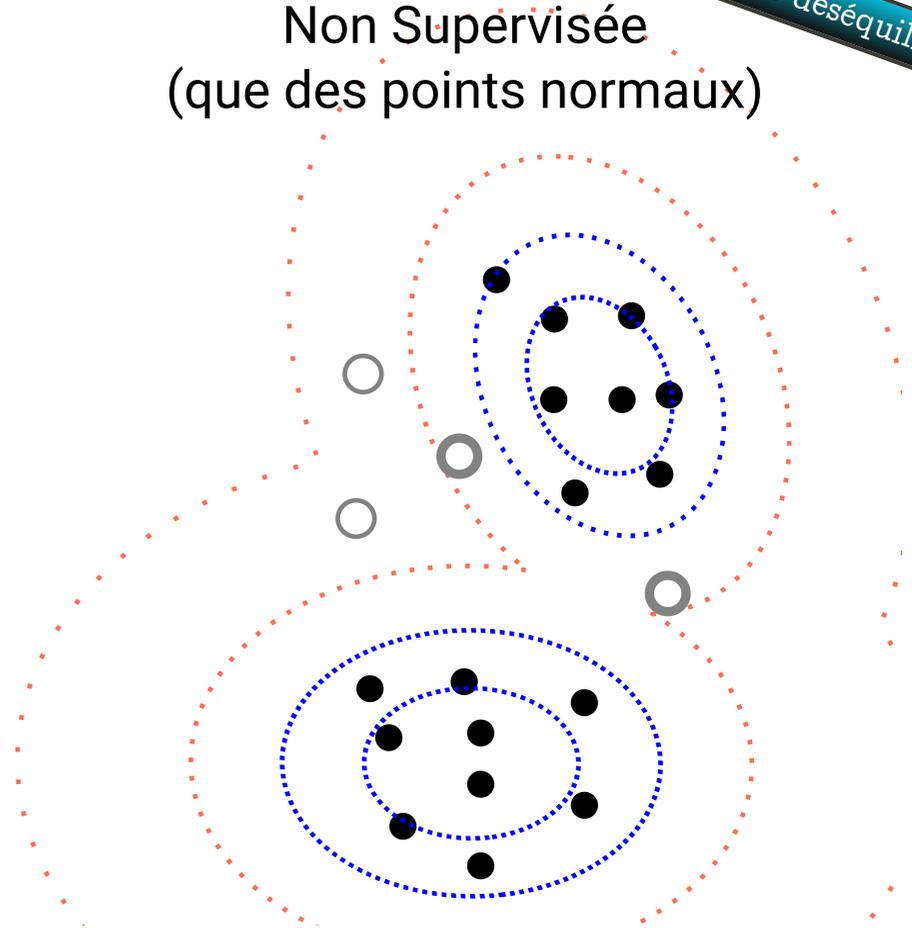
Non Supervisé
(sans étiquettes)



Supervisée
(quelques anomalies)



Non Supervisée
(que des points normaux)

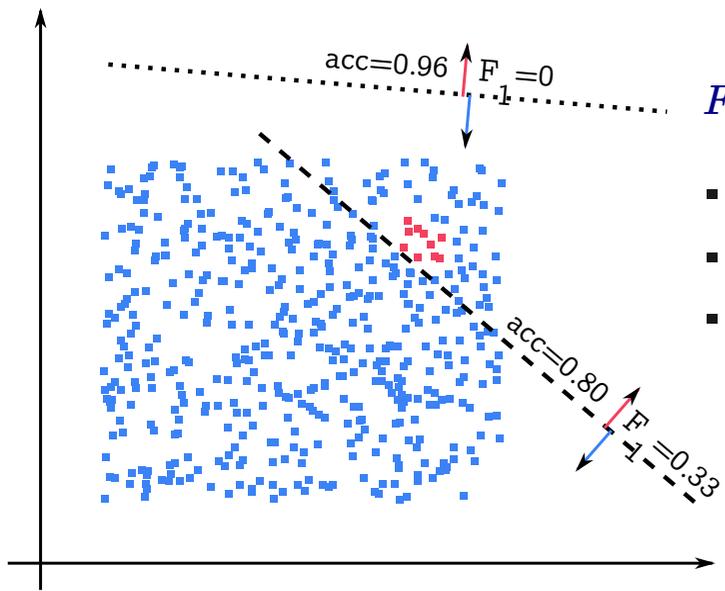


Limites de l'exactitude, utilité de la F-mesure (ou « F-score »)

FP (faux positifs) FN (faux négatifs) $TP = P - FN$

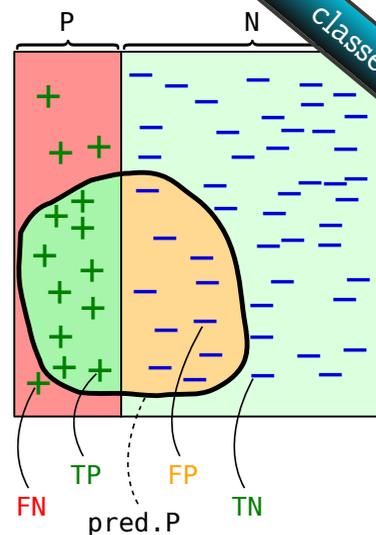
$$exactitude = 1 - \frac{FP + FN}{P + N}$$

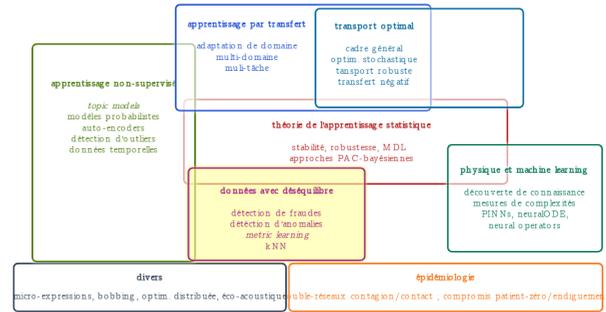
$$\text{Mesure } F_\beta = \frac{(1 + \beta^2) \cdot (P - FN)}{1 + \beta^2 P - FN + FP}$$



F_β -mesure : challenges

- discrète (comme l'*exactitude*)
- non convexe (même sa relaxation continue)
- **non-séparable**, i.e. $F_\beta \neq \sum_{(x_i, y_i) \in S} \dots$
 - ⇒ La perte d'un point dépend des autres
 - ⇒ Impossible à optimiser directement
 - ⇒ Impossible à optimiser sur des sous-ensembles (minibatches)



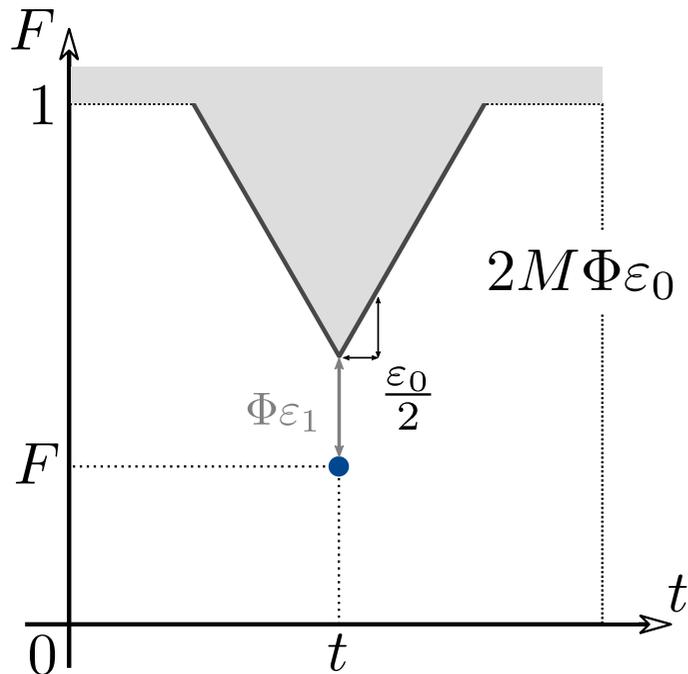


CONE

- Shameem P. Parambath, Nicolas Usunier, Yves Grandvalet

- NIPS 2014 (*Advances in Neural Information Processing Systems*)

- Observation : $F_\beta \geq t \iff (1 + \beta^2 - t) \cdot FN + t \cdot FP \leq -P(1 + \beta^2) + t(1 + \beta^2 P)$



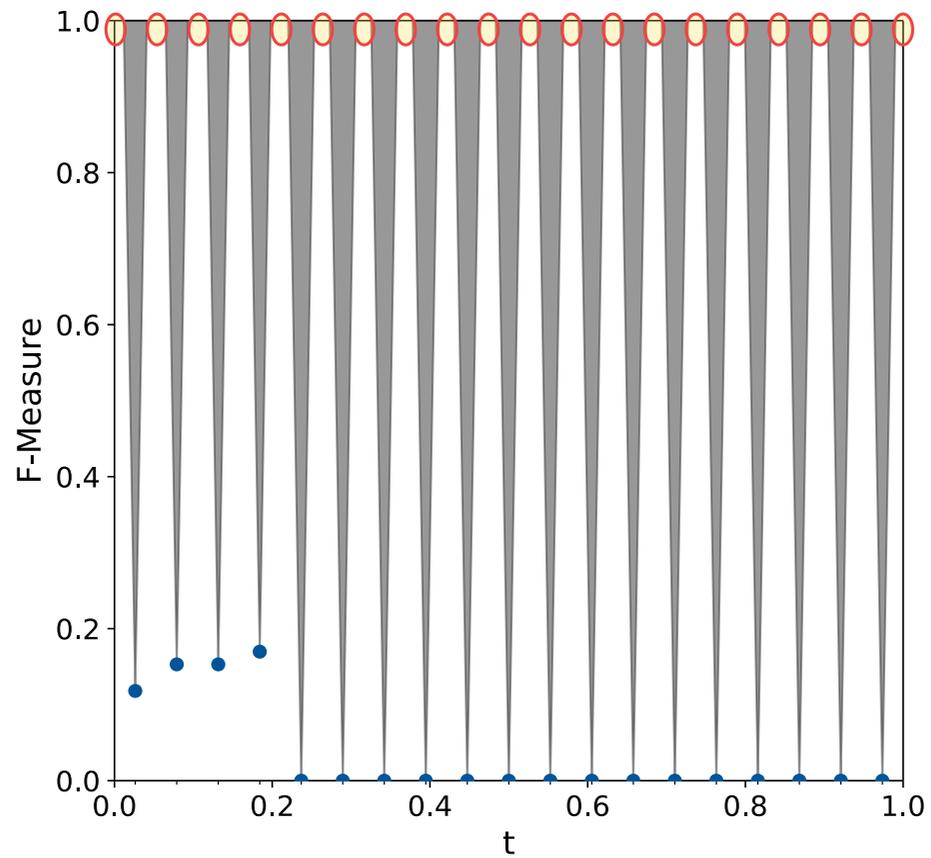
$$F_\beta^* \leq const_1 + const_2 |t - t'|$$

Étant donné

- une valeur de t (et des poids de classe $a(t)$)
- un classifieur pondéré h (optimal pour $a(t)$) donnant ... un profil d'erreur $e = (e_1, e_2) = (FN, FP)$
... une mesure $F = F_\beta(e)$
- (ε_1 , la sub-optimalité de l'apprentissage pondéré)

Visualisation de la borne de Parambath et al. (grille de 19 modèles)

classes déséquilibrées



⇒ on pourrait penser que des valeurs de 1.0 sont toujours possibles...

Aussi appelé CONE

- **Kevin Bascol**, Rémi Emonet, Élisabeth Fromont, Amaury Habrard, **Guillaume Metzler**, Marc Sebban
- AISTATS2019
 - revisite de Parambath2014
 - *interprétation géométrique*
 - bornes plus *fin*es
 - algorithme guidé par les bornes

Principes !

- passer de l'existence... à un lien précis
- garder les termes calculables
- éviter le « pire cas »
- casser la symétrie

⇒ bornes plus fines, asymétrique, utilisant $e = (e_1, e_2) = (FN, FP)$

Then for all $t' < t$:

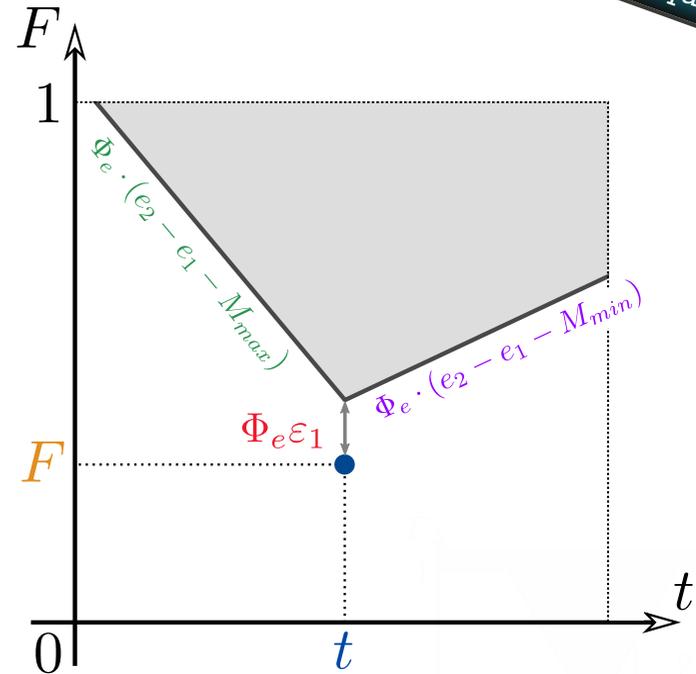
$$F(e') \leq F(e) + \Phi_{e\varepsilon_1} + \Phi_e \cdot (e_2 - e_1 - M_{max})(t' - t),$$

$$\text{where } M_{max} = \max_{\substack{e'' \in \mathcal{E}(\mathcal{H}) \\ \text{s.t. } F(e'') > F(e)}} (e_2'' - e_1'')$$

and, for all $t' > t$:

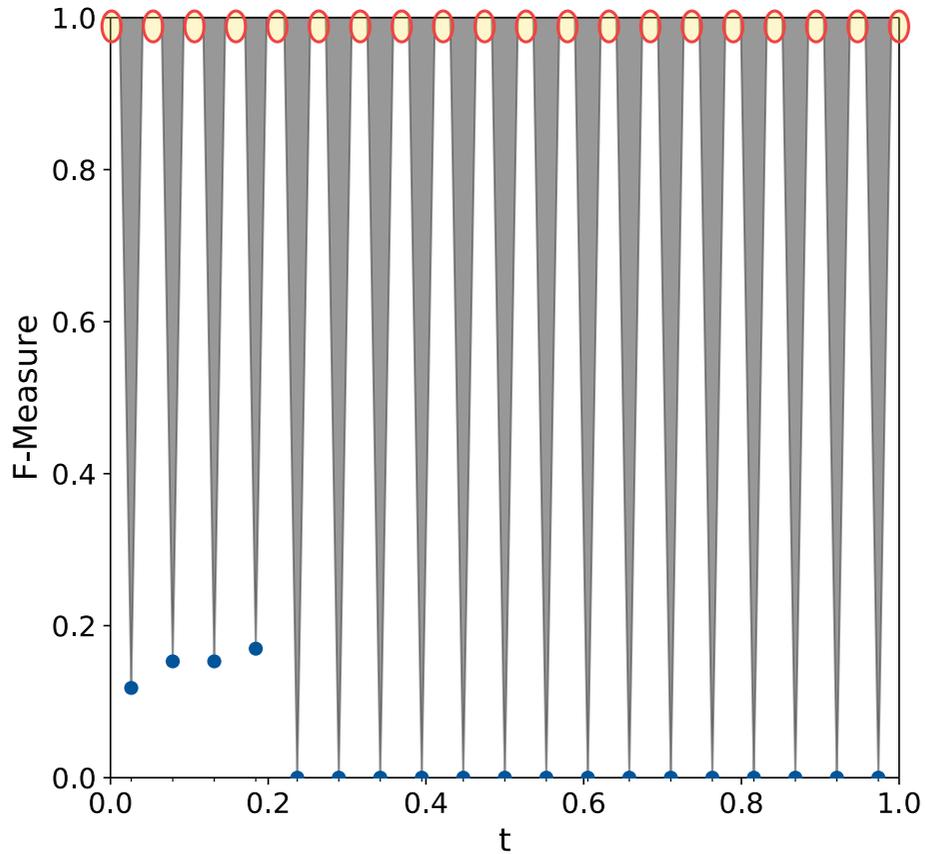
$$F(e') \leq F(e) + \Phi_{e\varepsilon_1} + \Phi_e \cdot (e_2 - e_1 - M_{min})(t' - t),$$

$$\text{where } M_{min} = \min_{\substack{e'' \in \mathcal{E}(\mathcal{H}) \\ \text{s.t. } F(e'') > F(e)}} (e_2'' - e_1'').$$



(re) Visualisation de la borne de Parambath et al. (grille de 19 modèles)

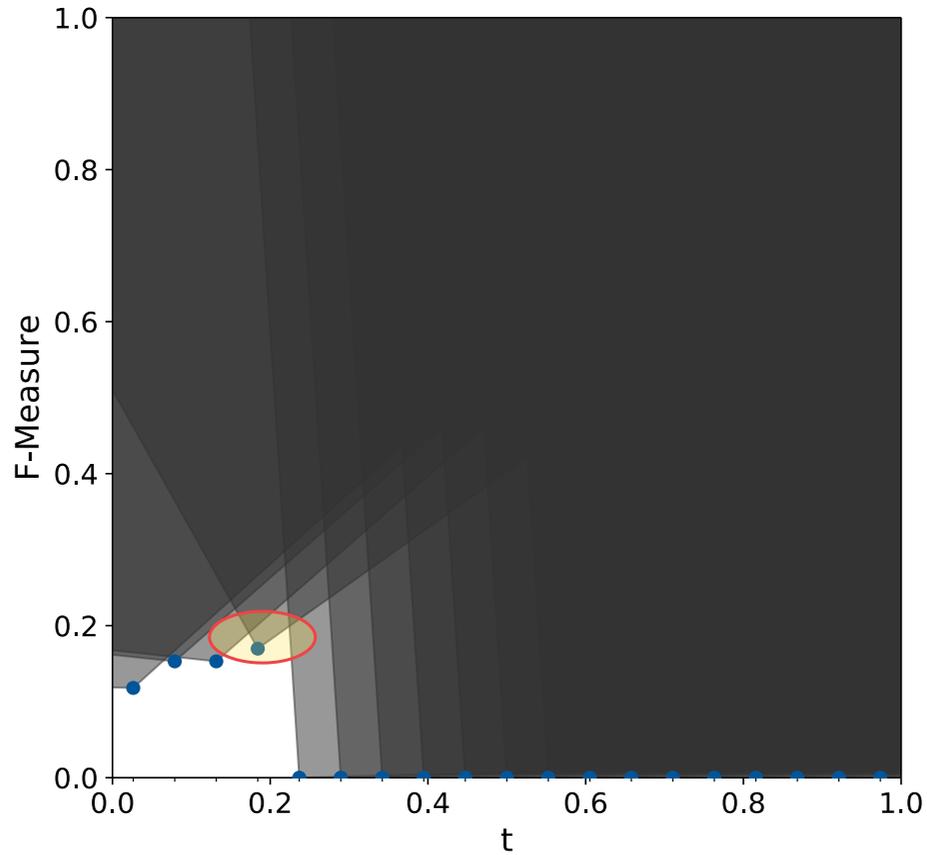
classes déséquilibrées



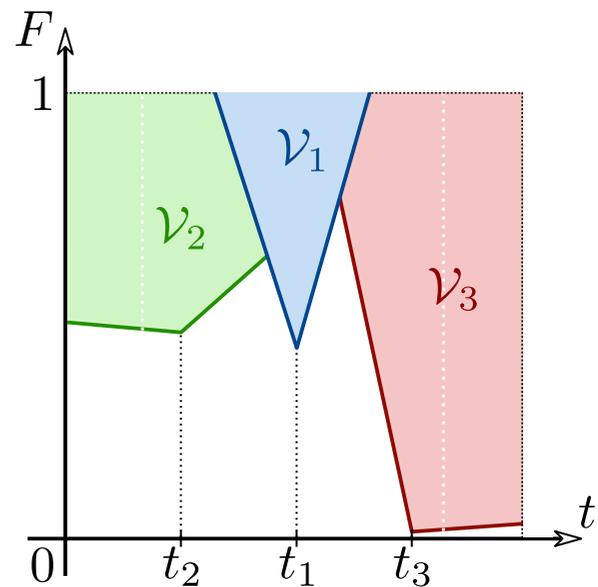
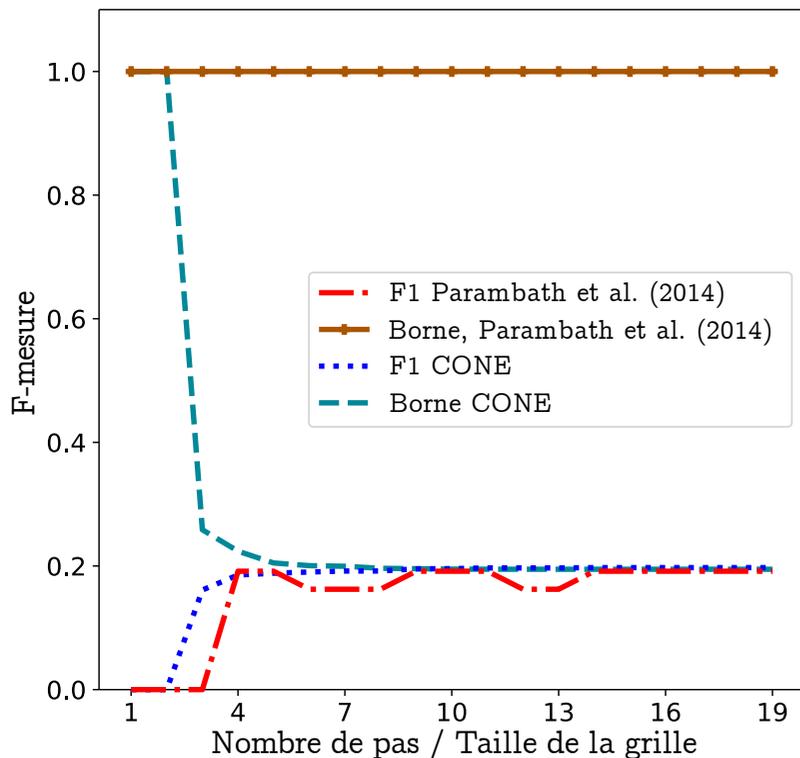
⇒ on pourrait penser que des valeurs de 1.0 sont toujours possibles...

Visualisation de notre borne (grille de 19 modèles)

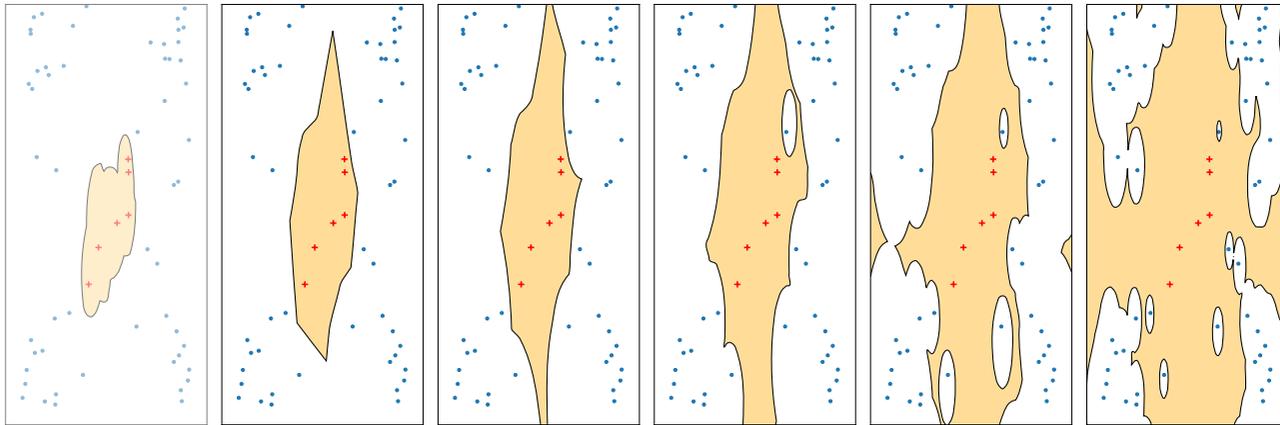
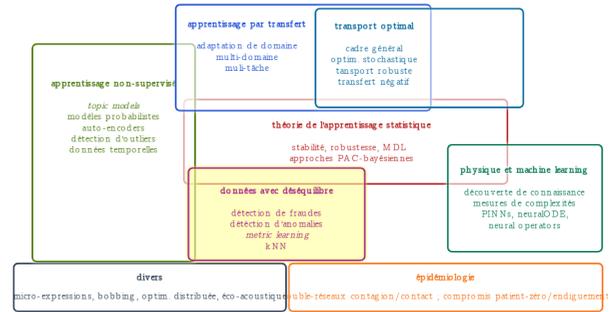
classes déséquilibrées



⇒ maximum théorique très proche de la meilleure valeur empirique !



⇒ Borne vraiment informative, algorithme incrémental, augmentation dès le début



γ kNN

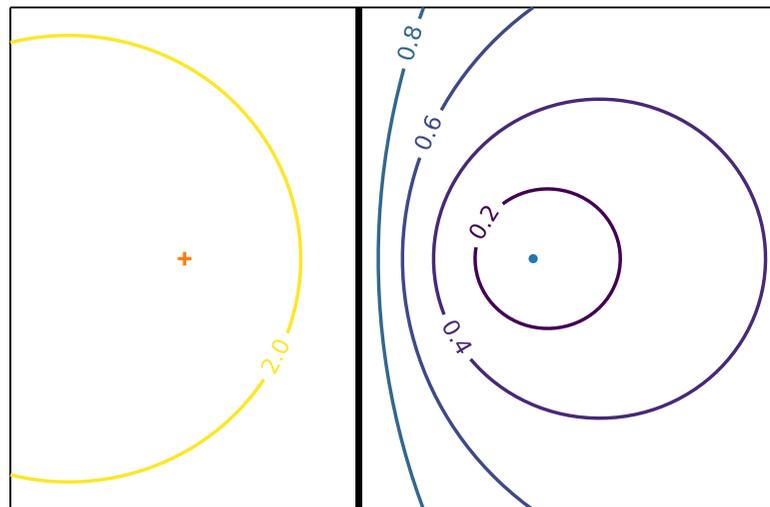
Objectif

- étendre la zone d'influence des positifs
- éviter la complexité des GANs
- reprendre le problème de la base

Idée : virtuellement rapprocher les positifs de la requête

Principe (avec une distance d et un paramètre γ)

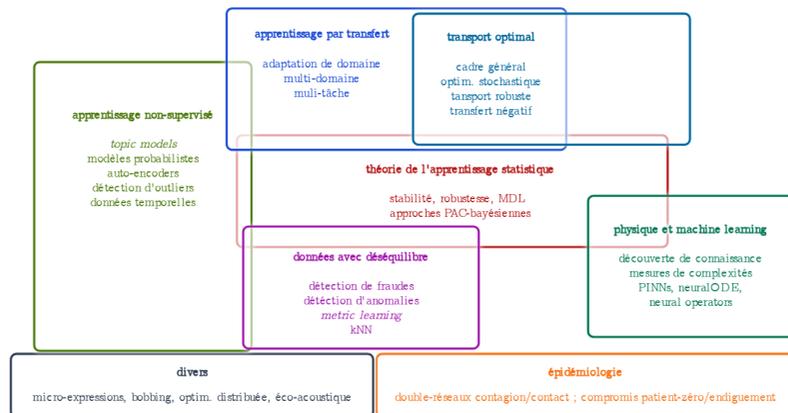
- distance à un négatif x_i : $d_\gamma(x, x_i) = d(x, x_i)$
- distance à un positif x_j : $d_\gamma(x, x_j) = \gamma \cdot d(x, x_j)$
- cross-valider le γ



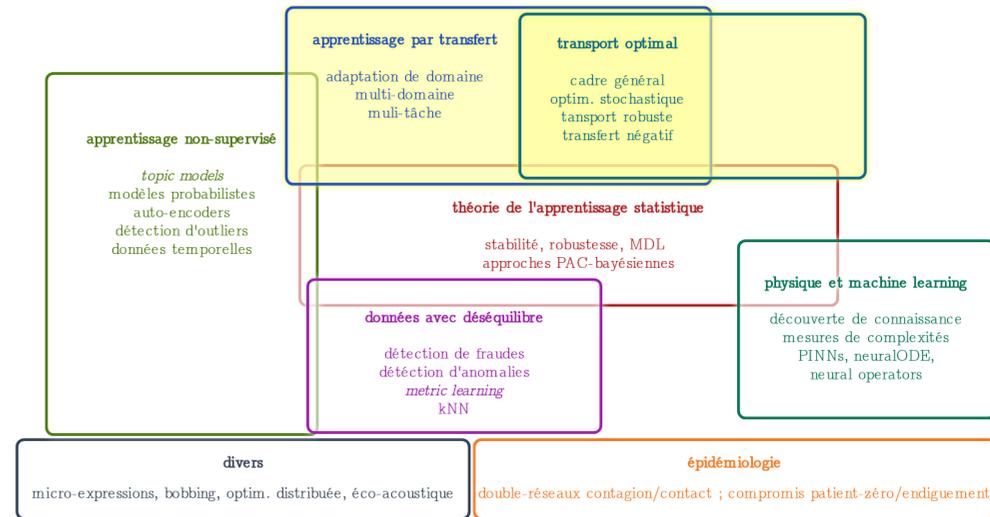
⇒ implémentation simple et efficace, garanties théoriques, extension métrique learning

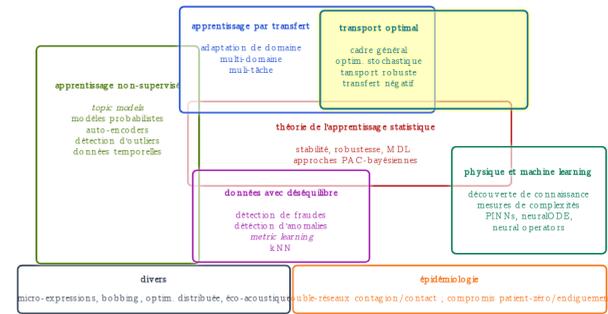
⇒ déployé à la DGFIP

1. Trajectoire
2. Focus 1 : approches pour la détection d'anomalies supervisée
3. Focus 2 : transfert hétérogène par *optimal tensor transport*
4. Projet de recherche



Focus 2 : transfert hétérogène par *optimal tensor transport*

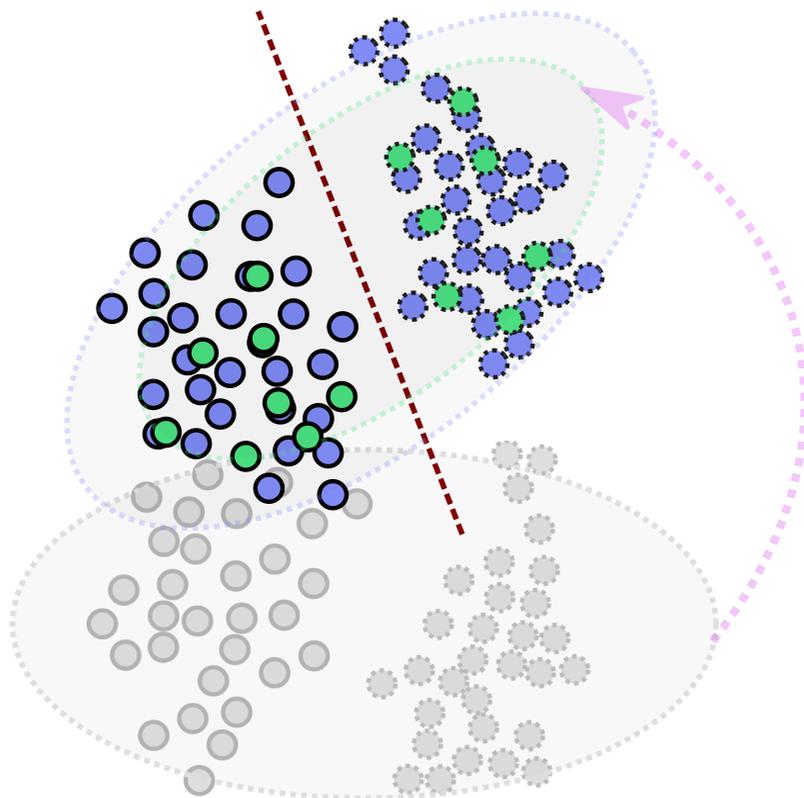


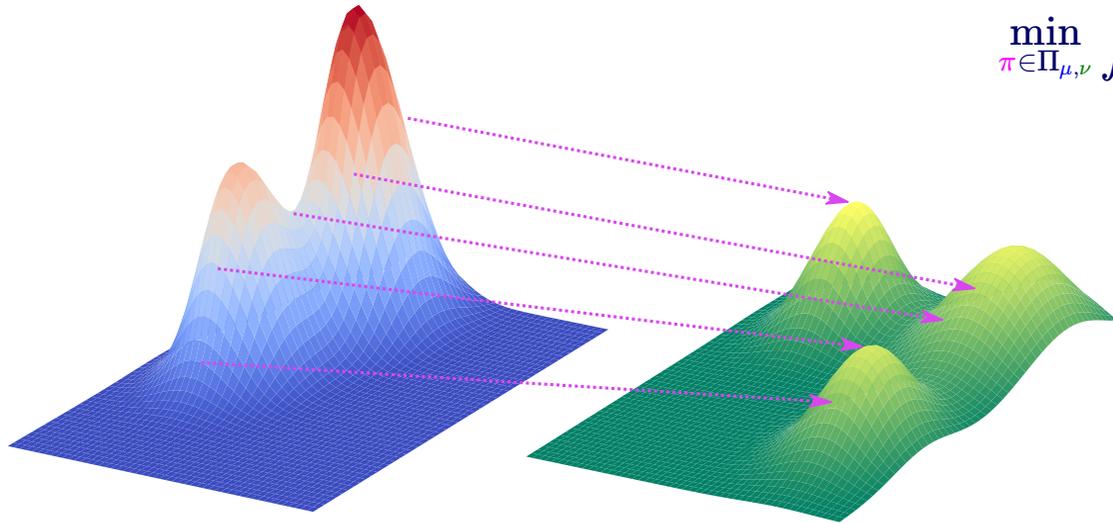
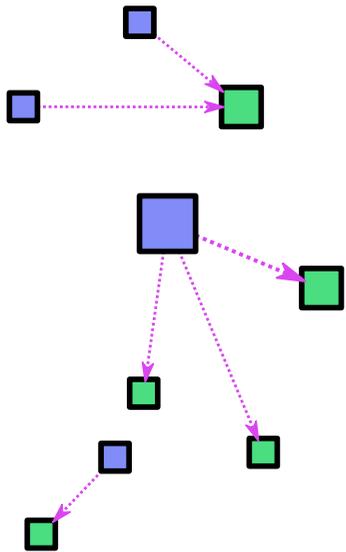


Apprentissage par transfert, adaptation de domaine, transport optimal

Domaine cible
(étiquettes ?)

Domaine source
(étiqueté)

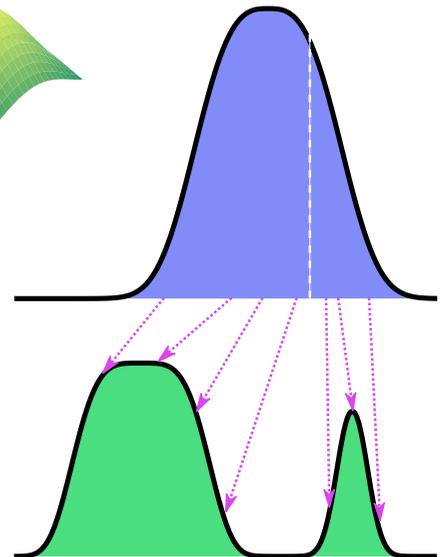


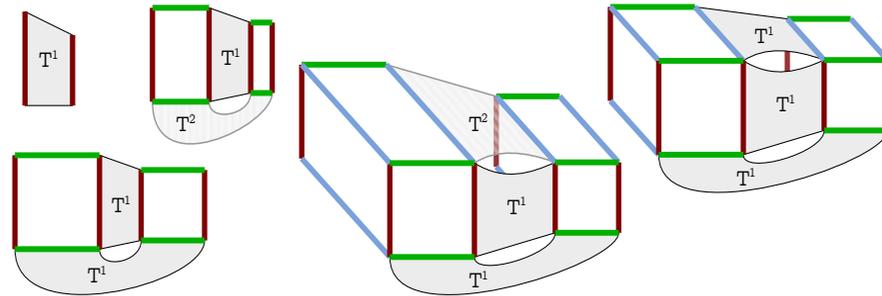
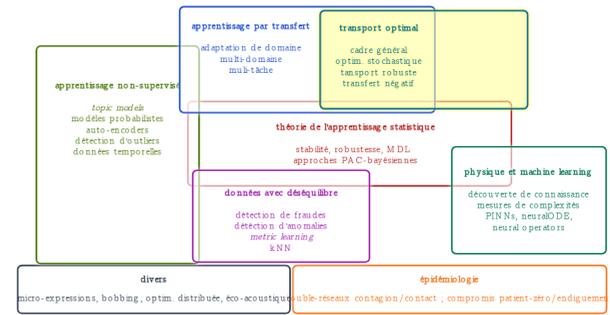


$$\min_{\pi \in \Pi_{\mu, \nu}} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$

$$\min_{T \in \Pi_{a, b}} \sum_{i, j} T_{ij} C_{ij}$$

$$\min_{T \in \Pi_{a, b}} \langle T, C \rangle$$





OTT : Optimal Tensor Transport

Transport entre domaines hétérogènes : Gromov Wasserstein (+ Fused)

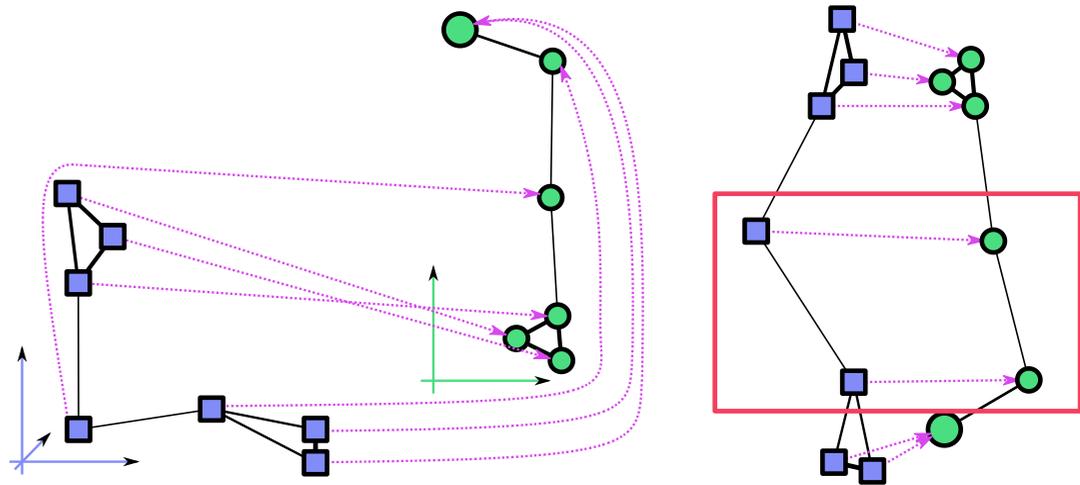
adaptation, transport

Rappel, problème de transport $OT(C) = \min_{T \in \Pi_{a,b}} \sum_{i,j} T_{ij} C_{ij} = \min_{T \in \Pi_{a,b}} \langle T, C \rangle$

⇒ **limité aux données dans un même espace** (pour définir C)

Problème de Gromov-Wasserstein $GW(C^X, C^Y) = \min_{T \in \Pi_{a,b}} \sum_{i,j} \sum_{i',j'} T_{ij} T_{i'j'} \mathcal{L}(C_{ii'}^X, C_{jj'}^Y)$

- apparie points sources et cibles
 - minimize la différence de distances
 - entre une paire source
 - et une paire cible
 - qui sont appariées
- ⇒ appariement de graphes



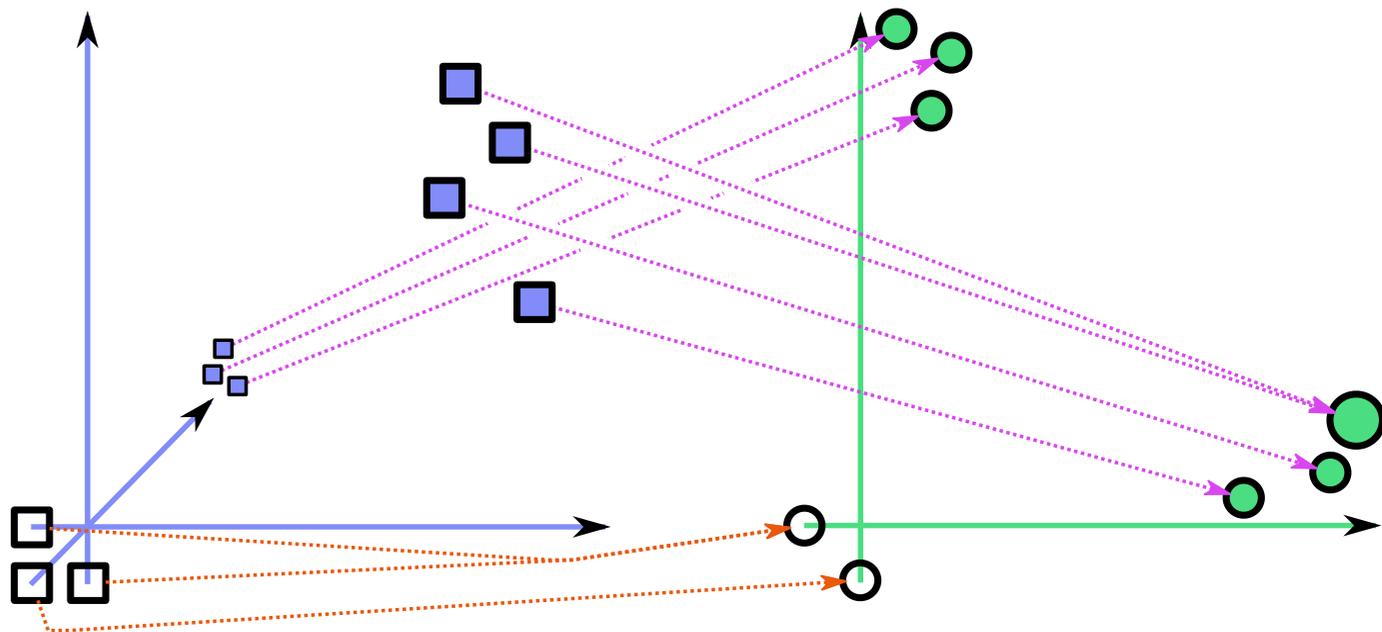
Fused Gromov Wassertein « $FGW(C, C^X, C^Y) = (1 - \beta) \cdot OT(C) + \beta \cdot GW(C^X, C^Y)$ »

Problème du Co-Optimal Transport

$$Co-OT(X, Y) = \min_{T^1 \in \Pi_{a^1, b^1}} \min_{T^2 \in \Pi_{a^2, b^2}} \sum_{i^1, j^1} \sum_{i^2, j^2} T^1_{i^1, j^1} T^2_{i^2, j^2} \mathcal{L}(X_{i^1, i^2}, Y_{j^1, j^2})$$

Apparie

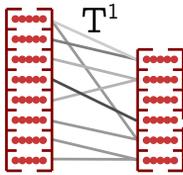
- points sources/cibles
- dimensions sources/cibles
- avec 2 plans de transport



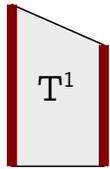
OTT* (*optimal tensor transport*)

adaptation, transport

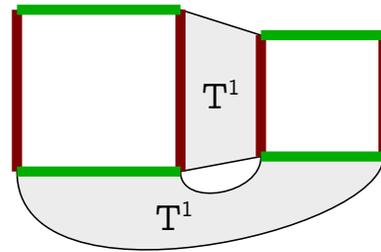
représentation avec features



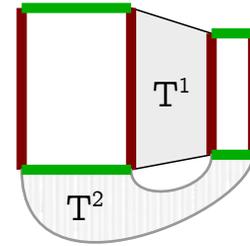
représentation compacte



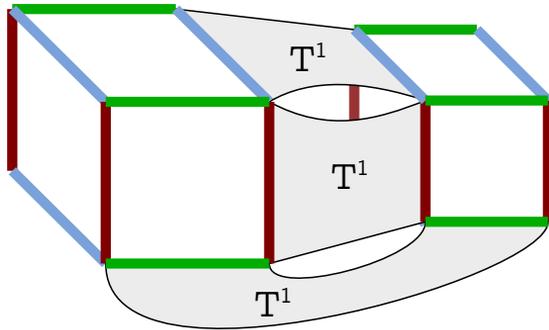
a) OTT₁ (OT) (Fs = Ft=5 features)



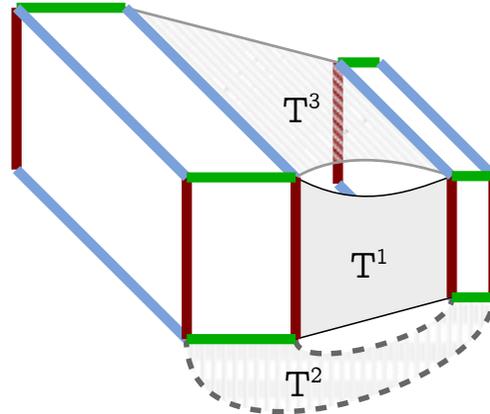
b) OTT₁₁ (GW)



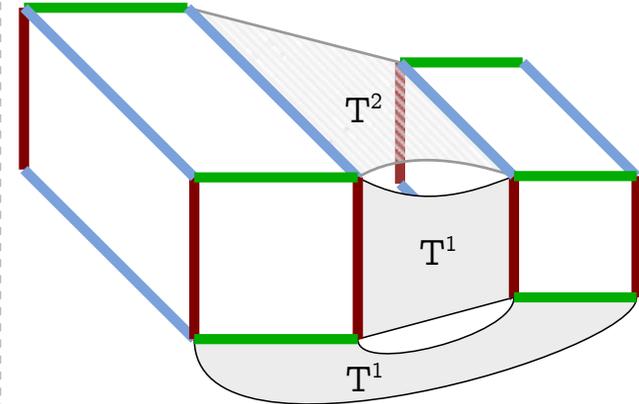
c) OTT₁₂ (Co-OT)



d) OTT₁₁₁ (triplets)



e) OTT₁₂₃ (triCo-OT)



f) OTT₁₁₂ (collections de GW)

Rappel, problème de Gromov-Wasserstein (ré-écrit)

$$GW(\mathbf{C}^X, \mathbf{C}^Y) = \min_{\mathbf{T} \in \Pi_{a,b}} \sum_i \sum_j \mathbf{T}_{ij} \sum_{i',j'} \mathbf{T}_{i'j'} \mathcal{L}(\mathbf{C}_{ii'}^X, \mathbf{C}_{jj'}^Y)$$

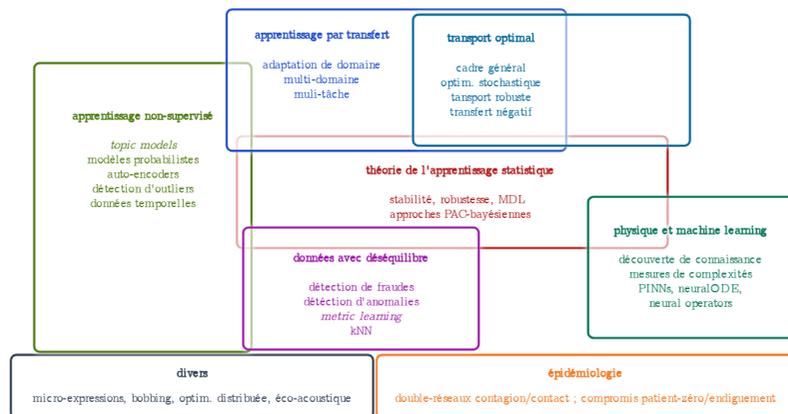
Algorithm SaGroW

Require: \mathbf{a}, \mathbf{b} (probability vectors of μ and ν), $\mathbf{C}^X, \mathbf{C}^Y$ (cost matrices), \mathcal{L} (loss function), M (number of samples), ϵ (entropy regularization), α (partial update weight)

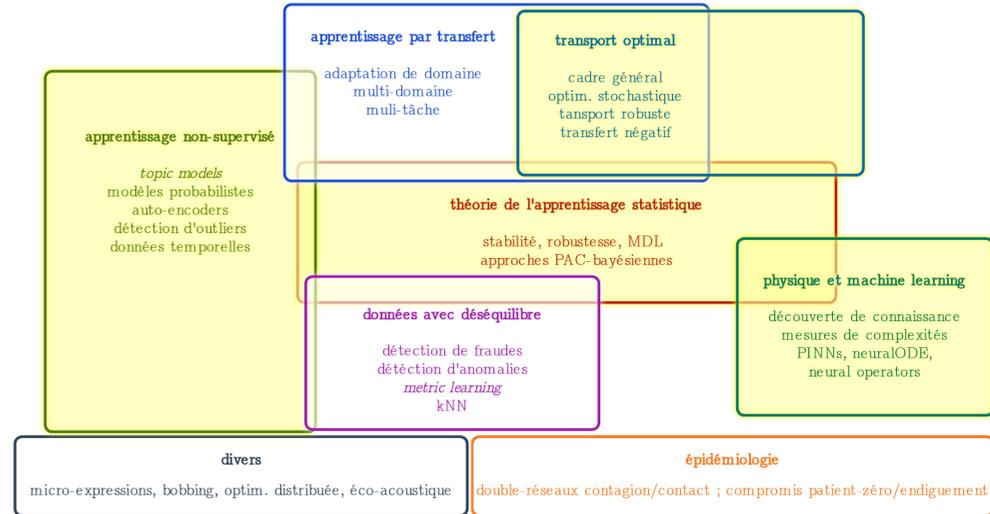
- 1: $\mathbf{T}_0 = \mathbf{a}\mathbf{b}^\top$
 - 2: **for** $s= 0$ to $S-1$ **do**
 - 3: $(j_m, l_m) \sim \text{Sample}(\mathbf{T}_s) \forall m \in \llbracket 1, M \rrbracket$
 - 4: $\hat{\Lambda}_{ik} = \frac{1}{M} \sum_{m=1}^M \mathcal{L}(\mathbf{C}_{i,j_m}^X, \mathbf{C}_{k,l_m}^Y) \forall i, k \in \llbracket 1, N \rrbracket$
 - 5: $\mathbf{T}'_s = \text{solve the regularized OT problem } (\mathbf{a}, \mathbf{b}, \hat{\Lambda}, \epsilon)$
 - 6: $\mathbf{T}_{s+1} = (1 - \alpha)\mathbf{T}_s + \alpha\mathbf{T}'_s$
 - 7: **end for**
 - 8: **return** \mathbf{T}_{S-1}
-

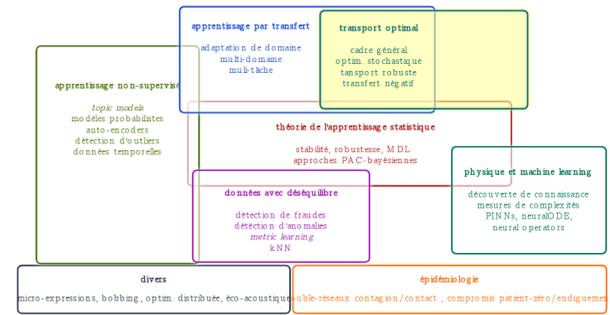
- termes vues comme des espérances mathématiques
- algorithme stochastique par échantillonnage
- complexité contrôlée et possibilité de transport 1d

1. Trajectoire
2. Focus 1 : approches pour la détection d'anomalies supervisée
3. Focus 2 : transfert hétérogène par *optimal tensor transport*
4. Projet de recherche



Projet de recherche





Autour du transport optimal

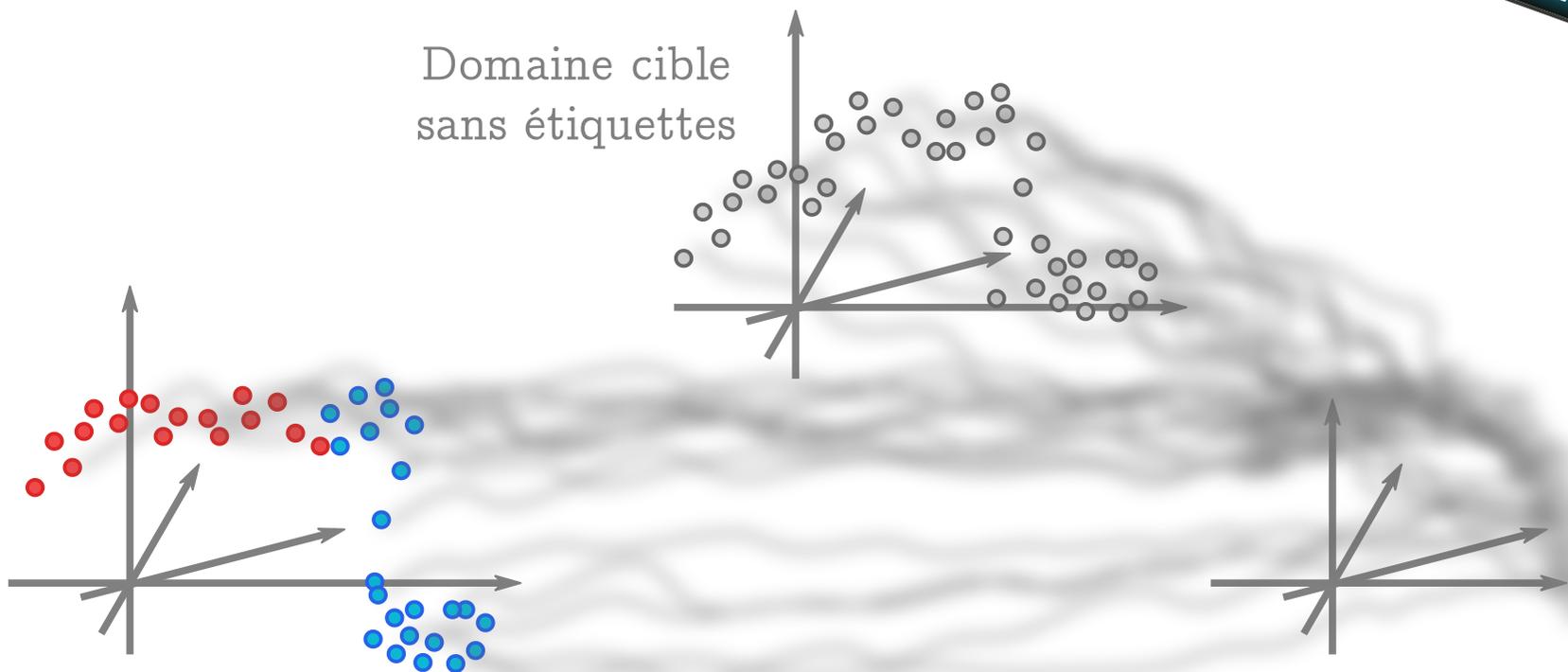
Cadre généralisé du transport optimal OTT++

- Approches pour l'optimisation
 - hiérarchiques
 - stochastiques avec agrégation/vote
- Garanties pour GW, OTT
 - généralisation
 - complexité d'échantillon
 - formulation probabiliste
 - bornes PAC-bayésiennes

Lien entre transport et diffusion

- dualité de Kantorovich-Rubinstein
- liens entre OT et diffusion (DDPM)

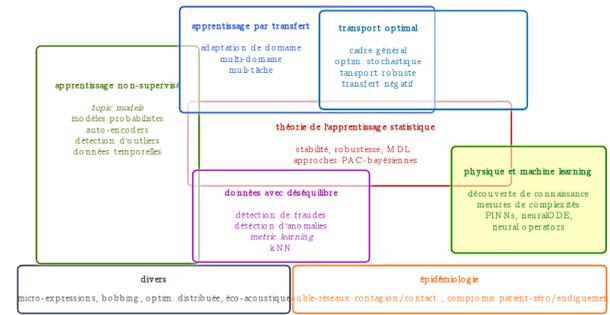




Domaine cible sans étiquettes

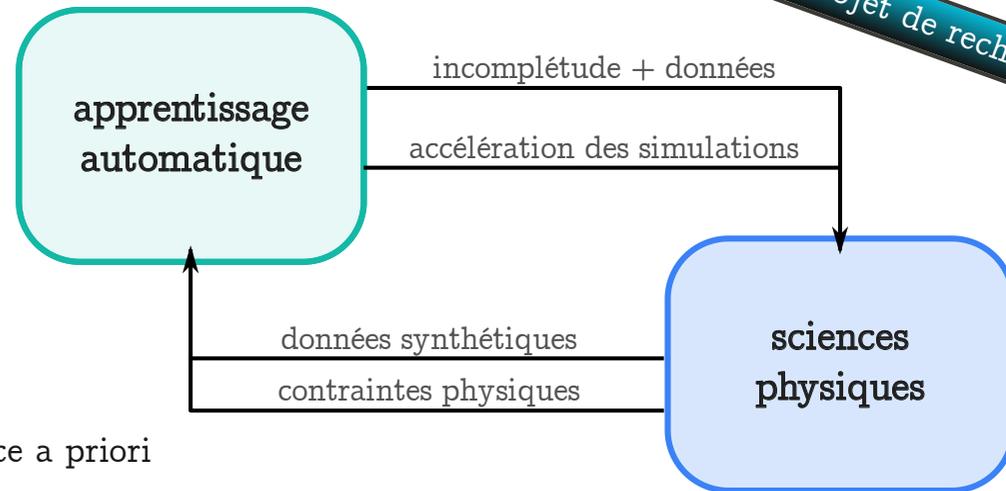
Domaine source avec étiquettes

Représentation pivot distribution de bruit



Physique et apprentissage automatique

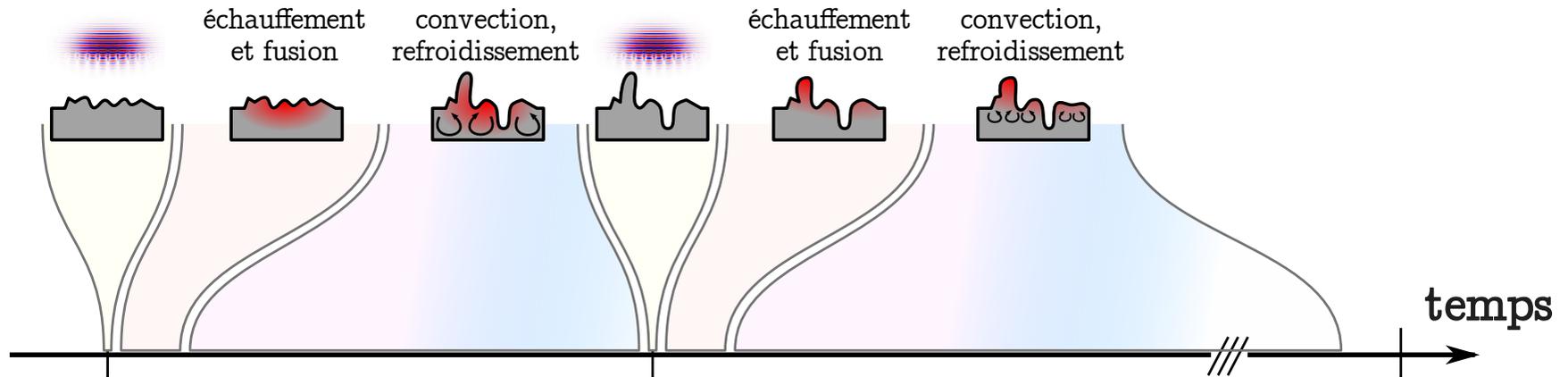
- Nouvelle équipe-projet Inria
- Projets et collaboration en cours
 - 2 thèses en collaborations
 - 1 postdoc
 - 1 ANR + dépôts
- Garanties des modèles
 - équivalence entre données et connaissance a priori
 - besoin de données réelles
 - besoin de données de simulation
 - gestion d'incertitude
- Nouvelles approches
- Découverte de connaissance



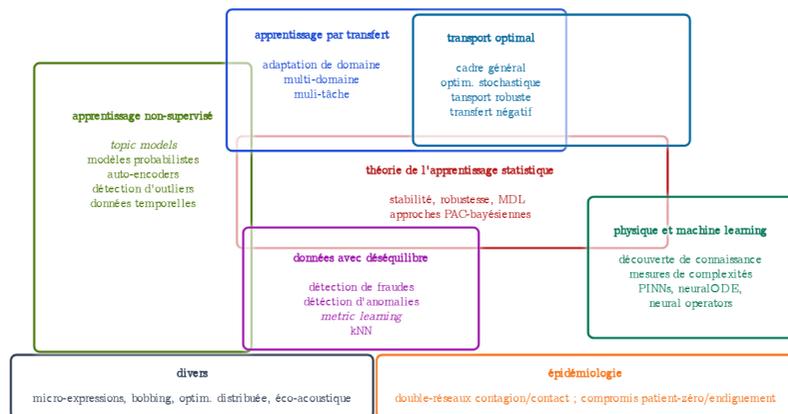
impulsion laser
femto-seconde
(Maxwell)

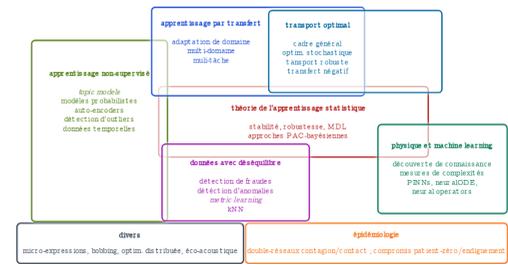
seconde impulsion
(Maxwell)

Observation
(MEB, AFM)

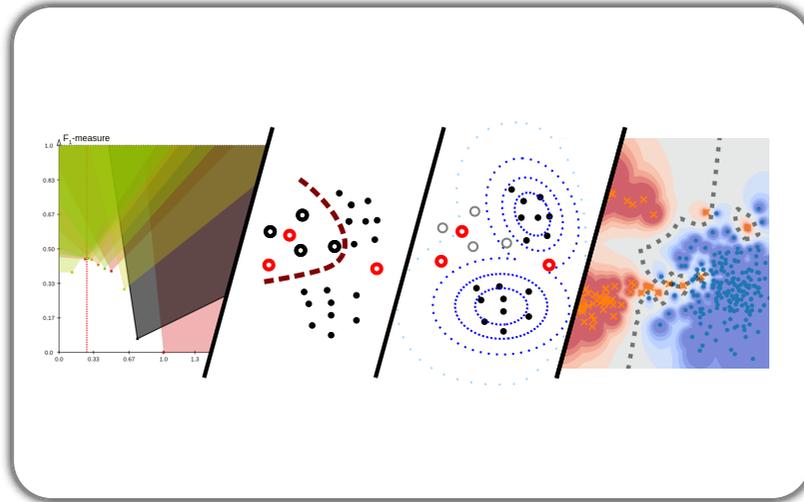


1. Trajectoire
2. Focus 1 : approches pour la détection d'anomalies supervisée
3. Focus 2 : transfert hétérogène par *optimal tensor transport*
4. Projet de recherche





Merci de votre attention ! Questions ?



Un grand merci à tous les collègues, collaborateurs et doctorants.

FIN

