

TITOUAN VAYER

7 rue Trarieux, 69003 Lyon, France

(+33)682325570 ♦ titouan.vayer@ens-lyon.fr ♦ <https://tvayer.github.io/>

PARCOURS UNIVERSITAIRE

Doctorat en informatique

2017 - 2020

Université Bretagne Sud, IRISA, Vannes

Titre: *Une contribution au Transport Optimal sur des espaces incomparables*

Supervisée par Nicolas Courty, Laetitia Chapel et Romain Tavenard

Défendue le 5 Novembre 2020 devant le jury composé de:

- *Rapporteurs*: Gabriel Peyré (DR CNRS DMA), Marc Sebban (PR Univ. Jean Monnet)
- *Examineurs*: Filippo Santambrogio (PR Univ. Claude Bernard), Julie Delon (PR Univ. Paris Descartes), Pierre Vanderghenst (PR EPFL), Rémi Flamary (MCF CMAP École Polytechnique)

Diplôme d'ingénieur

2013 - 2017

Ecole Centrale de Lille

Spécialités : Mathématiques, informatique et science des données.

Licence de Mathématiques

2013 - 2015

Université de Lille

Spécialités : Probabilités et statistiques.

EXPÉRIENCES PROFESSIONNELLES

Ecole Normale Supérieure, Lyon

nov 2020 - présent

Post-Doctorant.

Laboratoire de l'informatique et du Parallélisme (LIP, UMR 5668). Équipe DANTE

Thématiques de recherche: théorie de l'apprentissage, apprentissage compressif, problèmes inverses.

Université Bretagne Sud, Vannes

2017 - 2020

Doctorant.

Institut de recherche en informatique et systèmes aléatoires (IRISA, UMR 6074). Équipe OBELIX

Thématiques de recherche: théorie du transport optimal, graphes, séries temporelles, données hétérogènes.

Crédit Agricole Consumer Finance, Lille

2016 - 2017

Data Scientist dans le service modélisation

Création et industrialisation de modèles de machine learning pour le crédit.

Analyse de jeux de données massifs (web) avec Spark et Hadoop.

Régression logistique et forêt d'arbres décisionnels.

Michelin, Clermont-Ferrand

May 2015 - Dec 2015

Stagiaire

Analyse de données et statistiques.

Création d'un modèle de prévision des flux de pneus.

Réseau bayésien, analyse de séries temporelles.

Skins International Trading, Zürich, Suisse

Jan 2013 - Feb 2013

Stagiaire

Optimisation du stock de gestion de la société.

Programmation linéaire.

PUBLICATIONS

Voici mes différentes publications à jour. J'ai publié dans le domaine de l'apprentissage statistique où les auteurs sont habituellement classés par ordre de contribution et, sauf indication contraire, cet ordre est respecté dans les publications citées.

Revue internationale avec comité de lecture

1. Bonet, C., Vayer, T., Courty, N., Septier, F. & Drumetz, L. Subspace Detours Meet Gromov–Wasserstein. *Algorithms* **14**, 366. <http://dx.doi.org/10.3390/a14120366> (2021)
2. Flamary^{*}, R., Courty^{*}, N., Gramfort^{*}, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., Gautheron, L., Gayraud, N. T., Janati, H., Rakotomamonjy, A., Redko, I., Rolet, A., Schutz, A., Seguy, V., Sutherland, D. J., Tavenard, R., Tong, A. & Vayer, T. POT: Python Optimal Transport. *Journal of Machine Learning Research (JMLR)* **22**, 1–8. <http://jmlr.org/papers/v22/20-451.html> (2021)¹
3. Vayer, T., Chapel, L., Flamary, R., Tavenard, R. & Courty, N. Fused Gromov-Wasserstein distance for structured objects. *Algorithms* **13**, 212. <http://dx.doi.org/10.3390/a13090212> (2020)

Conférences internationales

4. Vincent-Cuaz, C., Flamary, R., Corneli, M., Vayer, T. & Courty, N. *Semi-relaxed Gromov-Wasserstein divergence and applications on graphs* in *International Conference on Learning Representations (ICLR)* (Online, 2022). <https://openreview.net/forum?id=RShaMexjc-x>
5. Marcotte, S., Barbe, A., Gribonval, R., Vayer, T., Sebban, M., Borgnat, P. & Gonçalves, P. *Fast Multiscale Diffusion on Graphs* in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Singapore, Singapore, 2022). <https://hal.archives-ouvertes.fr/hal-03212764>
6. Barbe, A., Gonçalves, P., Sebban, M., Borgnat, P., Gribonval, R. & Vayer, T. *Optimization of the Diffusion Time in Graph Diffused-Wasserstein Distances: Application to Domain Adaptation* in *IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)* (Online, 2021), 786–790
7. Vincent-Cuaz, C., Vayer, T., Flamary, R., Corneli, M. & Courty, N. *Online Graph Dictionary Learning* in *International Conference on Machine Learning (ICML)* **139** (PMLR, Online, 2021), 10564–10574. <https://proceedings.mlr.press/v139/vincent-cuaz21a.html>
8. Vayer^{*}, T., Redko^{*}, I., Flamary^{*}, R. & Courty^{*}, N. *CO-Optimal Transport* in *Neural Information Processing Systems (NeurIPS)* **33** (Online, Canada, 2020). <https://papers.nips.cc/paper/2020/hash/cc384c68ad503482fb24e6d1e3b512ae-Abstract.html>
9. Vayer, T., Flamary, R., Courty, N., Tavenard, R. & Chapel, L. *Sliced Gromov-Wasserstein* in *Neural Information Processing Systems (NeurIPS)* **32** (Vancouver, Canada, 2019). <https://papers.nips.cc/paper/9615-sliced-gromov-wasserstein>
10. Vayer, T., Courty, N., Tavenard, R., Chapel, L. & Flamary, R. *Optimal Transport for structured data with application on graphs* in *International Conference on Machine Learning (ICML)* **97** (Long Beach, USA, 2019), 6275–6284. <https://proceedings.mlr.press/v97/titouan19a.html>

¹ ^{*} = contributions égales. Le reste est trié par ordre alphabétique.

Conférences Nationales avec comité de lecture

11. Vayer, T., Chapel, L., Flamary, R., Tavenard, R. & Courty, N. *Transport Optimal pour les Signaux sur Graphes* in *Groupe d'Etudes du Traitement du Signal et des Images (GRETSI)* (Lille, France, 2019)

12. Vayer, T., Chapel, L., Flamary, R., Tavenard, R. & Courty, N. *Fused Gromov-Wasserstein distance* in *Conférence sur l'Apprentissage automatique (CAp)* (Rouen, France, 2018)

Rapports de recherche et articles soumis

13. Vayer, T. & Gribonval, R. *Controlling Wasserstein distances by Kernel norms with application to Compressive Statistical Learning* **Submitted** to Journal of Machine Learning Research (JMLR). 2021. <https://hal.archives-ouvertes.fr/hal-03461492>

14. Vayer, T., Chapel, L., Courty, N., Flamary, R., Soullard, Y. & Tavenard, R. *Time Series Alignment with Global Invariances* **Submitted** to IEEE Transactions on Signal Processing. 2020. <https://hal.archives-ouvertes.fr/hal-02473959>

ACTIVITÉS DE RECHERCHE

Je travaille sur les problématiques suivantes:

- Le transport optimal pour les données structurées.
- L'apprentissage à partir de données hétérogènes et l'adaptation de domaine.
- La compression de données pour l'apprentissage statistique.

Contexte de recherche et résumé succinct des activités

La science des données et l'apprentissage statistique sont de plus en plus confrontés à des défis à la fois méthodologiques et pratiques. D'une part les données disponibles sont en forte augmentation et d'autre part elles sont souvent de plus en plus diverses. Dans ce cadre complexe, mes travaux sont motivés par la création de méthodes d'apprentissage qui: **1) sont capables d'exploiter la diversité des données 2) permettent de passer à l'échelle**, c'est à dire qui sont économes en mémoire et temps de calcul. **Au cœur de mes travaux se trouvent les données structurées et hétérogènes.** Les premières modélisent des relations complexes entre des objets et peuvent être décrites par des graphes. Les secondes interviennent lorsqu'on veut construire des modèles à partir de données diverses ou lorsque les méthodes d'acquisition qui les génèrent varient au cours du temps ou que la représentation des données change. **Ces types de données sont omniprésents en apprentissage statistique et leur analyse requiert des méthodes statistiques avancées.** Pour faire face à ces défis, j'ai travaillé sur des **méthodes d'apprentissage basées sur le transport optimal et sur l'apprentissage à partir d'un sketch linéaire d'une base de données**, qui est une **représentation compressée efficace** basée sur des **plongements aléatoires** et calculable en une passe sur les données (appropriée pour le calcul en ligne ou distribué).

Projets de recherche

Je collabore sur plusieurs projets de recherche qui sont présentés ci-dessous:

• MASSILIA

début 2022-présent

Matrices spectral structures in graph learning and its applications

ANR Jeune Chercheur. Porteur du projet: Arnaud Breloy.

Expert externe pour le Working Package : *Graph signal processing with optimal transport on spectral features.*

Ce projet a pour but d'aborder les problèmes actuels liés à l'apprentissage sur graphes et à ses applications d'une manière unifiée centrée sur la décomposition spectrale de la matrice laplacienne des graphes.

• ACADEMICS

2020-présent

Machine learning & data science for complex and dynamical models.

Projet ANR. Ce projet a pour but de développer de nouveaux algorithmes d'apprentissage dans deux domaines scientifiques impliquant des données complexes : la modélisation du climat et la compréhension quantitative des systèmes sociaux.

• OATMIL

2017-2020

Optimal transport for machine learning.

Projet ANR. Ce projet a pour but de proposer de nouveaux concepts et de nouvelles méthodologies pour l'exploitation de grandes collections de données en croisant les outils et les idées issus du transport optimal et du machine learning.

Développements technologiques

J'ai contribué à plusieurs logiciels open-source. J'ai choisi d'évaluer mes contributions dans ces logiciels selon le modèle "Self-assessment software criteria" de l'INRIA.²

- **Python Optimal Transport (POT)**, Family = research; audience = universe; evolution = lts; duration >= 1; contribution = devel, softcont; url = <https://github.com/PythonOT/POT>. Librairie Python pour le transport optimal. Contributeur au code (méthodes FGW/GW et méthodes reliées) et co-auteur d'un papier JMLR décrivant la librairie [2]. Plus de 460k téléchargements (Pypi+conda). Coût de développement estimé: 555,862 \$. 152 citations, 43 contributeurs (je suis 7ème contributeur en terme de commits <https://github.com/PythonOT/POT/graphs/contributors>), intégration continue Travis.
- **FGW**, Family = vehicle; audience = partners; evolution = nofuture; duration >= 0.5; contribution = leader ; url = <https://github.com/tvayer/FGW>. Code et données de l'article [10]. 60 étoiles sur Github.
- **SGW**, Family = vehicle; audience = partners; evolution = nofuture; duration >= 0.5; contribution = leader ; url = <https://github.com/tvayer/SGW>. Code et données de l'article [9]. 42 étoiles sur Github.
- **PSCN**, Family = vehicle; audience = partners; evolution = nofuture; duration >= 0.5; contribution = leader ; url = <https://github.com/tvayer/PSCN>. Librairie Python pour l'article *Learning Convolutional Neural Networks for Graphs* par Mathias Niepert, Mohamed Ahmed and Konstantin Kutzkov. ICML. 2016. Je ne suis pas l'auteur du papier mais j'ai réimplémenté leur méthode en Python et l'ai mise à disposition. 39 étoiles sur Github.

Responsabilités collectives

- Rapporteur pour les conférences internationales: ICLR (2021), NeurIPS (2021/2020/2019, **Outstanding Reviewer Award** en 2020 et 2021), ICML (2021/2020), ECML-PKDD (2020), de la conférence nationale CAP en 2020 et 2022.
- Rapporteur pour les journaux *Information and Inference* (IMAI) et *European Journal of Applied Mathematics* (EJAM).
- Organisateur des séminaires *Machine Learning and Signal Processing (MLSP)* <http://www.ens-lyon.fr/PHYSIQUE/seminars/machine-learning-and-signal-processing> ($\approx 2 \times$ par mois, 26 séminaires depuis Janvier 2021). Ces séminaires pluridisciplinaires sont organisés en association avec des équipes de recherche du LIP, LMBC et UMPA, et sont hébergés au sein de l'ENS Lyon. Ils sont généralement présentés par des intervenants extérieurs et explorent des thématiques allant de l'apprentissage automatique au traitement du signal. Je m'occupe d'animer les séminaires, de coordonner les différentes équipes pour l'organisation du planning et du choix des intervenants (conjointement avec ces équipes).

²<https://www.inria.fr/sites/default/files/2021-01/Criteria%20software%20self%20assessment.pdf>