

# Transport Optimal sous Contrainte de Régularité pour l'Adaptation de Domaines entre Graphes avec Attributs

Dominique BARBE<sup>1,2,3</sup>, Pierre BORGNAT<sup>1</sup>, Paulo GONÇALVES<sup>2</sup>, Marc SEBBAN<sup>3</sup>

<sup>1</sup>Univ Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, LIP UMR 5668, F-69342, Lyon, France

<sup>2</sup>Univ Lyon, ENS de Lyon, UCB Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France

<sup>3</sup>Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France

dominique.barbe@ens-lyon.fr, pierre.borgnat@ens-lyon.fr  
paulo.goncalves@ens-lyon.fr, marc.sebban@univ-st-etienne.fr

**Résumé** – Cet article porte sur le problème d'adaptation de domaines par transport optimal entre deux graphes. La tâche visée est le transfert de connaissance d'un graphe source étiqueté pour aider la classification de nœuds d'un graphe cible non étiqueté. On s'intéresse à des scénarios où se combinent une structure de graphe et des attributs associés à chaque nœud. L'approche proposée vise à optimiser un mapping entre les deux graphes sous contraintes (i) de préservation des structures transportées et (ii) d'homogénéité des étiquettes transférées sur un même nœud.

**Abstract** – This paper addresses the problem of domain adaptation between two graphs by optimal transport. We aim at benefiting from the knowledge of a labeled source graph to improve the classification of nodes in an unlabeled target graph. We focus on the setting where a set of features is associated to each node of the graphs. The method presented in this paper optimizes a transportation plan from the source to the target that (i) preserves the structures transported between the graphs and (ii) prevents the mapping from transporting two source nodes with different labels to the same destination.

## 1 Introduction

De nombreux contextes applicatifs présentent des jeux de données structurés sous forme de graphe où les nœuds sont associés à un ensemble d'attributs. C'est le cas par exemple des pages Wikipédia, certaines structures moléculaires, mais également les réseaux de co-citations entre chercheurs, ou plus généralement les réseaux sociaux.

Une question classique est de savoir retrouver des sous-structures dans ces graphes, notamment dans le contexte spécifique où les nœuds sont étiquetés et disposent d'attributs liés aux classes [5]. Nous nous intéressons dans cet article au problème de transfert de connaissance d'un graphe source étiqueté à un graphe cible potentiellement issu d'une distribution différente. Pour ce faire, nous proposons un nouvel algorithme d'adaptation de domaine dédié au transfert entre graphes et qui exploite la théorie du transport optimal [3]. La spécificité de nos travaux est que nous contraignons la fonction de transport à prendre en compte la structure des graphes concernés. L'intérêt de cette fonction est de retrouver automatiquement dans le graphe cible des communautés qui auraient été identifiées (de façon automatique ou manuelle) dans un graphe source. Prenons l'exemple de Wikipédia qui évolue avec le temps. Des pages sont modifiées, ajoutées ou retirées et des

liens sont régulièrement créés ou disparaissent. Malgré ces changements, la structure générale du graphe évolue peu. Ainsi, on pourrait vouloir exploiter une classification du graphe opérée à un instant  $t$  pour identifier les étiquettes des nœuds du graphe ayant évolué à l'instant  $t + 1$ .

Le cadre de nos travaux est proche de [2] où le transport optimal est exploité pour réaliser une tâche d'adaptation de domaines, mais dans un contexte limité à des données non structurées. Nos travaux s'apparentent plus fortement à [7] où le transport optimal est utilisé dans le cas où les données sont structurées, et où une extension de la distance de Gromov-Wasserstein est introduite. Nous offrons ici une piste alternative, en proposant un nouveau terme de régularisation au problème de transport optimal qui repose sur le laplacien des graphes décrivant la structure des données. L'intérêt de ce terme provient ainsi de son interprétation en termes de traitement du signal sur graphe.

La suite de cet article est organisée comme suit: dans la section 2, on rappelle le cadre du transport optimal classique et de la régularisation sur les étiquettes transportées proposée dans [2] permettant de réaliser des tâches d'adaptation de domaine entre données non structurées; on introduit ensuite notre terme de régularisation pour exploiter de l'information structurelle sur les graphes. Dans la section 3, on présente tout d'abord le modèle des données

synthétiques utilisées pour tester expérimentalement notre contribution. Les résultats de nos expériences sont ensuite décrits. Nous concluons en section 4.

## 2 Adaptation de domaines par transport optimal

### 2.1 Transport optimal classique

Soient deux ensembles de données  $X^s \in \mathbb{R}^{N^s \times d}$  et  $X^t \in \mathbb{R}^{N^t \times d}$ , dits *source* et *cible*. Soit  $C \in \mathbb{R}^{N^s \times N^t}$  une *matrice de coûts*, contenant les coûts de transport des  $N^s$  données source vers les  $N^t$  données cible :

$$C = (\|X_i^s - X_j^t\|)_{\substack{1 \leq i \leq N^s \\ 1 \leq j \leq N^t}}. \quad (1)$$

Le transport optimal s'intéresse à la recherche du meilleur *couplage*  $\gamma \in \Pi$  (ou *plan de transport*) entre les deux distributions discrètes source et cible, où  $\Pi$  est l'ensemble des distributions conjointes respectant les distributions marginales source et cible (typiquement uniforme en l'absence d'a priori). Formellement, le couplage optimal est obtenu en résolvant le problème d'optimisation suivant, appelé problème de Kantorovich régularisé :

$$\operatorname{argmin}_{\gamma \in \Pi} \{ \langle \gamma, C \rangle_F + \lambda_e \cdot \Omega_e(\gamma) \}, \quad (2)$$

où  $\Omega_e(\gamma) = \sum_{\substack{1 \leq i \leq N^s \\ 1 \leq j \leq N^t}} (\gamma_{i,j} \cdot \ln(\gamma_{i,j}))$  est l'*entropie négative* de  $\gamma$ ,  $\lambda_e$  est un hyperparamètre de régularisation à régler, et  $\langle A, B \rangle_F = \sum_{i,j} A_{i,j} \cdot B_{i,j}$  est la norme de Frobenius.

Il s'agit d'un problème d'optimisation strictement convexe grâce au terme de régularisation  $\Omega_e$ . Pour le résoudre, on peut faire appel à l'algorithme itératif Sinkhorn [3] équivalent à résoudre le dual du problème (2) avec un algorithme *block coordinate ascent*. C'est cet algorithme que nous utilisons par la suite avec la particularité que les données manipulées sont désormais des graphes.

### 2.2 Régularisation sur les étiquettes transférées du graphe source

On suppose désormais que les données sources  $X^s$  et cibles  $X^t$  possèdent une structure de graphe. On les note  $G^s$  et  $G^t$  respectivement. Ces graphes peuvent ainsi être définis par leurs matrices d'adjacence respectives  $W^s \in \mathbb{R}^{N^s \times N^s}$  et  $W^t \in \mathbb{R}^{N^t \times N^t}$ . On parlera alors indifféremment des données (sources ou cibles) et des nœuds des graphes.

On suppose également que l'on a accès à une  $K$ -partition des nœuds du graphe source, notée  $B_1^s, \dots, B_K^s$  où chaque donnée est associée à une *étiquette*. On peut imaginer exploiter cette information supervisée pour résoudre le problème de transport optimal. Si on suppose qu'il existe une fonction de transport préservant l'homogénéité des étiquettes transférées, il peut être raisonnable de ne pas vouloir transporter de masse de deux nœuds source d'étiquettes

différentes vers le même nœud cible. Suivant l'idée proposée dans [2], nous utilisons ici le terme de régularisation group-lasso défini comme suit :

$$\Omega_c(\gamma) = \sum_{1 \leq j \leq N^t} \sum_{1 \leq c_l \leq K} \|\gamma_{B_{c_l}^s, j}\|_2. \quad (3)$$

Cette régularisation peut-être vue comme une contrainte supplémentaire au problème de transport optimal de base, eq. (2). A noter que le prix à payer pour cette régularisation discriminante est que le nouveau problème d'optimisation, défini ci-dessous, devient non convexe en  $\gamma$  :

$$\operatorname{argmin}_{\gamma \in \Pi} \{ \langle \gamma, C \rangle_F + \lambda_e \cdot \Omega_e(\gamma) + \lambda_c \cdot \Omega_c(\gamma) \}. \quad (4)$$

Dans [2], les auteurs proposent d'utiliser la méthode du gradient conditionnel généralisé pour résoudre ce problème.

Nous pourrions être tentés d'appliquer directement la solution du problème (4) pour transporter le graphe source  $G^s$  vers le graphe cible  $G^t$ . Cependant, un tel transport ferait totalement abstraction de la structure des graphes. Ainsi, une sous-structure de  $G^s$  (e.g. une communauté d'un réseau social) n'aurait aucune garantie d'être préservée après transport sur  $G^t$ . Si cette sous-structure est supposée corrélée à la densité locale des classes, il est donc nécessaire de contraindre le plan de transport à la préserver.

### 2.3 Régularisation sur la structure des graphes

Nous cherchons à ce que le plan de transport calculé non seulement minimise le coût de transfert des attributs des nœuds source mais également tienne compte des structures des graphes  $G^s$  et  $G^t$ .

Pour ce faire, nous proposons de rajouter au problème (4) le terme de régularisation suivant :

$$\Omega(\gamma) = \langle \gamma, C' \rangle_F, \quad \text{avec } C' = \left( \left\| (L^s \cdot X^s)_i - (L^t \cdot X^t)_j \right\|_p \right)_{i,j}, \quad (5)$$

où  $L^s$  et  $L^t$  sont les laplaciens des graphes source et cible, et  $p \geq 1$  est une constante à ajuster. Il est important de noter que  $C'$  définit une seconde matrice de coût. De ce fait, on peut la prendre en compte au même titre que la matrice  $C$  du problème de transport initial et réécrire le problème (4) comme suit :

$$\operatorname{argmin}_{\gamma \in \Pi} \{ \langle \gamma, C + \lambda \cdot C' \rangle_F + \lambda_e \cdot \Omega_e(\gamma) + \lambda_c \cdot \Omega_c(\gamma) \}, \quad (6)$$

où  $\lambda$  est le paramètre de régularisation associé à (5). Ce problème se résout alors exactement comme précédemment.

On peut noter que  $\Omega(\gamma)$  peut se réécrire ainsi :

$$\Omega(\gamma) = \sum_{i,j} \left[ \gamma_{i,j} \cdot \left\| \sum_{i' \sim i} [W_{i,i'}^s (X_i^s - X_{i'}^s)] - \sum_{j' \sim j} [W_{j,j'}^t (X_j^t - X_{j'}^t)] \right\|_p \right]. \quad (7)$$

Nous pouvons interpréter ce nouveau terme de régularisation de deux points de vue, selon que l'on utilise les expressions (5) ou (7) : la première formulation montre que minimiser ce terme revient à essayer de conserver la façon dont les vecteurs d'attributs se diffusent autour d'un nœud, le laplacien étant l'opérateur de diffusion. La seconde formulation montre que minimiser ce terme force à ce que, pour deux nœuds source et cible associés ( $\gamma_{i,j}$  grand), les vecteurs d'attributs dans leurs voisinages respectifs varient de manière similaire.

## 3 Étude expérimentale

### 3.1 Génération de données synthétiques

Afin d'évaluer l'efficacité de la méthode proposée et l'effet du nouveau terme de régularisation exploitant les structures des graphes, nous avons effectué des expérimentations sur des jeux de données synthétiques. Pour générer les graphes source, on utilise un modèle stochastique par blocs contextuel [5], qui permet de générer simultanément une structure de graphe et des vecteurs d'attributs.

Pour la structure de graphe, ce modèle est paramétré par sa matrice de connectivité, qui donne la probabilité que deux nœuds soient adjacents en fonction de leur appartenance à un bloc spécifique. Dans nos expériences, nous utilisons des graphes composés de deux blocs, avec une matrice de connectivité de la forme :

$$\begin{pmatrix} c_{00}/N & c_{01}/N \\ c_{10}/N & c_{11}/N \end{pmatrix}, \quad (8)$$

où  $N$  est le nombre de nœuds du graphe. Les entrées de la matrice sont de la forme  $c/N$  pour que le degré moyen d'un nœud soit indépendant de la taille des graphes.

Dans nos expériences, nous avons choisi un cadre simplifié où les termes hors de la diagonale sont tous fixés à la même valeur  $c_0/N$ , et les termes diagonaux tous fixés à  $c_1/N$ . Dans les expériences on choisit dans un premier temps  $c_0$  et  $c_1$  de telle sorte que les blocs soient identifiables (voir [1] pour les conditions exactes). Finalement, on s'assure d'avoir systématiquement un graphe connexe.

Pour les vecteurs d'attributs, ce modèle est paramétré par  $v \in \{-1,1\}^N$  le vecteur encodant la partition en blocs de nœuds,  $d$  la dimensionnalité des données,  $u \sim \mathcal{N}(0, I_d/d)$  un vecteur latent relié aux centres des blocs et  $\mu$ . Les vecteurs d'attributs  $z_i$  sont tirés ainsi :

$$z_i = \sqrt{\frac{\mu}{N}} \cdot v_i \cdot u + \frac{Z_i}{d}, \quad (9)$$

où les  $Z_i \in \mathbb{R}^d$  suivent une loi normale centrée réduite.

Pour la génération du graphe cible, on modifie plus ou moins le graphe source, en répétant un certain nombre de fois ces deux opérations :

- *délétion de nœud* : un nœud tiré au hasard uniformément est supprimé.
- *insertion de nœud* : un nœud est ajouté ; son bloc est choisi uniformément au hasard, ses arrêtes et son vecteur d'attributs sont tirés selon les paramètres du graphe source.

### 3.2 Plan d'expériences

On souhaite comparer ici cinq méthodes : (i) le transport optimal standard (problème (2)), (ii) l'utilisation seule du terme de régularisation sur les étiquettes (problème (4)), (iii) l'exploitation seule de notre contrainte sur les structures (problème (6) avec  $\lambda_c = 0$ ), (iv) l'ensemble des régularisations (problème (6)) et (v) la méthode de [7] qui se rapproche de nos travaux (qu'on appellera *Fused Gromov-Wasserstein*).

Pour les graphes sources,  $N^s$  prend successivement les valeurs  $\{200, 400, 600, 800\}$ . Pour les graphes cibles, on modifie le graphe source avec  $N^s$  opérations choisies uniformément (insertion ou délétion). Pour la matrice de connectivité, on choisit  $c_0 = 2$  (hors diagonale) et  $c_1 = 128$  (diagonale). Les vecteurs d'attributs sont générés en dimension 2. Les distributions marginales source et cible sont uniformes :  $\Pi = \left\{ \gamma \in \mathbb{R}^{N^s \times N^t} \mid N^s \cdot \gamma \cdot \mathbb{1} = \mathbb{1} \text{ et } N^t \cdot \mathbb{1}^T \cdot \gamma = \mathbb{1}^T \right\}$ . Les méthodes sont évaluées par un score, qui est le total de masse correctement transportée. Plus précisément, étant données les partitions  $v^s$  et  $v^t$  qui correspondent respectivement aux blocs des graphes sources et cibles, le score d'un plan de transport est  $s(\gamma) \in [0,1]$  :

$$s(\gamma) = \sum_{i=1}^{N^s} \sum_{j=1}^{N^t} \left[ \gamma_{i,j} \cdot \mathbb{1}_{v_i^s = v_j^t} \right]. \quad (10)$$

Pour chaque valeur de  $N^s$ , et pour chaque méthode, on calcule a posteriori la moyenne du score des plans de transport calculés pour 4 jeux de données différents. Les hyperparamètres ont été choisis au préalable, à partir de 100 jeux de données synthétiques pour chaque méthode.

Le code de notre algorithme est écrit en Python3 et utilise la librairie POT [6] pour résoudre les problèmes de transport optimal et `pygsp` [4] pour la génération des graphes.

### 3.3 Résultats

Les résultats sont décrits en Figure 1. La courbe bleue (•) correspond au transport optimal standard (avec une régularisation entropique). La courbe verte (▲) représente le terme de régularisation prenant en compte les étiquettes. La courbe orange (+) correspond au transport optimal régularisé avec le terme proposé dans ce papier. La courbe

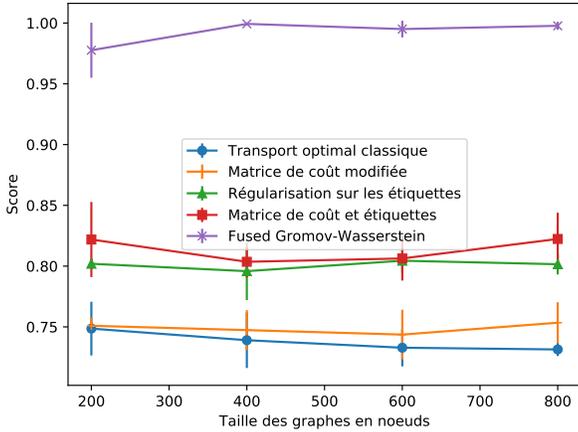


FIG. 1: Moyenne des scores des plans de transport calculés pour les cinq méthodes comparées.

rouge (■) décrit le transport optimal régularisé avec les deux termes précédents. Enfin, la courbe violette (✱) correspond à la méthode *Fused Gromov-Wasserstein*.

Nous pouvons faire les remarques suivantes. La méthode décrite dans [7] produit les meilleurs résultats. Pour les autres méthodes, l'utilisation conjointe des deux termes de régularisation (étiquettes + structures) produit les meilleurs résultats, quelle que soit la taille des graphes. Il est intéressant de noter que ces deux termes sont complémentaires, même si notre régularisation permet d'atteindre de meilleurs résultats que le terme seul sur les étiquettes. Ce constat est néanmoins assez logique puisque notre régularisation exploite de manière explicite non seulement des informations sur les structures mais également sur les étiquettes (liées aux blocs générés).

En termes de temps de calcul le terme de régularisation  $\Omega$  que nous avons introduit est en général très rapide à calculer (pour peu que le graphe ne comporte pas trop d'arêtes). La figure 2 donne une comparaison du temps de calcul moyen des trois algorithmes d'optimisation utilisés. La méthode *Fused Gromov-Wasserstein* est la plus lente (d'un facteur environ 10). Les méthodes qui utilisent le terme de régularisation sur les étiquettes  $\Omega_c$  sont les deuxième plus lentes, car elles doivent utiliser un algorithme d'optimisation coûteux.

## 4 Conclusion

Nous avons proposé dans cet article un nouveau terme de régularisation pour le transport optimal. Celui-ci permet de tirer profit de la structure des graphes quand la tâche vise l'adaptation de domaines entre un graphe source étiqueté et un graphe cible non étiqueté. Nous envisageons de faire évoluer notre modèle en exploitant des polynômes de Laplacien comme cela a déjà été proposé en réseaux

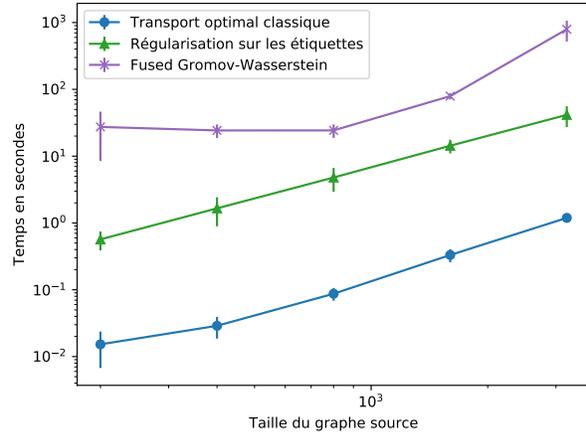


FIG. 2: Temps de calcul des trois problèmes de transport optimal, moyenné sur 8 expériences.

de neurones sur graphes, par exemple. Nous voulons aussi étudier la question du plongement des nœuds source dans l'espace cible avec la structure de graphe.

**Remerciements :** Ce travail a été réalisé dans le cadre du projet ACADEMICS soutenu par le Projet IDEXLYON de l'Université de Lyon dans le cadre du Programme Investissements d'Avenir (ANR-16-IDEX-0005).

## Références

- [1] Emmanuel Abbe. Community detection and stochastic block models. *Foundations and Trends in Communications and Information Theory*, 14(1-2):1–162, 2018.
- [2] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(9):1853–1865, 2017.
- [3] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS 2013, December 5-8, 2013, Lake Tahoe, Nevada, USA.*, pages 2292–2300, 2013.
- [4] Michaël Defferrard, Lionel Martin, Rodrigo Pena, and Nathanaël Perraudin. Pygsp: Graph signal processing in python. Oct 2017.
- [5] Yash Deshpande, Subhabrata Sen, Andrea Montanari, and Elchanan Mossel. Contextual stochastic block models. In *NeurIPS 2018, 3-8 December 2018, Montréal, Canada.*, pages 8590–8602, 2018.
- [6] Rémi Flamary and Nicolas Courty. Pot python optimal transport library, 2017.
- [7] Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Optimal transport for structured data. *CoRR*, abs/1805.09114, 2018.