**Author Feedback for paper "Partial Optimal Tranport with applications on Positive-Unlabeled Learning"**

We thank the reviewers for their thoughtful feedbacks. We are encouraged that reviewers found the algorithm to solve the partial-OT problem new (R1, R2, R4) and theoretically sound (R1, R3). Moreover, R3, R4 acknowledge that one contribution of the paper is the use of OT for PU learning problems, with ok performances (R1, R4) and with the first application of domain adaptation for PU (R3). We are glad that R2, R4 found the paper pedagogical, well written and that it clearly explains the contributions (R1). We finally thank the reviewers for their careful reading and we will perform a careful proof-reading to fix all mentioned typos. We now review the different comments raised.

**Other applications of exact partial-OT**

One primary concern of R2 and R4 is the lack of consideration of other applications such as color transfer, point clouds registration or deep learning that have already been tackled in the literature (e.g. `https://hal.archives-ouvertes.fr/hal-02111220`, `https://arxiv.org/abs/1607.05816`). We agree. We were constrained by space and we hence put the emphasis on PU learning, which is a novel application for OT. This also allows us to introduce domain adaptation for PU learning, which is, to our knowledge, a new task. Regarding deep learning models, one could consider using partial-OT to detect out-of-distributions examples such as in `https://arxiv.org/abs/1912.12510` or in open set domain adaptation (`https://arxiv.org/abs/1804.10427`). We propose to clarify this in the paper.

**Details about the running time**

R1 and R3 mention that time complexity is not discussed. It is true and we propose to include the running times in the supplementary and to discuss the complexity of the algorithms (cubic for partial-W, several iterations of partial-W for partial-GW). For information, considering all experiments, the maximum time for running one run of partial-W is less than 1.5s and 5.5s for partial-GW.

**Comments regarding the application of PU learning with partial-OT** We now discuss the comments of R3.

• $\#$**1** *The mass $\pi$ (positives) from the unlabeled sample gets transported to only mass $\pi$ of the labeled sample* and $\#$**5** *The text following equation 6, including the introduction of $\alpha$, regularization, is quite challenging to follow*. There is indeed a typo in line 175: $q_j = 1/m$ should read $q_j = \pi/m$ (as it is correctly formulated in line 182) – many thanks for spotting it! As such, it allows verifying proposition 2, and when choosing $\alpha = 0$, *all* the labeled samples are transported to a mass $\pi$ of the unlabeled sample. We believe that this misspecification of $q_j$ leads to the difficulty in reading the text and we propose to add in line 179 a discussion about $\alpha = 0$.

• $\#$**2** *An important PU method is not included in the comparisons [3]* and $\#$**3** *Some experiments on the sensitivity to class prior should be conducted.* Since many of the estimation methods of the prior are biased, we agree that it is important to evaluate the influence of the class prior $\pi$ in a biased setting. Note that, as stated in the conclusion, "we plan to derive an extension of this work to PU learning in which the proportion of positives in the dataset will be estimated in a unified optimal transport formulation" but leave it for future work. We then run some additional experiments by varying class prior as in [3] for the MNIST dataset and propose to add the results in the final version of the paper. Note that the method we propose is *transductive*, hence it leads to reduced performances (less than 1 point for all experiments, using $\pi' = 0.8\pi, 0.9\pi, ..., 1.2\pi$). It is true that we do not compare ourselves to [3]. Instead, we prefer relying on Kato et al. (2019), which is a more recent method and which itself builds upon [3]. As such, we strongly believe that the conclusions relative to Kato et al. will resemble the ones that could get by comparing to [3]. Nevertheless, if requested, the comparison could be added in the final version if the paper is accepted.

• $\#$**4** *What kind of bias does the proposed method handle?* and $\#$**6** *Would the partial-GW method still work if the negatives were obtained by adding a constant to the positives?* Actually, GW is rotation and translation invariant (or more generally invariant to *isometries*). This is a desirable property, as we would like to match data with similar geometry when working on different or unregistered spaces, but in the particular toy case mentioned in $\#$**6**, this behavior will indeed not allow the detection of the positives among the unlabeled. In the colored MNIST example, partial-GW identifies the unlabeled positives as they have the same "geometry" than the labeled dataset, even if the proportion of green samples in the unlabeled is the same than the prior of the positives (one could expect the method to wrongly label the green ones inside **Unl** as positives). We propose to clarify the invariances of GW in the final version of the paper, then the type of bias that can be handled, together with a deeper discussion of GW as suggested by R2.

**On the notations and clarity**

R3 states that *Authors should spend effort in improving notations* and *the paper will be easier to read if words are used to express the meaning of formulas*. We faced two difficulties: i) space limitation, ii) applying OT for PU learning, two communities which have different notations. As such, we choose to stick to the usual notations of OT (in which $\boldsymbol{p}$ and $\boldsymbol{q}$ usually denote the distributions of $n$ and $m$ bins with the same number of bins as points etc.) and we adjust the ones related to the PU learning, leading to potentially disturbing notations (see comment of R2 about $p$). We make our best to keep the notations consistent, and we try to avoid misunderstandings as much as possible (e.g. by stating **Pos** the labeled positive points rather than **P**). We believe this is not an obstacle to the paper comprehension. To ease the reading, we propose to add more details about the most important equations, and emphasize on the text when a careful reading must be done (e.g. when the source data $\boldsymbol{p}$ correspond to the unlabeled set). Comment $\#1$ of R1: line 103 indicates how to set the mass of the dummy point. Regarding the tuning of $\xi$, Proposition 1 illustrates that this parameter does not influence the solution of Partial-W. In practice we set it as zero as stated in Line 202.