We thank the reviewers for providing very thoughtful and helpful comments. In general, the reviews mainly focus on three concerns: motivation for low nuclear norm structure (from all three reviewers), originality (from reviewers #2 and #3), and experiment settings (from reviewers #2 and #3). We address these main concerns below.

Motivation for low nuclear norm structure. All three reviewers asked for us better relating the low nuclear norm structure in missingness patterns to real data (reviewer #2 also asks about relating it to synthetic data). Importantly, we want to reiterate that our assumptions of low nuclear norm (Assumption A1) and propensities scores being bounded from 0 and from 1 (Assumption A2) have special cases that include low-rank propensity score matrices, which are a broad class of matrices that the propensity scores can have (special cases include block-structured propensity score matrices that would, for instance, come from row/column clustering, or as a more elaborate example, topic modeling structure). We discuss why this low rank condition is a special case after stating Assumptions A1 and A2 in the paper. To be concrete, for the synthetic datasets we consider, we can easily verify how Assumptions A1–A3 are satisfied. For the real data we consider, we do not have the true propensity score matrices, but we can do the following: we can take the missingness mask matrices and check for whether they have block structure, which would suggest that how entries are revealed is approximately low rank. We discuss these in detail next; we will add these details to the paper.

Synthetic datasets. We verify that Assumptions A1-A3 hold for the synthetic datasets. Assumption A3 holds as both synthetic datasets have partially observed matrix X consist of ratings in a bounded interval. For `MovieLoverData`, the propensity score matrix P is a block matrix, so it is low-rank, and moreover it has three unique values that are all nonzero and less than 1; thus Assumptions A1 and A2 are both met. For `UserItemData`, the propensity score is $P_{u,i} = g(U_2[u]w_1 + V_2[i]w_2)$, where g is the standard logistic function. Hence parameter matrix A (in Assumptions A1 and A2) is low-rank and, moreover, in how we generate A, with high probability $\|A\|_{\max}$ is bounded (using standard concentration results for the maxima of a finite collection of Gaussians); thus Assumptions A1 and A2 are both satisfied. In summary, the synthetic datasets we consider do actually satisfy the assumptions of our theoretical analysis.

Real-world data. To demonstrate that real-world datasets we consider satisfy the low nuclear norm assumption, we run the following experiments: we use bi-clustering algorithms to rearrange rows/columns of the missingness mask matrix to reveal block structure; the results for `COAT` and `MovieLens` are presented in Figure 1.

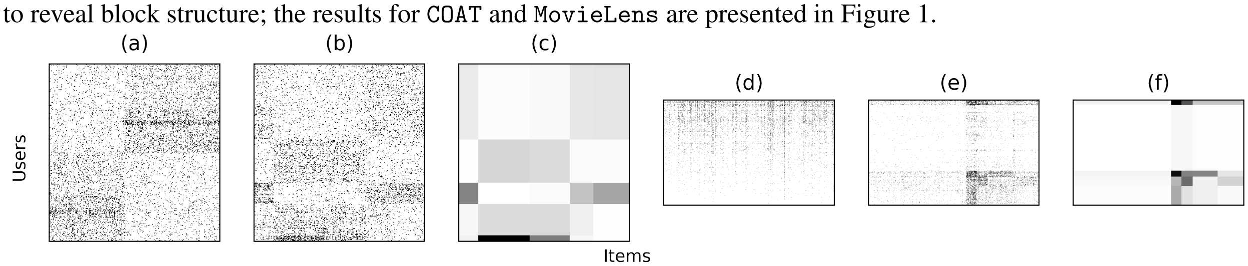

Figure 1: Biclustering the missingness mask matrices. (a,b,c): Raw, biclustered missingness mask matrices and average missingness for `COAT`; (d,e,f): Raw, biclustered missingness mask matrices and average missingness for `MovieLens`.

One can see that block structure appears in both datasets, which suggests the propensity score matrix can be well modeled as low-rank, which is a special case of the low nuclear norm (provided that propensity score is bounded from 0 and 1). We will add this figure to the paper.

Originality. Reviewer #2 mentioned that the propensity score estimation method 1BITMC is based on Davenport et al. [2014], which is not developed by the authors. We agree with reviewer #2 that 1BITMC is not new. We specialize their analysis to fully-observed matrices. This fully-observed setting allows us to arrive at a slightly tighter error upper bound. Even so, we consider our theoretical analysis to be an incremental contribution. Instead, we argue that the main contribution of our paper is to demonstrate that low nuclear norm structure in propensity score matrices appears to be reasonable for real-world data and to provide a theoretically-justified propensity estimation approach when MAR data and user/item features are not available. Though the idea is straightforward, we believe the proposed approach is novel and useful for estimating the propensity scores without using any extra information. Reviewer #3 commented that the novelty of this paper is undermined by insufficient motivation for low nuclear norm structure in missingness patterns. We hope that we addressed this concern with the clarifications in previous paragraphs.

Experiment settings. Reviewer #2 believes that using MNAR ratings as the test set might be biased, while we argue that the measure $L_{\text{IPS-MAE}}$ is unbiased even with MNAR test set. In both synthetic and real-world datasets, we compare different algorithms through $L_{\text{IPS-MAE}}$ to avoid bias. One major contribution of this paper is to estimate the propensity matrix without using MAR data or user/item features, hence we prefer to work on datasets that do not contain MAR ratings. Instead, we believe using $L_{\text{IPS-MAE}}$ is enough to conduct the fair and unbiased evaluation. We would like to thank reviewer #3 for the suggestion of using user/item features and MAR data together with 1BITMC to further improve performance. We consider this a promising future research direction. For simplicity, in this paper we focus on propensity score estimation without any extra information (but using nuclear norm structure) and benchmark against logistic regression and naive Bayes baeslines that do use auxiliary information.