We thank the reviewers for their valuable comments. Relevant points raised are consolidated and addressed together.

**Setting:**

- **Edge constraint (Reviewer 1):** Define $k_i$ as the number of edges in $\mathcal{G}_i$. $k_1$ and $k_2$ can be distinct and take arbitrarily different values in the range $[0, k]$. We analyze the minimax error rate, i.e., the error performance for the worst-case combination $(k_1, k_2)$. This represents structure learning in the most *difficult* combination. The minimax rates take different forms in different parameter regimes, as specified in the paper. Analysis in each regime hinges on identifying the worst-case combination in that regime.
- **Side Information (Reviewer 3):** We thank reviewer 3 for sharing their perspective on side information, and indeed their interpretation is also valid. We comment that the case of unilateral side information (e.g,. $\mathcal{G}_2$ serves as the side information for $\mathcal{G}_1$) is a special case of the bi-lateral scenario that we consider, in which each graph serves as the side information for the other one. We can recover unilaterial side information results by changing the metric in (7) from $\min_{i\in\{1,2\}}\{|E_i \Delta \hat{E}_i|\} \geq d$ to $|E_1 \Delta \hat{E}_1|\} \geq d$.

**Novelty in analysis (Reviewer 3):** We, respectfully, disagree about lack of novelty in analysis. To furnish a context:

1. Santhanam and Wainwright, 2012 [SW2012] focuses on *exact* recovery and provides both lower and upper bounds.
2. Scarlett and Cevher, 2016 [SC2016] focuses on *approximate* recovery and provides **only** a lower bound.

- **A hitherto un-investigated scenario:** Besides generalizing the regimes [SW2012] of [SC2016] for joint recovery, we also provide *upper bounds on the approximate recovery*, which is the missing scenario in [SW2012] and [SC2016]. Hence, as a special case of our results we can recover the results for this missing regime for single-graph structure learning as well. The way different parameter regimes for this scenario are constructed and the ensuing analyses are distinct from ensemble construction and proofs of both [SW2012] and [SC2016].
- **Generalization of other scenarios** Please note that generalizing the other scenarios from one graph to two graphs is non-trivial. Even though the ensemble selections are inspired by [SW2012] and [SC2016], their choices and the techniques for analyzing the minimax rate are different. The similarity in some of the approaches is inevitable (e.g., Fano's inequality is pivotal for proving converses in information theory).
- **New insights for ML decoding:** Finally, we note that [SC2016] provides the lower bounds identical or near-identical to those of exact recovery in [SW2012] for a wide range of edge-bounded Ising models, based on which it was conjectured that approximate recovery is as hard as exact recovery for the complete edge-bounded subclass. In this paper, we also establish that this conjecture is true for the edge-bounded subclass for an ML based decoder.

**Relevance to the ICASSP paper (Reviewer 2):** There are significant differences in settings, objectives, and results:

- **Settings** (parameter regimes): The ICASSP paper focuses on very specific Gaussian and Ising models (specific parameter regimes). In this paper we do *not* consider Gaussian, and focus on a much broader subclass of Ising models. Specifically, the ICASSP paper focuses on an Ising setting with $k \leq \frac{p}{4}$ and restrictions on the girth and separation criterion. The results are provided only for the specific regime $\lambda = O(\sqrt{k^{-1}})$ for the relative choices of $k \leq \frac{p}{4}$ and $\lambda$. In this paper, we consider all values of $k$ and all possible regimes for the relative choices of $k$ and $\lambda$.
- **Objective** (approximate vs. full recovery): The focus of the Ising model section of the ICASSP paper is only on approximate recovery, while we consider both approximate and full recovery objectives.
- **Results** (necessary & sufficient conditions): The ICASSP paper provides only necessary conditions (lower bounds) for the specific class mentioned, while we provide both necessary and sufficient conditions for the general settings.

**Sample complexity results:**

- **Shared cluster (Reviewer 1):** Even though not presented in the paper, we can readily show that in *most* parameter regimes the performance in the shared cluster is similar to single-graph recovery with $p\eta$ nodes and at most $k\gamma$ edges.
- **Results in Table 1 (Reviewer 3):**
  - **Tightness:** While we agree that the bounds are not very tight, we also would like to emphasize that that is the case even for the simpler problems in the literature (e.g., single graph recovery). Our bounds are not any looser than those for single-graph recovery. Also, in some regimes the results are tighter than others. For instance, when $k = O(p)$ the difference is only a factor $k$. The mismatch is more profound in denser graphs.
  - **Effect of $d$:** The analyses of the necessary conditions in Theorems 2 and 3 and sufficient conditions in Theorem 1 show that $d$ does not affect the asymptotic scaling rate of their respective bounds even when $d$ scales as fast as linearly with $k$. For instance, in Theorem 1, $d$ appears only in a logarithmic factor scaling at most at the rate of $\log k$ which is dominated by $k \log p$ in $A_1$ and $A_2$. That is why the summary results in Table 1 do not include $d$. We will add a comment to highlight this in the final version.

**Numerical results:** Reviewer 1 is correct (the solid curve represent ML in Fig. 2). We will clarify this in the final version. Also, as Reviewer 1 suggested, we will update Fig. 3 to showcase an average performance over an ensemble of random graphs. Also, we will provide more explanation on recovery accuracy as we increase $d$ and its interplay with the computational cost of ML. Specifically, the main observation is that as we *slightly* increase $d$, while the computational complexity improves *slightly*, the recovery declines *rapidly*. Regarding the question of Reviewer 3, we note that the numerical evaluations were carried out on a family of sparse graphs, for which the ML estimation was tractable and implemented via counting the number of instances a vertex has the same value as other vertices. In the settings we examined, we did not observe any phase transition in the error rate when varying $k$, $p$, $\eta$.

**Typos:** We thank the reviewers for noting the typo on $\gamma_k$, which we will fix. Also, the two appendices submitted as supplementary documents were generated independently. We will consolidate them for the final submission.