We thank the reviewers for their insightful comments and interest in the work, find below line by line responses. For space reasons we do not reply to comments on organization or typos but have incorporated them into the paper.

R1, R2, R3, R5: Limitation to trees. Several reviewers raised questions about our limitation to tree graphs. We apologize for not motivating our assumption more clearly, and will modify the paper to clarify that the algorithm works whenever the undirected components of the graph form a forest. As briefly mentioned in L51-52, after the observational stage, the undirected components of the essential graph are chordal. Orienting each of them provides no information on the others, thus they have to be handled separately by any algorithm. Our algorithm only requires these components be trees, but they do not need to be connected. Note we can identify all v-structures from observational data. For example this assumption is satisfied when the original graph is bi-partite, since chordal components of bi-partite graphs are forests. Examples of bi-partite causal graphs occur in system biology networks, e.g. gene-disease networks or gene-protein networks [https://www.ncbi.nlm.nih.gov/pubmed/27265032]. We will add this discussion to the paper.

R2: ...information greedy algorithm is suboptimal. (low significance) While we don't dispute this, we point out that the dramatic failure of information greedy puts our problem in a different class from many active learning scenarios.

[the case in which] $P(Y = y)$ is very similar to $P(Y = y|X_i = 1)$. As R2 correctly states, our algorithm may not be optimal if some edges are much weaker than others (i.e. when Condition 1 in Sec. 3.4 does not hold). Note that under Condition 1, we provide formal guarantees for a version of our algorithm. We also believe that Condition 1 is justified in real-world examples: this is because with finite observational data, we expect weak edges to be left out by methods that learn the graph skeleton.

...how such an information greedy algorithm works w.r.t. finite-sample case. In the finite sample case, the information greedy algorithm must take more interventions than in the infinite sample case. Hence the number of required interventions is still linear in the size of the graph n for our counter example structure. Meanwhile under Condition 1, the infinite-sample central node algorithm can be applied to the finite sample case by repeating each intervention until a branch has at least $1-\delta/(\log n)$ probability, which requires order $\log \log n$ repetitions in expectation by Prop. 2. By the union bound then, this algorithm finds the solution with $1 - \delta$ confidence in number of interventions nearly logarithmic in n, a major speedup over information greedy.

...incorporate a prior distribution... We only consider priors over the graph structure. Our approximation guarantees with respect to the optimal algorithm hold for any prior assumed over the source node locations (see Sec. 3.1).

R5: interventions are assumed to be single target. Is there any straightforward way to extend the updates...to simulations interventions? Lemma 1 can indeed be extended, we did not originally include it because it would complicate the paper and raise several corner cases which complicate the algorithms. We will add it to the supplement.

The type of interventions used in this work is Pearl's atomic intervention...it may be possible that the cause affects the effect only under certain outcomes. We wrote the paper with $do(X_i = 1)$ for notational simplicity, as long as there is any value a_i for which $do(X_i = a_i)$ affects the effect variables, we can find it from the observational data and the theory will still hold. We will expand and rigorize this discussion in the camera-ready.

the extension is not clear...In the experiments, the authors have restricted themselves to binary variables. We will add a discussion and new experiments for the non-binary variables, the results are qualitatively the same. Essentially, the extension hinges on the fact that we simply need the result of the do-intervention to provide information on the direction of the root via the posterior update (Lemma 1), all our results follow from this.

Connection with [Ghassami et al. 2017]: Thank you for pointing us to this paper, we will be sure to cite it and make the connection clear. Indeed, their Definition 8 and our central node definition are the same, and our algorithm can be seen as repeating the first step of ProBAL with evolving posteriors. In fact, this is true for any non-adaptive algorithm whose first step is intervening on the central node of the prior. However, ProBAL is a non-adaptive algorithm designed for a fixed budget setting (number of interventions). Interestingly, our results show that it is a good algorithm with a constant factor approximation guarantee even in the adaptive, fixed confidence noiseless setting. We also show an adaptation of this works well in case of restricted nodes and in the noisy case with an assumed bound on the noise level.

Comparison to [Hauser and Buhlmann, 2014]: While OptSingle is non-adaptive, it can be turned into an adaptive algorithm if one applies it repeatedly on the remaining essential graph after the intervention and application of Meek rules. The objective is to minimize the number of unlearned edges in the worst case graph. In the case of a tree with a uniform prior, this algorithm will also intervene on the central node, since only that will minimize the worst case number of unoriented edges. However, when the prior is non-uniform the two algorithms are not equivalent: OptSingle will only use the structural properties and not take into account the prior. In the noisy case, a single intervention will not change the support of the posterior distribution, hence OptSingle will continue intervening on the same node without making any progress. Hence we did not see a straightforward way to compare to it in our experiments, all of which consider the noisy case. We will add appropriate citations and discussion of these papers in the camera-ready.