We thank the reviewers for the helpful feedback. With the exception of Reviewer #1, the reviewers agree that our approach is novel and presents a solid contribution. In particular, Reviewer #2 says that "this is a solid contribution" and that "the method is novel and the results look impressive". Reviewer #3 comments that "the proposed approach is novel" and that "this is a very solid paper". Finally, Reviewer #4 highlights that "this is a good contribution that addresses a novel problem" and that "the advantage of the methodology is clearly reflected in the experiments".

Reviewer #1: *(1) Novelty is not significant*. We respectfully but strongly disagree. Our contribution is significantly different to standard practice as we consider a prior distribution over the *graph* structure and use it to address specific challenges. This has been largely overlooked in the previous literature. We respectfully refer the reviewer to our summary of the other reviewers' feedback above and their corresponding detailed comments, where they agree that our approach is novel and presents a solid contribution.

(2) Too many free parameters. We also present a compact low-rank parameterization (LRP) of the posterior. We discuss this in the supplement, §H.1. In some cases, the LRP achieves similar performance but exhibits much higher variance, hence requiring more epochs to converge. Furthermore, as mentioned in §5, our method can be combined with other graph neural network approaches, especially with those designed to improve the efficiency of GCNs. To illustrate this point, we have recently begun development of a more scalable extension based on Cluster-GCN (Ciang et al); early experiments on PUBMED indicate that our approach, with an order-of-magnitude less parameters, can perform similarly or better than scalable competing benchmarks, while other methods such as LDS threw out-of-memory errors.

(3) Scaling for KL too small and having a minor influence. We have analyzed the KL-dampening factor selected through cross-validation on the adversarial experiments across all datasets. It turns out that $\beta = \{1, 0.01, 0.001\}$ get selected $\{32\%, 35\%, 21\%\}$ of the time, while $\beta = 10^{-4}$ only 12%. In addition, the performance benefits in the main paper show that the KL regularization resulting from variational inference does have an effect. Lastly, we have found that even when β is very small, the KL term still has an effect as it can be several orders of magnitude larger than the ELL.

(4) Method not the best in Fig. 3. We believe ML papers should also showcase when a proposed approach does not work best and this was our intention in Fig. 3. This provides some guidance on when other methods might be preferable ("no free lunch"). Nevertheless, in this experiment our method is still very competitive and does not suffer from catastrophic performance degradation when node features are not available.

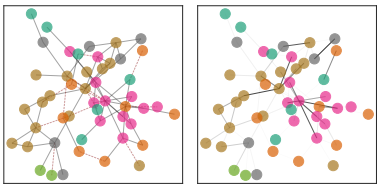
Reviewer #2: *(1) Prior work in intro*. We thank the reviewer for this suggestion and we will make the first part of the intro more consistent with the subsequent related work in Sec 1.1.

(2) Tuning of baselines. We would like to clarify that we tuned the baselines across all experiments including the featureless experiment. This was done to the degree to which the reference implementations allowed us to tune them and followed the recommendations of the original papers. For example, as described in the supplement, we tuned GCN similarly to our method and the LDS implementation adopts a very thorough cross-validation procedure. For other baselines such as graphSage we got around the high-variance estimate problem by using multiple predictions, which provided better performance.

(3) Prior knowledge in LDS. Although LDS includes a generative model, the initial probabilities use a deterministic distribution (their Algorithm 1) and do not play an explicit role in the objective function being optimized, i.e. there is no prior constraining the search-space over graphs. Using different initial probabilities will not achieve a similar effect to that obtained in our probabilistic framework. We will expand more on this in the final version.

Reviewer #3: *(1) Alternative choices for inductive setting*. This is indeed an exciting avenue of research that we are pursuing based on non-linear matrix factorization approaches using Gaussian processes (GPs). Incorporating GP priors over graphs can enable one to generalize to unseen nodes.

(2) Qualitative analysis. To qualitatively analyze our approach, we examine densely-connected subgraphs of the CITESEER graph used in the experiment shown in Fig. 7 of the supplement (i.e. adding edges). Preliminary results in the given figure (node colors indicate node labels; left: original corrupted graph with red dashed lines indicating added edges; right: learned graph shown with edge opacities proportional to posterior probabilities; best seen zoomed in on screen) indicate that, with a few exceptions, the posterior probabilities of the *added* edges are indeed attenuated. We will expand on this analysis in the final version.



(3) Neural relational inference (NRI). Thank you for the reference. NRI (Kipf et al, ICML 2018) addresses the problem of learning the interactions between components in a dynamical system given their trajectories, i.e., the entities evolve over time. Besides the similarities wrt using variational autoencoders during inference to learn these interactions, their focus is very different to ours but we are happy to cite it and expand on this in the final version.

Reviewer #4: *(1) Computational complexity*. Please see Reviewer #1, item (2) above.

(2) Using the model for inferences on the network itself. This is an interesting point. The posterior probabilities can, indeed, be used to make inferences about the network. However, one must be cautious when comparing these to the ground truth graph, as the loss being optimized focuses on the classification problem and does not explicitly include a suitable link-prediction loss. Thus, the given links enter the objective only as a soft regularizer through the KL term. Nevertheless, we refer the reviewer to our comment on qualitative analysis above, Reviewer #3, item (2).