Thank you all for all the positive comments and useful suggestions. We address each reviewer separately, breaking down into topics when possible. References are at the end.

Reviewer #1

Misunderstanding of experiments.

1. Our approach is **inductive**, while the baselines in the appendix are transductive. **Transductive** methods **are supposed to be better than inductive methods on transductive tasks**, since the transductive methods know that the examples in train and test data are the same. **To our great surprise, our approach is better in 13 out of 14 experiments** over transductive methods. This showcases the strength of our inductive method. Nevertheless, these positive results belong to the appendix since our work focuses on inductive tasks.

2. Recent theoretical results [2] (and older empirical ones [1]) show that GNN node representations cannot perform k-ary tasks. **The poor performance of GNNs that output node representations was no surprise.**

Misunderstanding of theory.

Line 37 in our paper shows that the challenge is overcoming the **intractability** of the **unormalized** probability function $\Phi(\mathbf{A}, \mathbf{X}; \mathbf{W})$, which is a challenge specific to k-ary tasks on graphs. Simply, there is no existing EBM method that deals with this issue. The "intractability" of the normalization factor Z is easy to solve using standard methods (NCE). We will further clarify this point in our paper.

Reviewer #2. Great suggestions!

New InfoGraph results.

Thank you for the great reference, we were not aware of InfoGraph! **We ran InfoGraph experiments** and found it comparable to our approach (with $k = n$) on IMDB tasks but significantly worse than our approach (with $k = 3, 4, 5$) on ENZYMES and PROTEINS (just for reference, the new results are: PROTEINS: 0.690 ± 0.04 ENZYMES: 0.278 ± 0.04 IMDB-BIN: 0.691 ± 0.04 IMDB-MULT: 0.466 ± 0.02). These do not affect our conclusions and we will add these results to the final version of the paper.

Why not k-GNNs.

The main paper (line 79) discusses why k-GNNs should not be used in our tasks (more at line 569 (appendix)). In short, adapting k-GNNs to our task would mean proposing an incorrect objective function for optimization, since each edge would appear $\binom{n-2}{k-1}$ times in the objective function with different (inconsistent) predictions for each possible subgraph. We fear this could be misconstrued as our endorsement for this optimization objective.

DGI and edge-base loss.

The main paper (line 72) discusses why DGI is not an edge-based model. Recent work [2] shows how single-node GNN representations are theoretically incapable of jointly representing k nodes, thus DGI cannot be trivially extended to joint k-node representations.

Reviewer #3. Thanks for the great suggestions!

We will add to the appendix a figure with a graph, its k-HON and its k-CNHON with an example tour. We will also move Figure 1 to Section 3 and incorporate your suggestions. Furthermore, we will add references to works on EBMs in the related work of our supplement, making it clear that existing work do not handle an intractable *unormalized energy* of our type.

Reviewer #4. Thanks for the great comments!

Connected Induced Subgraph (CIS).

There is a large body of empirical work in the complex networks literature evidencing how CIS'es are effective in representing graphs, for example see [3]. We will add those references as motivation for using CIS'es in section 4.

Why node-based GNNs underperform in k-ary tasks.

Recent theoretical results [2] (and older empirical ones [1]) show that GNN node representations cannot perform k-ary tasks. **The poor performance of GNNs that output node representations was no surprise.** The simple use of raw features performs better. We will emphasize the theory in the results section.

Future work.

The reviewer proposes challenging but exciting future work. Making k flexible by simply decomposing the energy into a sum of different k's does not guarantee that the model will not overfit on some specific k value, not learning with the other subgraph sizes. Coping with these challenges is an exciting area for future work.

References:

[1] Meng et al. "Subgraph pattern neural networks for high-order graph evolution prediction." AAAI. 2018.
[2] Srinivasan & Ribeiro. "On the Equivalence between Positional Node Embeddings and Structural Graph Representations." ICLR 2020.
[3] Milo et al. "Network motifs: simple building blocks of complex networks." Science 2002.