We thank all four reviewers for unanimous support for the paper and constructive comments. To recap, our submission is the first paper to study meta-path prediction as an auxiliary task to train GNNs on heterogeneous graphs. Our meta learning-based frameworks adaptively balance auxiliary tasks (meta-path prediction) with the primary task (link prediction or node classification). The proposed methods improve representational power without any additional data or labels. Overall, reviewers are positive about our contributions: [R5] "The proposed framework is very general and can serve as a plugin to improve the performance of any existing GNNs.", [R4] "They propose a pre-training strategy specifically designed for heterogeneous graphs. It is the 'first' paper to do so.", and [R1] "The problem is important. The overall quality of this paper is good." Also, we will address questions raised by the reviewers below hoping for more vigorous support.

Q1 [R1, R2, R5]. Is the proposed method applicable to existing heterogeneous GNNs such as GTNs [1]?
Yes, as [R4] pointed out, our framework can be applied to any GNNs in a plug-in manner. Table. 1 shows our framework consistently improves the representational power of Graph Transformer Networks (GTNs) for heterogeneous graphs in both link prediction (Last-FM) and node classification (IMDB). We have repeated the experiments three times as requested by **[R5]**. Also, we added the HINT column to evaluate the efficacy of HintNet itself as suggested by **[R2]**.

Table 1: Performance of GTN trained by various learning schemes.

Dataset	Vanilla	w/o meta-path	Ours			
			w/ meta-path	Hint	SELAR	SELAR+Hint
Last-FM	0.7836±0.0030	0.7744±0.0022	0.7865±0.0042	0.7883±0.0054	0.7978±0.0018	**0.8053**±0.0064
IMDB	0.5804±0.0073	0.5792±0.0017	0.5818±0.0088	0.5889±0.0112	0.5994±0.0097	**0.6063**±0.0046

Q2 [R4, R5]. Performance gain from the Hint Network.
Table. 1 in our supplementary materials show clearer improvement by the Hint Network rather than the results in the main paper. Our Hint Network with a regularizer improves the performance of GNNs along with SELAR in most cases. In 7 out of 8 cases, SELAR+HintNet outperforms SELAR. We found that imposing a regularizer on HintNet is helpful. We will move these results to the main paper in our final version.

Q3 [R4, R5]. Motivation and details of Hint Network.
We introduced Hint Networks in Section 3.3 in the main paper. Compared to primary tasks such as link prediction and node classification, the meta-path prediction might be more difficult to be learned. For instance, when training a GNN on a large graph, some important intermediate nodes and edges may not be available in a small mini-batch. Moreover, GNNs with a small number of layers are not capable to learn long meta-paths. Our Hint Network makes the challenging tasks more solvable by correcting the answer at learner's need with hub nodes which have rich connections with other nodes. Roughly speaking, meta-learning for SELAR allows GNNs to learn which task to learn (meta-path selection) and meta-learning for HintNet adjusts the difficulty level of a given task. In our experiments in the main paper, HintNet f_H^t is the exactly same model as learner f^t but it uses the augmented mini-batch with hub nodes. Our method can learn from either original topology or augmented topology with hub nodes at learner's needs.

Q4 [R1, R2, R4]. Why meta-path prediction? More discussion about self-supervised learning on graphs.
Great question! We found that meta-path prediction is an effective self-supervision task to train GNNs on heterogeneous graphs. Our experiments support that meta-path prediction itself improves the representational power of GNNs significantly. As reviewers pointed out, self-supervision and pre-training of GNNs have recently been studied in the literature [2, 3]. Weihua Hu et al. [2] have introduced effective strategies for pre-training GNNs such as attribute masking and context prediction. Separated from the pre-training and fine-tuning strategy, [3] studies multi-task learning and analyzes why the pretext tasks are useful for GCNs and its close variants. However, one problem with both pre-training and multi-task learning strategies is that all the auxiliary tasks are not beneficial for the downstream applications. Motivated by this problem, we studied 'auxiliary learning' for GNNs that explicitly focuses on the primary task. Auxiliary learning is similar to multi-task learning but it is not obliged to improve the performance of any auxiliary tasks. The auxiliary tasks just assist the primary task. We proposed frameworks that utilize multiple auxiliary tasks without harming the primary task. This can be formulated as a general meta-learning problem and our methods can incorporate all the self-supervision tasks mentioned above as auxiliary tasks. Also, we believe that the same idea can be used in any representation learning. We will add more discussion about related work in the final version if accepted.

References

[1] Seongjun Yun et al. Graph transformer networks. In *NeurIPS*, pages 11983–11993, 2019.
[2] Weihua Hu et al. Strategies for pre-training graph neural networks. In *ICLR*, 2020.
[3] Yuning You et al. When does self-supervision help graph convolutional networks? *ICML*, 2020.