Response to the reviewers

We thank the reviewers for their critical assessment of our work. In the following we address their concerns point by point.

Reviewer 1

Reviewer Point P 1.1 — There are lots of symbols in this paper. Some symbols are reused and confusing, such as s denotes sub-layers or edge features.

Reply: TODO

Reviewer Point P 1.2 — Some typical applications of GNNs should be included, such as video object segmentation [ref1], human-object interaction [ref2] and human-parsing [ref3].[1] Zero-shot video object segmentation via attentive graph neural networks,iccv 2019 [2] Learning human-object interactions by graph parsing neural networks, eccv 2018. [3] Hierarchical human parsing with typed part-relation reasoning, cvpr 2020.

Reply: Thank you for pointing out our shortcomings. Computer vision is indeed another important application area of graph neural networks. We have added the mentioned references in the INTRODUCTION section in the revised manuscript. We quote the related sentence below:

> The powerful expression ability makes GNNs achieve good accuracy in not only graph analytical tasks [8, 9, 10] (like node classification and link prediction) but also computer vision tasks (like human-object interaction [11], human parsing [12], and video object segmentation [13]).

Minor

Reviewer Point P 1.3 — Typo in line xy.

Reply: Fixed.

Reviewer 2

Reviewer Point P 2.1 — This is the first point of Reviewer 2. With some more words foo bar foo bar ...

Reply: Our reply to it with reference to one of our points above using the LATEX's label/ref system (see also **??**).