Thanks to all the reviewers for their insightful, and constructive feedback! The reviewers raise important concerns regarding presentation, and other technical concerns. First of all, it is encouraging to see the reviewers find that

1. The unified formulation is strong, and tries to capture the most complicated graph-structured data - **Reviewer 1 (R1)**
2. The work is theoretically grounded, and the empirical results are promising - **Reviewer 2 (R2)**
3. The methods are well-motivated, and the experiments are well-designed and detailed with ablation studies - **Reviewer 3 (R3)**
4. The experimental evaluation is the biggest strength, and the proposed models properly generalise a wide variety of models in literature - **Reviewer 4 (R4)**

Response to Common Concern on the Proof of Proposition 1:

R1, **R2**, **R3** wanted clarity on T_u and T_v in the proof of proposition 1. T_u and T_v represent vertex types of the vertices u and v in the heteorogeneous graph (V, E, S), and they are the same as S_u and S_v respectively. This is certainly a typo (also acknolwedged by **R3**), and we will replace T_u, and T_v by S_u, and S_v respectively.

Response to Reviewer 1: Thanks for taking the time to review the paper!

Storyline is unclear: As stated in the introduction, our main contributions are to unify exisiting MPNNs on different structures into G-MPNN, demonstrate the strong inductive capability of G-MPNN, and extend MPNN to recursive structures. All three contributions have a common motivation e.g. Fig. 1 (acknowledged by **R2**, **R3**).

Response to Reviewer 2: Thanks for thoroughly reviewing the paper!

Why not run [66,16,24,51,44,23] as baselines?: It is not straightforward to extend those methods to the inductive setting (ability to handle unseen entities at test time). All those methods assume that all entities are seen during training (transductive), and hence they are not baselines in the inductive experiments. Furthermore, we do use some of them as baselines in the transductive setting (Appendix Section 2.6, Table 7) where G-MPNN performs comparably.

How do these models compare?: The baselines of G-MPNN have fewer parameters than G-MPNN. The additional parameters of G-MPNN are relational embeddings, and positional embedings. The ablation study in the appendix (Section 2.1, Table 4) demonstrates the additional information are required for G-MPNN's effectiveness.

Multiple entries have the same properties in Table 1: Yes, in fact, multiplex networks are typically special multi-relational graphs with homophilic relations (e.g. facebook friends, twitter followers, linkedin connections in a 3-plex network). Proposition 1 in the paper means that heterogenous graphs and multi-relational graphs are very similar. These two points only emphasise the point that tens of recent GNNs / MPNNs have recently been proposed with small modifications, and our G-MPNN generalises and captures the key idea in all these models (also acknowledged by **R1**).

Unclear how datasets are setup: Datasets for MPNN-R / recursive hypergraphs are setup from their raw sources (from their source websites such as aminer, arXiv, etc.). For academic networks, authors are depth-1 hyperedges, and documents are vertices, and depth-0 hyperedges (please see Example 1, line 162 for details). Datasets for G-MPNN / multi-relational ordered hypergraphs are obtained from popular multi-ary benchmarks used in existing literature [1, 2].

Response to Reviewer 3: Thanks for the review and the useful pointers!

Comparison with GAAT: Thanks for bringing attention to the GAAT paper. We will include this as a baseline in the transductive experiments (with proper credits to GAAT). One interesting weakness / extension of our paper is G-MPNN-R (Figure 1 in the paper shows a use case).

Comparison with HAN and MAGNN: Thanks for the suggestion, we include them as baselines in the Table shown on the right.

Classification error (lower is better) on recursive hypergraph datasets used in the paper

Method	Cora	DBLP	ACM	arXiv
HAN	27.24 ± 1.9	22.18 ± 1.4	23.21 ± 2.1	25.02 ± 2.2
MAGNN	26.78 ± 1.5	$\mathbf{21.68 \pm 1.8}$	22.29 ± 1.9	24.23 ± 1.8
MPNN-R	$\mathbf{25.34 \pm 1.5}$	$\mathbf{21.45 \pm 1.7}$	$\mathbf{20.32 \pm 2.1}$	$\mathbf{22.34 \pm 1.7}$

Citeseer, PubMed datasets: MPNN-R is most effective for datasets containing at least depth-1 hyperedges (and higher depths). We did try to get authorship information (depth-1 hyperedges) for Citeseer, PubMed datasets but they seem to be proprietary (not publicly accessible).

Response to Reviewer 4: Thanks for a very careful review, and the detailed comments!

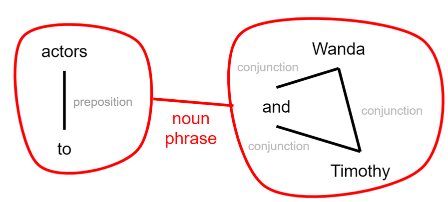
All the implementation details are compressed in Equation 3: For G-MPNN, in addition to equation 3, we have given more implementation details in experimental section 5.2, and appendix section 2.2. Still, as suggested, we will add more details (e.g. pseudocode) in the appendix to ensure that the paper is fully reproducible.

Motivation for typed hyperedges over typed edges: While we agree that typed edges are popular / common knowledge representations in triple stores, typed hyperedges, on the other hand, have their own benefits over typed edges (e.g. more flexible organisation of multi-ary relational facts, more representative than binary relations, etc.) and have been a recent research topic of interest esp. in practice [1,2]. In a separate work [3], it has been shown that a few sentence types such as claims about claims in natural language (e.g. A claimed that B claimed C) can flexibly be represented by recursive hypergraphs with typed hyperedges rather than graphs with typed edges. As suggested, we will expand the introduction section with more motivation.

Is there a specific reason why node types are not permitted directly in this framework?: Given the G-MPNN formulation, and a large body of work on MPNNs with node types, it is quite straightforward to permit node types (without the vehicle of edge types) in a generalised MPNN (e.g. use a separate function for node types).

Statistical tests to validate the significance of gains: Thanks for the suggestion (we will add statistical tests). **R3** has suggested two more recent baselines for MPNN-R (please see the table above). Based on a Welch t-test, the p-value on DBLP is 0.35, and a small p-value on the other three (less than 0.0005) validates the significance.

Figure 1 is insufficiently clear: Thanks for the important comment on improving the understanding of the Figure. Due to space limitations, we have attempted to express a subgroup of words (*to actors Timothy and Wanda*) through typed edges (shown on the right). As suggested, we will include the representation of the example in terms of a graph (with typed edges).



Explain boldface letters in Equation 3, $*$, and $e - v$: The detailed comments are highly appreciated! We will fix them. Yes, $*$ represents element-wise product.

Motivation should be provided for product over sums: This follows directly from knowledge representation literature where multiplicative interactions are preferred to additive interactions (e.g. [4]). Our specific multiplicative choice of G-MPNN also generalises existing formulations shown in lines 147-149.

Definition of recursive hypergraphs could be made clearer: Thanks for the alternative definition. We agree that it has important benefits. A small caveat is that R is already used for MPNN readout, and $\mathcal{R}$ is used for relations (we will see how best to use the alternative with intuitive notations without overloading existing notations).

Depth-0 hyperedges as vertices in HetGNN: We have explaned a general method of modelling a recursive hypergraph with a heterogenous graph. For the specific example of academic networks used in the paper, we agree that both vertices and depth-0 hyperedges represent the same entities with the same type (documents).

Thanks to all the reviewers again for the comments! All typos, concerns on ethics, and other minor concerns will be fixed.

[1] Bahare Fatemi, Perouz Taslakian, David Vazquez, and David Poole. Knowledge hypergraphs: Prediction beyond binary relations. In IJCAI'20.

[2] Jianfeng Wen, Jianxin Li, Yongyi Mao, Shini Chen, and Richong Zhang. On representation, embedding of knowledge beyond binary relations. In IJCAI'16.

[3] Telmo Menezes, Camille Roth. Semantic Hypergraphs, 2019

[4] Embedding Entities and Relations for Learning and Inference in Knowledge Bases, In ICLR'15