Table 1: Quantitative Results Compared with Additional Graph Based Methods **(R3)**

Dataset	H3D				NBA				SDD
Time	1.0s	2.0s	3.0s	4.0s	1.0s	2.0s	3.0s	4.0s	4.8s
STGAT (ICCV 2019)	0.24 / 0.33	0.34 / 0.48	0.46 / 0.77	0.60 / 1.18	0.62 / 1.04	1.26 / 1.91	1.73 / 3.05	2.47 / 3.86	18.8 / 31.3
STGCNN (CVPR 2020)	**0.23** / 0.32	0.36 / 0.52	0.49 / 0.89	0.73 / 1.49	0.68 / 1.12	1.34 / 2.13	1.85 / 3.10	2.56 / 4.03	20.6 / 33.1
EvolveGraph	**0.23 / 0.29**	**0.31 / 0.44**	**0.39 / 0.58**	**0.48 / 0.86**	**0.56 / 0.80**	**0.82 / 1.19**	**1.20 / 1.92**	**1.76 / 3.04**	**13.9 / 22.9**

The authors thank the reviewers for their helpful and valuable feedback. We are encouraged that the reviewers pointed out the strengths of our paper: novelty and significance (R1,R4); strong motivation (R1,R4); correctness and reproducibility (R1,R2,R3,R4); clarity (R1,R3,R4); extensive experiments and analysis (R1,R2,R4); broader impact (R1-R4). We will respond to the feedback from all reviewers below and incorporate all feedback in the revised version.

R1: The authors sincerely express our acknowledgement for your helpful suggestions and affirmation on the novelty and significance of our work. We agree that inferring the graph structure is definitely crucial especially for very large interacting systems with many nodes, and we will add the suggested reference to the main paper. Some comments on the predicted structures on physics system can be found in Fig. 2 and L247-250. Due to the limit of computational resources and time, we were not able to provide comprehensive results on very large systems. However, it will definitely be a meaningful future research direction to apply the proposed framework to large interacting systems such as very complex physics systems. Moreover, we will figure out the most appropriate sub-fields for this paper as you mentioned.

R2: The authors are happy to refine Sec. 3 and 4 if R2 could provide detailed advice, since all the other reviewers think the paper is easy to follow. Some intuitive explanation and empirical evidence can be found in L32-42 and supp L26-39.

R2,R3,R4: **Multi-modality and dynamic relations** The multi-modality is discussed in detail from three aspects in L188-203. In particular, we would like to highlight the multi-modality driven by dynamic relation modeling. Different evolution of relations recurrently leads to more diverse outcomes, which in return establishes the value of dynamic relations in multi-modality. Therefore, we would say that the performance improvements come from both dynamic relations and multi-modality, which are highly correlated. We believe dynamically evolving the interaction graph structures and inherently encouraging diversity and multi-modality, to our best knowledge, is novel and significant for trajectory prediction as mentioned by R1 and R4.

R3: **Comparison with graph-based methods** The authors thank the reviewer for pointing out additional related work. In fact, we have compared EvolveGraph with two graph-based methods in Table 2-4, i.e. NRI (ICML2018) and Trajectron++ (ECCV2020). EvolveGraph can reduce 4.0s ADE/FDE by 40.5%/42.2% on NBA (L292-295), and reduce 4.8s ADE/FDE by 34.4%/36.6% on SDD (L300-301). We also add requested comparison results of another two graph-based methods STGAT (ICCV2019) and Social-STGCNN (CVPR2020) in **Table 1 in the rebuttal**. Compared to STGAT, EvolveGraph can reduce 4.0s ADE/FDE by 20.0%/27.1% on H3D, 28.7%/21.2% on NBA, and reduce 4.8s ADE/FDE by 26.1%/26.8% on SDD. We will add these results and discussion in the revised version.

Dynamic interaction graph The authors would like to clarify that both DG (single stage) and DG (double stage) have dynamic interaction graphs, while SG only has static interaction graph (supp L128-142). The difference between DG (single stage) and DG (double stage) mainly lies in the training procedures. Detailed analysis of ablative results was provided in supp L4-39 and we highlight some facts and statements here. In Table 2-4, compared to SG, DG (double stage) can reduce 4.0s ADE/FDE by 21.3%/24.6% on H3D, 31.3%/33.3% on NBA, and reduce 4.8s ADE/FDE by 32.5%/37.1% on SDD. The prediction horizons are standard in trajectory prediction tasks and we believe empirical results on all datasets already validate the significant superiority of dynamic graph.

R4: **Uncover dynamic relations** We appreciate the suggestion of additional synthetic experiments. (a) With "out of sync", we observed that smaller re-encoding gap is beneficial to evaluate the relation changes. Although there is a trade-off between the performance and computational cost, prediction error decreased as re-encoding gap becomes smaller. (b) Due to the time limit, we were not able to conduct other experiments. However, we believe that our three real-world prediction tasks can be used to evaluate the model performance with multiple aperiodic relation changes to some extent. For example, in NBA dataset the relations among players change frequently and aperiodically due to intensive cooperations and competitions. Our significant improvements in accuracy demonstrate the efficacy of dynamic relation modeling. **Attention coefficients** The authors would like to clarify that the attention coefficient α_{ij} is not a simple learnable constant, but it is computed based on the features of involved agents i and j similar to GAT [A]. As the encoding horizon slides to future steps, the coefficients are re-computed at each time. Therefore, these coefficients are not data-agnostic and they can figure out relative significance of impacts during each re-encoding process. We will make it clear in the revised version. [A] Veličković, et al. "Graph Attention Networks." ICLR, 2018.

Modes and dynamic relations The authors would like to clarify that we do not either pre-define a fixed number of modes, or explicitly differentiate multiple modes. Instead, at each time our model outputs Gaussian mixture distributions and the predicted trajectories are sampled from these distributions. Together with interaction graph evolution, this process may naturally result in multiple modes without having an explicit mode selection process. Please also refer to L14-19 above for more explanations. **Loss function** The camera-ready version allows an additional content page and we will move the loss function to the main paper. **Prior** Since the underlying relations in real-world heterogeneous interacting systems are complex and diverse, it is not easy to select a proper prior of dynamic relations for specific tasks. Some prior with preference on "non-edge" may encourage sparsity for easier visualization or interpretability. Exploring prior incorporation and its detailed impact will be our future work.