This paper proposes a novel pre-training method for DGNN - the Contrastive Pre-training DGNN Algorithm (CPDG). Taking into account the Pre-training+Fine-tuning method that has been deeply developed in NLP, it is applied to the DGNN scenario. The algorithm in this paper is a further optimization on the TGN algorithm, which solves the problem that DGNN cannot be efficiently applied to the actual industrial recommendation system, and has great practical value.

S1: The memory module of TGN can capture long-term information of nodes, but for DGNN scenarios, short-term information is also very important. Therefore, this paper proposes two sampling methods, sampling two subgraphs with temporal and structure respectively, at the same time, designing two opposite sampling strategies in each sampling method as contrastive training, capturing short-term node information for Temporal Contrast (TC) and Structural Contrast (SC). Finally, in the fine-tuning stage(EIE), the memory module is combined to make the algorithm capable of mining the long-short term information. This method is very innovative.

S2: The experiments were very rich. The performance of the algorithm was tested under different settings and different datasets. After that, the experiments were set up to test the pre-training method proposed in this paper is batter than the algorithm using the ordinary pre-training method without the pre-training method. The comparison further illustrates the generalization ability of this method. Finally, experiments are set up to illustrate the effectiveness of the SC, TC and EIE proposed in this article.

W1: In TABLE VII，there is a lack of column comparison of the three methods of Time Transfer, Field Transfer and Time+Field Transfer under the same dataset, and recommending to readers the best method under a certain type of data set. Because it is impossible to implement three settings at the same time in actual application scenarios.

W2: In the data of TABLE IX，the performance of CPDG on the MOOC dataset is lower than that of TGN. The explanation of the paper is that the structure and temporal patterns are not obvious in the MOOC dataset as in other datasets. I would like to ask what is the standard of this “obvious” and whether a clear test can be provided.

千阁师哥，这篇论文的第15页中SC那块的负样本是使用不等于node i 的顶点做随机游走生成的，我没看懂为什么这么做，有什么好处