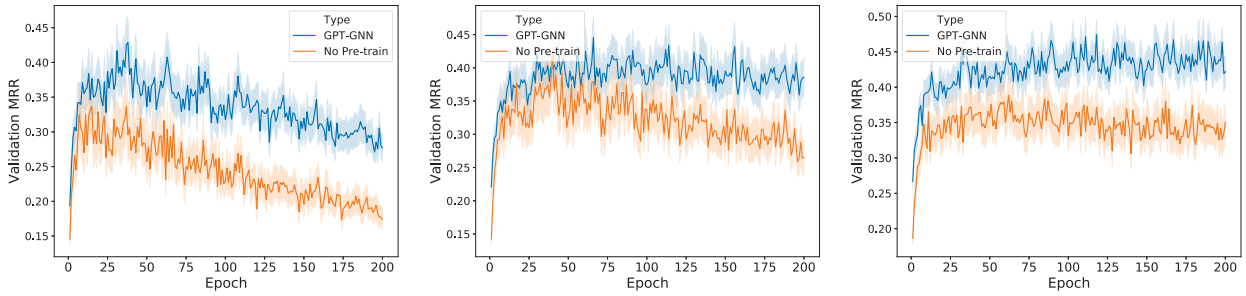
Predict Paper Title	Groundtruth Paper Title
person recognition system using automatic probabilistic classification	person re-identification by probabilistic relative distance comparison
a novel framework using spectrum sensing in wireless systems	a secure collaborative spectrum sensing strategy in cyber physical systems
a efficient evaluation of a distributed data storage service storage	an empirical analysis of a large scale mobile cloud storage service
parameter control in wireless sensor networks networks networks	optimal parameter estimation under controlled communication over sensor networks
a experimental system for for to the analysis of graphics	an interactive computer graphics approach to surface representation

Table 4: Generated paper title samples. The left column is generated by GPT-GNN, and the right column is the groundtruth.



(a) Data Percentage: 10% (b) Data Percentage: 20% (c) Data Percentage: 50%

Figure 4: Fine-tuning convergence comparison of GPT-GNN with no-pretrain, under different training data percentage.

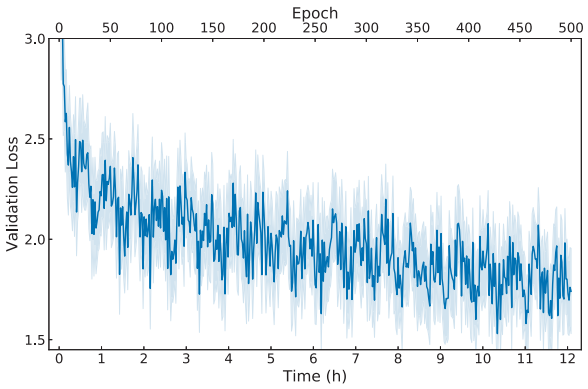


Figure 5: The pre-training convergence curve of GPT-GNN on OAG. It took about 10 hours (400 epochs) to converge.

instead of just finding a trivial solution very fast. This to some extent shows that the generative pre-training task is hard enough and can thus can guide the model to really capture the intrinsic structure of the graph data. It took about 12 hours for GPT-GNN to converge. For downstream tasks, we show the convergence curve utilizing our GPT-GNN with no-pretrain, with different data percentage. As is illustrated in Figure 4, GPT-GNN can always get a more generalized model than no-pretrain, and is more robust to over-fitting since a good initialization from pre-training.

D PAPER TITLE GENERATION EXAMPLES

For OAG, since our attribute generation task is oriented on the paper title, we'd like to see how well our GPT-GNN can learn to generate the title. The results are shown in table 4. We can see that the model can capture the main meaning of each paper to be predicted, only by looking at partial neighborhoods (note that we use Attribute Generation Node for this task, which replace the input attribute as a share vector). For example, for the first sentence, our model successfully predict the key words for this paper, including 'person recognition', 'probabilistic', etc. This shows that the graph itself contains rich semantic information, and explains why a pre-trained model can generalize well to downstream tasks.