1 Supplemental Material

1.1 Input node embedding optimization

As discussed in the paper, we observed that overfitting happens in RGCN, when optimizing the input node embeddings with the same learning rate as the model parameters, especially with the decrease in sample size and batch size. This problem becomes more severe if the sample size is low. As a result, the training accuracy goes up to 100%, and the validation loss starts to increase after a few epochs resulting in a decrease in the validation accuracy. Figure 1 shows the validation accuracy trend of ReWise-LDRN for `amplus` with two different learning rates for the node embeddings. This behavior can be because the number of neighbors for each node decreases with a lower sample size. Consequently, the network overfits the node embeddings instead of learning from the neighbors. Therefore, we chose to optimize the node embeddings with a lower learning rate than the model parameters.

1.2 Unbiased Estimation

In this section We evaluated the effect of multiplying the unbiased estimation term $1/q(v_j|\mathcal{V})$ mentioned in Eq. (4) to the message-passing equation on the test accuracy. Table 1 shows the result of this addition and the comparison with original ReWise-LDRN (without the unbiased term). As the results indicate, although the unbiased estimation term is a correct estimation of the case with no sampling, it slightly reduces the test accuracy. We, therefore, eliminated this term in our experiments.

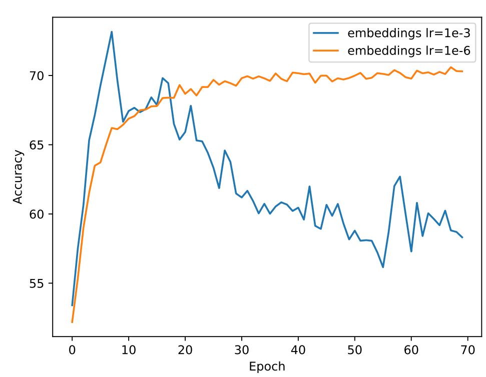


Figure 1: The effect of node embeddings' learning rate on the performance of ReWise-LDRN for `amplus` for the batch size of 64 and sample size of 16 per relation. The model exhibits overfitting when using an embedding learning rate of $1e-3$, but reducing the learning rate to $1e-6$ resolves the problem.

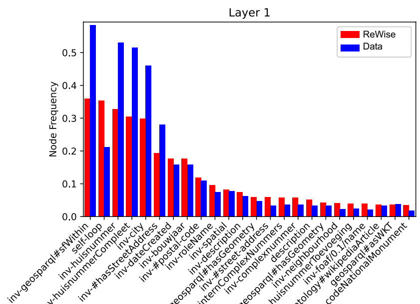
Table 1: The effect of unbiased estimation on the accuracy of RGCN with ReWise-LDRN. The sampling rate indicates the number of nodes being sampled in each relation. The results are the average of five repeats with standard deviation.

Dataset	Method	Batch Size	Sampling Rate	Accuracy (%)
`amplus`	ReWise-LDRN (unbiased)	2048	-	76.95 ± 0.86
	ReWise-LDRN (original)	2048	2048	77.59 ± 1.35
`dmgfull`	ReWise-LDRN (unbiased)	2048	-	71.00 ± 0.26
	ReWise-LDRN (original)	2048	2048	71.35 ± 0.32
`mdgenre`	ReWise-LDRN (unbiased)	512	-	63.71 ± 0.90
	ReWise-LDRN (original)	512	512	63.79 ± 0.83

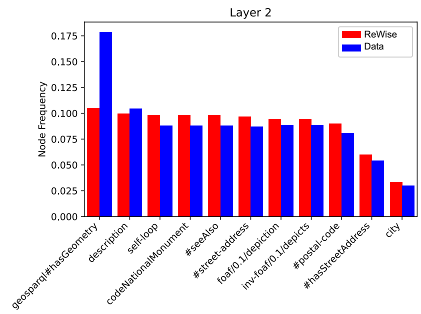
1.3 Semantic analysis of `dmfgull` and `mdgenre`

Figure 2 shows the result of semantic analysis of `dmgfull`. As Figures 2a and 2b show, similar to `amplus` ReWise-LDRN reduces the node frequency of the highly appearing relations, and therefore, the ratio of the less appearing relations increases. For instance, the relation `geosparql#sfWithin` connects monuments to their corresponding municipality. As the number of monuments is more than the number of municipalities, the inverse of this relation which connects the municipalities to the monuments has more different objects and appears to have the greatest number of objects in layer 1. Moreover, Figures 2c and 2d indicate the match between the top k relations in the RGCN and ReWise-LDRN in the two layers, as the blue line is above the orange line, which compares a random ordering of the relations with RGCN. In layer 1, the RGCN has the highest weight norm for the relation `rnaSubject` where ReWise-LDRN samples the most nodes from `inv-geosparql#sfWithin`, and in layer 2, the RGCN has `location` as the most important relation and ReWise-LDRN `geosparql#hasGeometry`.

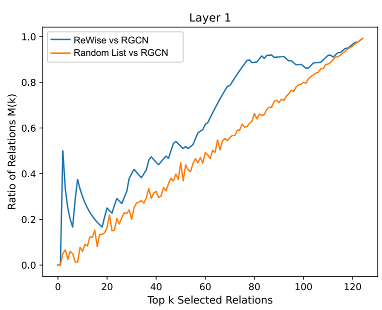
Figure 3 shows the semantic analysis of the relations for `mdgenre`. Similar to the other two datasets, the top relations for ReWise-LDRN matches mostly with RGCN. However, this is more visible in layer 2 as the blue line has more distance from the random list indicated by orange. For `mdgenre`, the most important relation for the RGCN in layer 1 is `nominated for`, where for ReWise-LDRN is `inv-country of citizenship`. in layer 2, the RGCN find `work period (start)` the most important relation and the ReWise-LDRN samples the most nodes from `cast member`.



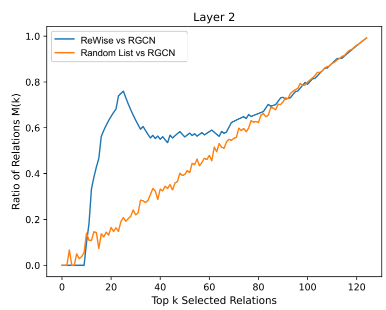
(a)



(b)

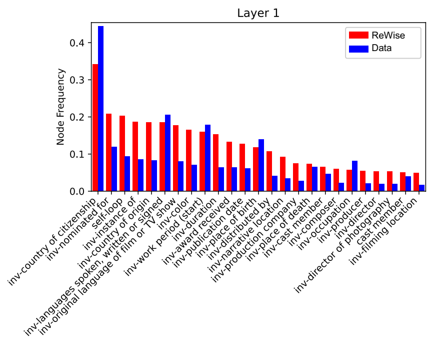


(c)

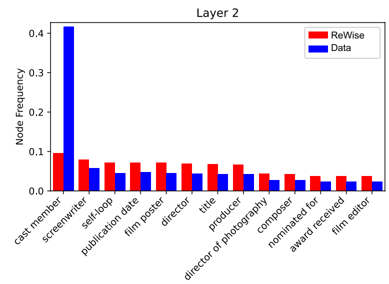


(d)

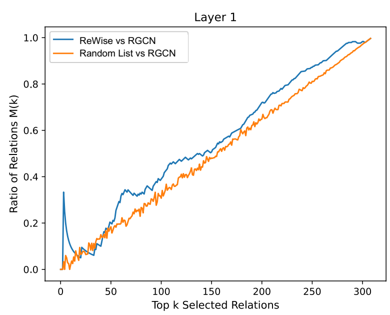
Figure 2: Semantic analysis of the relations in `dmgfull` dataset with the batch size and sampling rate of of 2048. The RGCN in graph (c) and (d) is trained in full batch. 'inv' in the start of the relation label indicates the inverse of that relation. For readability, we only display the top 24 relations for layer 1.



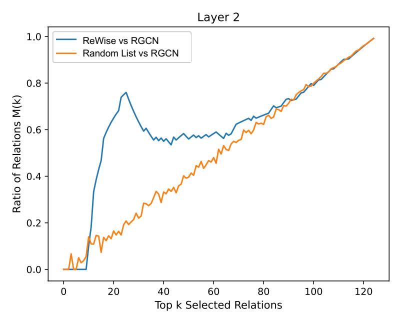
(a)



(b)



(c)



(d)

Figure 3: Semantic analysis of the relations in `mdgenre` dataset with the batch size and sampling rate of of 512. The RGCN in graph (c) and (d) is trained in full batch. ‘inv’ in the start of the relation label indicates the inverse of that relation. For readability, we only display the top 25 relations for layer 1.