Supplementary Material: Towards Understanding the Geometry of Knowledge Graph Embeddings

Chandrahas
Indian Institute of Science
chandrahas@iisc.ac.in

Aditya Sharma
Indian Institute of Science
adityasharma@iisc.ac.in

Partha Talukdar
Indian Institute of Science
ppt@iisc.ac.in

1 Hyperparameters

For experimenting with different models, we use the best reported values for hyper-parameters. We have listed these hyper-parameters and their values in Table 1.

Params	Additive Models		
	TransE	**TransR**	**STransE**
Rate	0.01	0.001	0.0001
Norm	L^1	L^1	L^1
#Epochs	1000	1000	1000
Loss	Pair-Loss	Pair-Loss	Pair-Loss
Margin	1	1	1
	Multiplicative Models		
	DistMult	**HolE**	**ComplEx**
L^2 **Reg.**	0.01	0	0.01
Rate	0.5	0.1	0.5
#Epochs	1000	500	1000
#Batches	100	100	100
Opt. algo.	AdaGrad	SGD	AdaGrad
Loss	Log-Loss	Pair-Loss	Log-Loss
Margin	-	0.2	-

Table 1: Hyper-parameter for all methods. Here Log-Loss and Pair-Loss refer to logistic loss and pairwise ranking loss respectively.

2 Relation Vector Analysis for FB15k

In this section, we present the analysis of the relation vectors for FB15k with respect to number of negative samples and dimensions. They show very similar behaviour as entity vectors and described in following sections.

2.1 Effect of Number of Negative Samples

Figure 1 (left) shows how the conicity of relation vectors varies with number of negative samples used. As observed in case of entity vectors, the conicity for relation vectors is invariant for additive models. However, the multiplicative models are sensitive to number of negative samples and the conicity of relation vectors show a small drop while increasing number of negative samples.[1] Similarly, as shown in Figure 1(right), average vector length of relation vectors is not sensitive to the number of negative samples for additive models. Except HolE, average vector length is invariant to the number of negative samples for multiplicative models.

2.2 Effect of Vector Dimension on Geometry

Figure 2 demonstrates the effect of vector dimensions on the conicity (left) and average vector length (right) of relation vectors. As we can see from the figure, relation vectors behave very similar to entity vectors. The conicity of vectors generated from additive methods is almost invariant to increase in vector dimension. In contrast, multiplicative models show a decreasing pattern with increase in vector dimesion.

Similar to conicity, the average vector length of relation vectors generated from additive models is almost invariant to increase in vector dimension, except for STransE which shows a small increase for 200 dimension. However, average vector length of relation vectors generated from multiplicative models show a clear increasing pattern with increasing vector dimension.

[1]Please note that all of these methods use negative sampling only for entities, not relations.

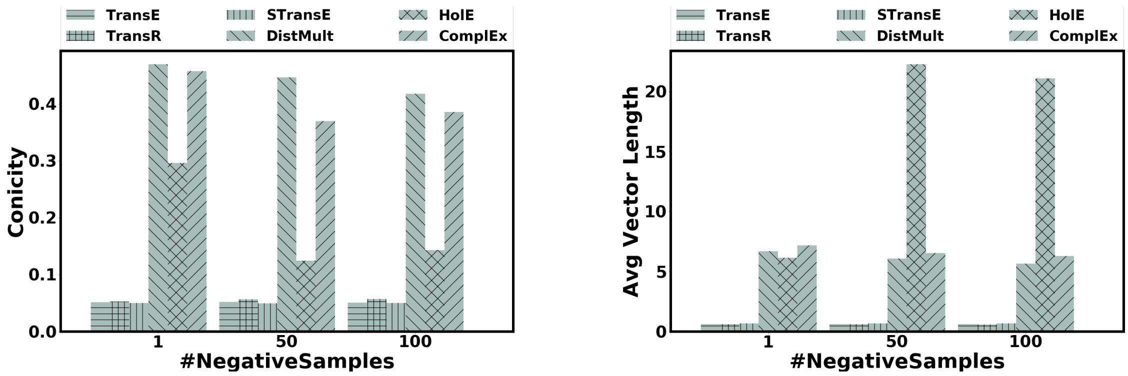


Figure 1: Conicity (left) and Average Vector Length (right) vs number of negative samples for relation vectors learned using various KG embedding methods on FB15k dataset. In each bar group, first three models are additive, while the last three are multiplicative.

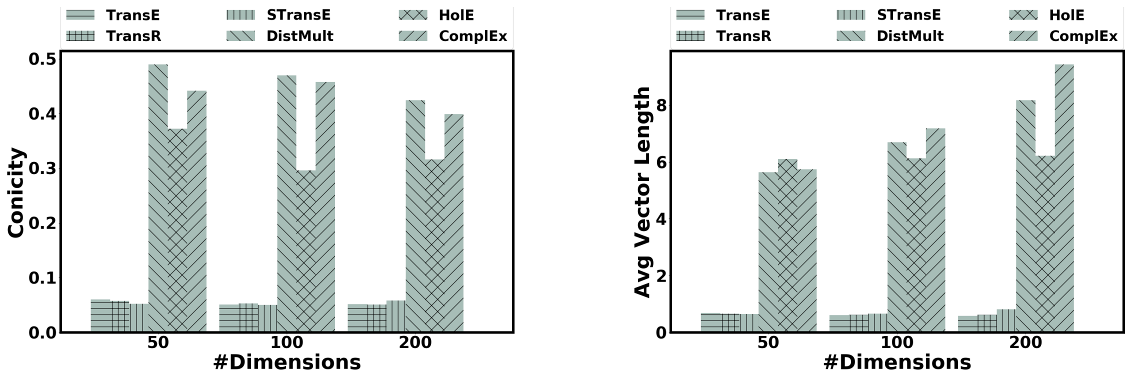


Figure 2: Conicity (left) and Average Vector Length (right) vs number of dimensions for relation vectors learned using various KG embedding methods on FB15k dataset. In each bar group, first three models are additive, while the last three are multiplicative.

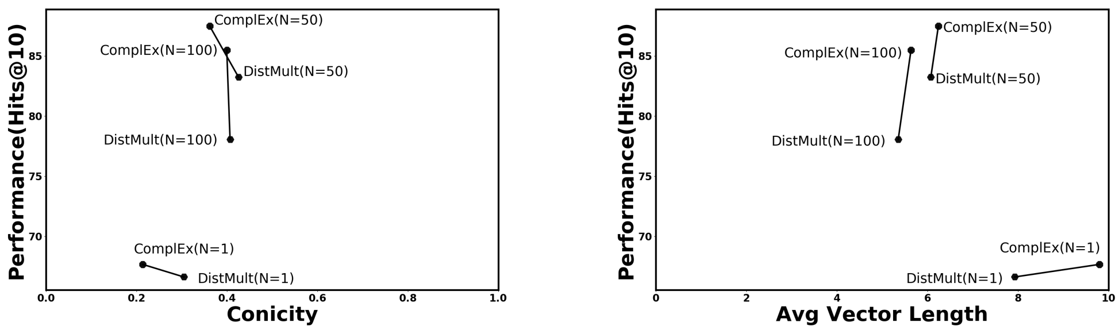


Figure 3: Relationship between Performance (HITS@10) on a link prediction task vs Conicity (left) and Avg. Vector Length (right) on FB15k dataset. For each point, N represents the number of negative samples used. Models with same number of negative samples are connected by line segment. This demonstrates that model performance has negative correlation with Conicity while positive correlation with average vector length for fixed number of negatives. Main findings are summarized in Section 3.

3 Correlating Geometry with Performance for multiplicative models

In this section, we present a subset of results already presented in Section 6.4 of main paper focusing on multiplicative models.[2] Figure 3 shows the correlation between model performance and

[2]We have excluded HolE for clarity of the plots. The correlation between performance and conicity holds for HolE as well.

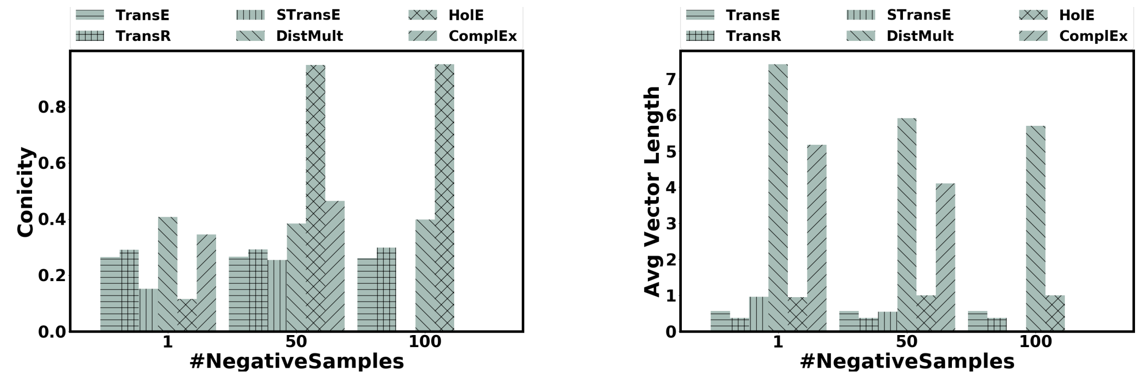


Figure 4: Conicity (left) and Average Vector Length (right) vs number of negative samples for entity vectors learned using various KG embedding methods on WN18 dataset. In each bar group, first three models are additive, while the last three are multiplicative.

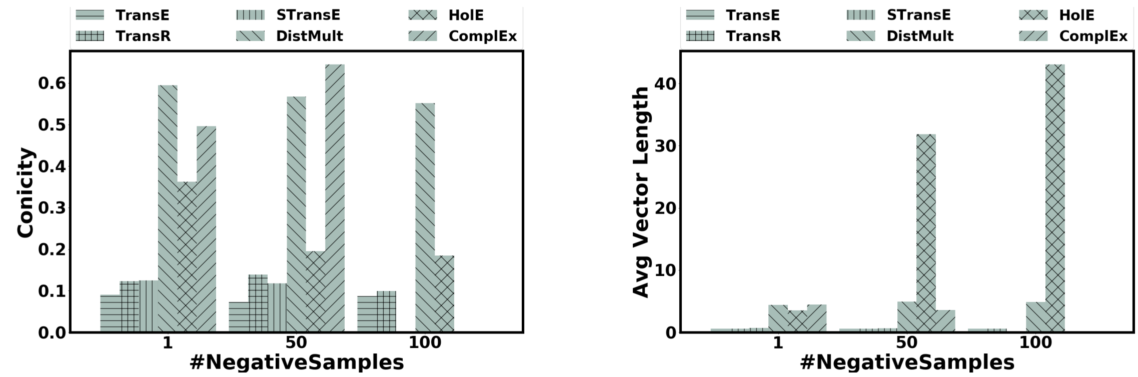


Figure 5: Conicity (left) and Average Vector Length (right) vs number of negative samples for relation vectors learned using various KG embedding methods on WN18 dataset. In each bar group, first three models are additive, while the last three are multiplicative.

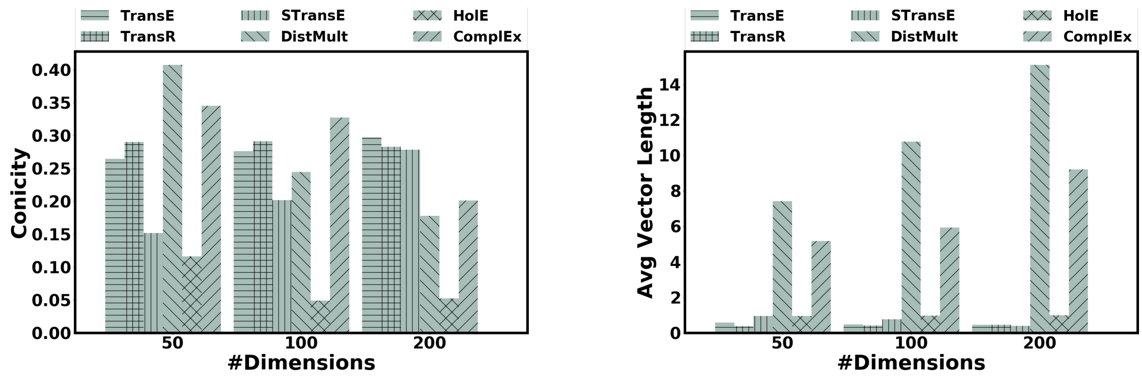


Figure 6: Conicity (left) and Average Vector Length (right) vs number of dimensions for entity vectors learned using various KG embedding methods on WN18 dataset. In each bar group, first three models are additive, while the last three are multiplicative.

geometry. Models which use same number of negative samples are connected with a line segment. In the left figure, the line segments have negative gradients. This suggests a negative correlation between model performance and Conicity for fixed number of negatives. In contrast, the line segments in the right figure have positive gradients which suggests a positive correlation between model performance and average vector length for fixed number of negatives. In both the cases, the magnitudes of the gradients of the line segments are larger for higher number of negative samples. This suggests that performance gain is more sensitive to geometry for higher number of negative samples.

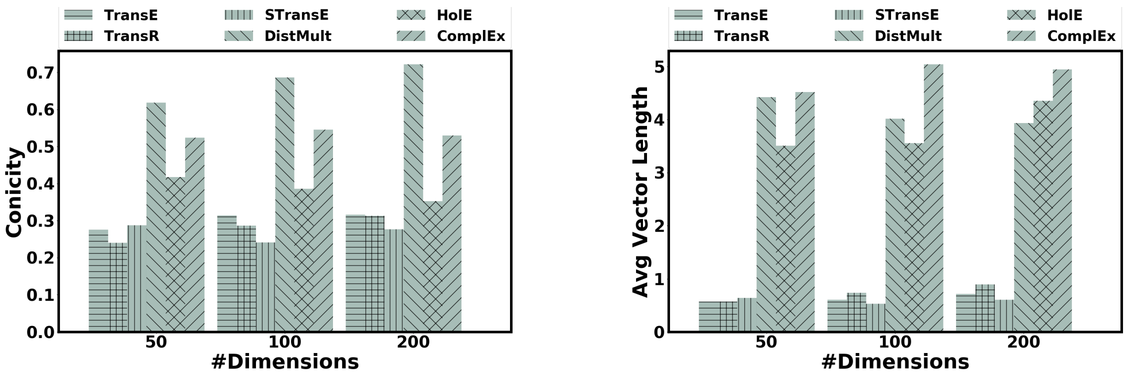


Figure 7: Conicity (left) and Average Vector Length (right) vs number of dimensions for relation vectors learned using various KG embedding methods on WN18 dataset. In each bar group, first three models are additive, while the last three are multiplicative.

4 Analysis for WN18

In this section, we present the analysis of entity and relation embeddings for WN18 dataset. The observations are very similar to FB15k dataset and described in following sections.

4.1 Effect of Number of Negative Samples on Geometry

For experiments in this section, we keep the vector dimesion constant at 50.

Entity Embeddings: The effect of number of negative samples on entity vector conicity (left) and average vector length (right) is shown in Figure 4. As seen from these figures, the conicity of entity vectors increases while average vector length decreases with number of negative samples for multiplicative models. Additive models, however, are not affected by increasing number of negative samples. These observations are in agreement with observations from FB15k dataset.

Relation Embeddings: As seen from Figure 5 (left), the conicity of relation vectors is invariant to number of negative samples for additive models. Except ComplEx, the conicity of relation vectors generated from multiplicative models show a decreasing pattern with increasing number of negative samples. Similar to FB15k, the average vector length of relation vectors is invariant to number of negative samples for both set of methods except HolE. HolE shows increase in average vector length with increasing number of negative samples, which is again consistent with FB15k.

For this analysis, we have skipped STransE and ComplEx models with 100 negative samples as these models reached system memory limit.

4.2 Effect of Vector Dimension on Geometry

For analysis in this section, we fixed the number of negative samples to 1.

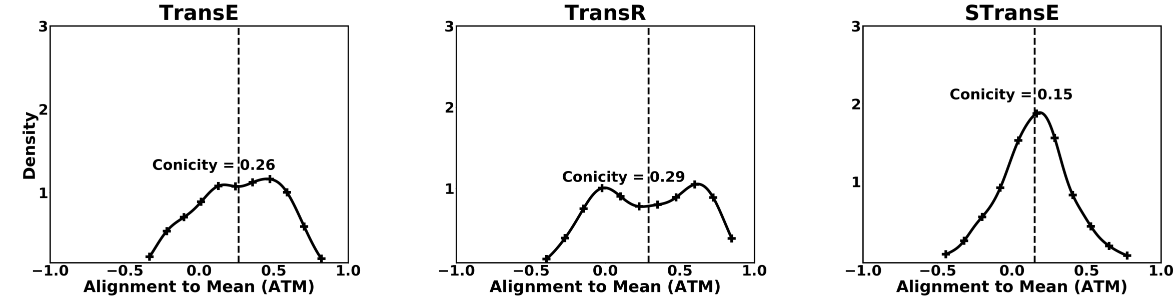
Entity Embeddings: The effect of increasing vector dimension on conicity (left) and average vector length (right) is shown on Figure 6. Similar to FB15k, the conicity of entity vectors show a decreasing pattern for multiplicative models. Unlike TransR, TransE and STransE show an increasing pattern. In case of average vector length, all additive models are invariant to increase in vector dimension. Except for HolE, average vector length of entity vectors increase with vector dimension for multiplicative models.

Relation Embeddings: Figure 7 shows the effect of increasing vector dimension on conicity (left) and average vector length (right). We do not see any discernible pattern here. The conicity increases slightly with vector dimensions for TransE, TransR and DistMult while it decreases for HolE. In case of average vector length, a slight increase is observed for TransE, TransR and HolE while DistMult shows a decreasing pattern.

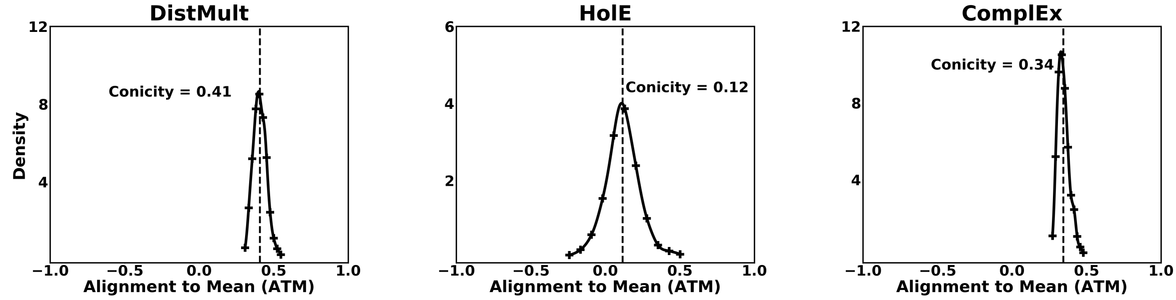
4.3 Effect of Model Type on Geometry

Entity Embeddings: Figure 8 shows the distribution of ATM for entity vectors. Additive models (top row) exhibit high vector spread while multiplicative models (bottom row) show low vector spread. Except HolE, multiplicative models have higher conicity than additive models. This reinforces our observation that the vectors generated from multiplicative models tend to lie inside a narrow cone.

Relation Embeddings: Similar behaviour is observed for relation vectors in Figure 9 where all

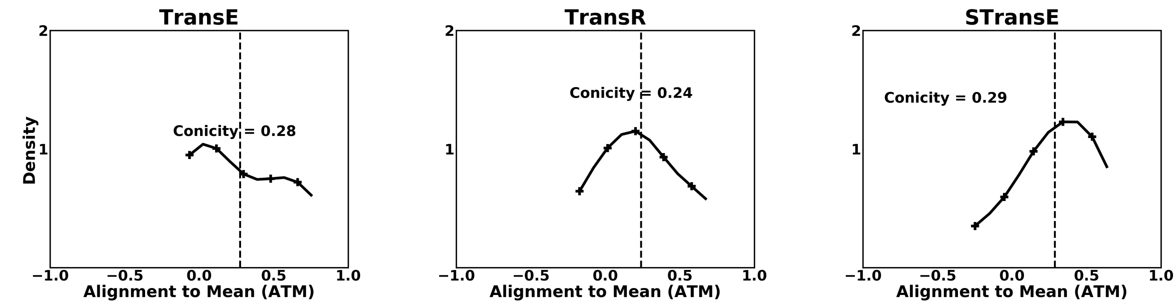


(a) Additive Models

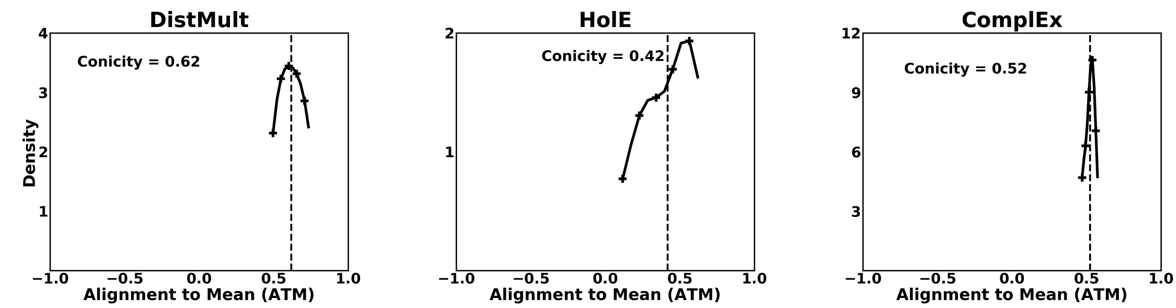


(b) Multiplicative Models

Figure 8: ATM vs Density plots for entity embeddings learned by various additive (top row) and multiplicative (bottom row) KG embedding methods on WN18 dataset. For each method, a plot averaged across entity frequency bins is shown.



(a) Additive Models



(b) Multiplicative Models

Figure 9: ATM vs Density plots for relation embeddings learned by various additive (top row) and multiplicative (bottom row) KG embedding methods on WN18 dataset. For each method, a plot averaged across relation frequency bins is shown.

the multiplicative models have higher conicity and lower vector spread than additive models. Also, these observations are consistent with observations on FB15k dataset.