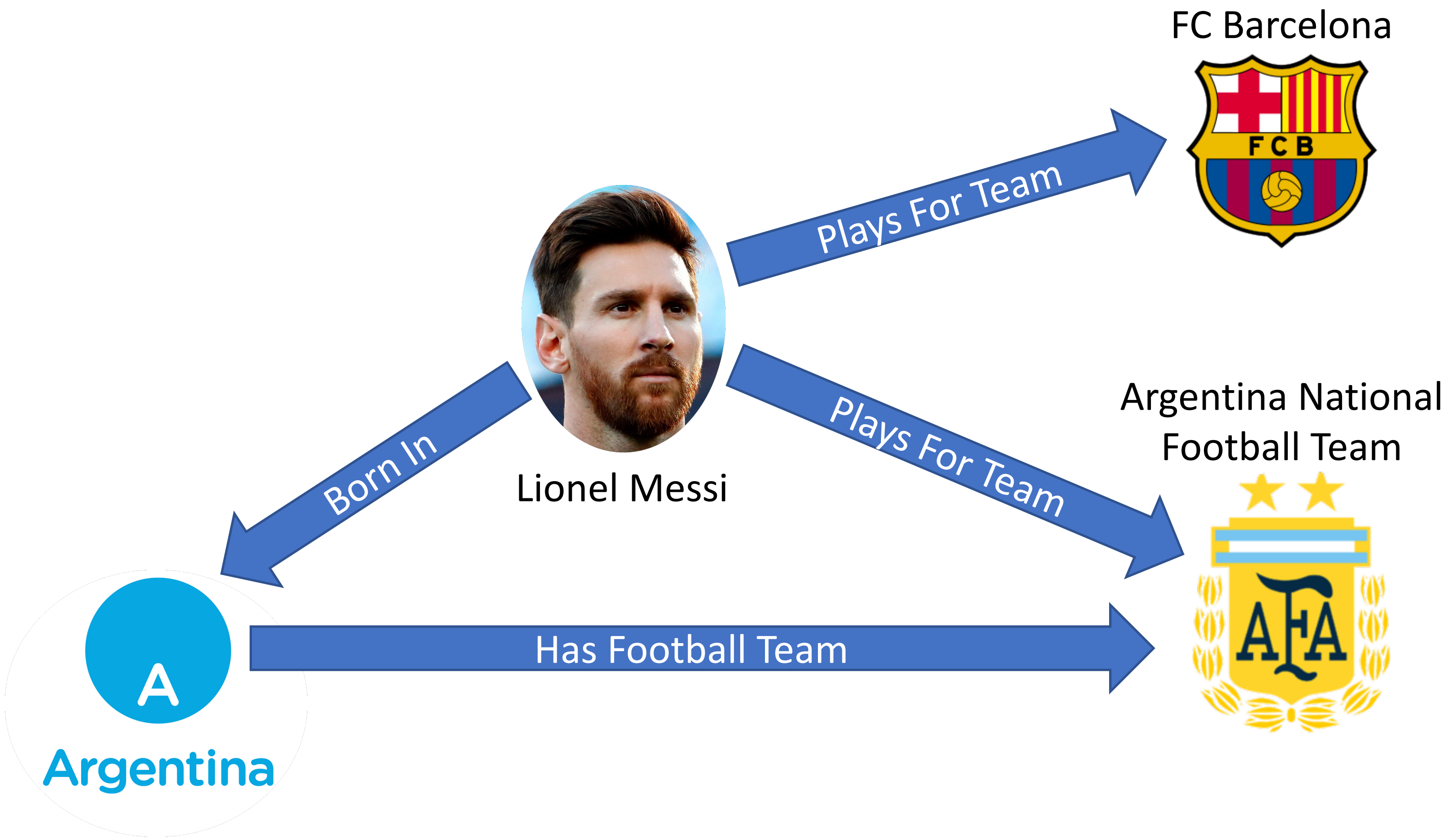
Towards Understanding the Geometry of Knowledge Graph Embeddings

Chandrahas[1], Aditya Sharma[2], Partha Talukdar[1,2]

[1]Computer Science and Automation
[2]Computatational and Data Sciences
Indian Institute of Science, Bangalore

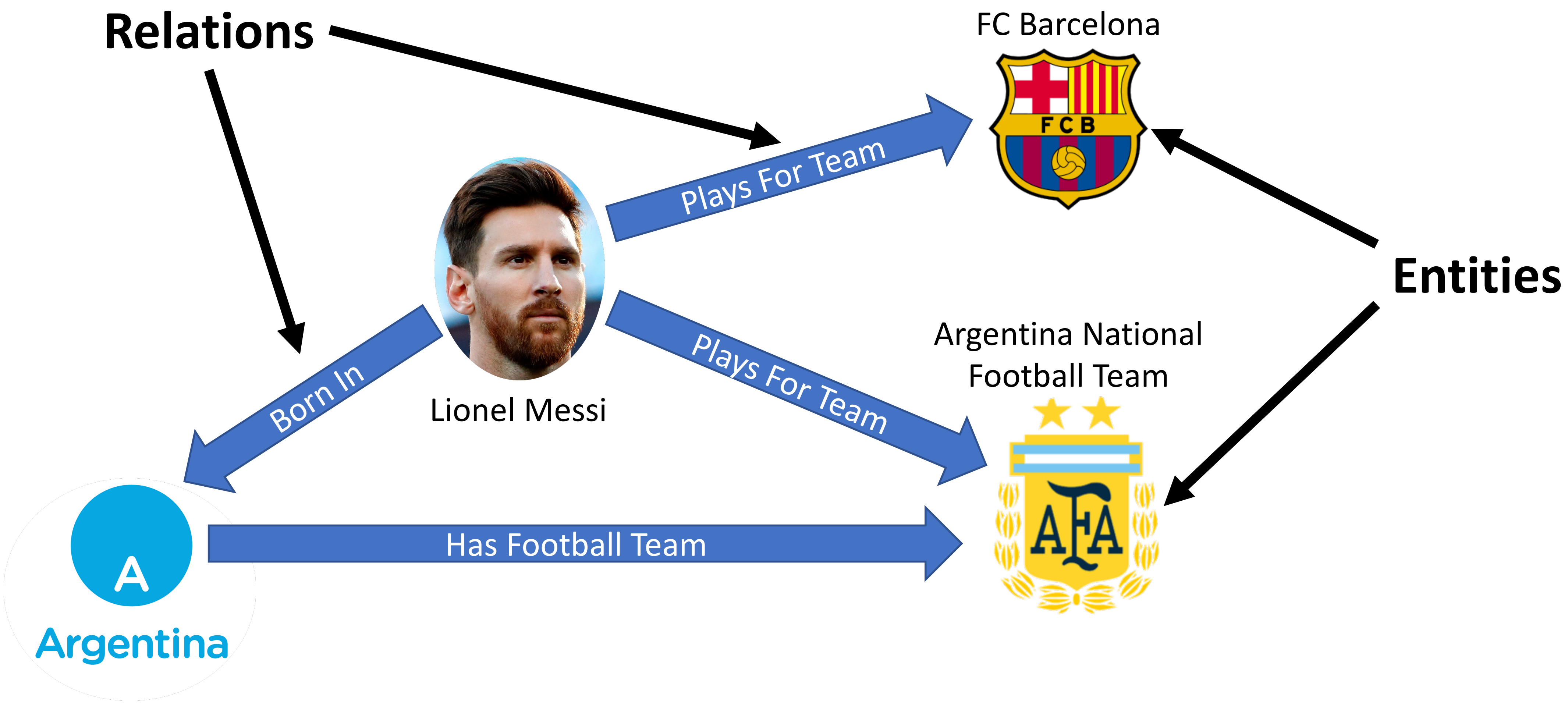
Department of Computational and Data Sciences

Knowledge Graphs (KG)



Knowledge Graphs (KG)

Knowledge Graphs (KG)



Knowledge Graphs (KG)

- Example KGs

Freebase™

yago
select knowledge

DBpedia

Knowledge Graphs (KG)

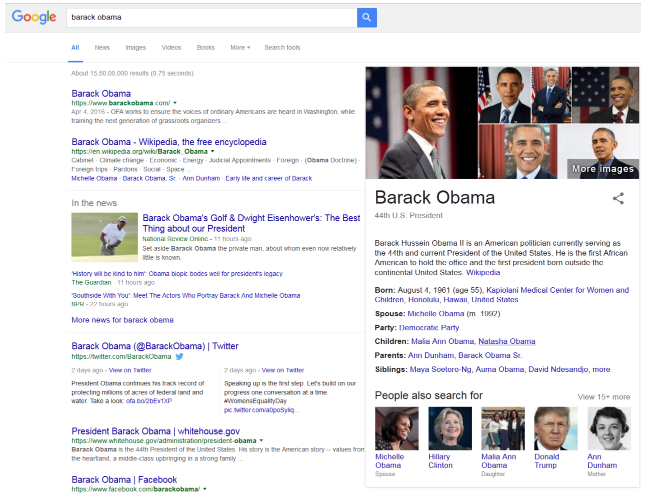
- Example KGs

Freebase™

yago select knowledge

DBpedia

- Applications
  - Search



Knowledge Graphs (KG)

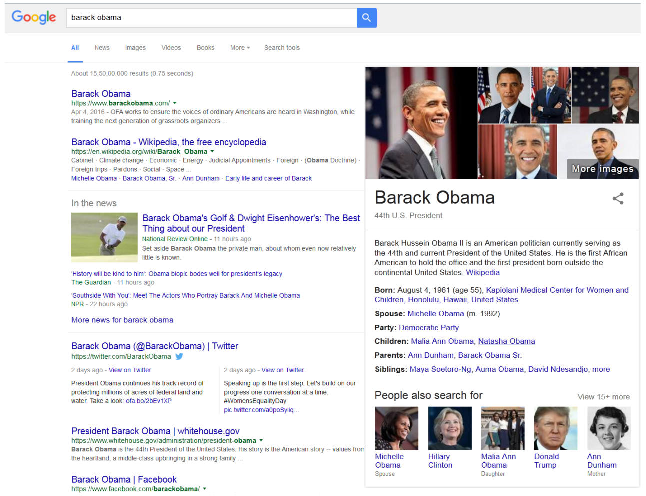
- Example KGs

Freebase™

yago
select knowledge

DBpedia

- Applications
  - Search

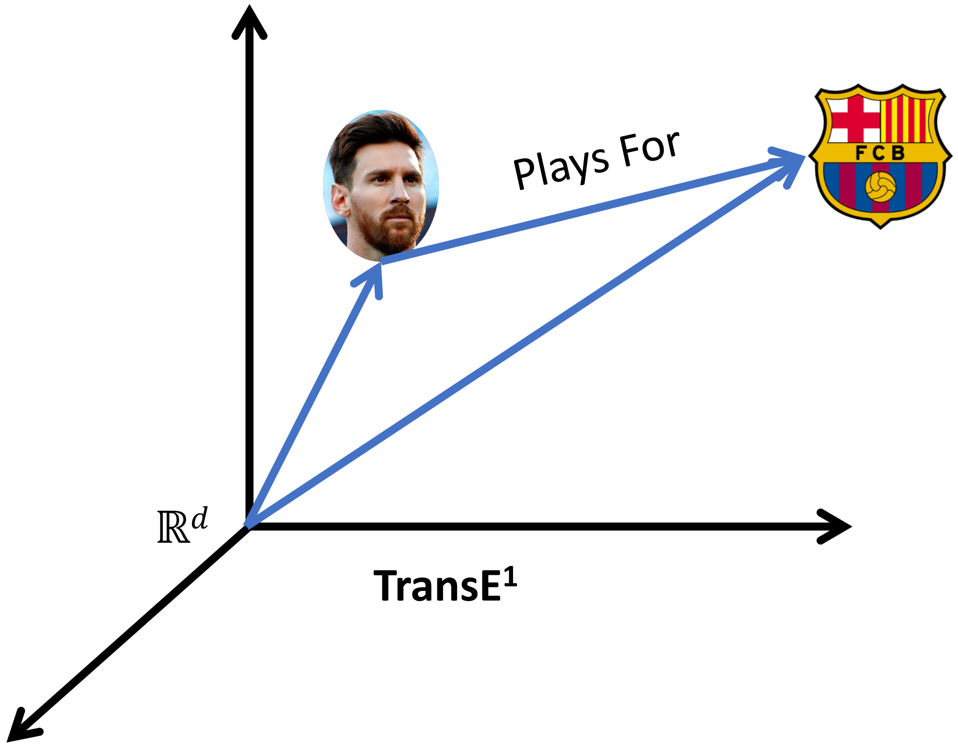


  - Question Answering

IBM Watson

KG Embeddings

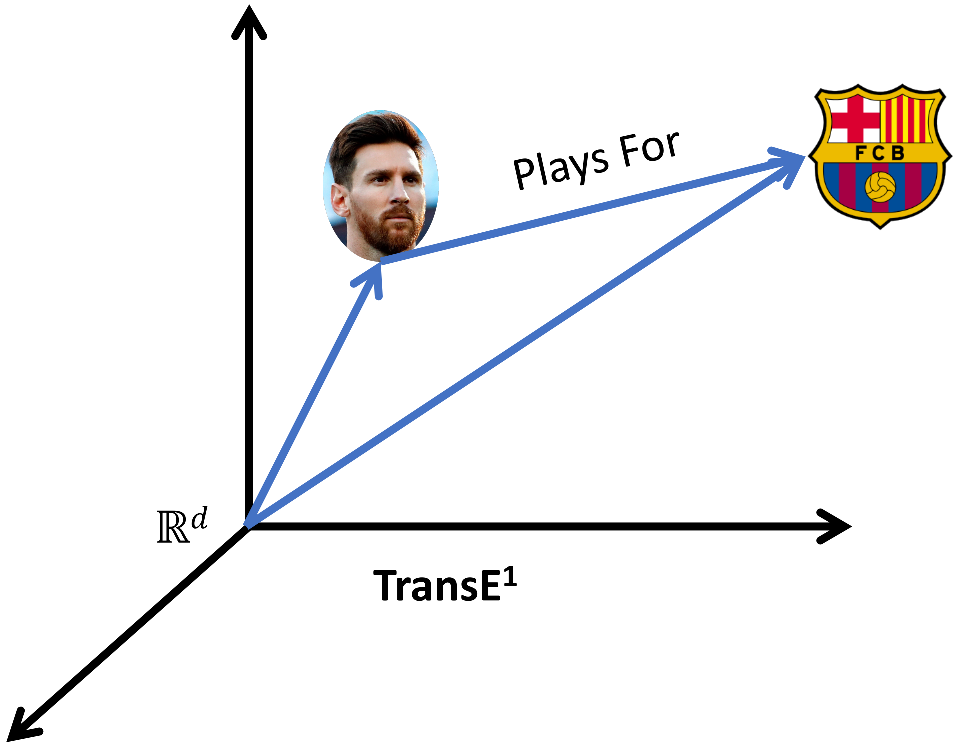
- Represents entities and relations as vectors in a vector space

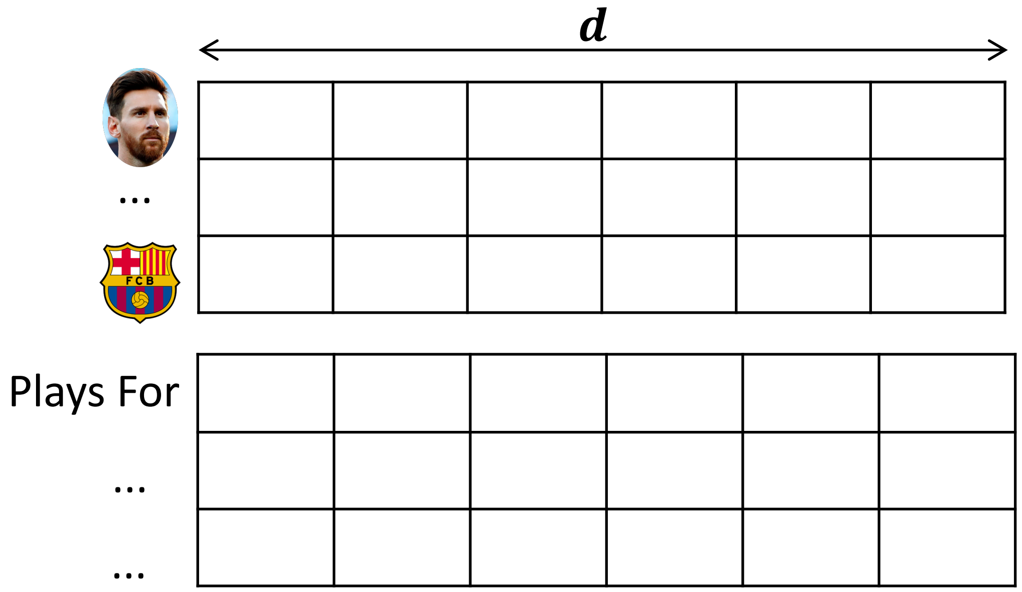


1. Translating Embeddings for Modeling Multi-relational Data, Bordes et al.

KG Embeddings

- Represents entities and relations as vectors in a vector space

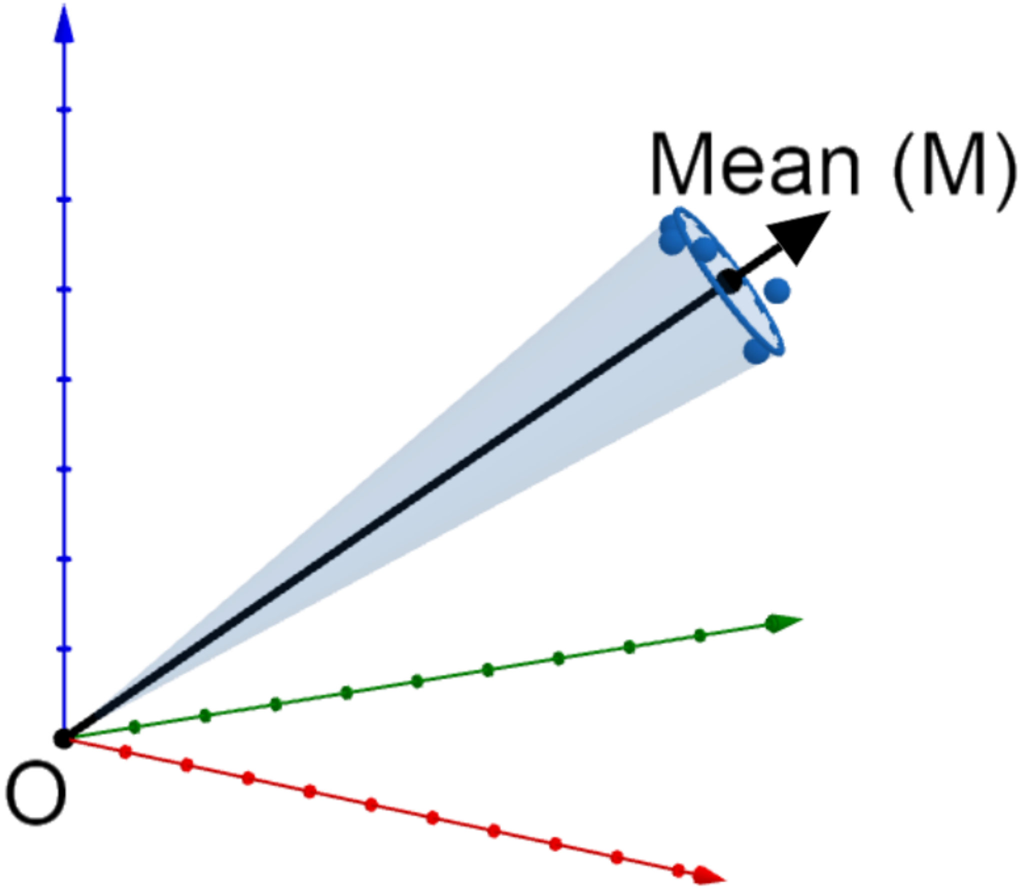


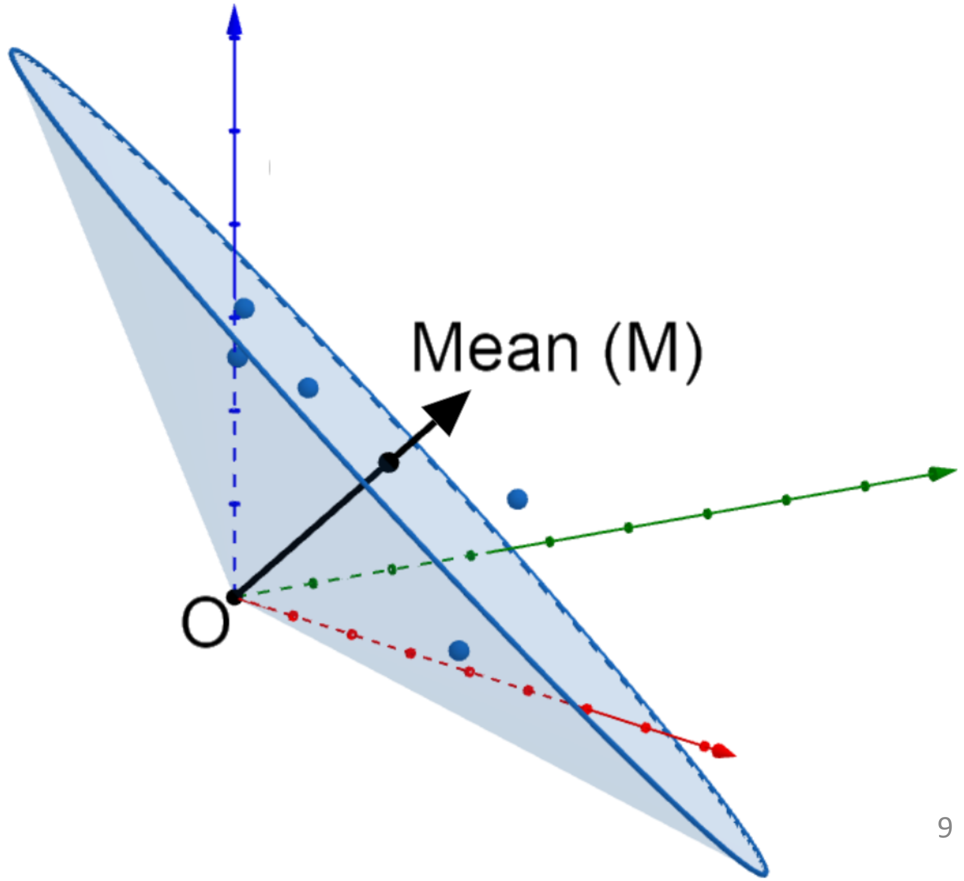


1. Translating Embeddings for Modeling Multi-relational Data, Bordes et al. NIPS 2013.

Geometry of Embeddings

- Arrangement of vectors in the vector space.





Geometry of Embeddings

- A recent work by (Mimno and Thompson, 2017)[1] presented an analysis of the geometry of word embeddings and revealed interesting results.

- However, geometrical understanding of KG embeddings is very limited, despite their popularity.

1. The strange geometry of skip-gram with negative sampling, Mimno and Thompson, EMNLP 2017

Problem

- Study the geometrical behavior of KG embeddings learnt by different methods.
- Study the effect of various hyper-parameters used during training on the geometry of KG embeddings.
- Study the correlation between the geometry and performance of KG embeddings.

KG Embedding Methods

- Learns d-dimensional vectors for entities $\boldsymbol{\mathcal{E}}$ and relations $\boldsymbol{\mathcal{R}}$ in a KG.

KG Embedding Methods

- Learns d-dimensional vectors for entities $\boldsymbol{\mathcal{E}}$ and relations $\boldsymbol{\mathcal{R}}$ in a KG.

- A score function $\boldsymbol{\sigma} : \mathcal{E}\times\mathcal{R}\times\mathcal{E}\rightarrow\mathbb{R}$ distinguishes correct triples T^{+} from incorrect triples T^{-}. For example,
  $\boldsymbol{\sigma}($Messi, plays-for-team, Barcelona$)$ > $\boldsymbol{\sigma}($Messi, plays-for-team, Liverpool$)$

KG Embedding Methods

- Learns d-dimensional vectors for entities $\boldsymbol{\mathcal{E}}$ and relations $\boldsymbol{\mathcal{R}}$ in a KG.

- A score function $\boldsymbol{\sigma} : \boldsymbol{\mathcal{E}} \times \boldsymbol{\mathcal{R}} \times \boldsymbol{\mathcal{E}} \rightarrow \mathbb{R}$ distinguishes correct triples T^{+} from incorrect triples T^{-}. For example,
  $\boldsymbol{\sigma}$(Messi, plays-for-team, Barcelona) > $\boldsymbol{\sigma}$(Messi, plays-for-team, Liverpool)

- A loss function $L(T^{+}, T^{-})$ is used for training the embeddings (usually logistic loss or margin-based ranking loss).

KG Embedding Methods

KG Embedding Methods

- Additive Methods

$$\sigma(h, r, t) = -\left\| M_r^1 \mathbf{h} + \mathbf{r} - M_r^2 \mathbf{t} \right\|_1$$

- Multiplicative Methods

$$\sigma(h, r, t) = \mathbf{r}^\top f(\mathbf{h}, \mathbf{t})$$

- Neural Methods

KG Embedding Methods

Type	Model	Score Function $\sigma(h, r, t)$
Additive	TransE (Bordes et al., 2013)	$-\lVert \mathbf{h} + \mathbf{r} - \mathbf{t} \rVert_1$
	TransR (Lin et al., 2015)	$-\lVert M_r\mathbf{h} + \mathbf{r} - M_r\mathbf{t} \rVert_1$
	STransE (Nguyen et al., 2016)	$-\lVert M_r^1\mathbf{h} + \mathbf{r} - M_r^2\mathbf{t} \rVert_1$
Multiplicative	DistMult (Yang et al., 2014)	$\mathbf{r}^\top(\mathbf{h} \odot \mathbf{t})$
	HolE (Nickel et al., 2016)	$\mathbf{r}^\top(\mathbf{h} \star \mathbf{t})$
	ComplEx (Trouillon et al., 2016)	$\mathbf{Re}(\mathbf{r}^\top(\mathbf{h} \odot \bar{\mathbf{t}}))$

$\odot$ Entry-wise product $\star$ Circular correlation

Geometrical Metrics

- Average Vector Length

$$\mathrm{AVL}(\mathbf{V}) = \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \|\mathbf{v}\|_2$$

Geometrical Metrics

- Average Vector Length

$$\mathrm{AVL}(\mathbf{V}) = \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \|\mathbf{v}\|_2$$

- Alignment to Mean

$$\mathrm{ATM}(\mathbf{v}, \mathbf{V}) = \mathrm{cosine}\left(\mathbf{v}, \frac{1}{|\mathbf{V}|} \sum_{\mathbf{x} \in \mathbf{V}} \mathbf{x}\right)$$

Geometrical Metrics

- Conicity

$$\mathrm{Conicity}(\mathbf{V}) = \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \mathrm{ATM}(\mathbf{v}, \mathbf{V})$$

Geometrical Metrics

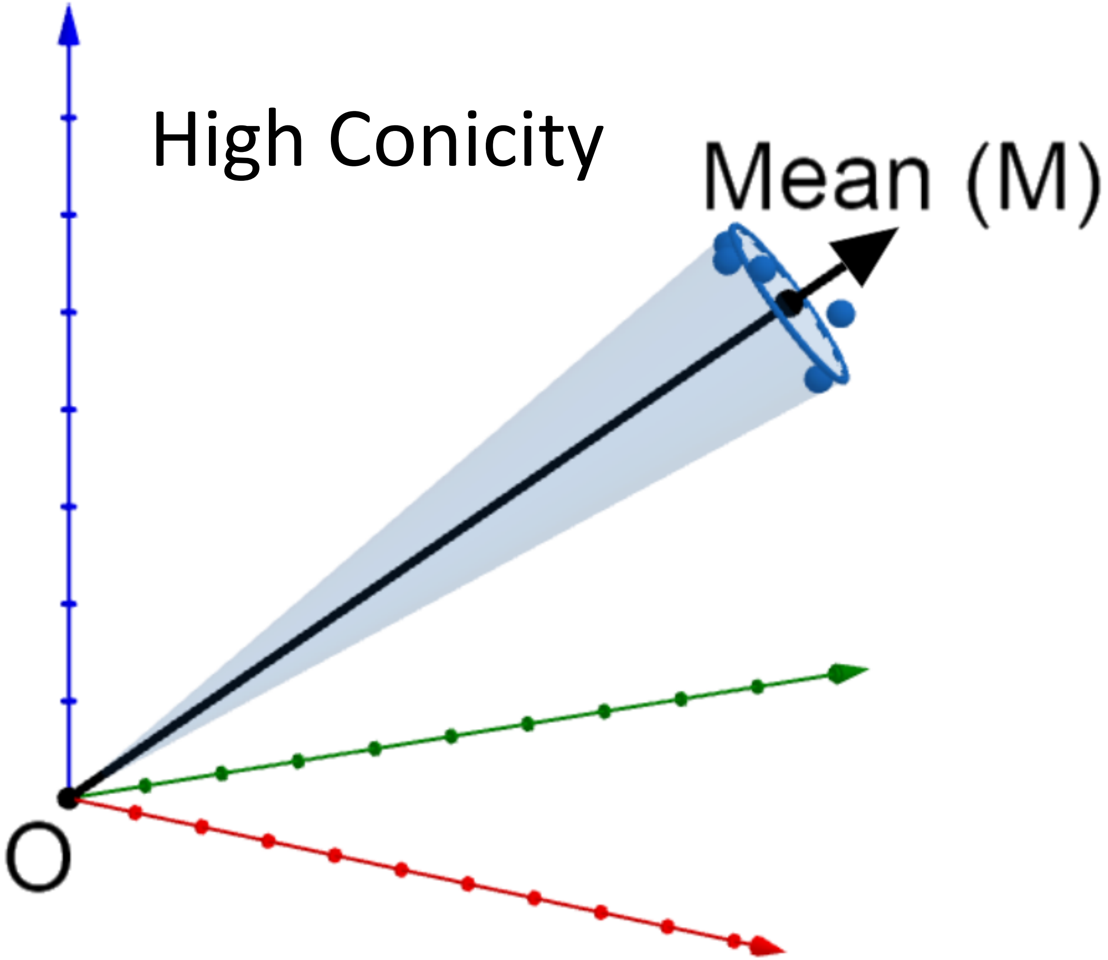
- Conicity

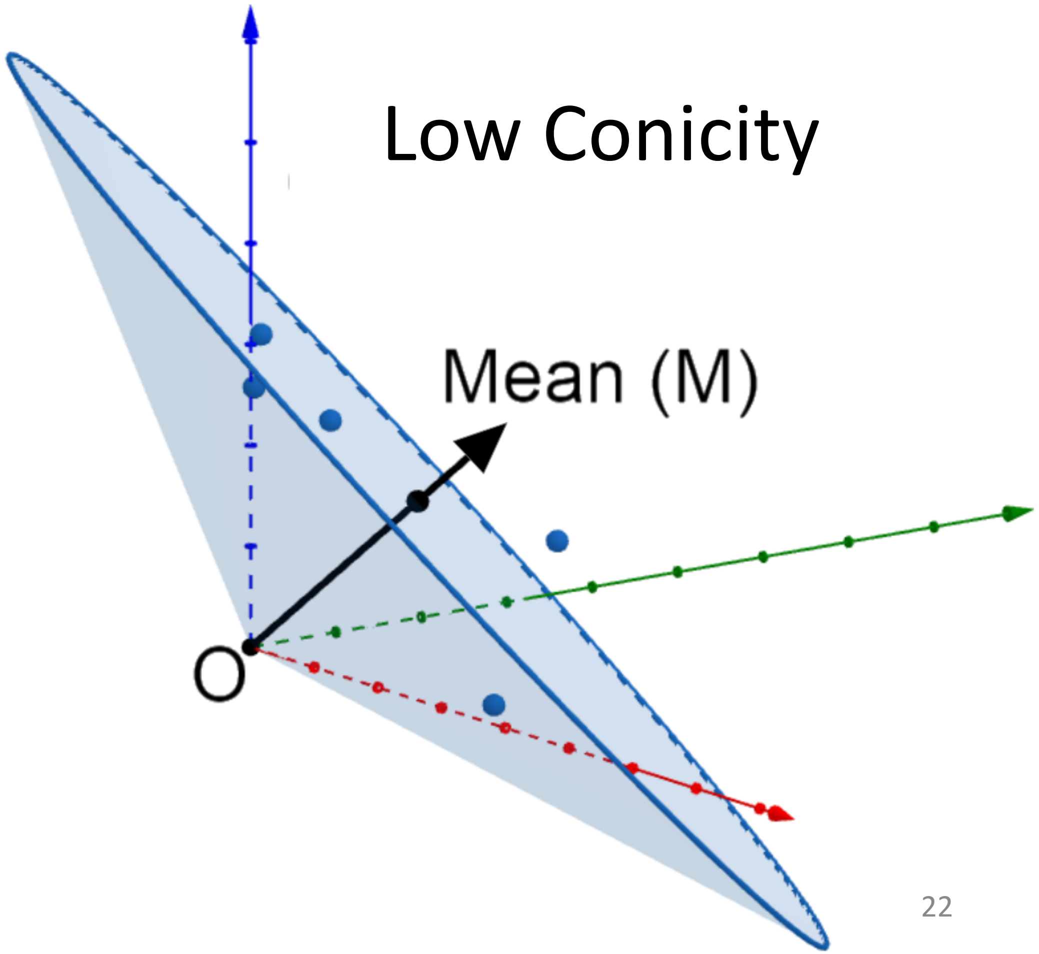
$$\mathrm{Conicity}(\mathbf{V}) = \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \mathrm{ATM}(\mathbf{v}, \mathbf{V})$$

- Vector Spread

$$\mathrm{VS}(\mathbf{V}) = \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \left(\mathrm{ATM}(\mathbf{v}, \mathbf{V}) - \mathrm{Conicity}(\mathbf{V}) \right)^2$$

Geometry of Embeddings





Experiments

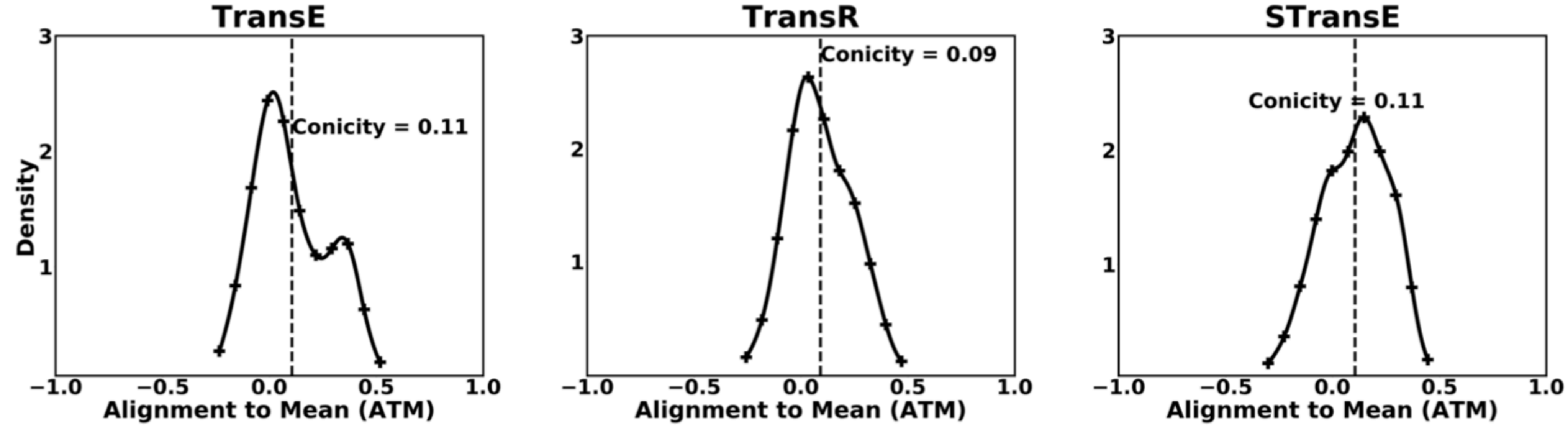
- We study the effect of following factors on the geometry of KG Embeddings
  - Type of method (Additive or Multiplicative)
  - Number of Negative Samples
  - Dimension of Vector Space

- We also study the correlation of performance and geometry.

- For experiments, we used FB15k dataset.

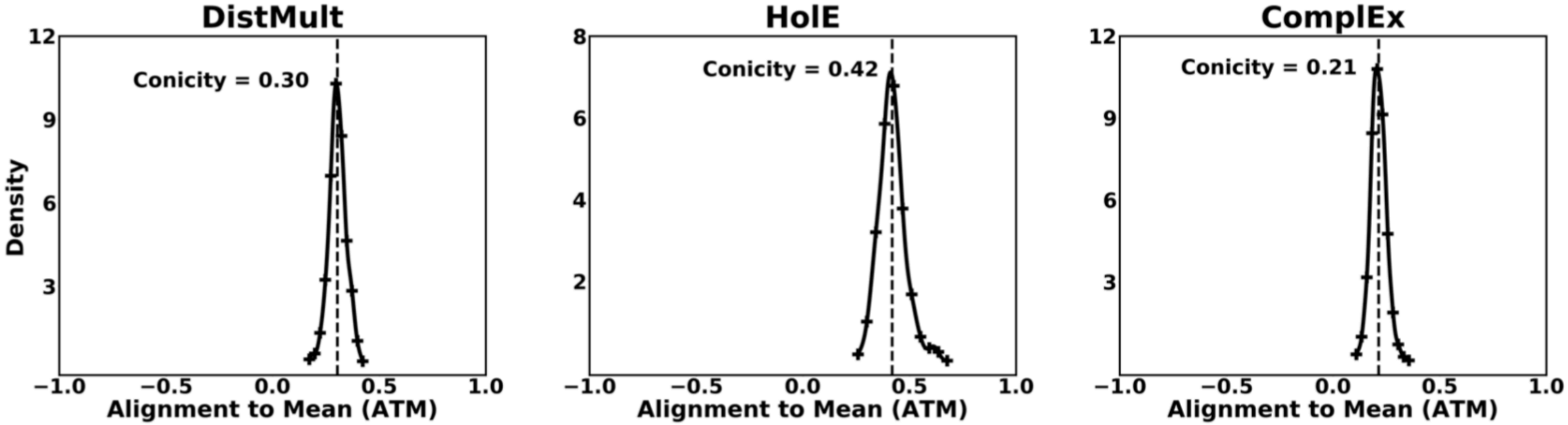
Additive vs Multiplicative (Entity Vectors)

Additive



(a) Additive Models

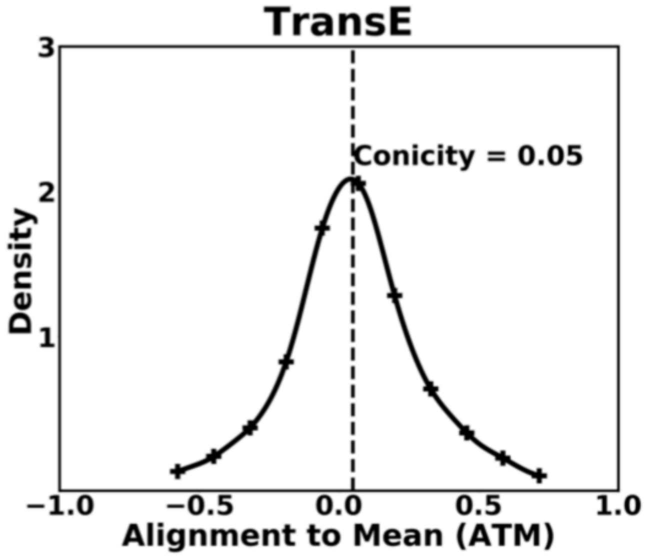
Multiplicative



(b) Multiplicative Models

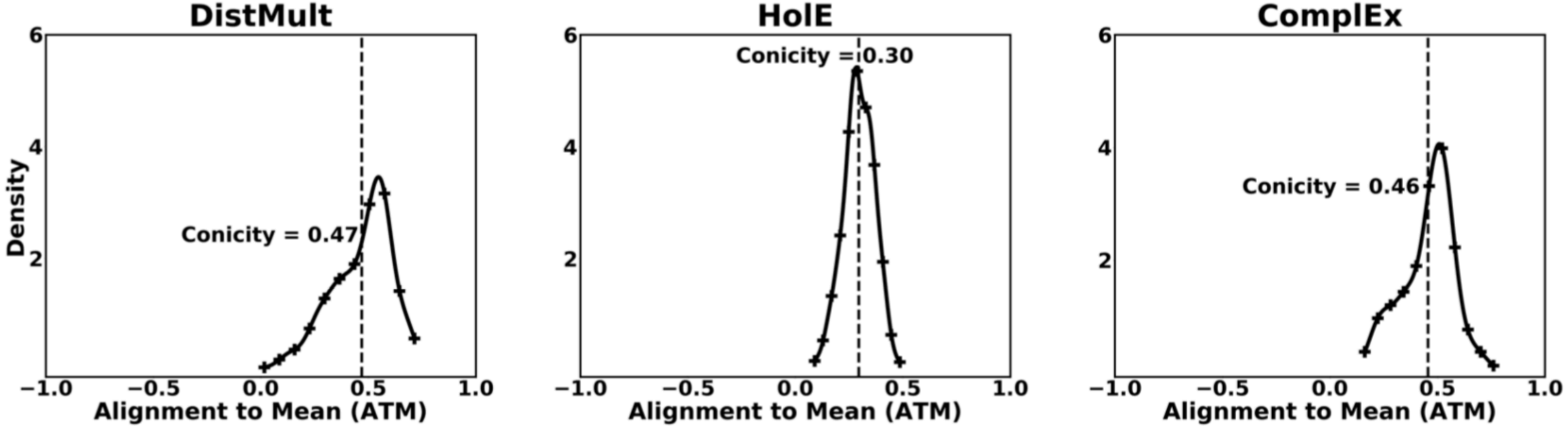
Additive vs Multiplicative (Relation Vectors)

Additive



(a) Additive Models

Multiplicative

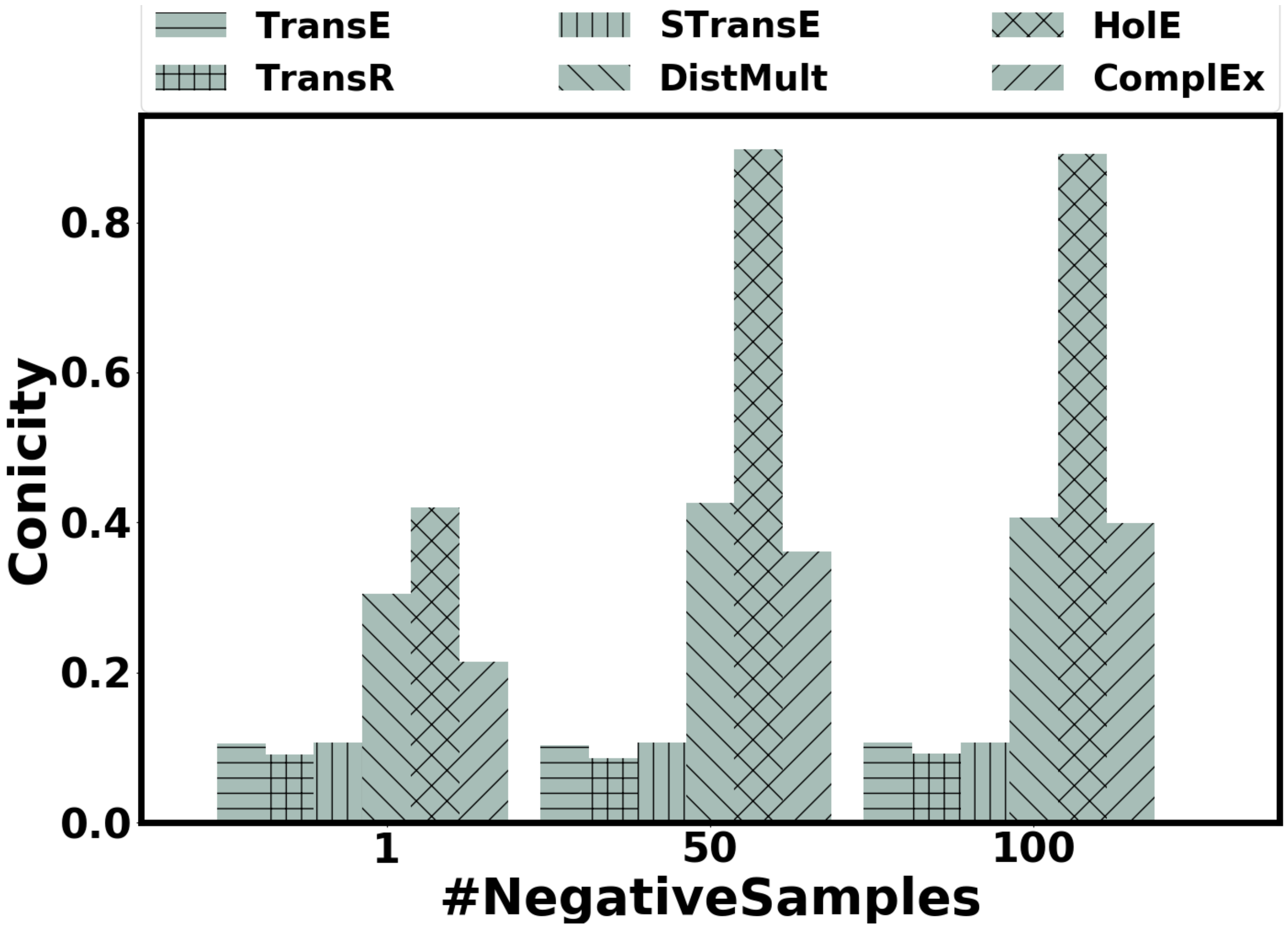


(b) Multiplicative Models

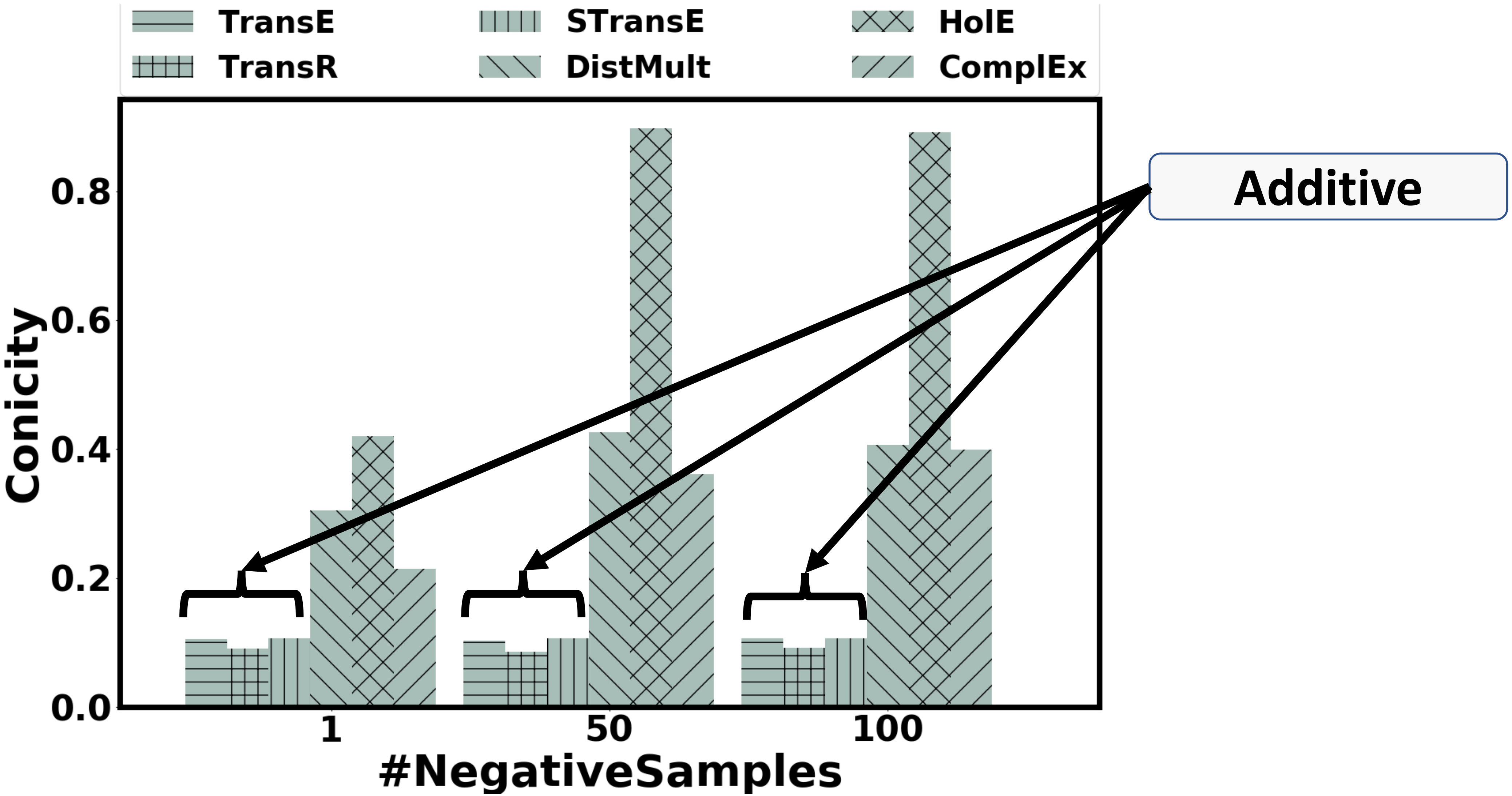
Additive vs Multiplicative

Model Type	**Conicity**	**Vector Spread**
Additive	Low	High
Multiplicative	High	Low

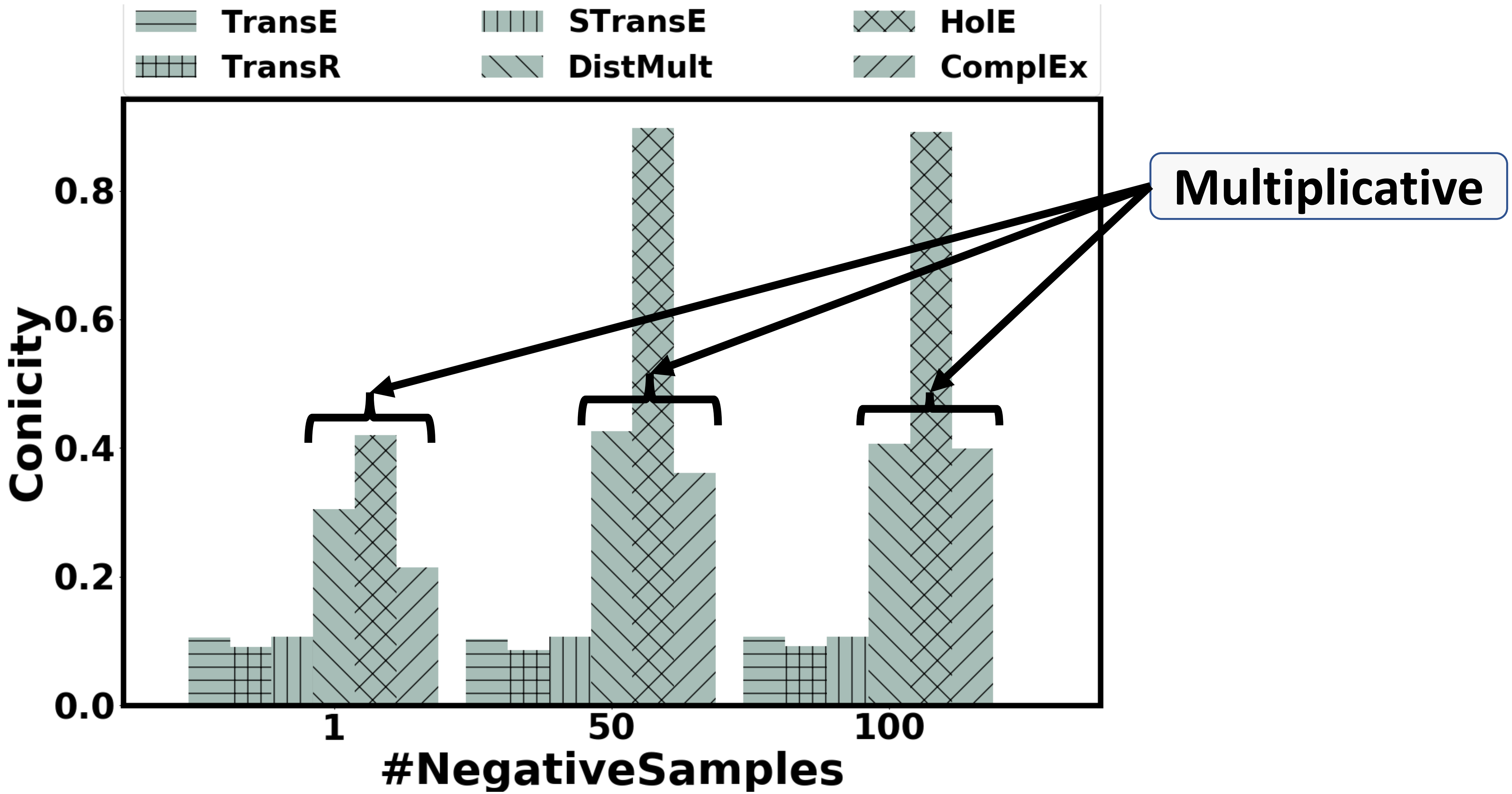
Effect of #Negative Samples (Entity Vectors)



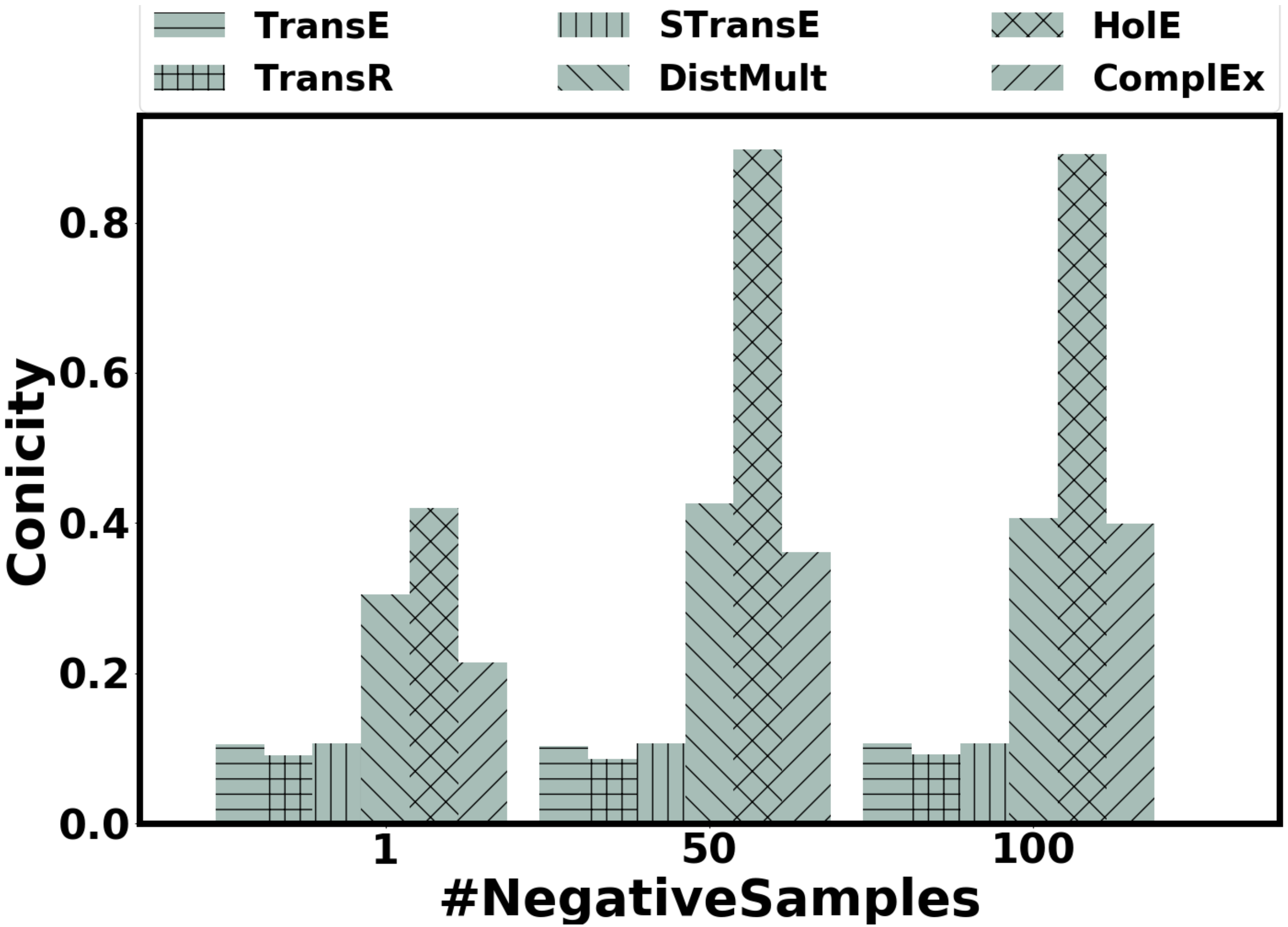
Effect of #Negative Samples (Entity Vectors)



Effect of #Negative Samples (Entity Vectors)

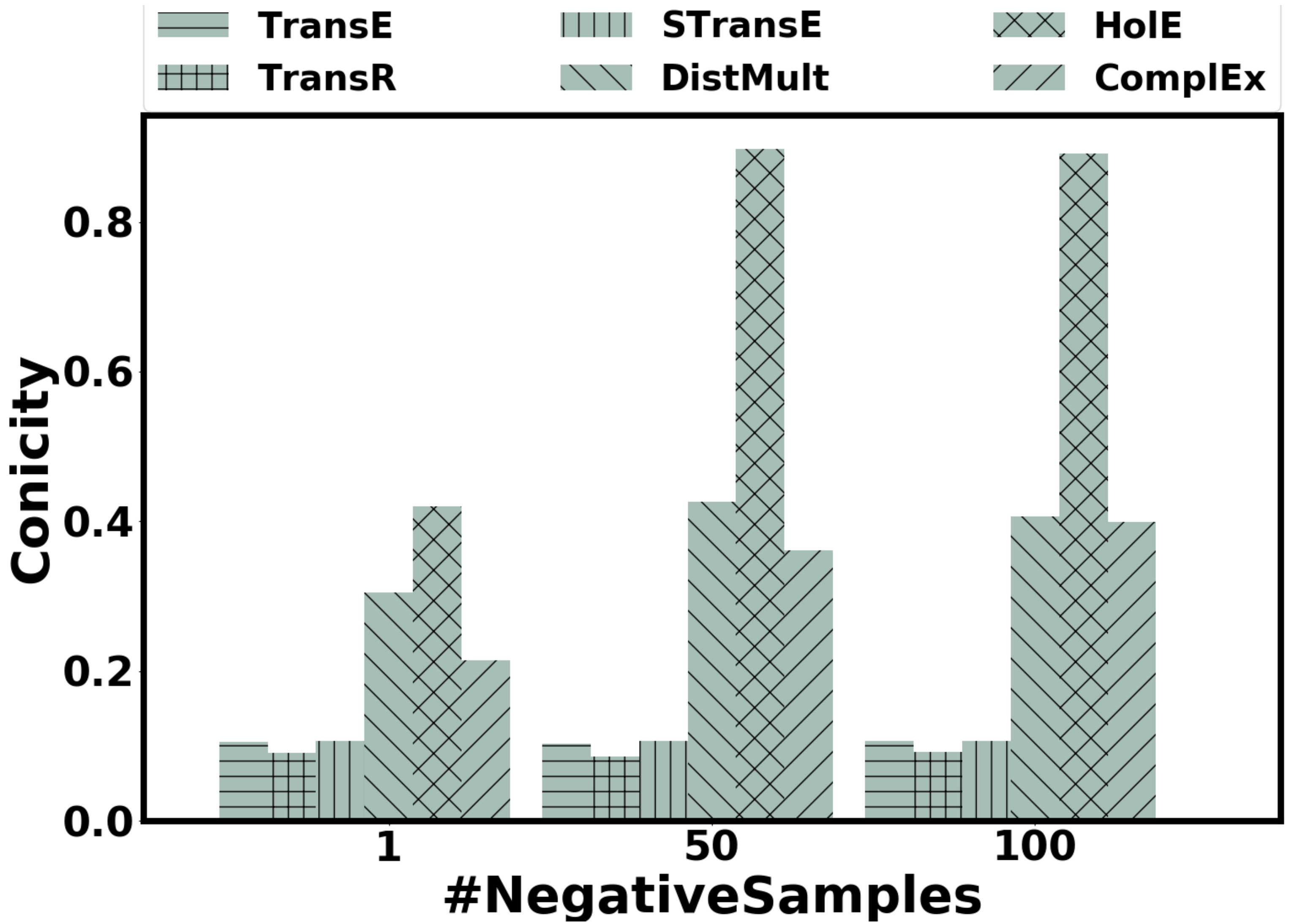


Effect of #Negative Samples (Entity Vectors)



Additive
No change

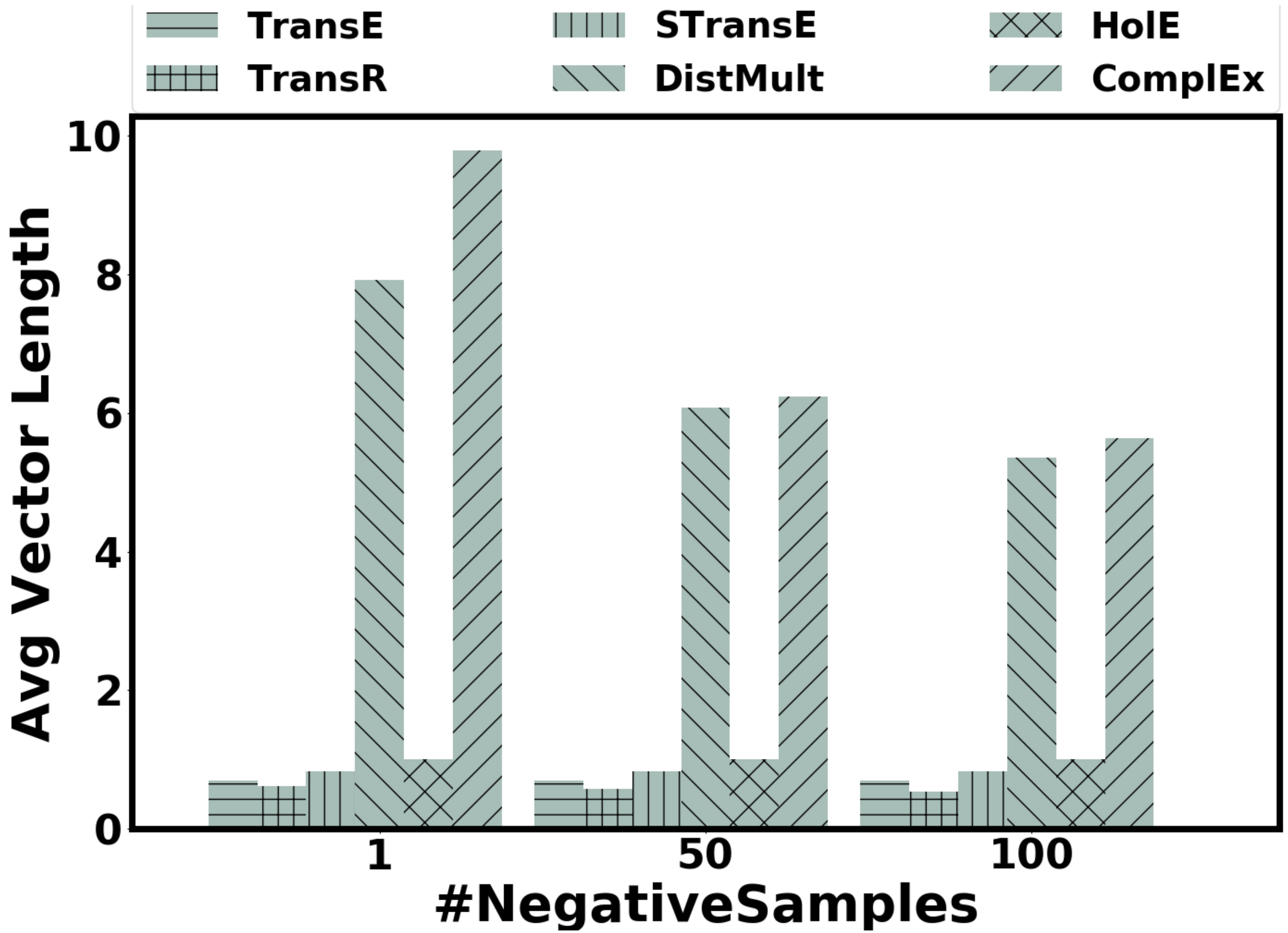
Effect of #Negative Samples (Entity Vectors)



Additive
No change

Multiplicative
Conicity Increases

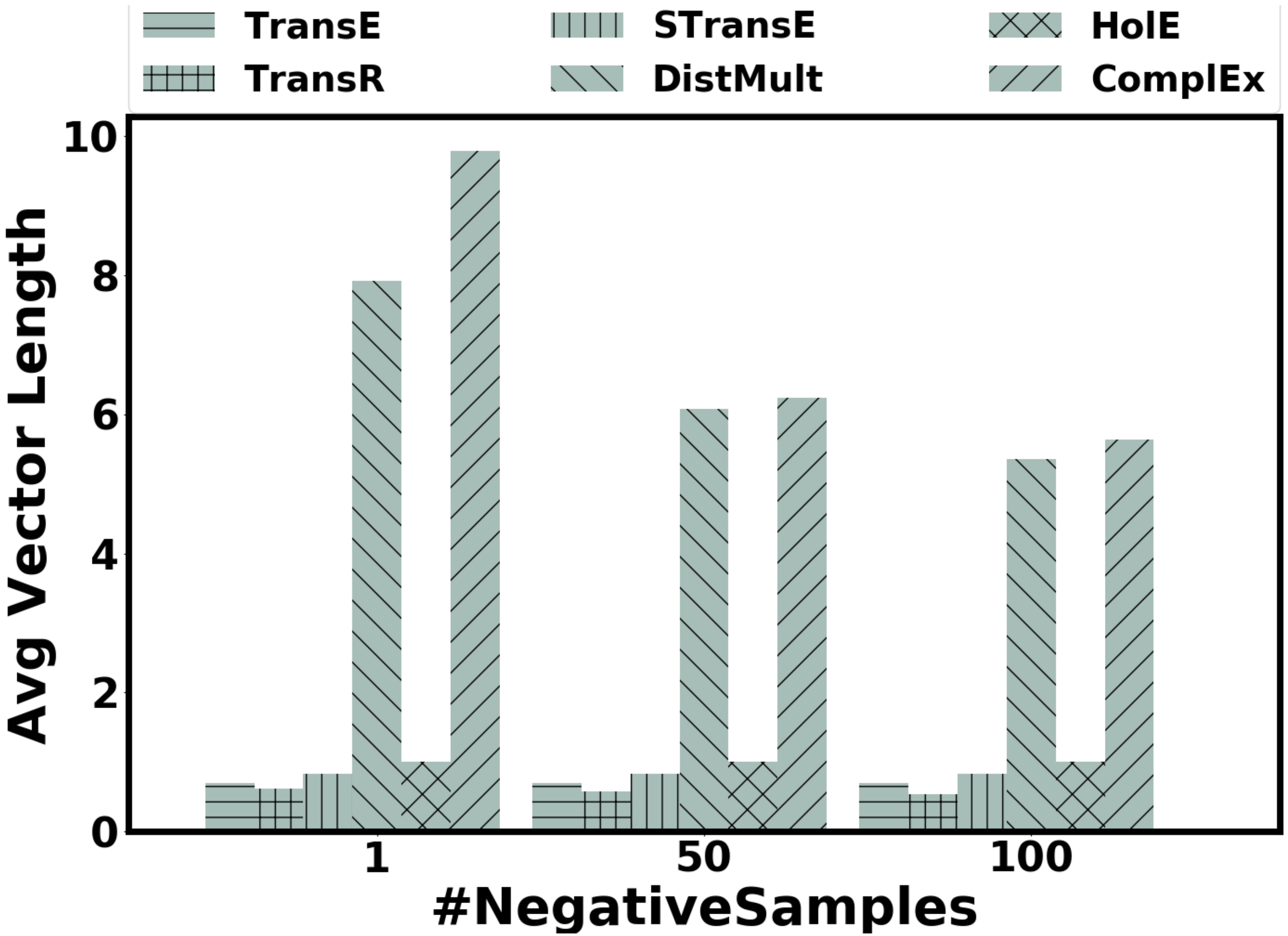
Effect of #Negative Samples (Entity Vectors)



Additive

No change

Effect of #Negative Samples (Entity Vectors)



Additive
No change

Multiplicative
AVL decreases

Effect of #Negative Samples

Model Type	Vector Type	Conicity	AVL
Additive	Entity	No Change	No Change
	Relation	No Change	No Change
Multiplicative	Entity	Increases	Decreases
	Relation	Decreases	No Change except HolE

SGNS (Word2Vec[1]) as Multiplicative Model

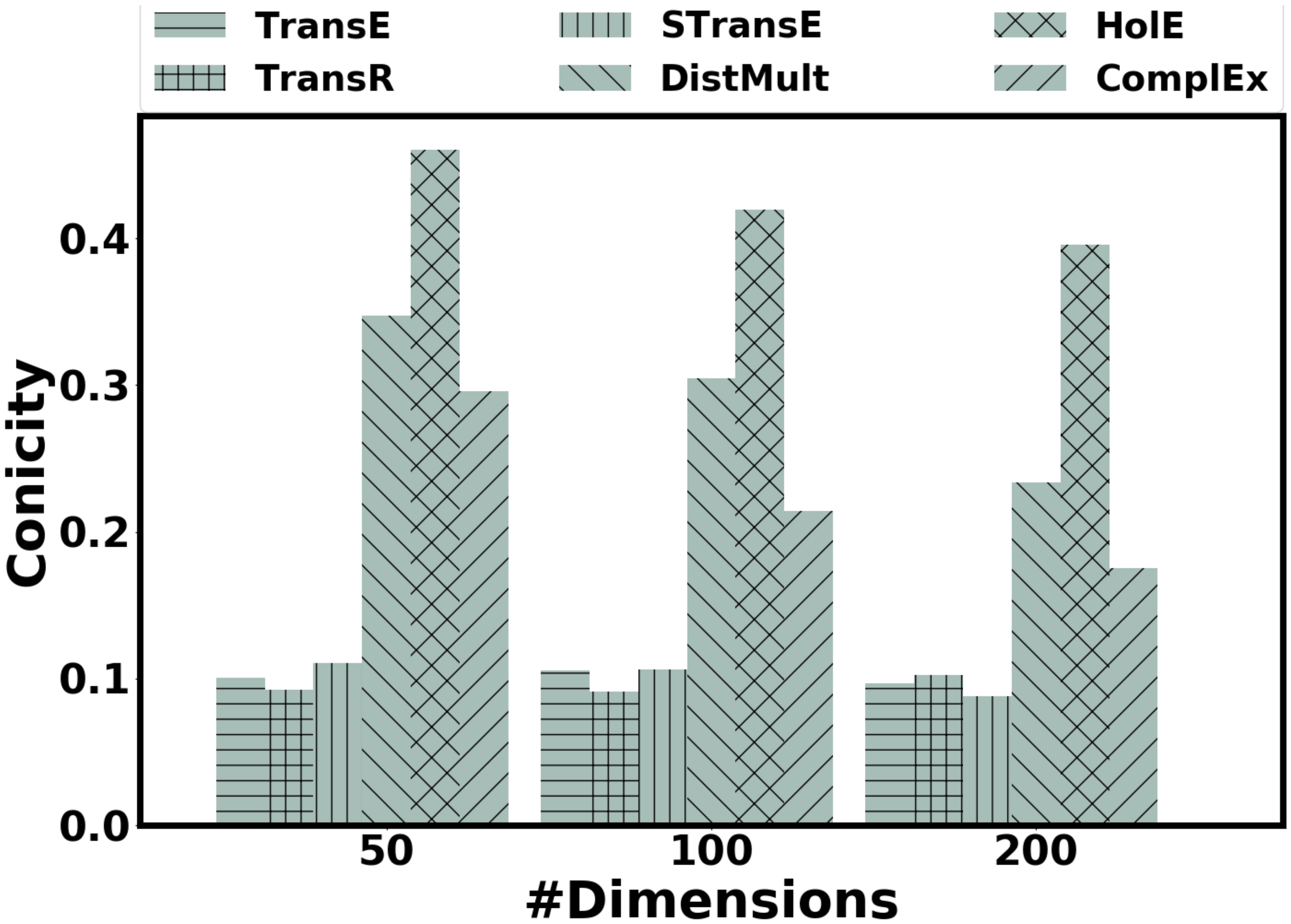
- Similar observation was made by (Mimno and Thompson, 2017)[2] for SGNS based word embeddings where higher #negatives resulted in higher conicity.

- Word2Vec[1] maximizes word and context vector dot product for positive word-context pairs.

- This behavior is consistent with that of multiplicative models.

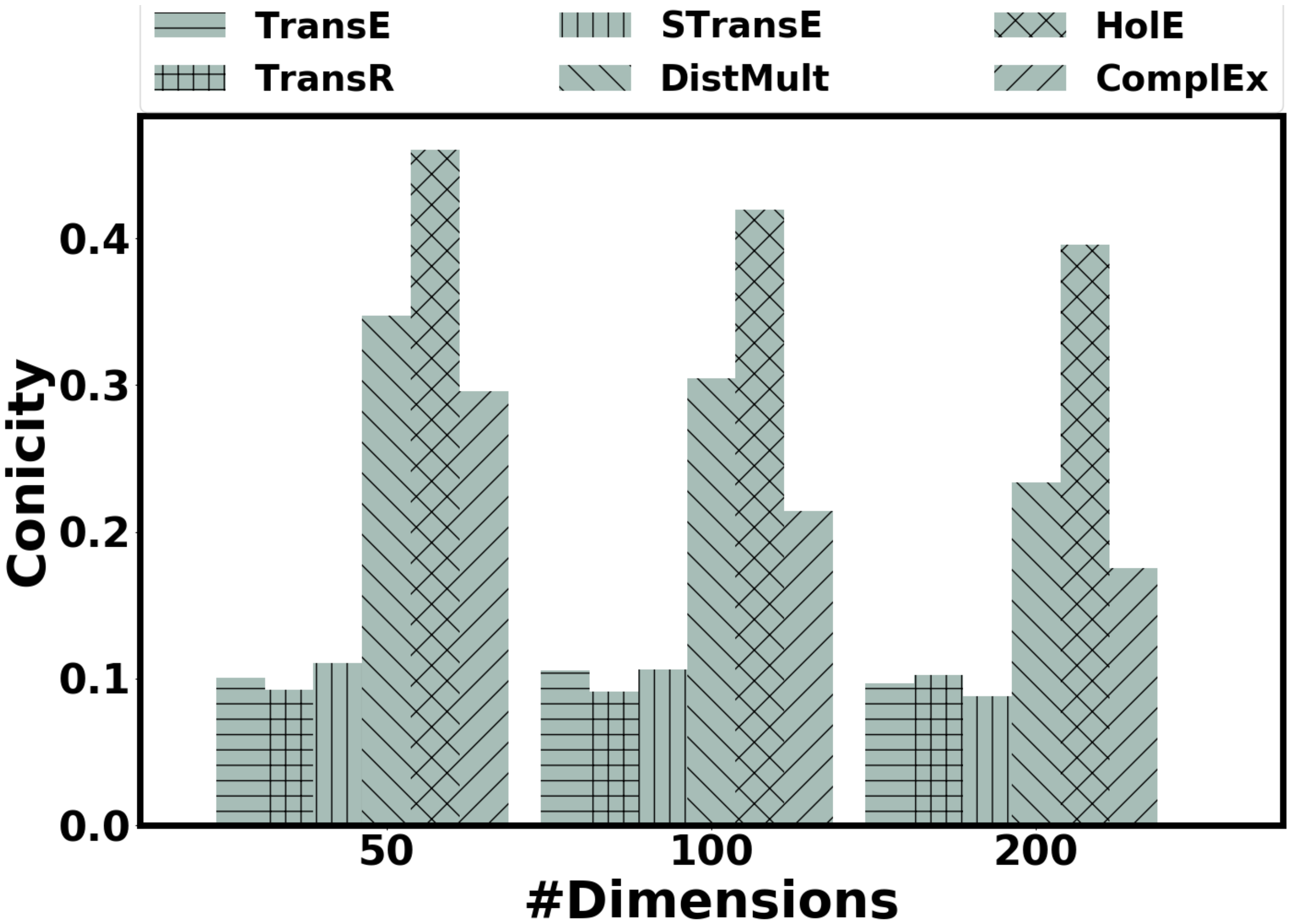
1. Distributed representations of words and phrases and their compositionality, Mikolov et al. NIPS 2013
2. The strange geometry of skip-gram with negative sampling, Mimno and Thompson, EMNLP 2017

Effect of #Dimensions (Entity Vectors)



Additive
No change

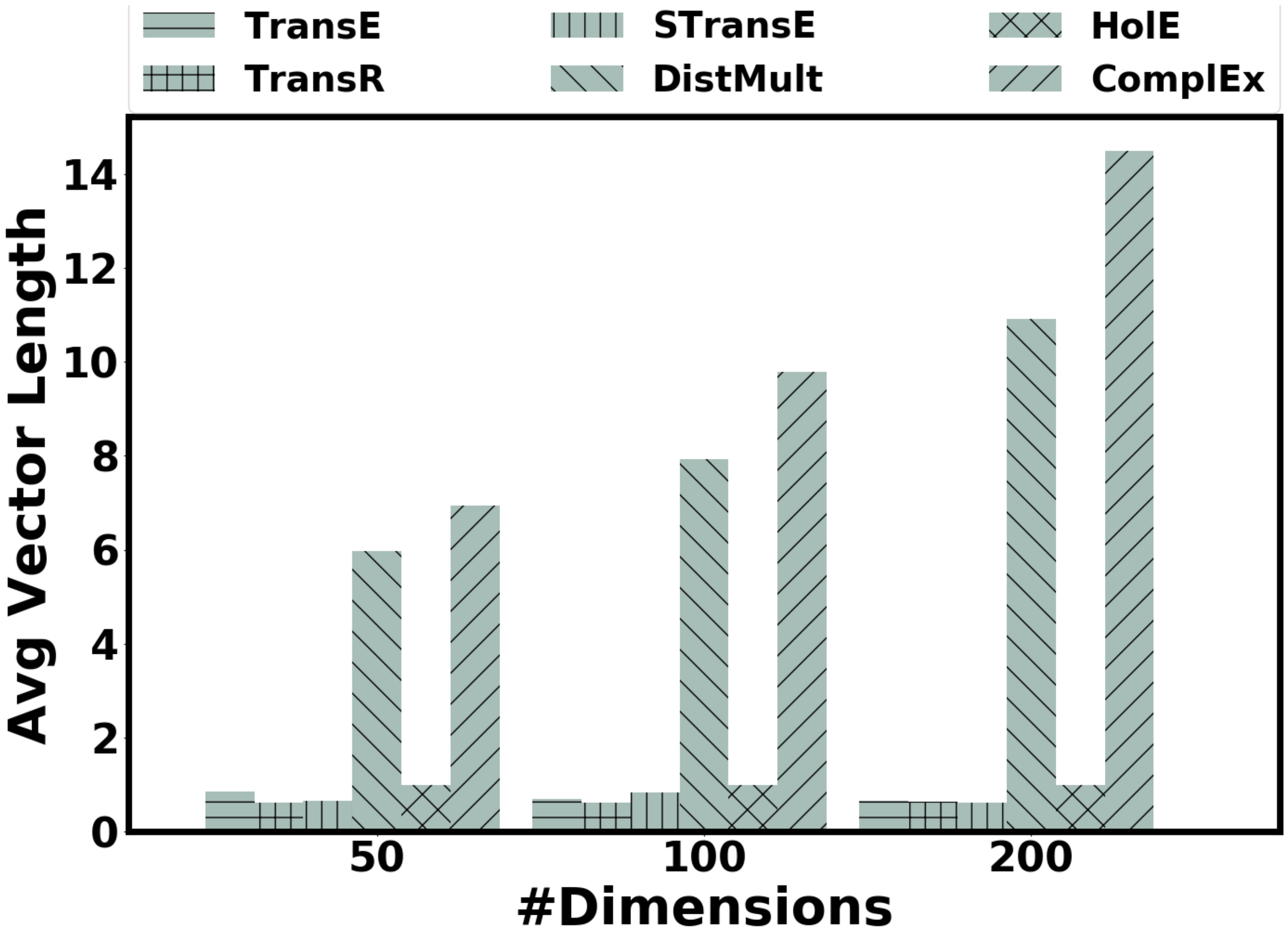
Effect of #Dimensions (Entity Vectors)



Additive
No change

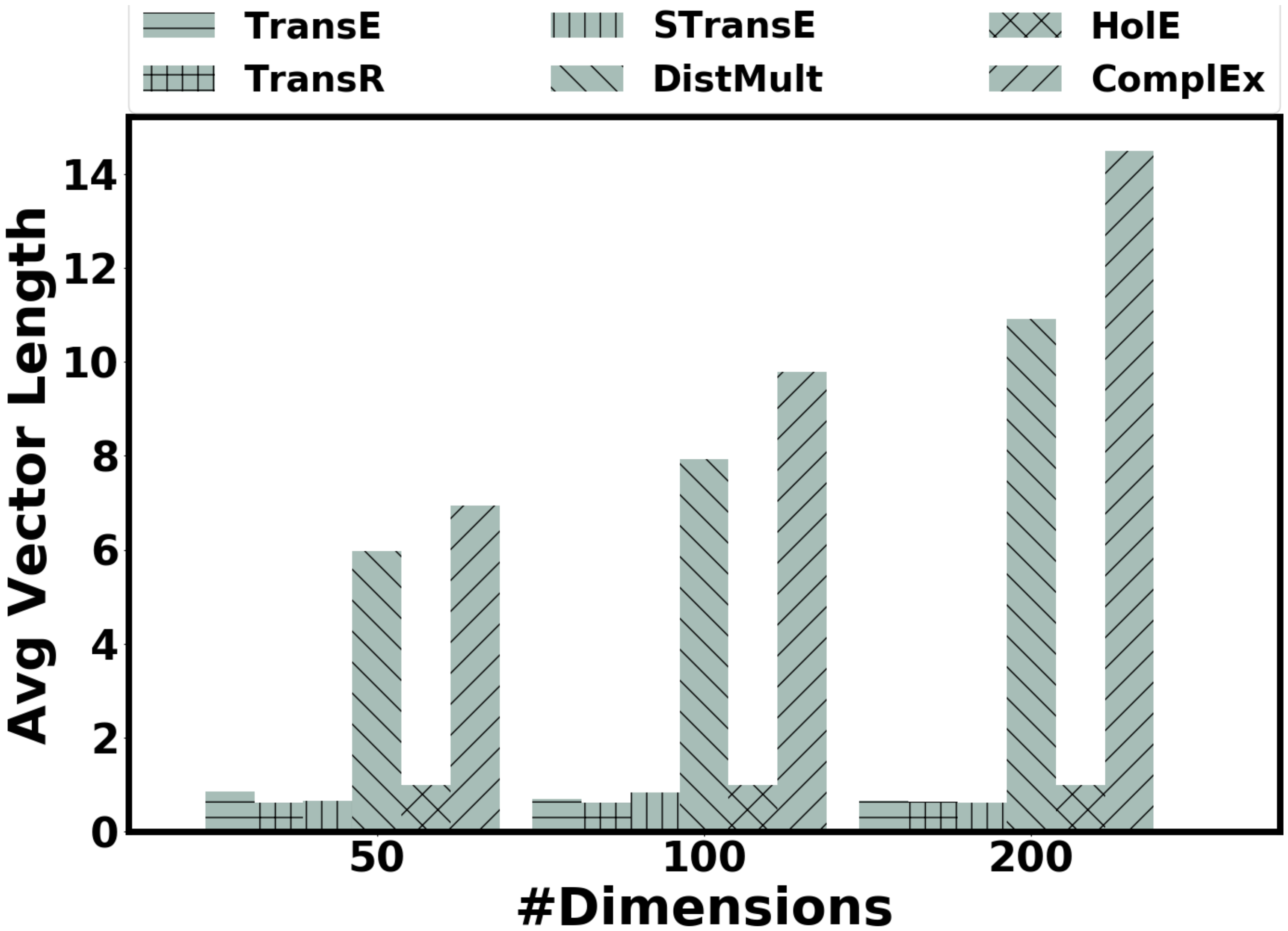
Multiplicative
Conicity decreases

Effect of #Dimensions (Entity Vectors)



Additive
No change

Effect of #Dimensions (Entity Vectors)



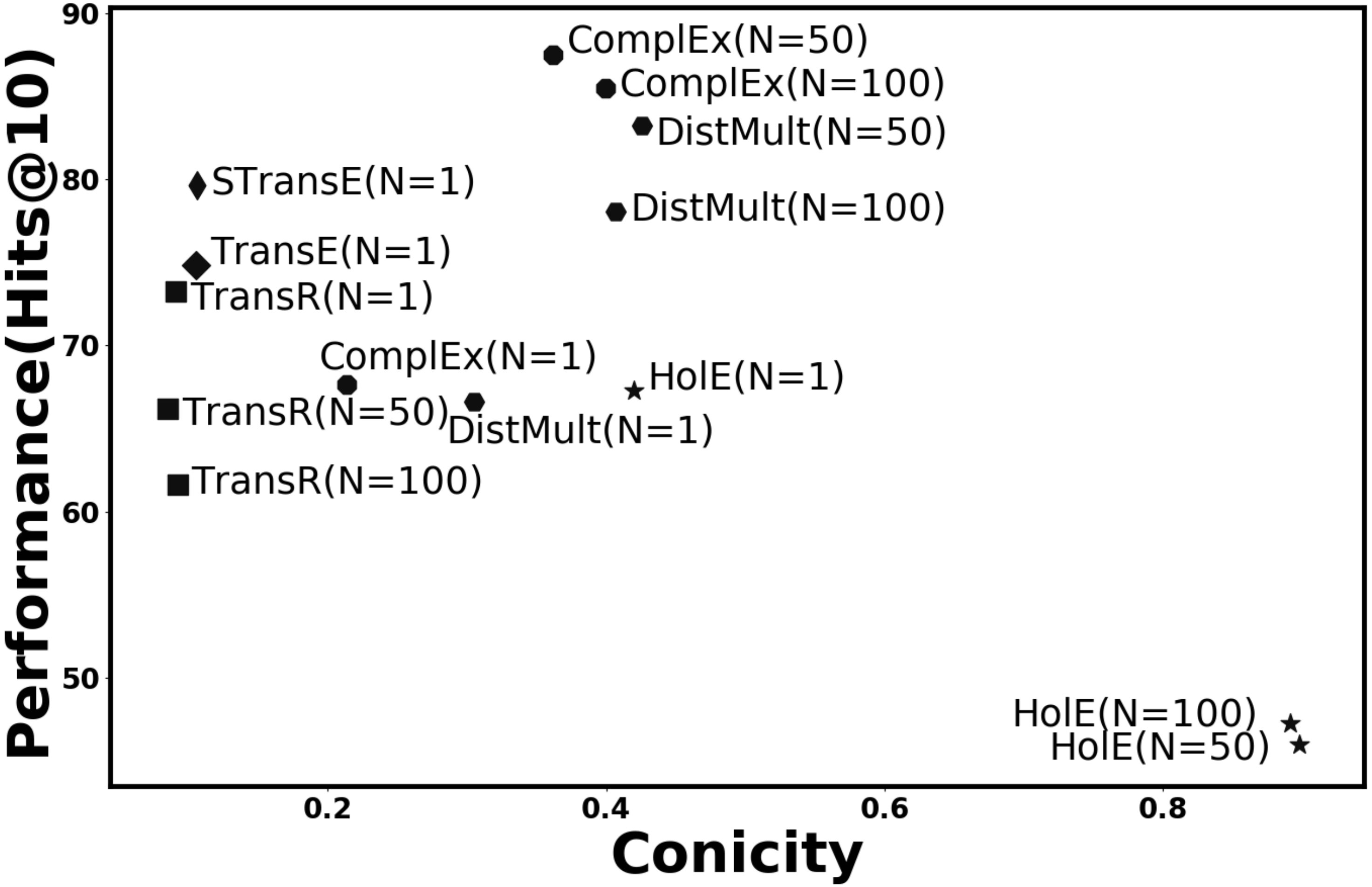
Additive
No change

Multiplicative
AVL Increases

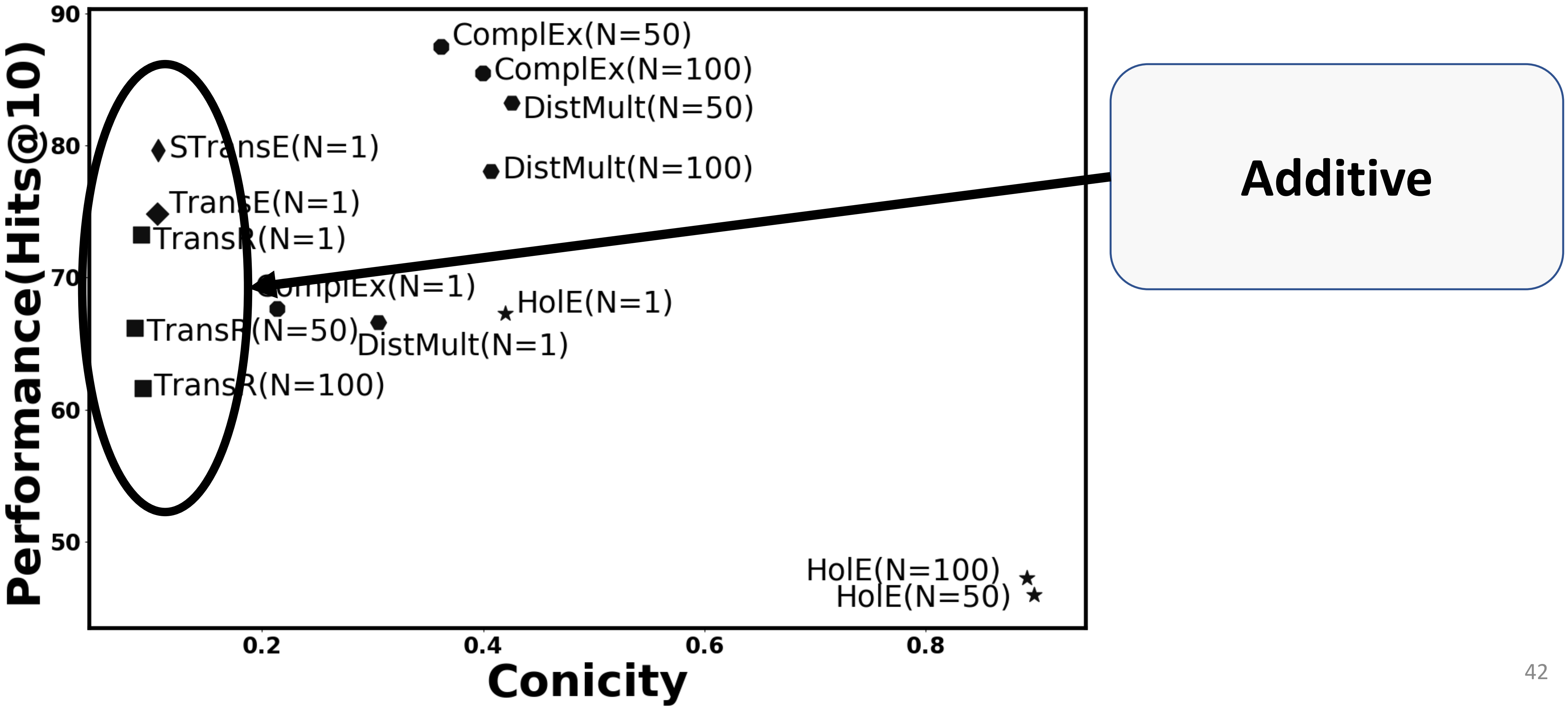
Effect of #Dimensions

Model Type	Vector Type	Conicity	AVL
Additive	Entity	No Change	No Change
	Relation	No Change	No Change
Multiplicative	Entity	Decreases	Increases
	Relation	Decreases	Increases

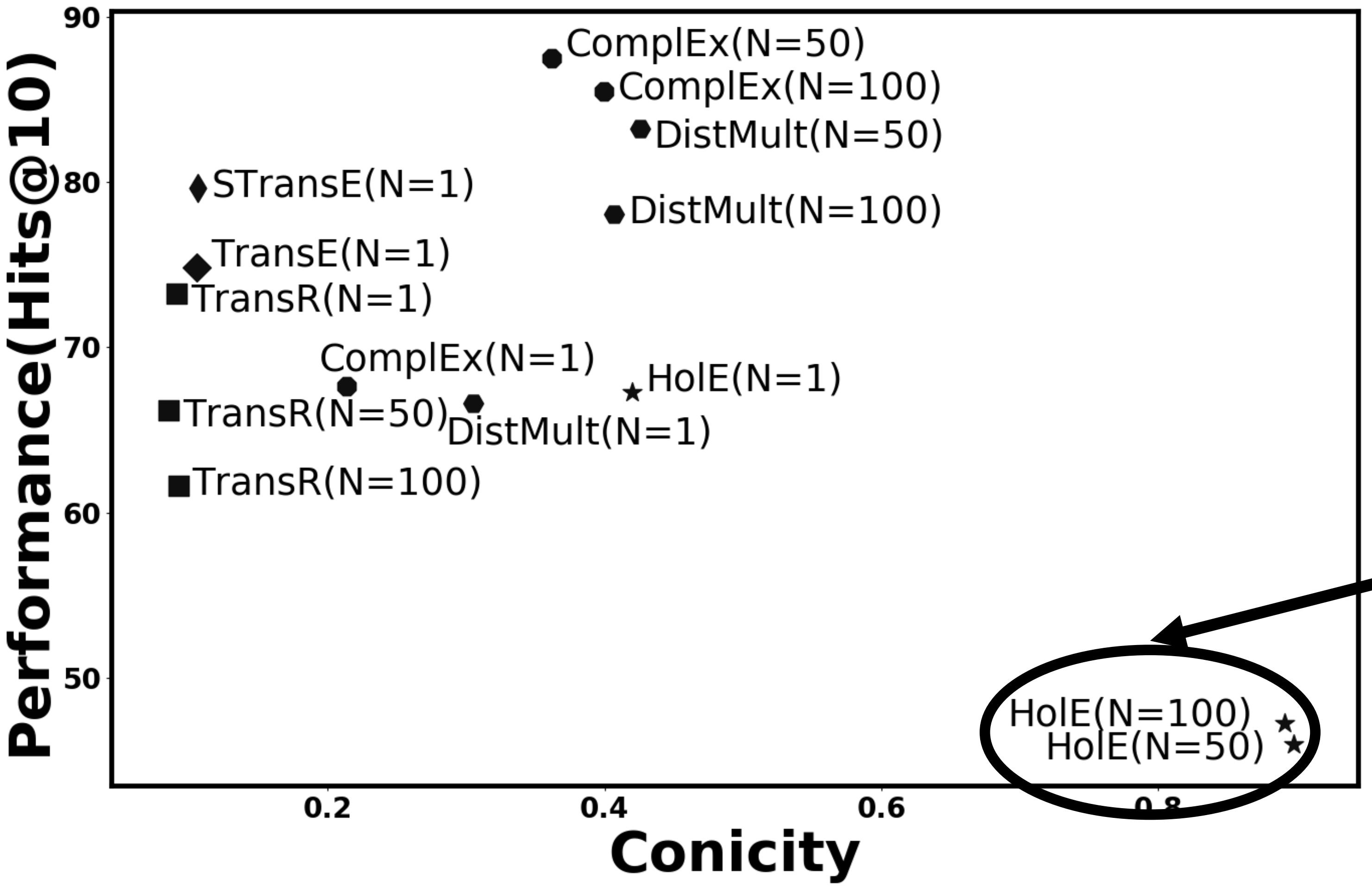
Correlation b/w Geometry and Performance



Correlation b/w Geometry and Performance

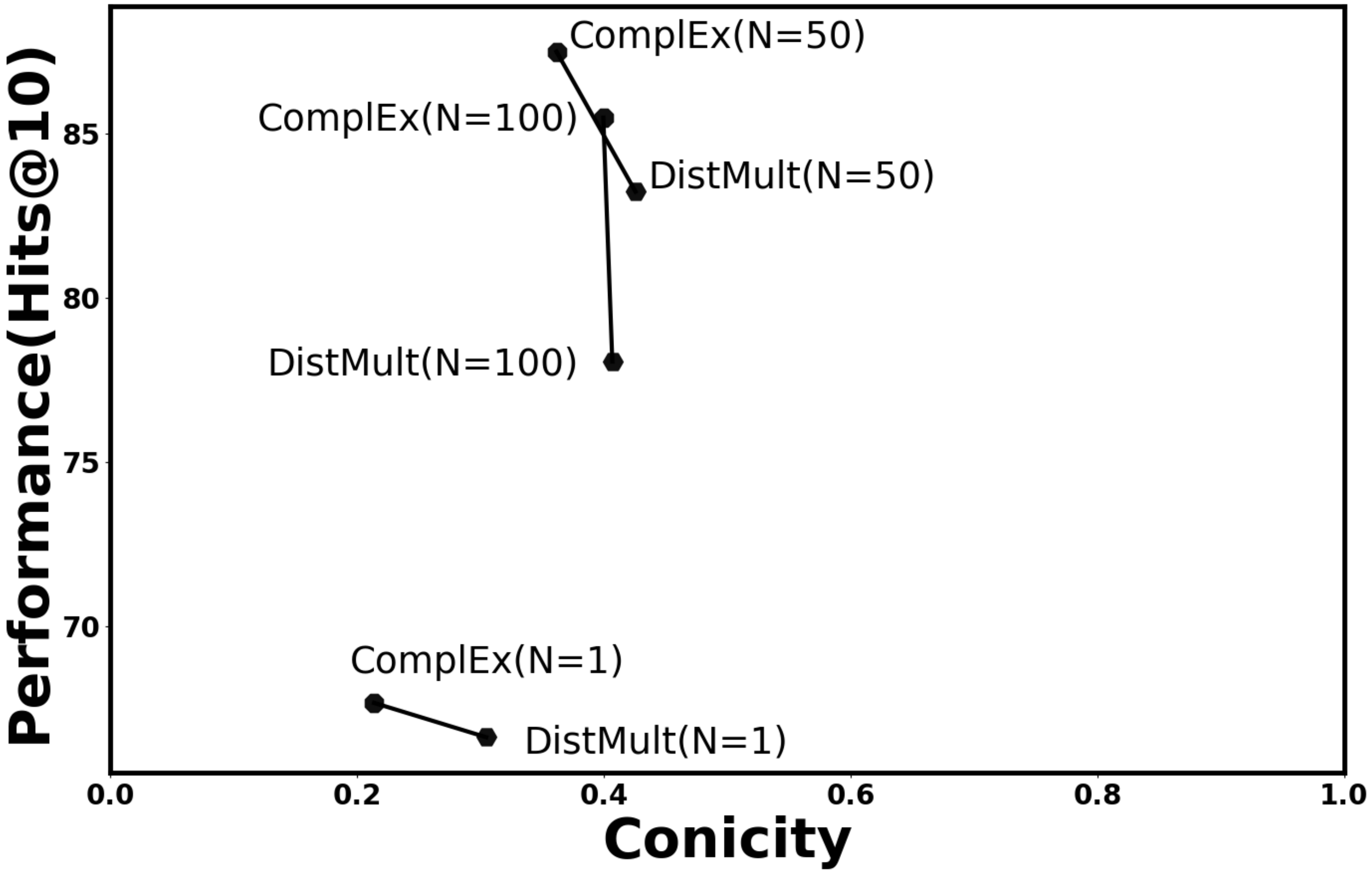


Correlation b/w Geometry and Performance



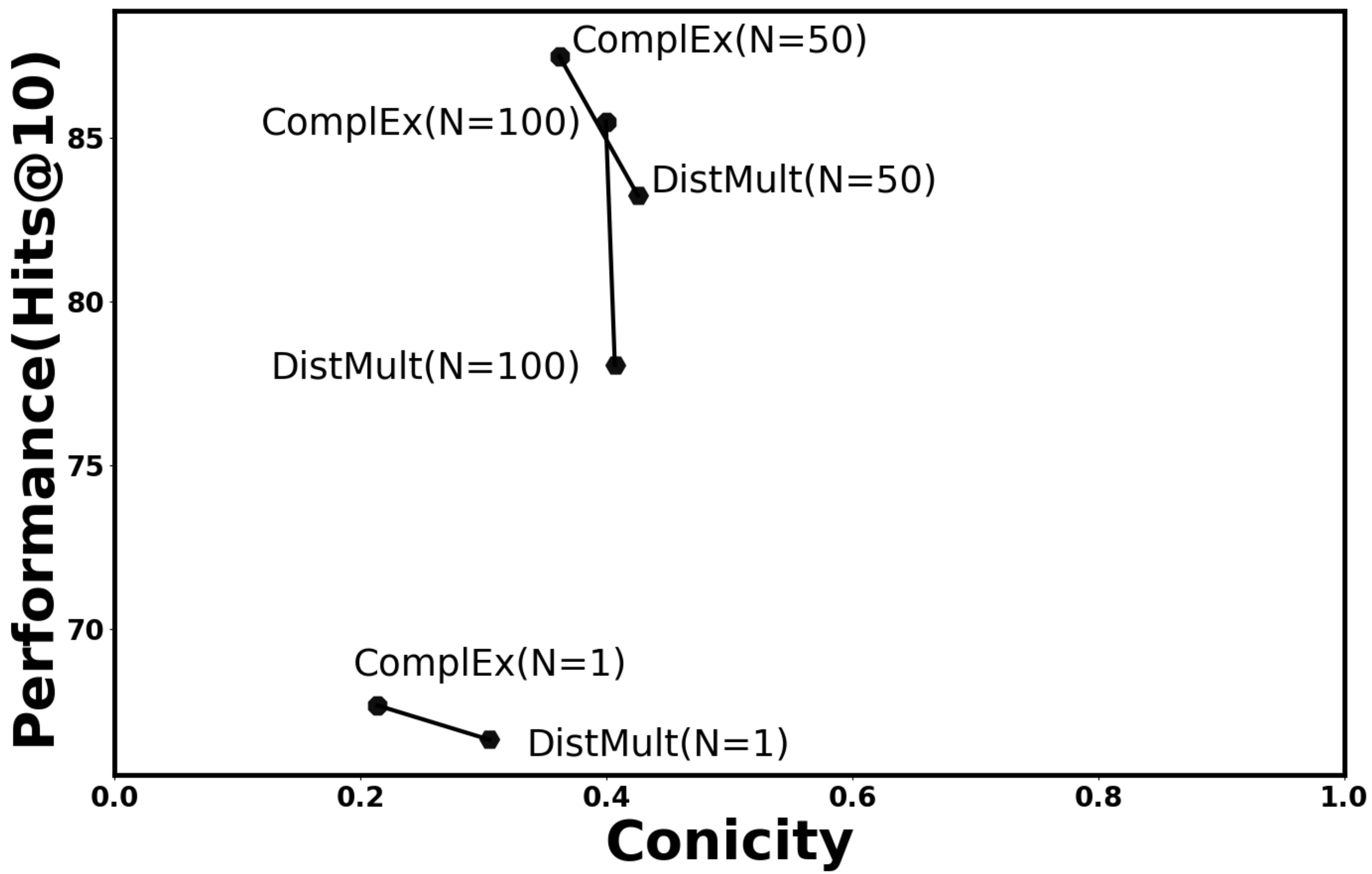
HolE performs bad with higher negatives

Correlation b/w Geometry and Performance



Negative Slope-
Negative Correlation

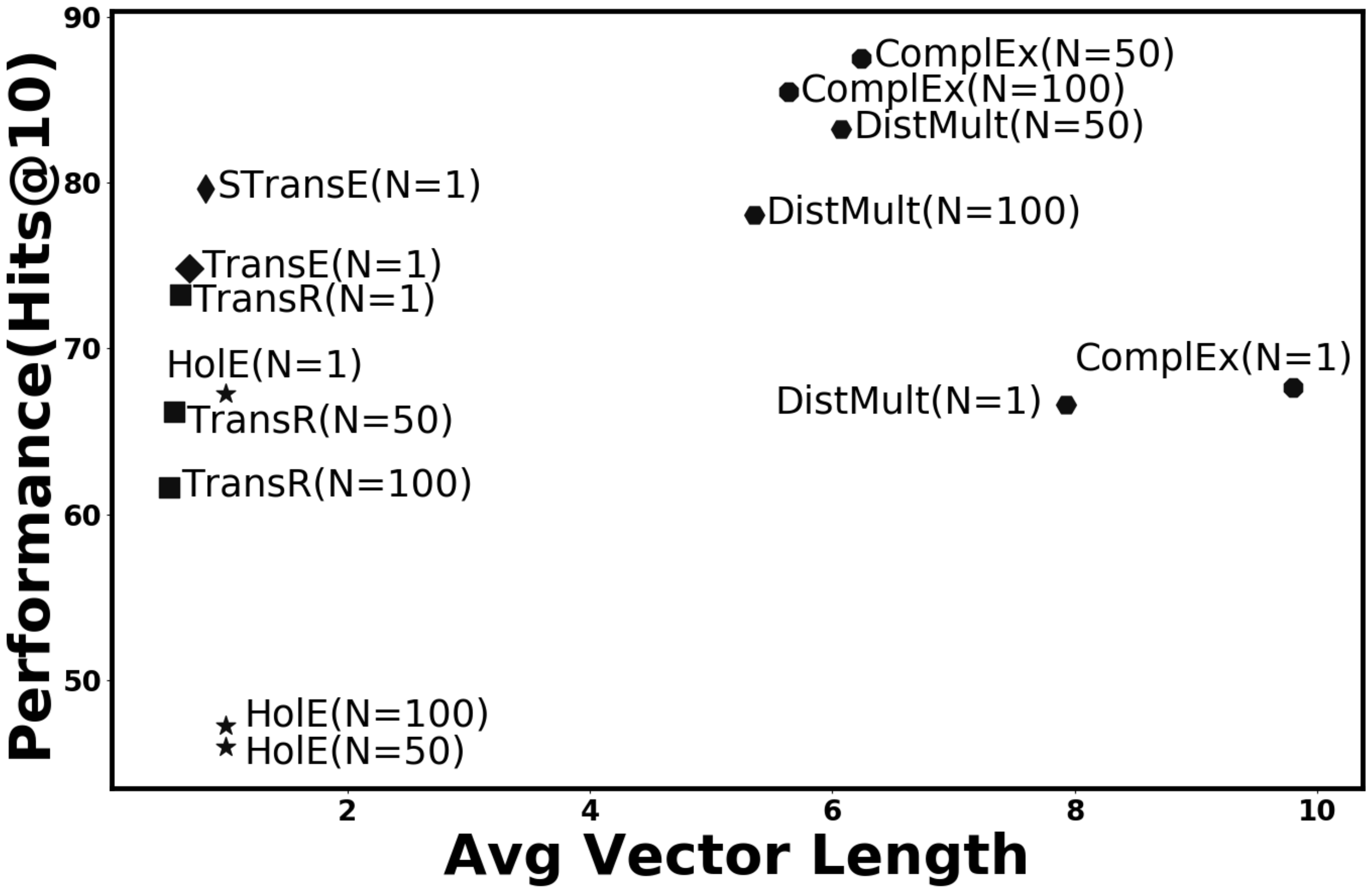
Correlation b/w Geometry and Performance



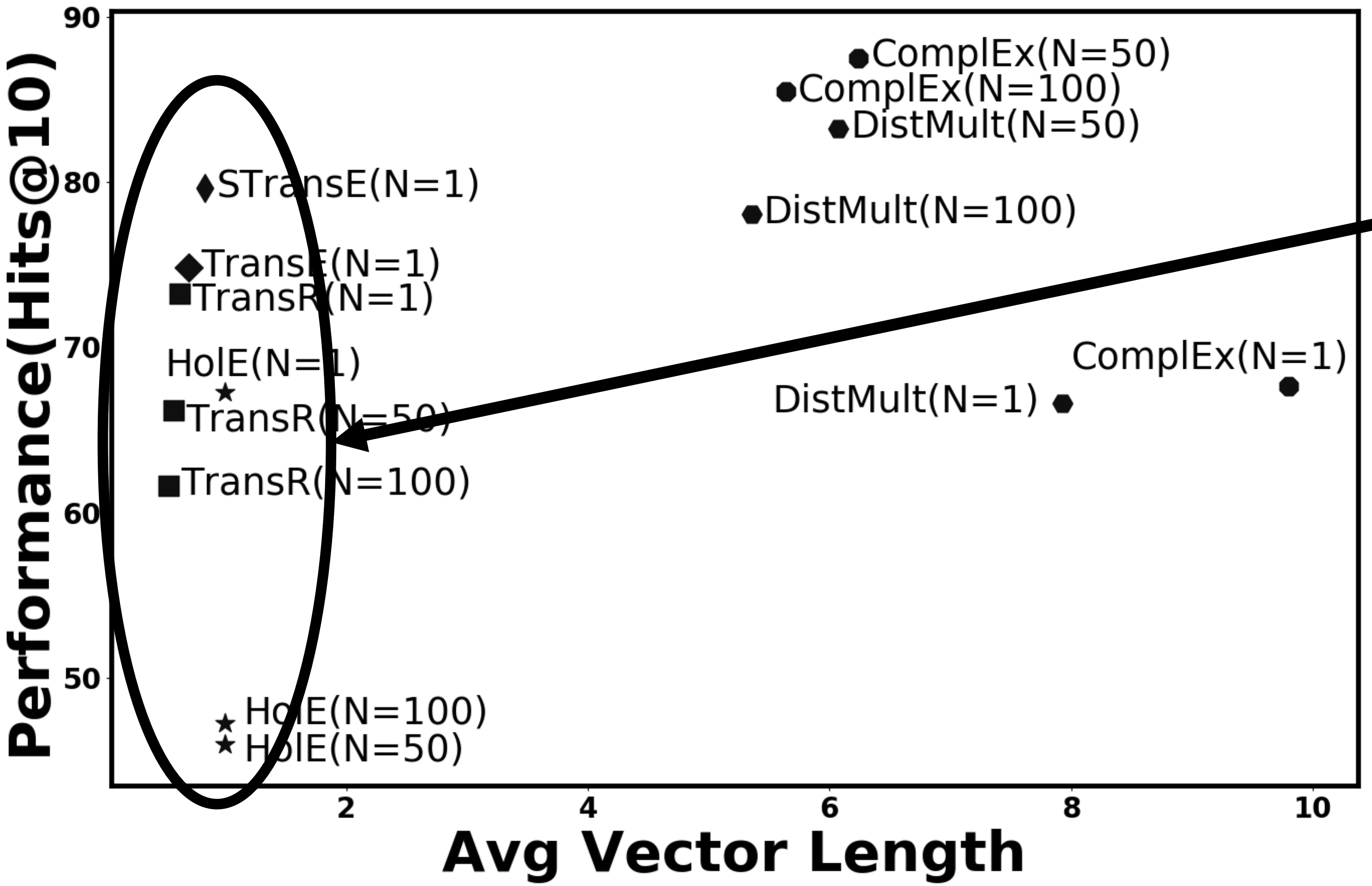
Negative Slope-
Negative Correlation

Higher Negatives-
Higher Slope
Magnitude

Correlation b/w Geometry and Performance

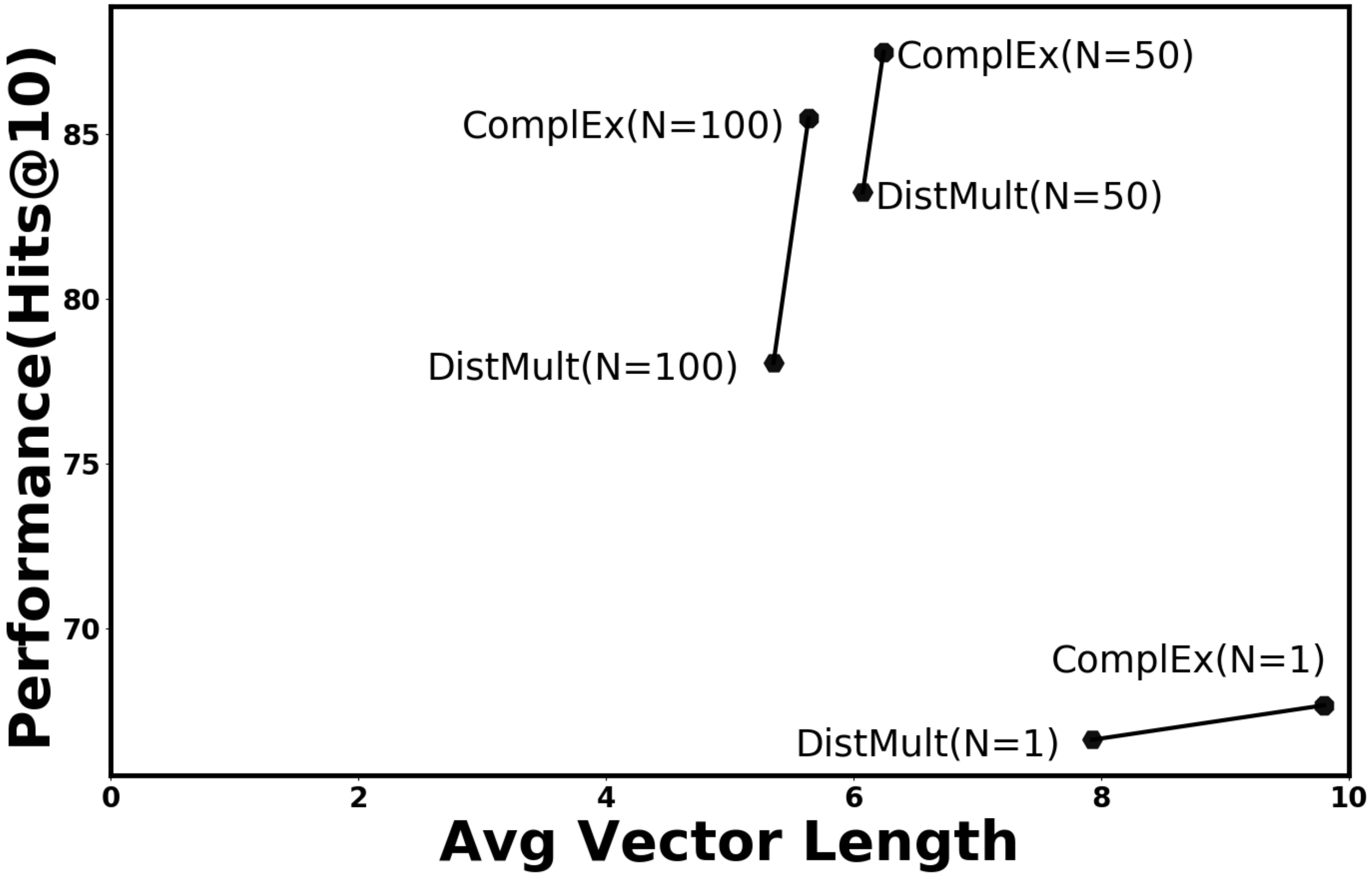


Correlation b/w Geometry and Performance



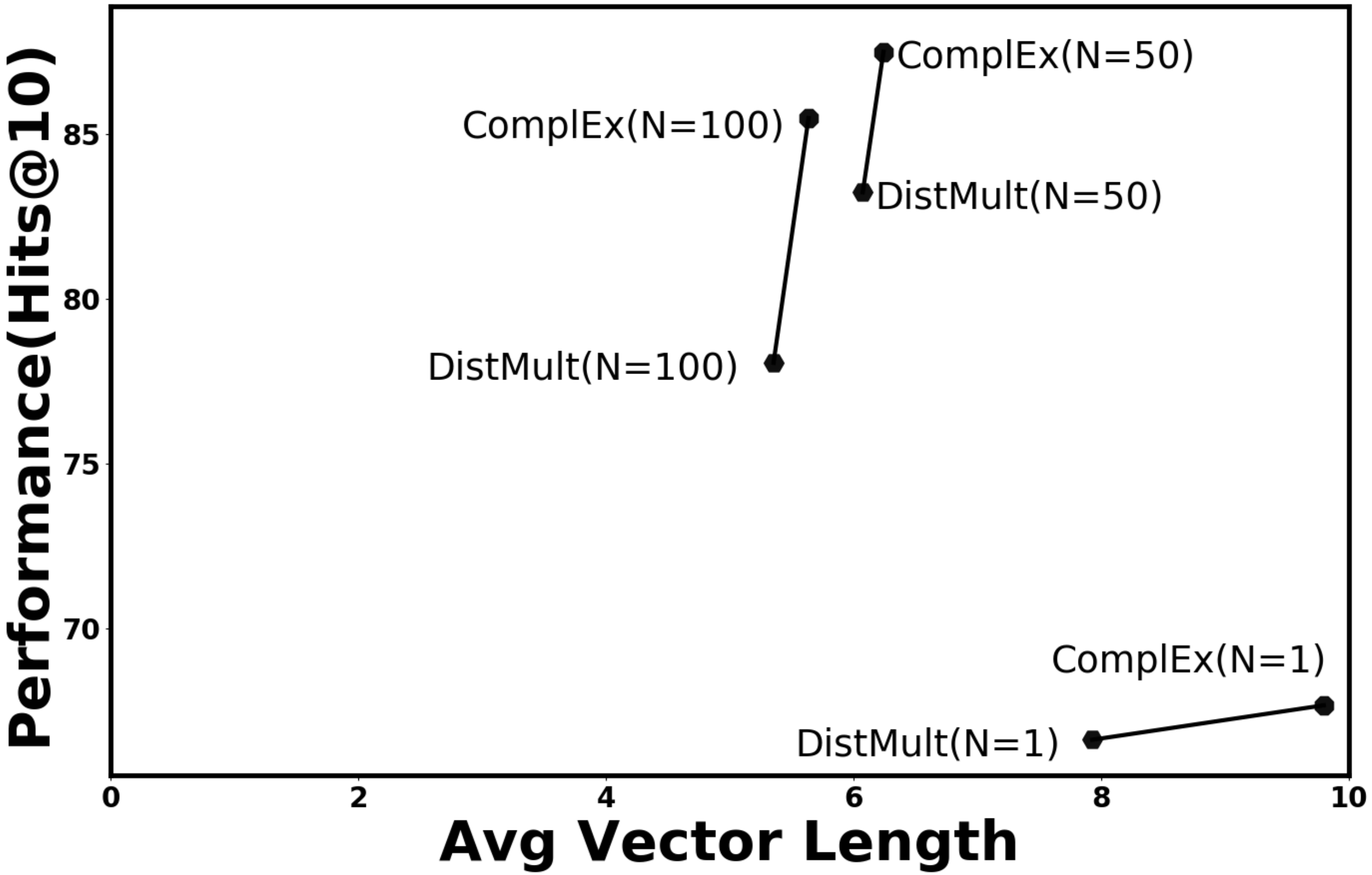
Additive and HolE

Correlation b/w Geometry and Performance



Positive Slope-
Positive Correlation

Correlation b/w Geometry and Performance



Positive Slope-
Positive Correlation

Higher Negatives-
Higher Slope
Magnitude

Correlation b/w Geometry and Performance

- **Additive:**
  No correlation between geometry and performance.

- **Multiplicative:**
  For fixed number of negative samples,
  - Conicity has negative correlation with performance
  - AVL has positive correlation with performance

Conclusion and Future Works

- We initiated the study of geometrical behavior of KG embeddings and presented various insights.
- Explore whether other entity/relation features (eg entity category) have any correlation with geometry.
- Explore other geometrical metrics which have better correlation with performance and use it for learning better KG embeddings.

Acknowledgements

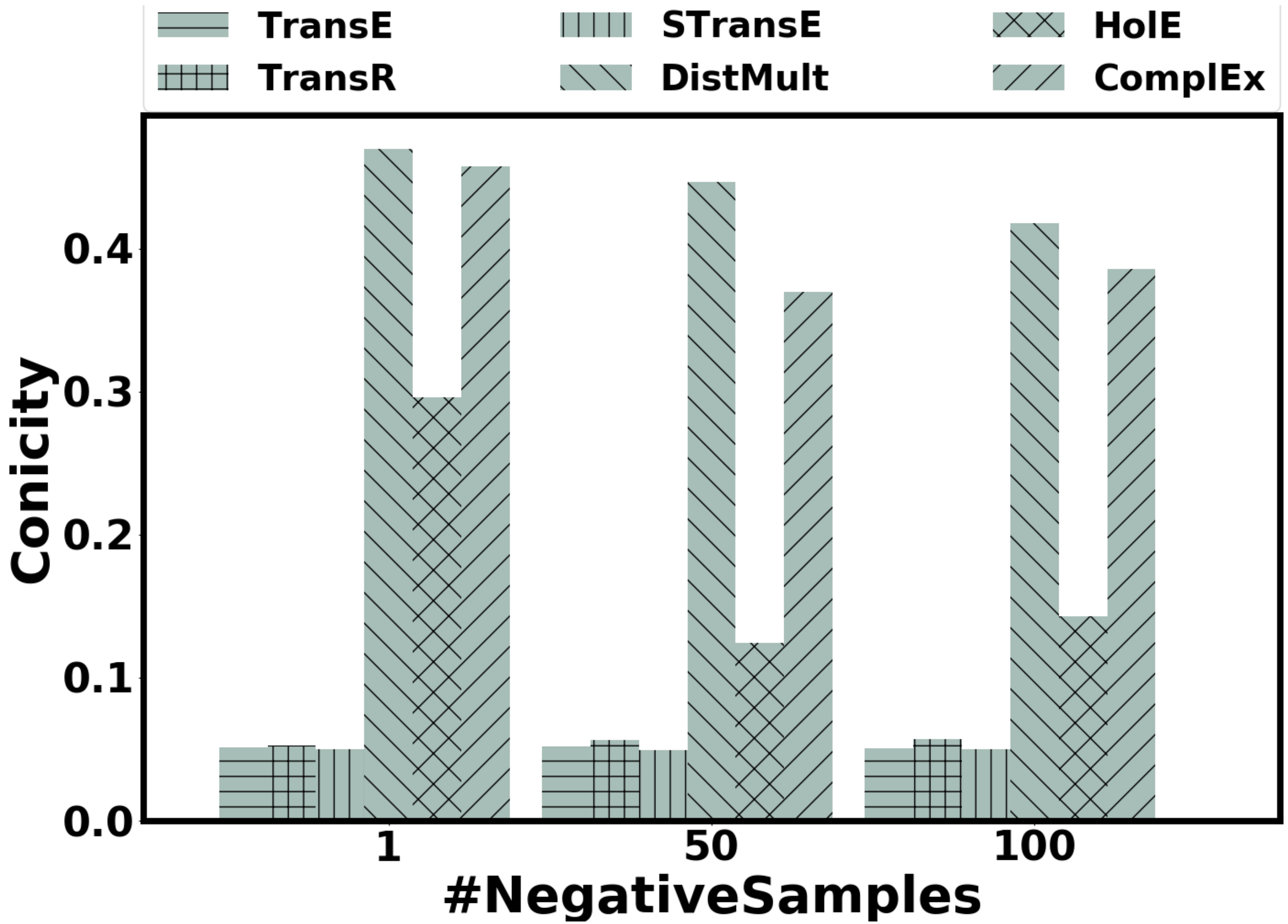
- We thank Google for the travel grant for attending ACL 2018.

- We thank MHRD India, Intel, Intuit, Google and Accenture for supporting our work.

- We thank the reviewers for their constructive comments.

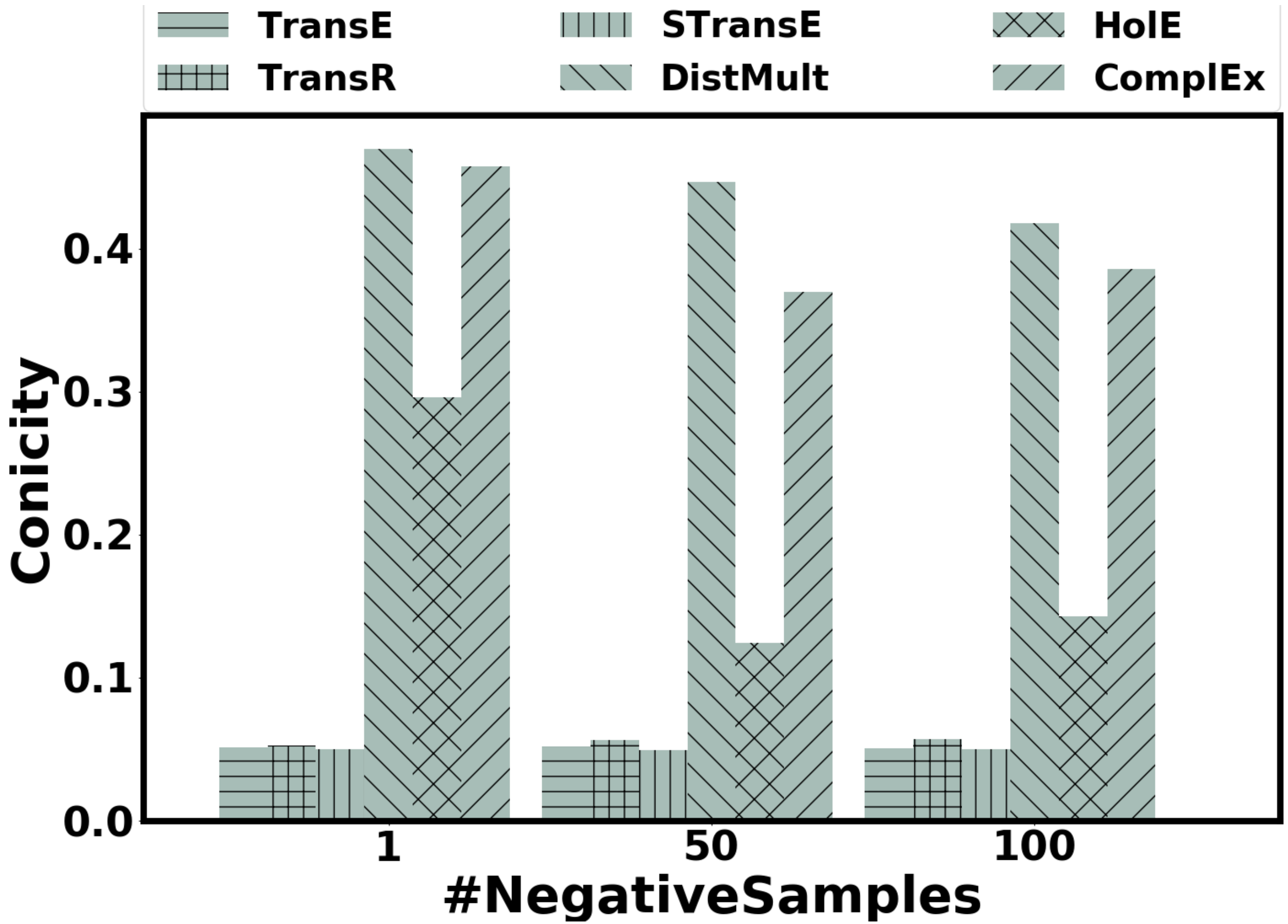
Thank you

Effect of #Negative Samples (Relation Vectors)



Additive
No change

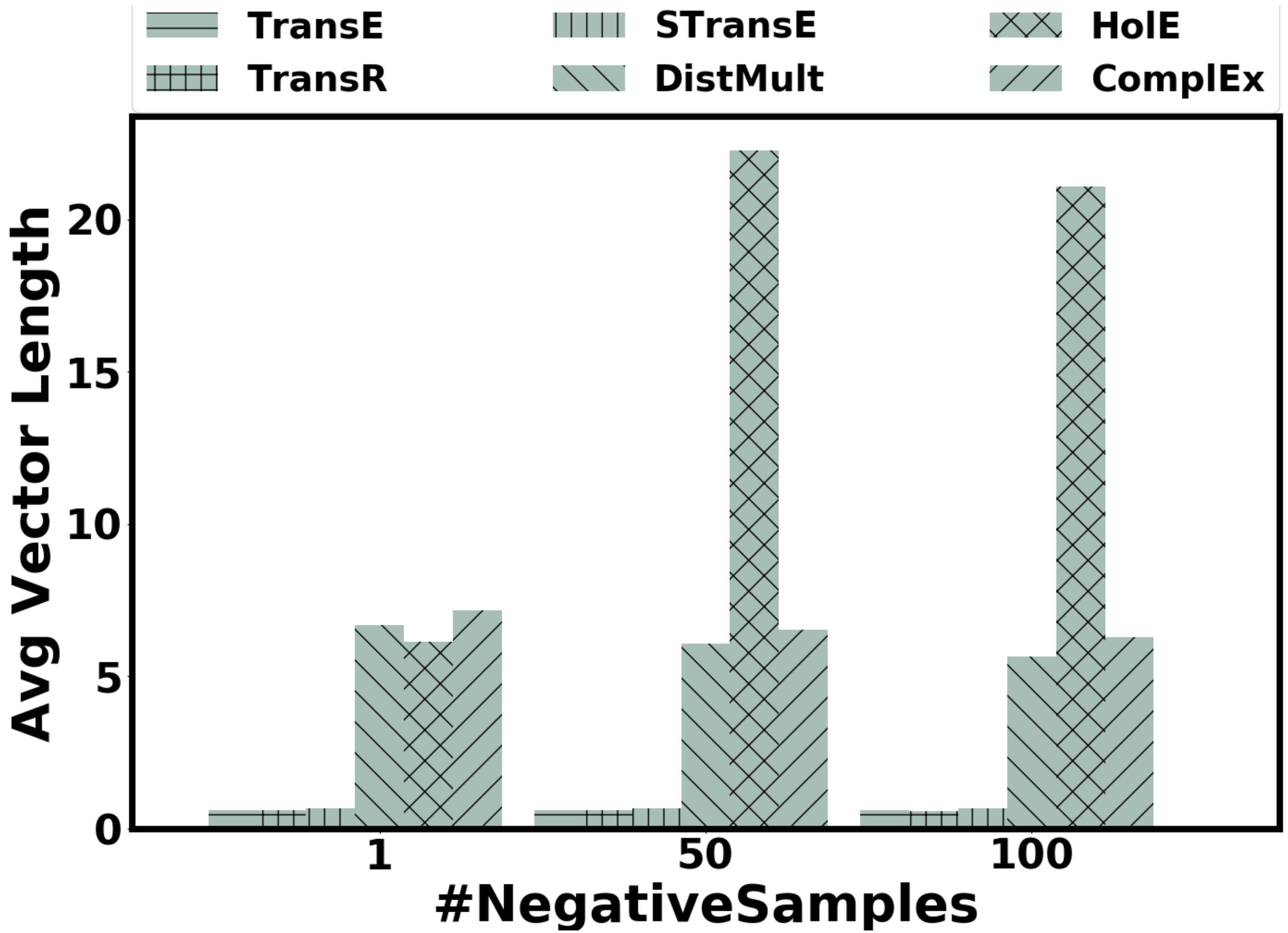
Effect of #Negative Samples (Relation Vectors)



Additive
No change

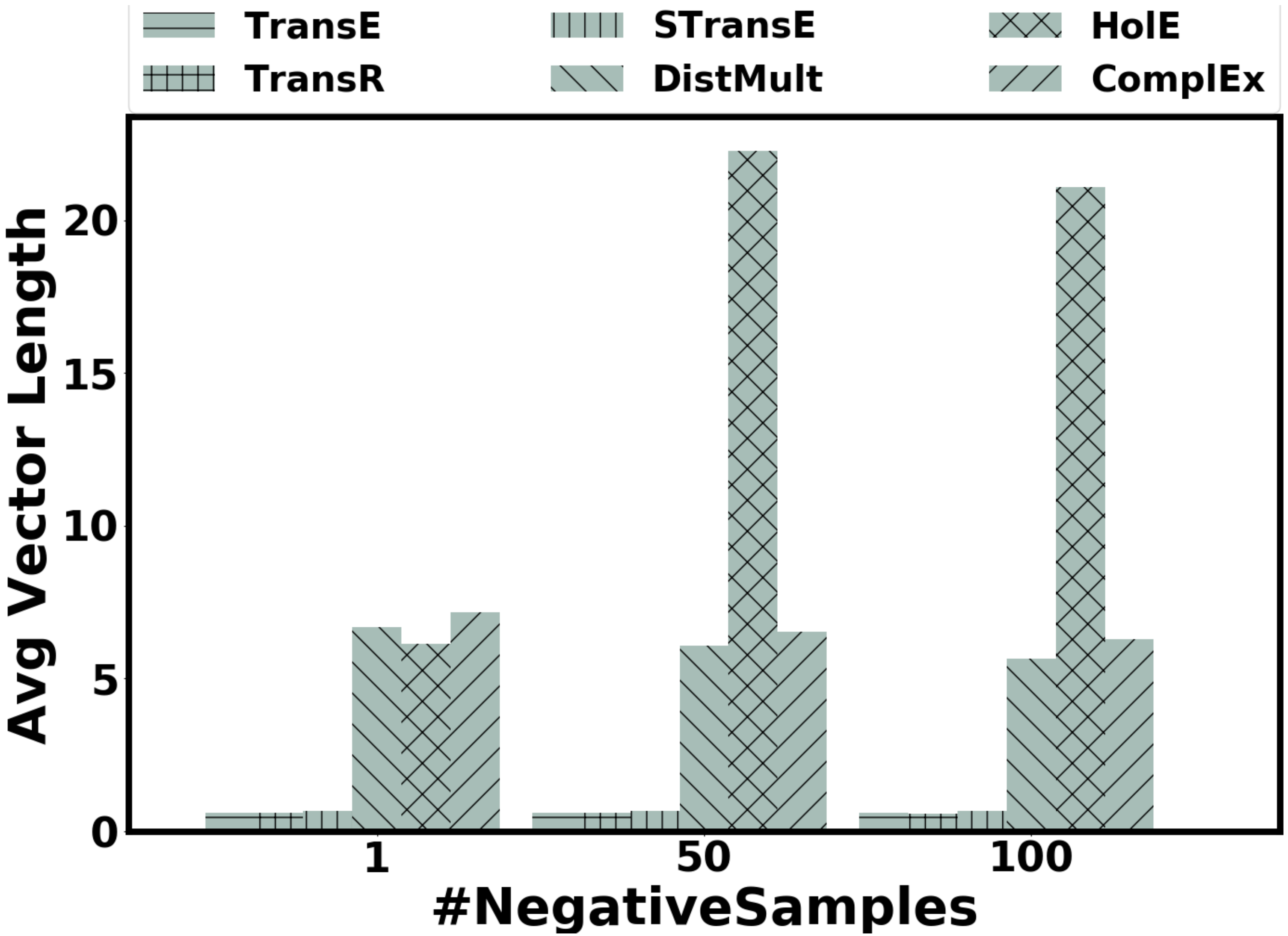
Multiplicative
Conicity decreases

Effect of #Negative Samples (Relation Vectors)



Additive
No change

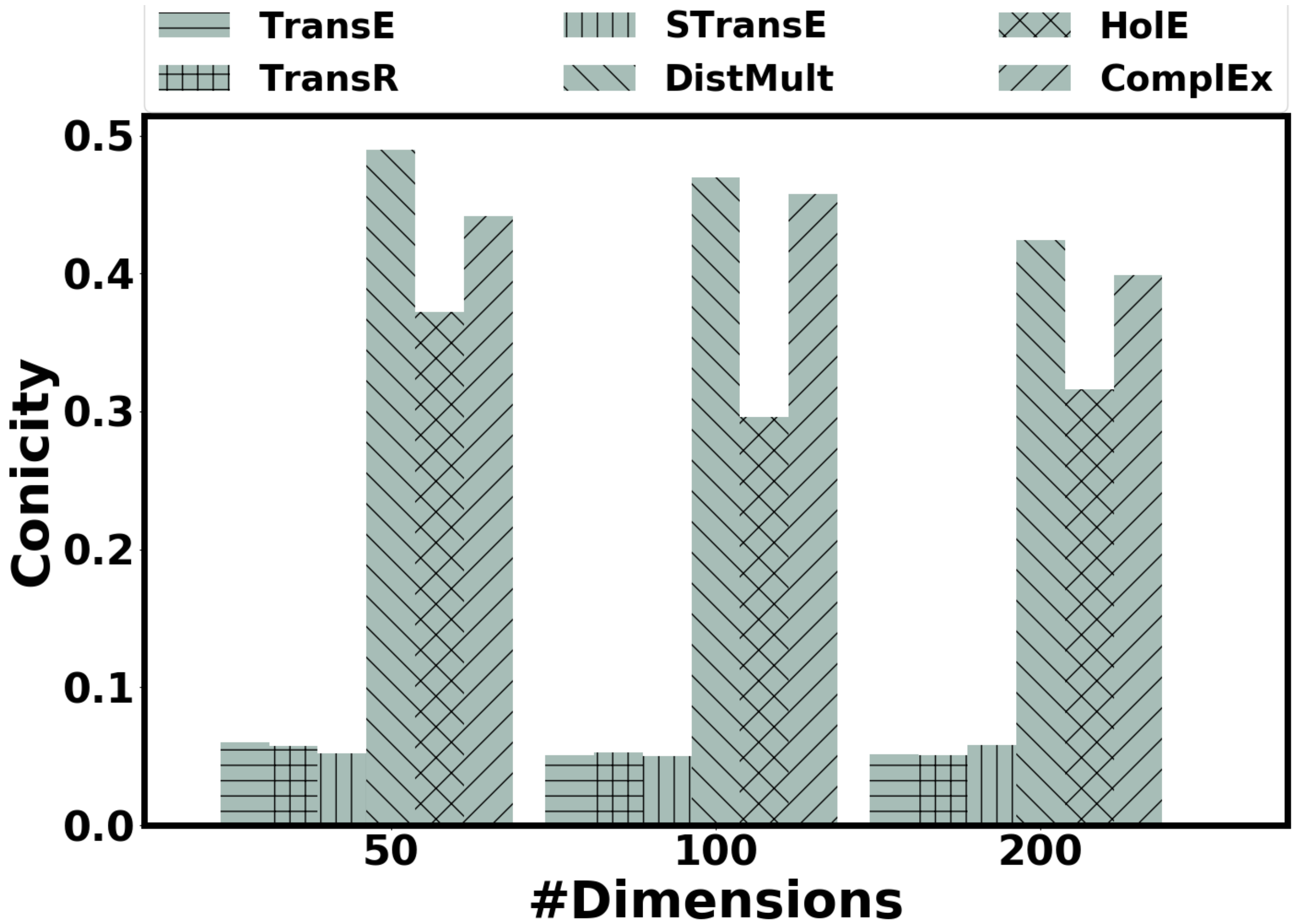
Effect of #Negative Samples (Relation Vectors)



Additive
No change

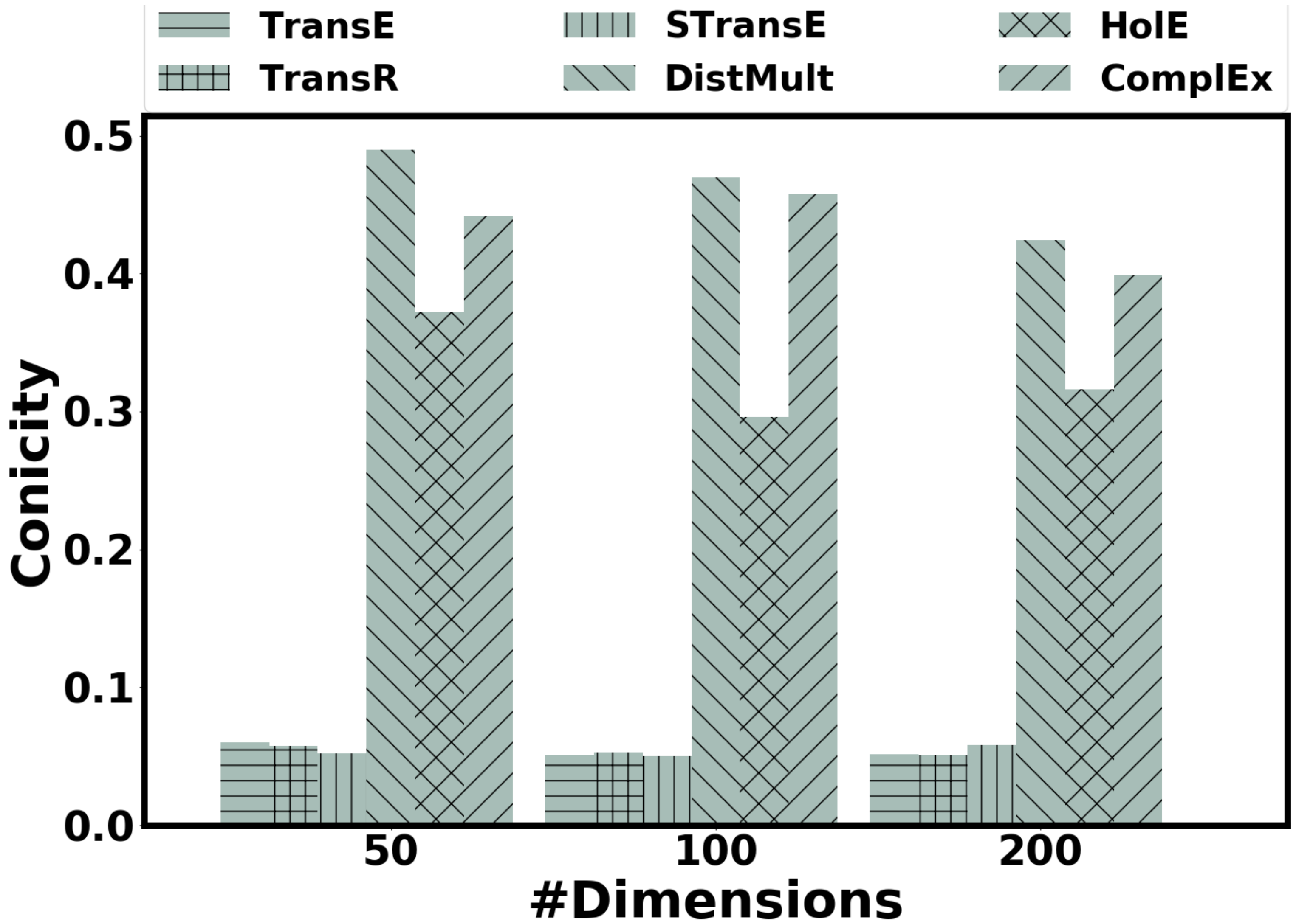
Multiplicative
No change except HolE

Effect of #Dimensions (Relation Vectors)



Additive
No change

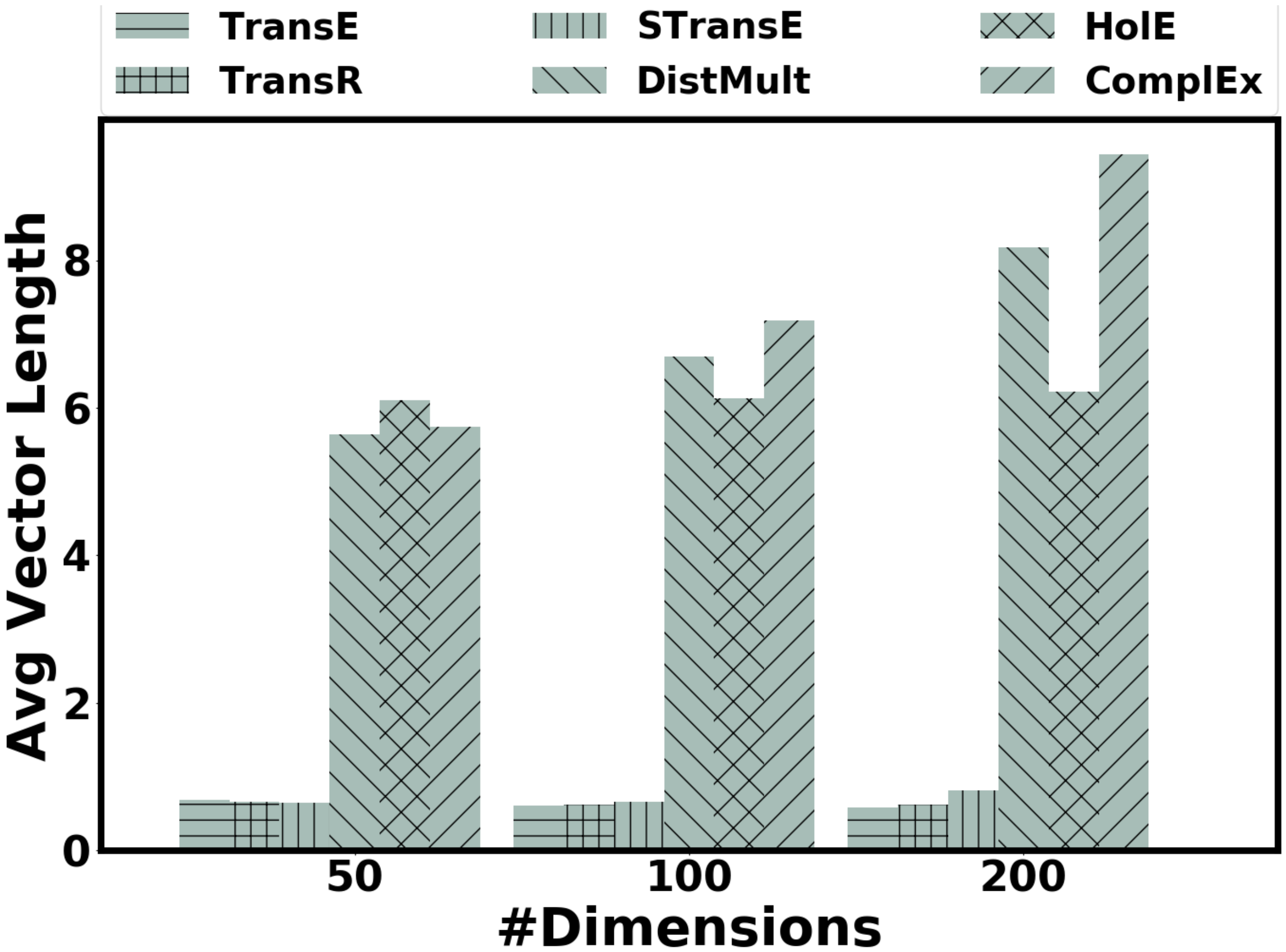
Effect of #Dimensions (Relation Vectors)



Additive
No change

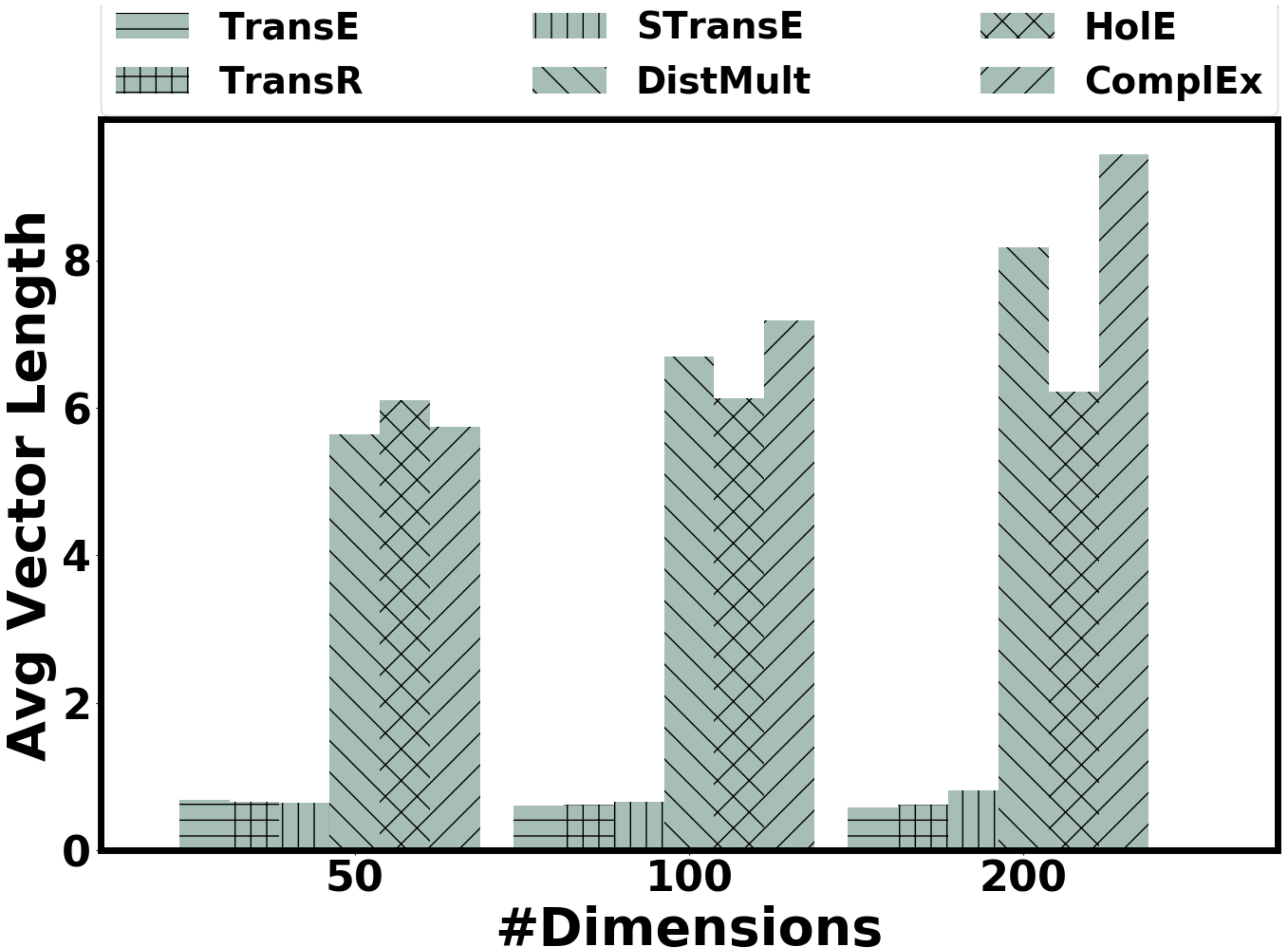
Multiplicative
Conicity decreases

Effect of #Dimensions (Relation Vectors)



Additive
No change

Effect of #Dimensions (Relation Vectors)



Additive
No change

Multiplicative
AVL Increases