APPENDIX A
DETAILS FOR SRKBE

Analysis of the geometric matrix. Recall Eq.(9), and here we will discuss some properties about it. The transformation matrix can be represented as follows:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \tag{14}$$

where $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$, $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$, $\boldsymbol{B} \in \mathbb{R}^{2\times 1}$, $\boldsymbol{C} \in \mathbb{R}^{1\times 2}$ and $\boldsymbol{D} \in \mathbb{R}^{1\times 1}$.

Then its inverse can be written as follows:

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{BWCA^{-1}} & -\boldsymbol{A}^{-1}\boldsymbol{BW} \\ -\boldsymbol{WCA}^{-1} & \boldsymbol{W} \end{bmatrix} \tag{15}$$

where $\boldsymbol{W} = (\boldsymbol{D} - \boldsymbol{CA}^{-1}\boldsymbol{B})^{-1}$.

Notice $\boldsymbol{A}^{-1} = \boldsymbol{A}^T$ since $\boldsymbol{A}$ is an orthogonal matrix. Then we have:

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{A}^T + \boldsymbol{A}^T\boldsymbol{BWCA^T} & -\boldsymbol{A}^T\boldsymbol{BW} \\ -\boldsymbol{WCA}^T & \boldsymbol{W} \end{bmatrix} \tag{16}$$

where $\boldsymbol{W} = (\boldsymbol{D} - \boldsymbol{CA}^T\boldsymbol{B})^{-1} = \frac{1}{\boldsymbol{D}-\boldsymbol{CA}^T\boldsymbol{B}}$ Since $\boldsymbol{D} - \boldsymbol{CA}^T\boldsymbol{B} \in \mathbb{R}^{1\times 1}$ is a scalar. Therefore, we can see that Eq.(17) does not involve matrix inverse operation, which greatly speeds up the calculation and saves the number of model parameters.

Here we discuss some special cases of Eq.(17).

(1) If $\boldsymbol{A} = \boldsymbol{I}$, then Eq.(17) can be written as follows:

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{I} + \boldsymbol{BWCA^T} & -\boldsymbol{BW} \\ -\boldsymbol{WC} & \boldsymbol{W} \end{bmatrix} \tag{17}$$

where $\boldsymbol{W} = (\boldsymbol{D} - \boldsymbol{CB})^{-1} \in \mathbb{R}^{1\times 1}$.

(2) If $\boldsymbol{B} = \boldsymbol{0}$, then Eq.(17) can be written as follows:

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{A}^T & \boldsymbol{0} \\ -\boldsymbol{WCA}^T & D^{-1} \end{bmatrix} \tag{18}$$

where $\boldsymbol{W} = \boldsymbol{D}^{-1} \in \mathbb{R}^{1\times 1}$.

(3) If $\boldsymbol{C} = \boldsymbol{0}$, then Eq.(17) can be written as follows:

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{A}^T & -\boldsymbol{A}^T\boldsymbol{BD}^{-1} \\ \boldsymbol{0} & D^{-1} \end{bmatrix} \tag{19}$$

Connection to existing models. The following proposition shows that some widely used models can be regarded as special cases of SRKBE in both binary and n-ary settings. The proof can be found in Appendix D.

Proposition 2 *The following statements hold:*

(a) *For the n-ary relational data learning, the m-cp model [17] can be regarded as the special case of SRKBE.*
(b) *For the binary relation data learning, the TransE [8], DisMult [72], and RotatE [53] can be generalized by the appropriate modification to SRKBE.*

In addition to SRKBE, other competitive models such as RAM and BoxE have shown remarkable performance. Specifically, RAM introduces a latent space for roles, allowing for the capture of semantically related roles. On the other hand, BoxE represents relations as explicit regions in the embedding space, scoring facts based on the positions of entity embeddings relative to relation boxes. However, both RAM and BoxE have a limitation in that they project all role-entity pairs onto a unified relational space, which hinders their ability to capture the interactions between relations and role-entity pairs effectively. Considering the benefits of semantic-based embeddings in SRKBE, introducing the paradigm into these models could be a promising and feasible solution. By doing so, these models could enhance their performance by more effectively capturing complex interactions between relations and role-entity pairs.

Algorithm 1: Training procedure for SRKBE

Input: N-ary relational KB $\mathcal{B} = (\mathcal{E}, \mathcal{R}, \mathcal{O}, \mathcal{F})$.
Output: Entity, role and relation embeddings

```
Initialize E_i for e_i ∈ ℰ, R_r^i for r ∈ ℛ, and α_i^r for o_i ∈ 𝒪
for t = 1, ··· , n_iter do
    Sample a mini-batch ℱ_batch ∈ ℱ of size m
    for each τ = r{o_1 : e_1, ··· , o_n : e_n} do
        Construct negative samples for fact x in Eq.(13)
        p_i^r ← compute role embeddings using Eq.(3)
        E_i ← compute entity embeddings using Eq.(5)
        R_r^i ← compute relation embeddings using Eq.(9)
        E'_i ← compute transformed entity embeddings using Eq.(8)
    end
    Computing the score function using Eq.(11)
    Update learnable parameters w.r.t gradient based on the whole objective in Eq.(12)
end
```

Apply SRKBE to other type KBs. Here we show how to apply SRKBE to other type KBs as follows:

Example 2 *As for the single-relation version formulated as* $\{r, e_1, e_2, \cdots, e_i, \cdots, e_n\}$*, noticing that it does not involve the information of role* o_i *for entity* e_i*, we use the position* s_i *of entity* e_i *to take the place of the role* o_i*. In this way, it can be formulated as* $r\{s_1 : e_1, s_2 : e_2, \cdots, s_i : e_i, \cdots, s_n : e_n\}$*, and we can conduct KBE for this n-ary relational version following the algorithm in the entity-role-relation version.*

Example 3 *As for relation-entity pairs version represented as* $\{e_1, r_1, \cdots, e_i, r_i, \cdots, e_n, r_n\}$*, it can be formulated as* $\{(r_1, e_1), \cdots, (r_i, e_i), \cdots, (r_n, e_n)\}$*. Due to the absence of a role, we use the position* s_i *of entity* e_i *to replace the role* o_i*. In this way, it can be formulated as* $\{(r_1, s_1 : e_1), (r_2, s_2 : e_2), \cdots, (r_i, s_i : e_i), \cdots, (r_n, s_n : e_n)\}$*, and we can conduct KBE for this n-ary relational version following the algorithm in the entity-role-relation version.*

Example 4 *As for triple-pairs version, which is represented as* $\{h, r, t, \cdots, r_i, e_i, \cdots, r_{n-2}, e_{n-2}\}$*, we reformulated it as* $\{(h, r, t), \cdots, (r_i, e_i), \cdots, (r_{n-2}, e_{n-2})\}$*, and further formulated as* $\{(r_1, h), (r_2, t), \cdots, (r_i, e_i), \cdots, (r_{n-2}, e_{n-2})\}$ *by decomposing the* r *into two subrelations* r_1 *and* r_2*. Noticing the absence of a role, we also use the position* s_i *of entity*

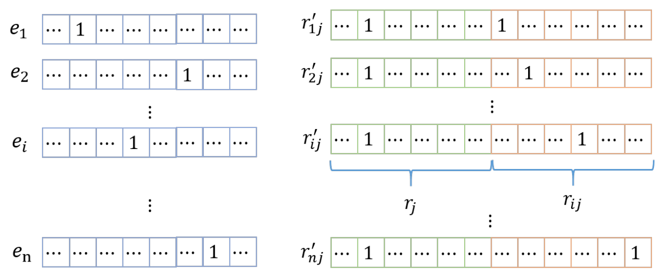


Fig. 6. An illustration of the proof process for the expressivity ability. $\boldsymbol{e}_i$ is the embeddings of the entity e_i, $\boldsymbol{r}_j$ is the embedding of relation r_j and $\boldsymbol{r}_{ij}$ represents the embedding of subrelation r_{ij} generated by r_j.

e_i to replace the role o_i. In this way, it can be formulated as $\{(r_h, s_h : e_h), (r_t, s_t : e_t), \cdots, (r_i, s_i : e_i), \cdots, (r_n, s_n : e_n)\}$, and we can conduct KBE for this n-ary relational version following the algorithm in the entity-role-relation version.

Appendix B
Proofs for Expressivity Ability

Proof *Let $\mathcal{F}$ be the set of all true facts in the n-ary relational KB. Noticing that the role is an attribute of an entity, here we prove the theorem by mainly considering entities and relations. Before we prove the theorem, we use an assignment of embedding values for each entity and relation in τ such that the scoring function of SRKBE is as follows:*

$$\phi(\tau) = \begin{cases} 1 & if\ \tau \in \mathcal{F} \\ 0 & otherwise \end{cases}$$

where τ is the fact. We begin the proof by first describing the embeddings of each of the entities and relations in SRKBE; we then move on to show that with such an embedding, SRKBE can represent any fact accurately. Without loss of generality, suppose that $|\mathcal{F}| > 0$. Let each entity $e \in \mathcal{E}$ be represented with a vector of length $\mathcal{E}$.

Given an n-ary relational fact $r_j(x_{i_1}, x_{i_2}, \cdots, x_{i_n})$ as shown in Figure 6, let $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \cdots, \boldsymbol{x}_{i_n}$ be the $|\mathcal{E}|$-dimensional one-hot binary vector representation of entities $x_{i_1}, x_{i_2}, \cdots, x_{i_n}$, and $\boldsymbol{r}_j$ be the $|\mathcal{R}|$ dimensional one-hot binary vector representation of a relation r_j. We let the j-th, i_1-th, i_2-th, $\cdots$ i_n-th element respectively of the corresponding vectors $\boldsymbol{r}_j$, $\boldsymbol{x}_{i_1}$, $\boldsymbol{x}_{i_2}$, $\cdots$, $\boldsymbol{x}_{i_n}$ be 1 and all other elements be 0.

Afterward, one can notice that the order of entities in the n-tuple plays an important part in the semantics, which may lead to the opposite meaning after changing the order of entity positions. In this sense, to model the position for each entity, we consider the semantic-based embedding $\boldsymbol{r}'_j$ for relation r_j, which can generate n sub-relations whose embedding is denoted as $\boldsymbol{r}_{ij}$ for each entity. Specifically, $\boldsymbol{r}'_j$ is composed by $\boldsymbol{r}_j$ ($|\mathcal{R}|$ dimensional one-hot binary vector) and $\boldsymbol{r}_{ij}$ (n dimensional one-hot binary vector). As shown in Figure 6, $\boldsymbol{r}_{ij}$ aims to represent the position of the entity in the n-ary relational fact, where the k-th position is 1 if x_i is in the k-th position and 0 otherwise.

In this way, given any n-ary relational fact, we can represent it with the $|\mathcal{E}|$ dimensional one-hot binary vector for entity and $n(|\mathcal{R}| + n)$ dimensional one-hot binary vector for relation in SRKBE.

Appendix C
Proofs for Modeling Three Patterns

Recall the score function Eq.(11), to simplify the denotations, we set the weighted parameters as 1, then we have

$$\begin{aligned} \phi(\tau) &= \sum_{i=1}^{n} \langle \boldsymbol{E}_1, \cdots, \boldsymbol{E}'_i, \cdots, \boldsymbol{E}_n \rangle \\ &= \sum_{i=1}^{n} \langle \boldsymbol{E}'_i, f(\boldsymbol{E}_1, \cdots, \boldsymbol{E}_n) \rangle \end{aligned} \tag{20}$$

where $f(\boldsymbol{E}_1, \cdots, \boldsymbol{E}_n)$ represents the multilinear product for $\boldsymbol{E}_1, \cdots, \boldsymbol{E}_{i-1}, \boldsymbol{E}_{i+1}, \cdots, \boldsymbol{E}_n$. In this way, we can view the score function as the superposition for the inner product of the entity (operated by the relation-care mechanism) and other entities. Here, we borrow the idea of proving the relation patterns in binary relational methods, *e.g.*, [53], [74]. Under such circumstances, $\boldsymbol{E}'_i$ should be equal to $f(\boldsymbol{E}_1, \cdots, \boldsymbol{E}_n)$ as much as possible since our goal is to maximize Eq.(20). For convenience, we use the case that x represents the entity to prove all the statements, and the conclusion for the case that x represents the role-entity pair can also be analogized accordingly. Consider the relation-care mechanism, then we will have the following formula for n-ary relational fact $r(e_1, \cdots, e_n)$:

$$\begin{gathered} r^1 \circ e_1 = e_2 \cdots e_n \\ \vdots \\ r^n \circ e_{a_r} = e_1 \cdots e_{n-1} \end{gathered}$$

where r^i is the sub-relation assigned by the model to the i-th role-entity pair under the relation r, $\circ$ represents the transformation operation of the relation on the entity, and the right-hand side of the equation represents the multilinear product of the elements. It can be observed that the system of equations satisfies the conditions for having a solution, and the right-hand side of the equations is non-zero, such that the nontrivial solution exists in the system of equations theoretically [50]. Then, we prove the propositions as follows.

Statement 1. SRKBE can infer symmetric (antisymmetric) pattern for n-ary relations.

Proof *We take e_i and e_j as examples to prove the theorem, and the conclusion for other entities can be analogized accordingly.*

If $r(e_1, \cdots, e_i, \cdots, e_j, \cdots, e_n)$ and $r(e_1, \cdots, e_j, \cdots, e_i, \cdots, e_n)$ hold, we have $r^i \circ e_i = e_1 \cdots e_{i-1} e_{i+1} \cdots e_n$, $r^j \circ e_j = e_1 \cdots e_{j-1} e_{j+1} \cdots e_n$, $r^i \circ e_j = e_1 \cdots e_{j-1} e_{j+1} \cdots e_n$, $r^j \circ e_i = e_1 \cdots e_{i-1} e_{i+1} \cdots e_n$, then we have

$$r^i \circ e_i = r^j \circ e_i$$

$$r^j \circ e_j = r^i \circ e_j$$

Therefore, we have $r^i = r^j$, which means r is symmetry for e_i and e_j. In conclusion, we have proved that SRKBE can infer the symmetric pattern.

If $r(e_1, \cdots, e_i, \cdots, e_j, \cdots, e_n)$ holds but $r(e_1, \cdots, e_j, \cdots, e_i, \cdots, e_n)$ does not hold, we can prove SRKBE can infer the antisymmetric pattern accordingly like above analysis.

Statement 2. SRKBE can infer inverse patterns for n-ary relations.

Proof *We take e_i and e_j as examples to prove the theorem and the conclusion for other entities can be analogized accordingly.*

If $r_1(e_1, \cdots, e_i, \cdots, e_j, \cdots, e_n)$ and $r_2(e_1, \cdots, e_j, \cdots, e_i, \cdots, e_n)$ hold, we have $r_1^i \circ e_i = e_1 \cdots e_{i-1} e_{i+1} \cdots e_n$, $r_1^j \circ e_j = e_1 \cdots e_{j-1} e_{j+1} \cdots e_n$, $r_2^i \circ e_j = e_1 \cdots e_{j-1} e_{j+1} \cdots e_n$, $r_2^j \circ e_i = e_1 \cdots e_{i-1} e_{i+1} \cdots e_n$, then we have

$$r_1^i \circ e_i = r_2^j \circ e_i$$
$$r_1^j \circ e_j = r_2^i \circ e_j$$

Therefore, we have $r_1^i = r_2^j$ and $r_1^j = r_2^i$, which means r_1 is reverse to r_2 for e_i and e_j. In conclusion, we have proved that SRKBE can infer the inverse pattern.

Statement 3. SRKBE can infer composition pattern for n-ary relations.

Proof *If $r_2(e_1, \cdots, e_i)$ holds, we have*

$$\begin{aligned} r_2^1 \circ e_1 &= e_2 \cdots e_i \\ &\vdots \\ r_2^i \circ e_i &= e_1 \cdots e_{i-1} \end{aligned} \tag{21}$$

If and $r_3(e_i, e_{i+1})$ holds, we have $r_3(e_i, e_{i+1})$:

$$r_3^i \circ e_i = e_{i+1} \tag{22}$$

Then we have $e_i = (r_3^i)^{-1} \circ e_{i+1}$, and take it into Eq.(23). Here we have

$$\begin{aligned} r_2^1 \circ e_1 &= e_2 \cdots e_{i-1} (r_3^i)^{-1} \circ e_i \\ &\vdots \\ (r_2^i \circ (e_3^i)^{-1}) \circ e_{i+1} &= e_1 \cdots e_{i-1} \end{aligned}$$

Then, we have

$$\begin{aligned} r_2^1 \circ e_1 ((r_3^i)^{-1} \circ e_i)^{-1} e_i &= e_2 \cdots e_{i-1} e_i \\ &\vdots \\ (r_2^i \circ (e_3^i)^{-1}) \circ e_{i+1} &= e_1 \cdots e_{i-1} \end{aligned}$$

Recall the geometric transformations in Eq.(9) can transform a point to any other point on a plane [50]. Then we denote $r_1^1 \circ e_1 = r_2^1 \circ e_1 ((r_3^i)^{-1} \circ e_i)^{-1} e_i$, where r_1^1 is a transformation matrix. We denote $r_1^i = r_2^i \circ (e_3^i)^{-1}$. Then we have

$$\begin{aligned} r_1^1 \circ e_1 &= e_2 \cdots e_i \\ &\vdots \\ r_1^i \circ e_i &= e_1 \cdots e_{i-1} \end{aligned} \tag{23}$$

We can see $r_1(e_1, \cdots, e_{i-1}, e_{i+1})$ holds, which means the relation r_1 is composed of relation r_2 and relation r_3. In conclusion, we have proved that SRKBE can infer the composition pattern.

Here we give some discussion about the hierarchy pattern, and the definition is as follows:

Definition 8 *An n-ary relation r_1 is hierarchy to relation r_2 for x_i if $\forall x_1, \cdots, x_N$*

$$r_2(x_1, \cdots, x_i, \cdots, x_N) \Rightarrow r_1(x_1, \cdots, x_i, \cdots, x_N)$$

A clause with such form is a hierarchy pattern.

For convenience, we use the case that x represents the entity to prove the statements, and the conclusion for the case that x represents the role-entity pair can also be analogized accordingly. Then, we have the statement that SRKBE can infer the hierarchy pattern for n-ary relations.

Proof *We take e_i as the example to prove the statement, and the conclusion for other entities can be analogized accordingly.*

If $r_1(e_1, \cdots, e_i, \cdots, e_n)$ and $r_2(e_1, \cdots, e_i, \cdots, e_n)$ hold, we have $r_1^i \circ e_i = e_1 \cdots e_{i-1} e_{i+1} \cdots e_n$, and $r_2^i \circ e_i = e_1 \cdots e_{i-1} e_{i+1} \cdots e_n$, then we have

$$r_1^i \circ e_i = r_2^i \circ e_i$$

Here we represent the r_1^i in Eq.(9) as the translation, and r_2^i Eq.(9) as the combination of translation and rotation [50]. Then we can see that transformations represented by r_2^i involve the transformations represented by r_2^i. According to the analysis above, we can also derive that transformations represented by r_2^i $(i = 1, \cdots, n)$ involve the transformations represented by r_2^i $(i = 1, \cdots, n)$. Then we have $r_2(e_1, \cdots, e_i, \cdots, e_n) \Rightarrow r_1(e_1, \cdots, e_i, \cdots, e_n)$. In conclusion, we have proved that SRKBE can infer the hierarchy pattern.

Appendix D
Proofs for Connection to Existing models

Statement 1. For the n-ary relational data learning, the m-cp model can be regarded as a special case of SRKBE.

Proof *Recall the score function Eq.(11); if we remove the relation-care mechanism, then we set the following score function:*

$$\phi(\tau) = \langle \boldsymbol{R} \circ \boldsymbol{E_1}, \boldsymbol{E_2}, \cdots, \boldsymbol{E_n} \rangle$$

Here we let the geometric transformations represented by r degenerate into the inner product of $\boldsymbol{R}$ and $\boldsymbol{E}_1$, then we have:

$$\begin{aligned} \phi(\tau) &= \langle \boldsymbol{R E_1}, \boldsymbol{E_2}, \cdots, \boldsymbol{E_n} \rangle \\ &= \langle \boldsymbol{R}, \boldsymbol{E_1}, \boldsymbol{E_2}, \cdots, \boldsymbol{E_n} \rangle \end{aligned}$$

Then we have proved that the m-cp model can be regarded as the special case of SRKBE.

Statement 2. For the binary relation data learning, the TransE, DisMult, and RotatE can be generalized by the appropriate modification to SRKBE.

Proof *Recall the score function Eq.(11); if we remove the relation-care mechanism and maintain two entities, the n-ary relation will degenerate into a binary relation. We have the following score function:*

$$\begin{aligned} \phi(\tau) &= \langle \boldsymbol{R} \circ \boldsymbol{E_1}, \boldsymbol{E_2}, \cdots, \boldsymbol{E_n} \rangle \\ &= \langle \boldsymbol{R} \circ \boldsymbol{E_1}, \boldsymbol{E_2} \rangle \end{aligned}$$

TABLE IX
DATASET STATISTICS. "ARITY" DENOTES THE INVOLVED ARITIES OF RELATIONS. "$\geq$ 5-ARY" DENOTES THE NUMBER OF FACTS WITH 5-ARY RELATIONS AND BEYOND. NOTICING THAT WN18, FB15K, WN18RR, AND FB15K237 ARE BINARY RELATIONAL DATASETS, HERE "-" DENOTES THAT THEY DO NOT INVOLVE THE HIGHER-ARY RELATIONAL FACTS.

Dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	Arity	Train	Valid	Test	2-ary	3-ary	4-ary	$\geq$ 5-ary
WikiPeople	47,765	707	2-9	305,725	38,223	38,281	337,914	25,820	15,188	3,307
JF17K	28,645	322	2-6	61,104	15,275	24,568	54,627	34,544	9,509	2,267
FB-AUTO	3,388	8	2,4,5	6,778	2,255	2,180	3,786	0	215	7,212
WN18RR	40,943	11	2	86,835	3,034	3,134	93,003	-	-	-
FB15k237	14,541	237	2	272,115	17,535	20,466	310,116	-	-	-

Noticing the Eq.(9), if we set T_1, T_3 to be $\mathbf{0}$ *and T_4 to be* 1, *the relation represents the translation. Here, we can see that TransE can be regarded as a special case of SRKBE.*

Noticing the Eq.(9), if we set T_2, T_3 to be $\mathbf{0}$, *T_4 to be* 1 *and T_1 to be a diagonal matrix, the relation can capture the symmetry pattern. Here, we can see that DisMult can be regarded as a special case of SRKBE.*

Noticing the Eq.(9), if we set T_2, T_3 to be $\mathbf{0}$, *T_4 to be* 1, *and T_1 to be a givens matrix, the relation represents the rotation. Here, we can see that RotatE can be regarded as a special case of SRKBE.*

APPENDIX E
EXPERIMENTAL DETAILS

In experiments, we performed a grid search, and the hyper-parameters are detailed in Table X.

TABLE X
HYPER-PARAMETERS OF SRKBE.

Dataset	learning_rate	batch_size	dimension	epoch
JF17K	0.005	64	50	70
WikiPeople	0.003	64	30	50
FB-AUTO	0.005	64	50	70
WN18	0.001	128	100	100
FB15K	0.005	128	100	100
WN18RR	0.001	64	100	100
FB15K237	0.005	64	100	100

TABLE XI
STATISTICS OF THE DATASETS USED IN THIS PAPER. (NUM_e REPRESENTS THE NUMBER OF ENTITIES AND NUM_r REPRESENTS THE NUMBER OF RELATIONS.)

Dataset	Num_e	Num_r	Training	Validation	Test	Scale
WN9-IMG	7k	9	12k	1k	1k	Small
FB-IMG	11k	1231	285k	29k	34k	Large

The impact of the embedding dimensions. To explore the impact of dimensionality, we varied the dimensional settings within the range of $\{4, 8, 10, 20, 30, 40, 50, 80, 100\}$, and the results are presented in Figure 7. From the figure, we can observe a clear trend: as the dimension number increases, the model's performance improves. This suggests that increasing the dimensionality of the embeddings allows the model to capture more intricate patterns and relationships within the data. However, it is noteworthy that the performance growth curve starts to flatten when the dimension exceeds 50. This indicates that there is a diminishing marginal return in performance improvement beyond this point. In this sense, our model demonstrates the ability to achieve satisfactory performance while maintaining a lower space complexity.

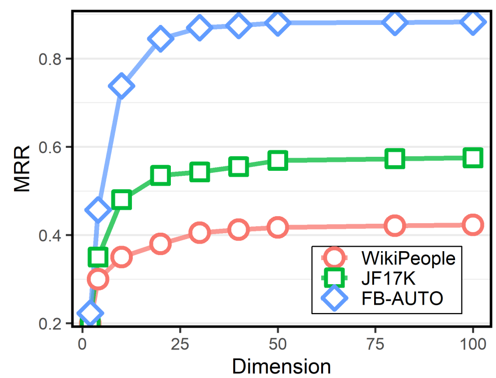


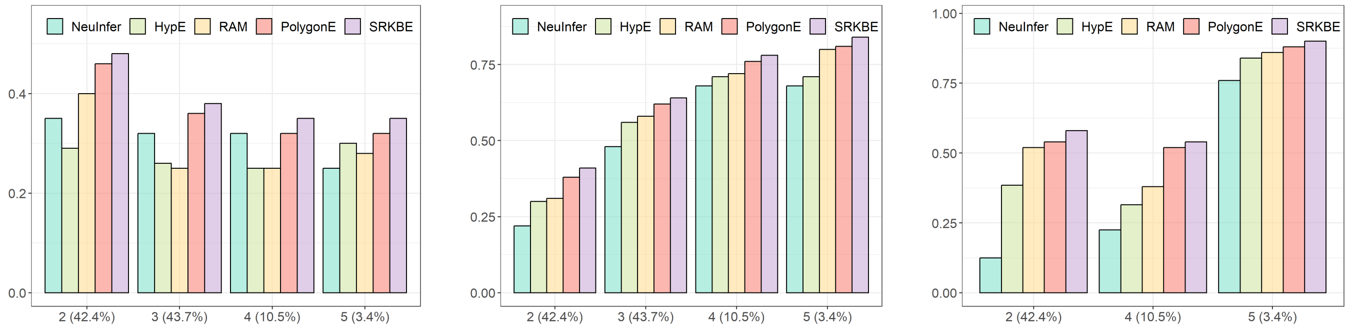
Fig. 7. The impact of the embeddings dimensions for SRKBE on WikiPeople, JF17K, and FB-AUTO datasets.

TABLE XII
THE LINK PREDICTION RESULTS FOR ABLATION STUDY OF ROLE AND POSITOION INFORMATION ON WIKIPEOPLE DATASET.

Model	MRR	Hits@1	Hits@3	Hits@10
SRKBE w/o r	0.428	0.557	0.455	0.328
SRKBE w/o p	0.432	0.565	0.461	0.332
SRKBE	**0.449**	**0.573**	**0.473**	**0.346**

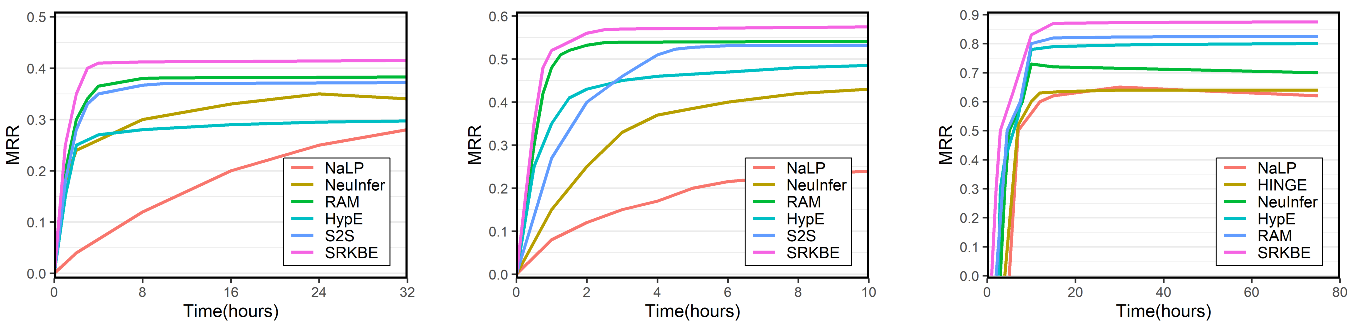
Training efficiency. Here, we study the training efficiency of SRKBE. As shown in Figure 9, we can see that the neural network models, *i.e.*, NaLP, and NeuInfer, require quite a long time to converge. That is because these two models utilize complex neural networks for training. Among all models, SRKBE can converge after 6 hours in the WikiPeople dataset and converge after 3 hours in the JF17K dataset, while other models need more time. Therefore, we can see that SRKBE achieves the fastest convergence, demonstrating that it requires less complexity.

The ablation study for positional information and role information. Recall we have combined the role information and position information $\boldsymbol{\beta}_i^r = \boldsymbol{o}_i^r + \boldsymbol{\alpha}_i^r$ as shown in Eq.(3). To



(a) WikiPeople. (b) JF17K. (c) FB-AUTO.

Fig. 8. Breakdown Performance of MRR for single-arity on WikiPeople, JF17K, and FB-AUTO datasets, where the $2(42.4\%)$ on the horizontal coordinate indicates that the proportion of binary relational facts to all facts in the dataset is 42.4%.



(a) WikiPeople. (b) JF17K. (c) FB-AUTO.

Fig. 9. Comparison of clock time of model training vs. testing MRR metric between SRKBE and other baselines on WikiPeople, JF17K, and FB-AUTO datasets.

study the effectiveness of different components, we denote the version removed the role information as SRKBE w/o r, *i.e.,* $\boldsymbol{\beta}_i^r = \boldsymbol{\alpha}_i^r$; denote the version removed the position information as SRKBE w/o p, *i.e.,* $\boldsymbol{\beta}_i^r = \boldsymbol{o}_i^r$. We conduct the ablation study in WikiPeople dataset as shown in Table XII. One can observe that positional information and role information are all essential to the model.

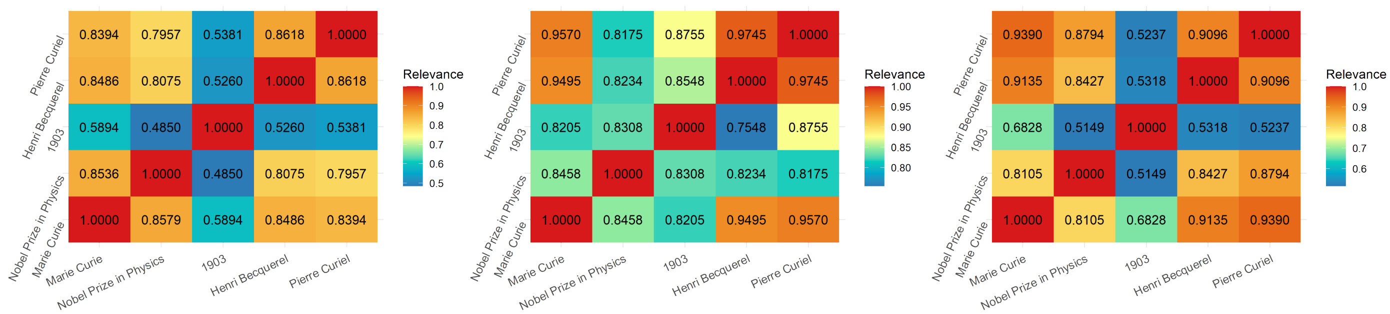
The ablation study of n-ary. To evaluate the performance of SRKBE in modeling different n-ary relational tuples, we conducted experiments on the WikiPeople, JF17K, and FB-AUTO datasets. Specifically, we trained SRKBE, PolygonE, RAM, and HypE on n-ary relational data and reported the results for each arity. As depicted in Figure 8, we observed that SRKBE outperformed all other models on all single-arity relational knowledge bases. This finding confirms the effectiveness of SRKBE in capturing relational information. Furthermore, these results highlight the advantages of studying n-ary relational knowledge base embedding methods that can handle mixed arities. It is important to note that SRKBE demonstrated significant improvements over these competitive models on higher-arity data. This can be attributed to the complexity of higher-arity facts, which shows that learning KB embeddings, in this case, is more challenging for other models.

Comparison with GNN-based methods To study the semantic learning representations for our method, we compare SRKBE with some popular GNN-based models on WikiPeople dataset. Specifically, we substitute the original semantic-based embedding learning process in SRKBE with a GNN-based approach, employing models such as GCN [30], GAT [57], GIN [70], and GraphSAGE [22]. These variations are referred to as SRKBE w GCN, SRKBE w GAT, SRKBE w GIN, and SRKBE w GraphSAGE, respectively. As illustrated in Table XV, it is evident that the original SRKBE outperforms these adaptations. This superior performance can be attributed to two main factors. Firstly, conventional GNN-based models generally aggregate features from neighboring nodes without adequately considering the holistic semantic relevances among these nodes in the n-tuple. Secondly, unlike these models, SRKBE additionally produces semantic-based embeddings for relations, thereby enhancing the nuanced interplay between entities and their relationships. This comparison underscores the efficacy and distinct advantages of SRKBE in comparison to conventional GNN-based methods.

Binary relational data. We investigate the effectiveness of SRKBE in learning binary relation data within N-ary relational datasets. We conduct experiments on the JFD17K, WikiPeople, and FB-AUTO datasets, focusing solely on learning and predicting binary relational tuples. As shown in Table XIII, the results demonstrate that SRKBE achieves state-of-the-art performance. For instance, on the JF17K dataset, SRKBE significantly improved MRR from 0.333 to 0.359.

TABLE XIII
BINARY RELATIONAL KNOWLEDGE BASE EMBEDDING RESULTS ON WIKIPEOPLE, JF17K, AND FB-AUTO DATASETS. THE BEST RESULTS ARE IN **BOLD**, AND THE SECOND-BEST RESULTS ARE UNDERLINED

Model	WikiPeople MRR	WikiPeople Hits@10	WikiPeople Hits@1	JF17K MRR	JF17K Hits@10	JF17K Hits@1	FB-AUTO MRR	FB-AUTO Hits@10	FB-AUTO Hits@1
TransE	0.312	0.574	0.146	0.276	0.495	0.167	0.313	0.562	0.132
DistMult	0.275	0.388	0.193	0.228	0.411	0.144	0.494	0.566	0.444
ComplEx	0.326	0.461	0.232	0.308	0.498	0.219	0.487	0.568	0.442
SimplE	0.326	0.449	0.249	0.313	0.502	0.224	0.493	0.577	0.440
RotatE	0.422	0.519	0.285	0.304	0.496	0.210	0.470	0.577	0.408
TuckER	0.429	0.538	0.365	0.333	0.512	0.244	0.510	0.621	0.450
BoxE	0.441	0.558	0.370	0.328	0.509	0.241	0.527	0.638	0.475
RAM	0.445	0.562	**0.374**	0.324	0.508	0.234	0.518	0.604	0.468
SRKBE	**0.449**	**0.574**	0.373	**0.359**	**0.535**	**0.259**	**0.551**	**0.662**	**0.490**



(a) SRKBE w/o relevance subrelations.
(b) SRKBE w/o subrelations.
(c) SRKBE.

Fig. 10. The visualization on addressing the problem of relational semantic imbalance in SRKBE.

TABLE XIV
THE LINK PREDICTION RESULTS ON THE JF17K-3 AND JF17K-4. THE BEST RESULTS ARE IN **BOLD** AND THE SECOND-BEST RESULTS ARE UNDERLINED.

Model	JF17K-3 MRR	JF17K-3 Hits@10	JF17K-3 Hits@3	JF17K-3 Hits@1	JF17K-4 MRR	JF17K-4 Hits@10	JF17K-4 Hits@3	JF17K-4 Hits@1
RAE	0.505	0.644	0.532	0.430	0.707	0.835	0.751	0.636
NaLP	0.515	0.679	0.552	0.431	0.719	0.805	0.742	0.673
HINGE	0.587	0.738	0.621	0.509	0.745	0.842	0.775	0.700
NeuInfer	0.622	0.770	0.658	0.533	0.765	0.871	0.808	0.722
n-TuckER	0.727	0.852	0.761	0.664	0.804	0.902	0.841	0.748
PolygonE	0.735	0.846	0.753	0.668	0.815	0.923	0.850	0.755
GETD	0.732	0.856	0.764	0.669	0.810	0.913	0.844	0.755
BoxE	0.738	0.854	0.776	0.673	0.818	0.915	0.846	0.753
RAM	0.739	0.852	0.778	0.674	0.788	0.887	0.822	0.733
S2S	0.740	0.860	0.770	0.676	0.822	0.924	0.853	0.761
SRKBE	**0.753**	**0.872**	**0.788**	**0.694**	**0.830**	**0.935**	**0.868**	**0.787**

TABLE XV
THE LINK PREDICTION RESULTS FOR COMPARISON BETWEEN SRKBE AND GNN-BASED MATHODS.

Model	MRR	Hits@1	Hits@3	Hits@10
SRKBE w GCN	0.410	0.542	0.438	0.319
SRKBE w GAT	0.418	0.543	0.446	0.324
SRKBE w GIN	0.425	0.553	0.458	0.335
SRKBE w GraphSAGE	0.428	0.556	0.461	0.339
SRKBE	**0.449**	**0.573**	**0.473**	**0.346**

N-ary relational data with a fixed arity. To further assess the performance of SRKBE, we compare it with other models using fixed arity conditions on n-ary relational datasets. Specifically, we evaluate SRKBE on JF17K-3 (which only includes 3-ary relational data) and JF17K-4 (which only includes 4-ary relational data). It is worth noting that these datasets contain a wide range of relation patterns [15]. From the results presented in Table XIV, we observe that SRKBE consistently outperforms other models across both datasets. This suggests that SRKBE's ability to model relation patterns and enable reasoning contributes to its superior performance. By effectively capturing these patterns, SRKBE can learn more accurate embeddings, leading to improved performance in knowledge base completion tasks. The comparison with other models on fixed arity n-ary relational datasets further reinforces the effectiveness of SRKBE in addressing the challenges of entity relevance and relation imbalance. Its capability to capture and reason with relation patterns is crucial in achieving superior performance in these scenarios.

Case study. We select a typical case from the validation set of WikiPeople, which is predicted correctly by SRKBE, to carry out an overall relatedness analysis. Given a 5-ary relational fact: *Win_Award_In_With*{PERSON: Marie Curie, AWARD: Nobel Prize in Physics, TIME: 1903, PERSON: Henri Becquerel, PERSON: Pierre Curie}, it implies that PERSON : Marie Curie *Win* AWARD: Nobel Prize in Physics in TIME: 1903 *with* PERSON: Henri Becquerel and PERSON: Pierre Curie. We analyze the overall relatedness learned by SRKBE using a correctly predicted 5-ary fact from the WikiPeople validation set. The fact states that Marie Curie won the Nobel Prize in Physics in 1903 with Henri Becquerel and Pierre Curie. As shown in Figure 10(a), the model assigns a higher relevance score between Marie Curie and the Nobel Prize in Physics compared to Henri Becquerel and Pierre Curie. This is counterintuitive since all three are classified as PERSON category and the Nobel Prize in Physics belongs to AWARD category. Figure 10 demonstrates the effectiveness of the semantic relevance mechanism in addressing the issue above. However, another drawback is evident that the embeddings learned for entities within n-ary relations are very similar, making it difficult to capture fine-grained semantic differences. As shown in Figure 10(c), subspaces can effectively address this limitation by promoting interaction between relations and individual entities, allowing for the full capture of fine-grained semantics. In this way, it demonstrates the effectiveness of SRKBE in reasoning for KBs.

APPENDIX F
PROOFS OF THEOREM 2

Definition 9 *$f \in \mathcal{F}$ is called K-Lipschitz continuous, $\forall \boldsymbol{a}, \boldsymbol{b} \in \mathcal{D}$ (where $\mathcal{D} \in \mathbb{R}^n$) if $|f(\boldsymbol{a}) - f(\boldsymbol{b})| \leq Kd(\boldsymbol{a}, \boldsymbol{b})$.*

Here are the proof for Theorem 1:

Proof *First of all, we prove that $|f - f'|$ is $2K$-Lipschitz continuous given K-Lipschitz continuous hypotheses $f, f' \in \mathcal{F}$. we can derive the following formula with using the triangle inequality:*

$$\begin{aligned} |f(x) - f'(x)| &\leq |f(x) - f(y)| + |f(y) - f'(x)| \\ &\leq |f(x) - f(y)| + |f(y) - f'(y)| + |f'(y) - f'(x)| \end{aligned} \tag{24}$$

Suppose $d(x, y)$ represents a function to measure the distance between x and y, for every $x, y \in \mathcal{X}$, then we have:

$$\begin{aligned} \frac{|f(x) - f'(x)| - |f(y) - f'(y)|}{d(x,y)} &\leq \frac{|f(x) - f(y)| + |f'(x) - f'(y)|}{d(x,y)} \\ &\leq 2K \end{aligned} \tag{25}$$

In this step, we can find that for every hypothesis f, f', given two distributions μ_s and μ_t (here μ_s is the multi-modal distribution while μ_t is the structural distribution), here we have

$$\begin{aligned} \epsilon_t(f, f') - \epsilon_s(f, f') &= \mathbb{E}_{x \sim \mu_t}[|f(x) - f'(x)|] - \mathbb{E}_{x \sim \mu_s}[|f(x) - f'(x)|] \\ &\leq \sup_{\|f\|_L \leq 2K} \mathbb{E}_{\mu_t}[f(x)] - \mathbb{E}_{\mu_s}[f(x)] \\ &\leq 2Kd(\mu_s, \mu_t) \end{aligned} \tag{26}$$

where $d_(\mu_s, \mu_t)$ is the distributional distance such as 1-Wasserstein distance. Then we can derive the following formula:

$$\epsilon_t(f) \leq \epsilon_s(f) + 2Kd(\mu_s, \mu_t) \tag{27}$$

By changing s,t, we have:

$$\begin{aligned} \epsilon_I(f) &\leq \epsilon_F(f) + 2Kd_1(\mu_I, \mu_F) \\ \epsilon_G(f) &\leq \epsilon_F(f) + 2Kd_1(\mu_G, \mu_F) \\ \epsilon_F(f) &\leq \epsilon_I(f) + 2Kd_1(\mu_I, \mu_F) \\ \epsilon_F(f) &\leq \epsilon_G(f) + 2Kd_1(\mu_G, \mu_F) \end{aligned}$$

Then the proof is completed.