Paper # 2995
@ NeurIPS'19

# Quantum Embedding of Knowledge for Reasoning

Dinesh Garg[1*], Shajith Ikbal[1*], Santosh K Srivastava[1], Harit Vishwakarma[2†], Hima Karanam[1], L Venkata Subramaniam[1]

[1]IBM Research AI, India
[2]Dept. of Computer Sciences, University of Wisconsin-Madison, USA
garg.dinesh, shajmoha, sasriva5@in.ibm.com, hvishwakarma@cs.wisc.edu, hkaranam, lvsubram@in.ibm.com

[*] The first two authors contributed equally.
[†] This work was done when the author was with IBM Research, New Delhi, India.

# What and Why?

What are we trying to do?

1. Design a technique to embed Symbolic Knowledge Base (KB) into a vector space while preserving its logical structure.

2. One should be able to perform logical operations on such embeddings in a manner similar to the Boolean Logical operations on a symbolic KB.

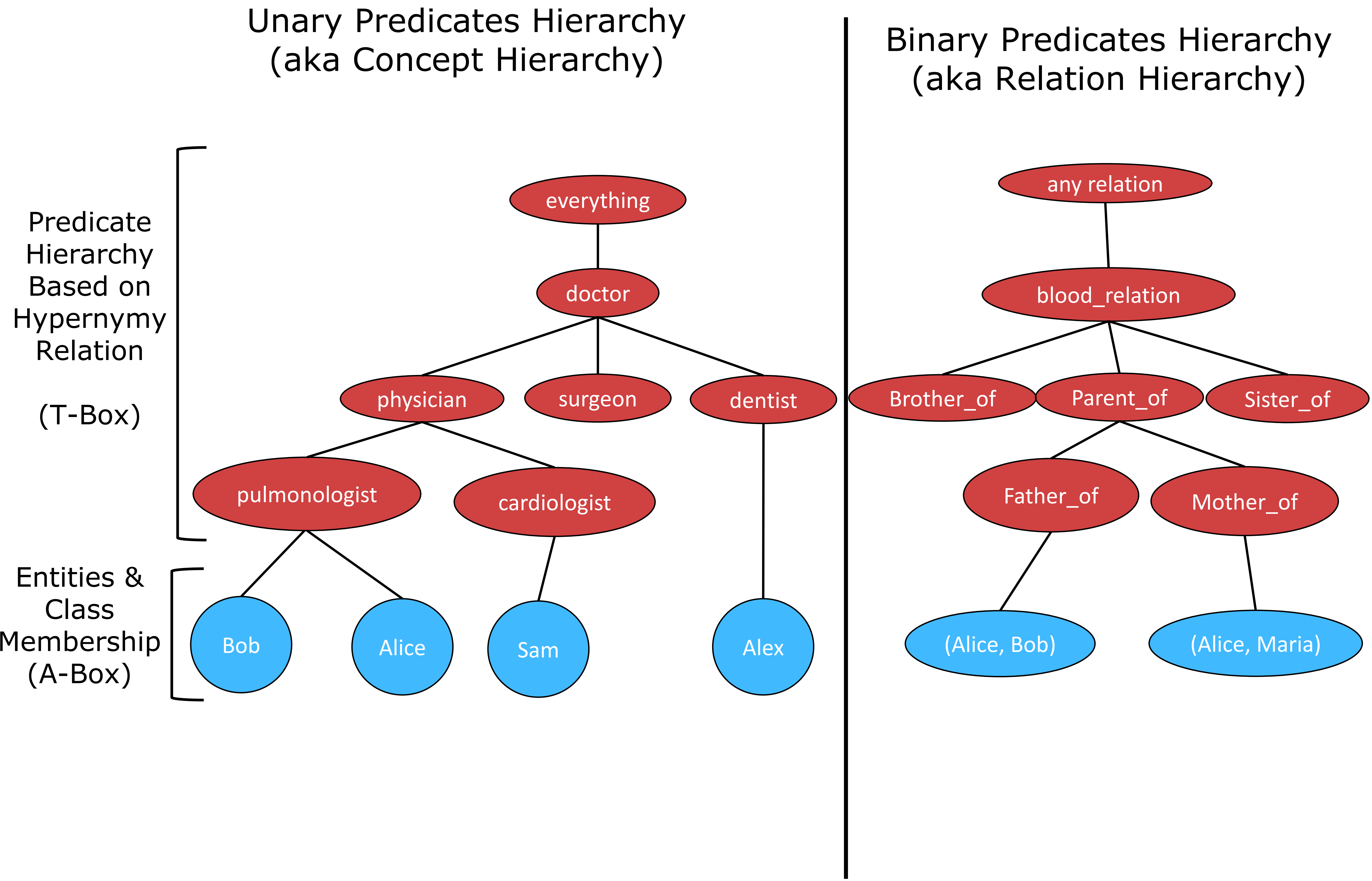
Why are we doing this? What’s the use?

1.Such embeddings can be leveraged by several non-symbolic (e.g. neural and vector) methods to accomplish various reasoning tasks, such as

A. Knowledge completion (Inductive Reasoning)

B.Complex Membership Queries (Deductive Reasoning)

# Input Knowledge: Entities, Predicates, & Hierarchy



Unary Predicates Hierarchy
(aka Concept Hierarchy)

Binary Predicates Hierarchy
(aka Relation Hierarchy)

Predicate
Hierarchy
Based on
Hypernymy
Relation

(T-Box)

Entities &
Class
Membership
(A-Box)

everything

doctor

physician

surgeon

dentist

pulmonologist

cardiologist

Bob

Alice

Sam

Alex

any relation

blood_relation

Brother_of

Parent_of

Sister_of

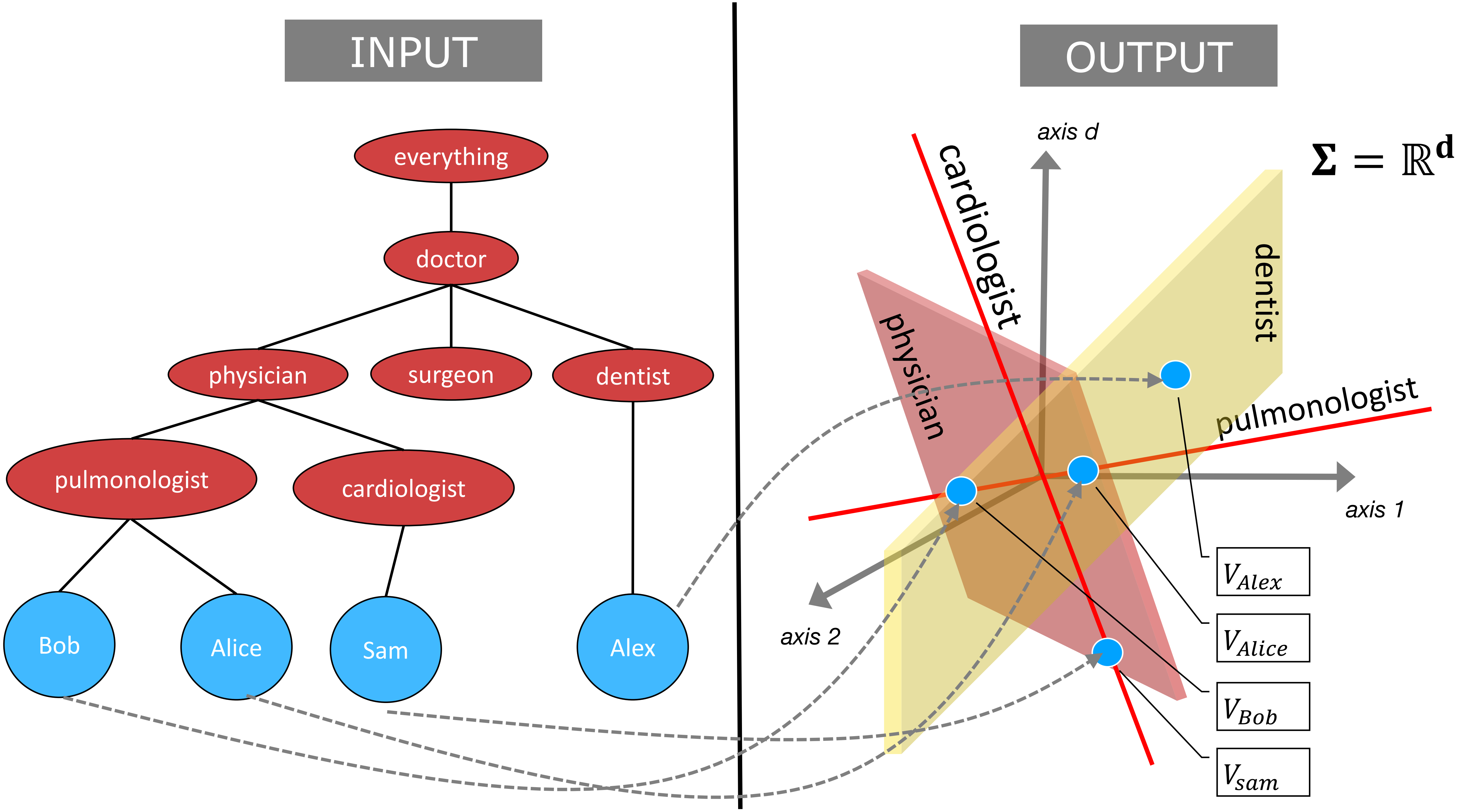
Father_of

Mother_of

(Alice, Bob)

(Alice, Maria)

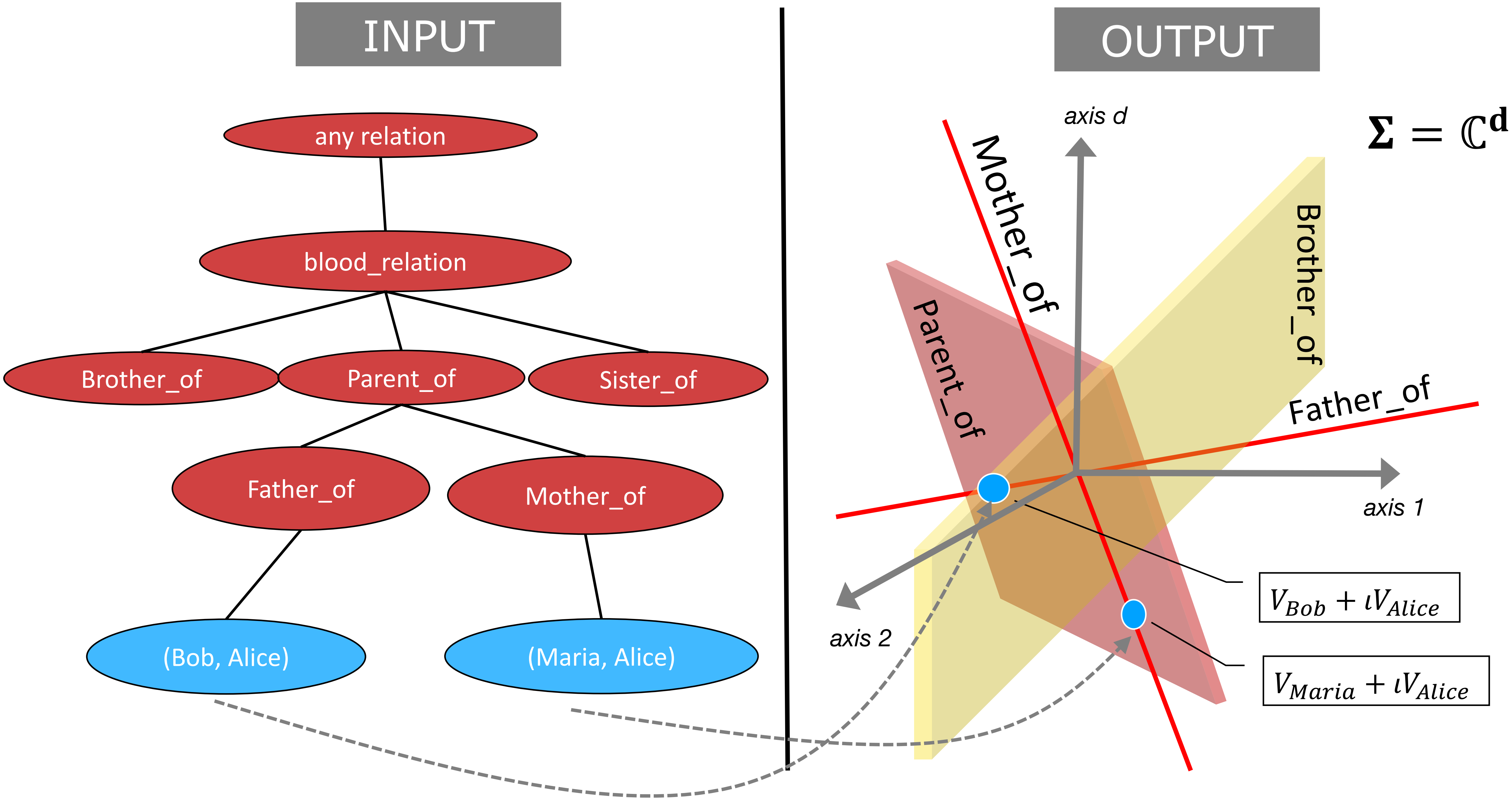
# Quantum Embedding of Concept Hierarchy



- A concept is represented by the subspace of the real vector space
- An Entity is represented by a real vector

*In the above (as well as subsequent) diagram, the diagram of 3D space is just for graphical illustration sake and should not be understood  as 3D space literally. We are using it just to as a cartoon illustration of any d-dimensional vector space.*

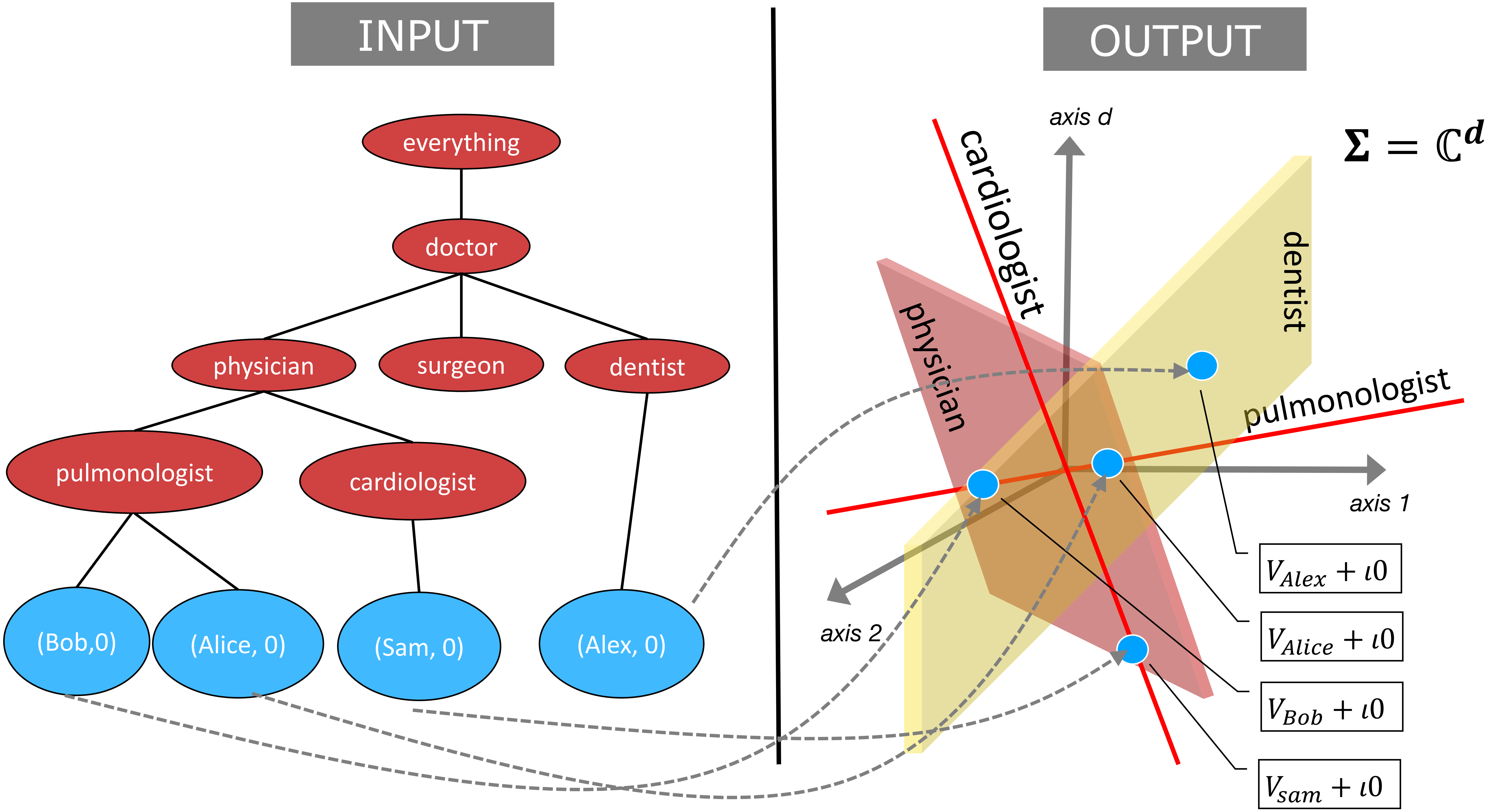
# Quantum Embedding of Relation Hierarchy



- A relation is represented by the subspace of the complex vector space
- An entity pair is represented by a complex vector

# Canonicalization

## Generalize Concepts → Relations, Entities → Entity Pairs



- In general, quantum embedding space Σ could be any finite/infinite dimensional Hilbert space
- An entity can be viewed as a pair of entities by pairing it with a dummy entity called 0

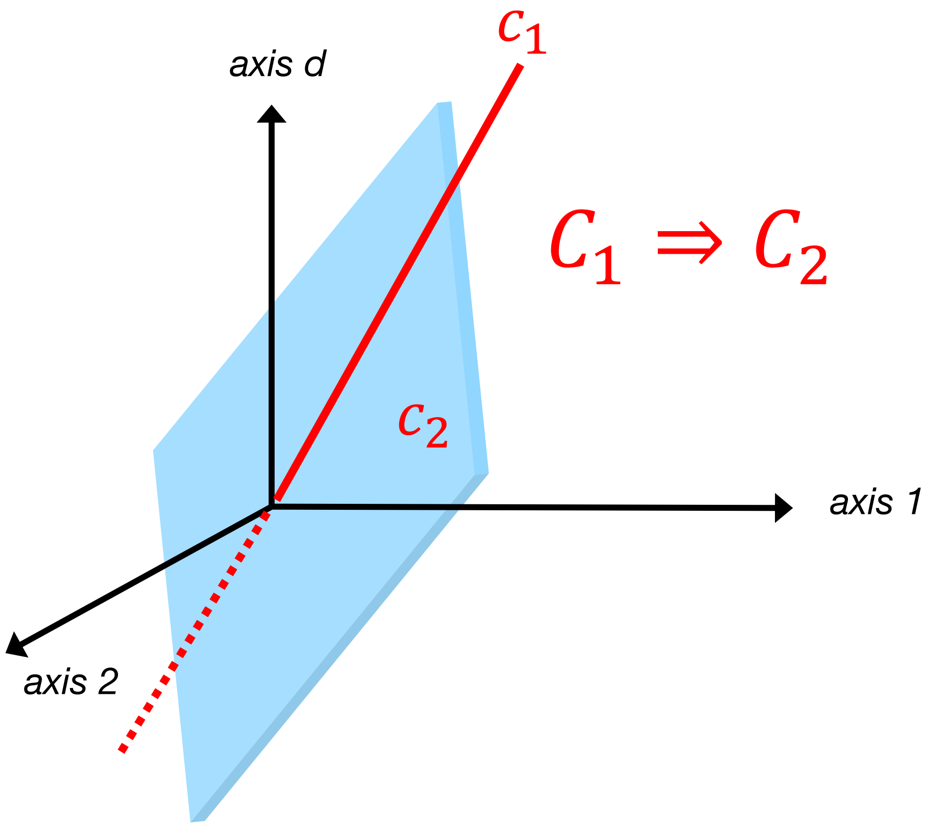
# Axioms of Quantum Logic

Axiom 1 [Logical Implication]

A concept $c_1$ implies concept $c_2$ iff $c_1$ itself is a subspace of $c_2$. Formally,

$$C_1 \Rightarrow C_2 \quad \Leftrightarrow \quad subspace(C_1) \subseteq subspace(C_2)$$

G. Birkhoff and J. von Neumann, *The Logic of Quantum Mechanics, *Annals of Mathematics, Vol. 37, pp. 823–843, 1936.*



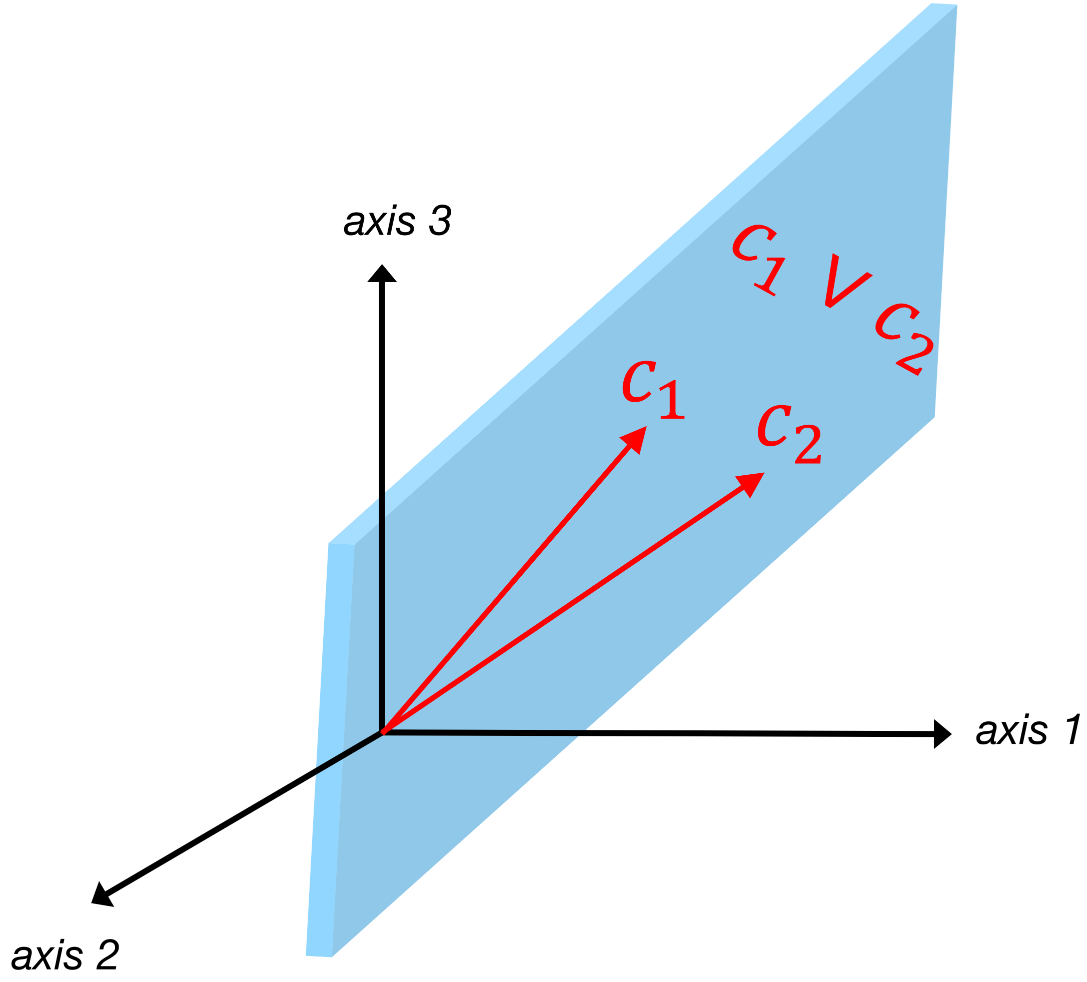
[Image Sources]
https://commons.wikimedia.org/wiki/File:George_David_Birkhoff.gif
http://www.lanl.gov/history/atomicbomb/images/NeumannL.GIF

# Axioms of Quantum Logic

Axiom 2 [Logical OR]

The logical OR of the concept $c_1$ and $c_2$ is given by the vector sum of the corresponding subspaces. More formally,

$$C_1 \vee C_2 \;=\; subspace(C_1) + subspace(C_2)$$

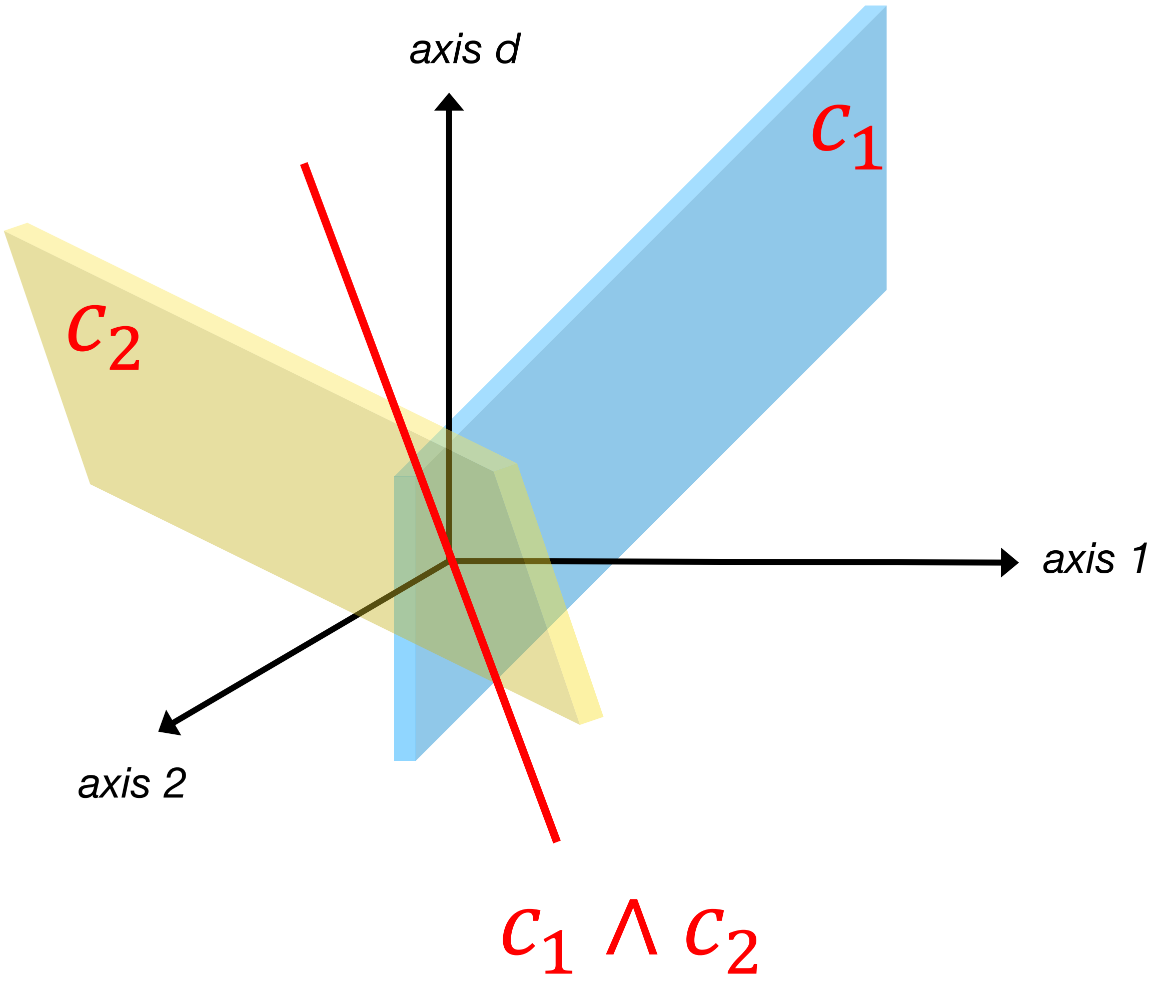


# Axioms of Quantum Logic

Axiom 3 [Logical AND]

The logical AND of the concept $c_1$ and $c_2$ is given by the intersection of the corresponding subspaces. More formally,

$$C_1 \wedge C_2 \quad = subspace(C_1) \cap subspace(C_2)$$



# Axioms of Quantum Logic

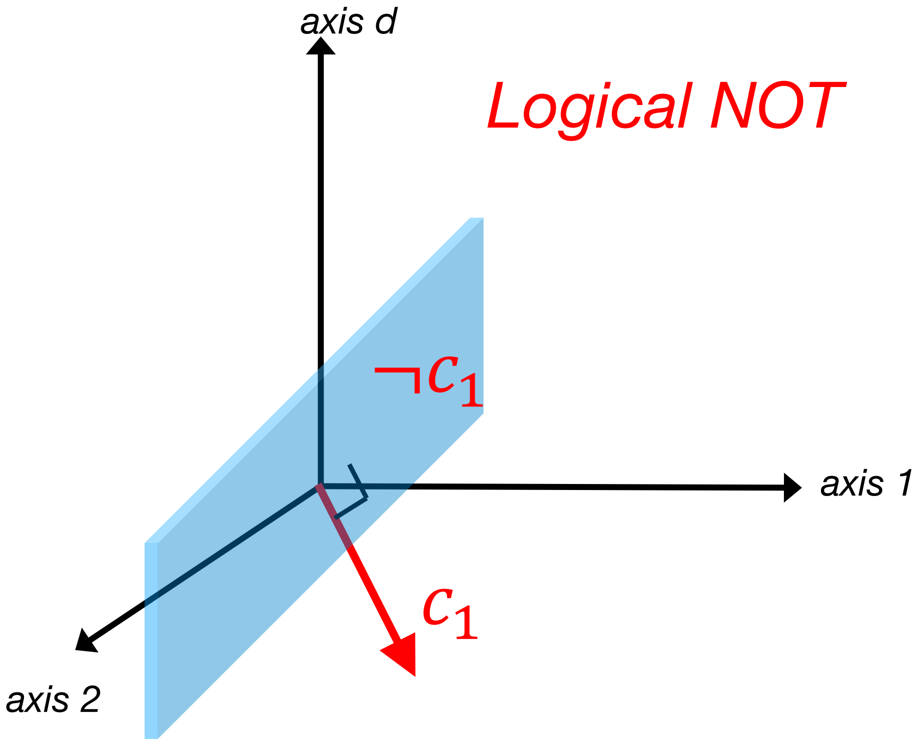
Axiom 4 [Logical NOT]

- The logical NOT of the concept $c_1$, denoted by $\neg c_1$, is a concept whose join and meet points are two extreme ends of the lattice.

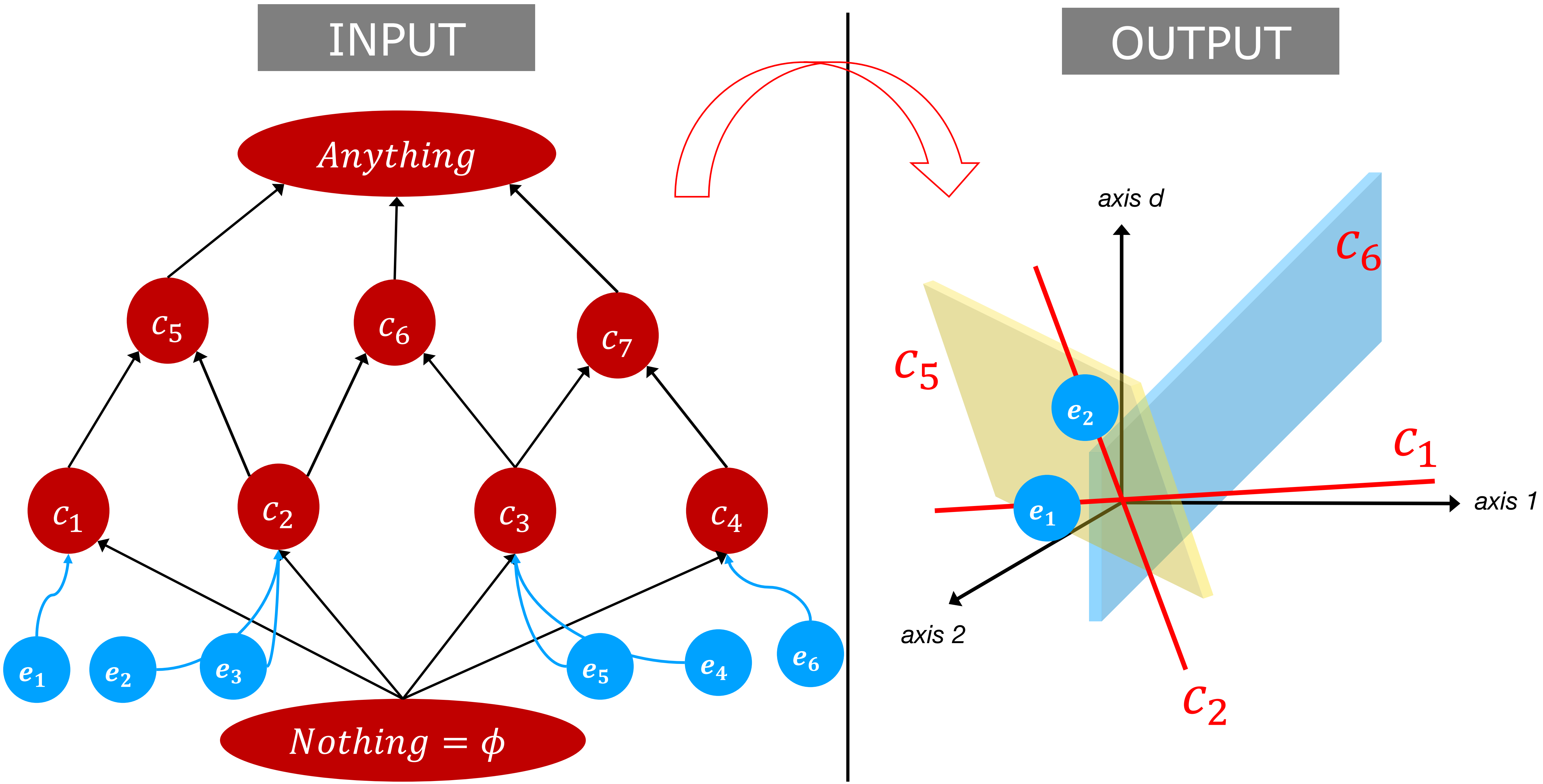
$$C_1 \vee \neg C_1 = \Sigma$$
$$C_1 \wedge \neg C_1 = 0$$

- It immediately follows from the above conditions that the subspace of $\neg c_1$ is an orthogonal complement of the subspace corresponding to $c_1$.

- From this, it also follows that any two vectors that are orthogonal to each other are unrelated concepts.



# Formulating Quantum Embedding as a CSP



[Subspace Constraints] For each $C_i$, we must have $\text{subspace}(C_i) = \text{span}\{\text{b}_{i1}, \dots, b_{ik_i}\}$

[Inclusion Constraints] If $C_i \Rightarrow C_j$ then we must have $\text{basis}(C_i) \subseteq \text{basis}(C_j)$

[Membership Constraints] If $e \in C_i$, then we must have $V_e \in \text{subspace}(C_i)$

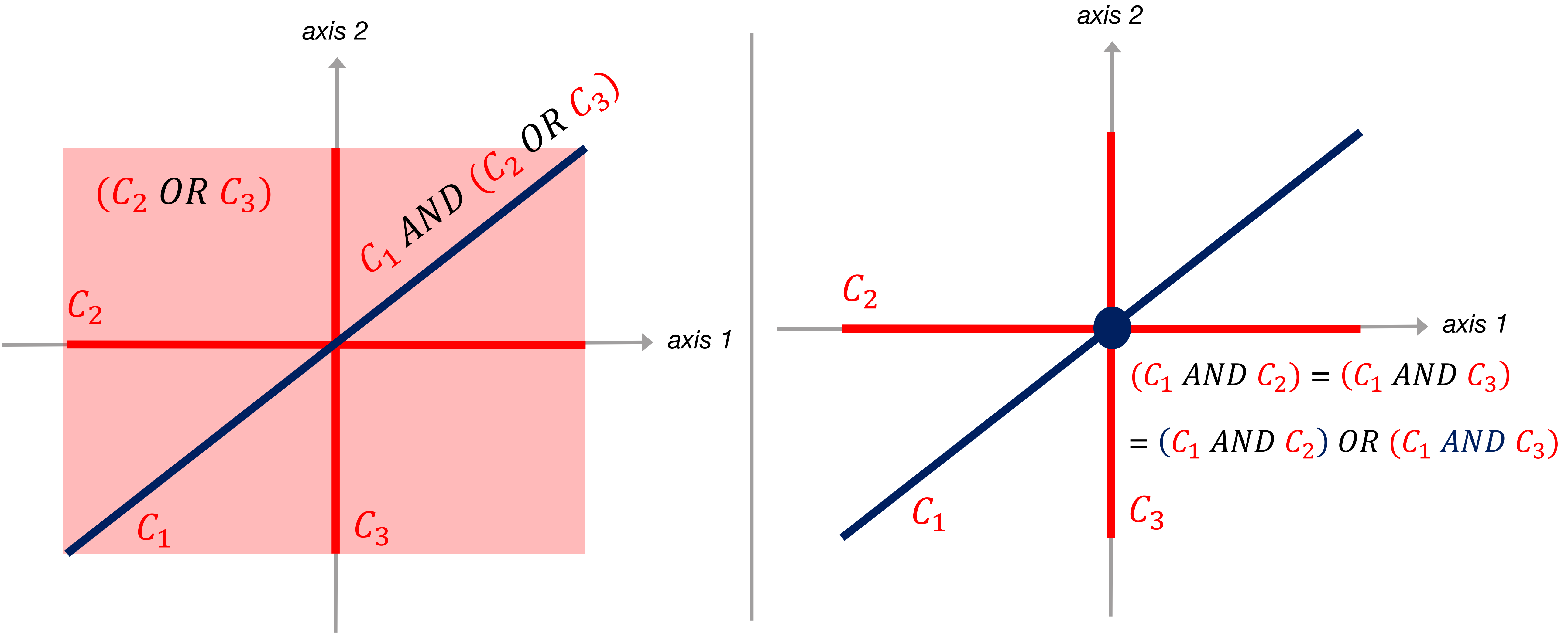
# But Wait, There is a Twist

The Distributive Law Doesn’t Hold True in Quantum Logic - A Known and Serious Limitation of Quantum Logic

$$C_1\ AND\ (C_2\ OR\ C_3) \neq (C_1\ AND\ C_2)\ OR\ (C_1\ AND\ C_3)$$

The diagram below is an example for the above inequality. Here, we have assumed all the concept subspaces are lines in a plane

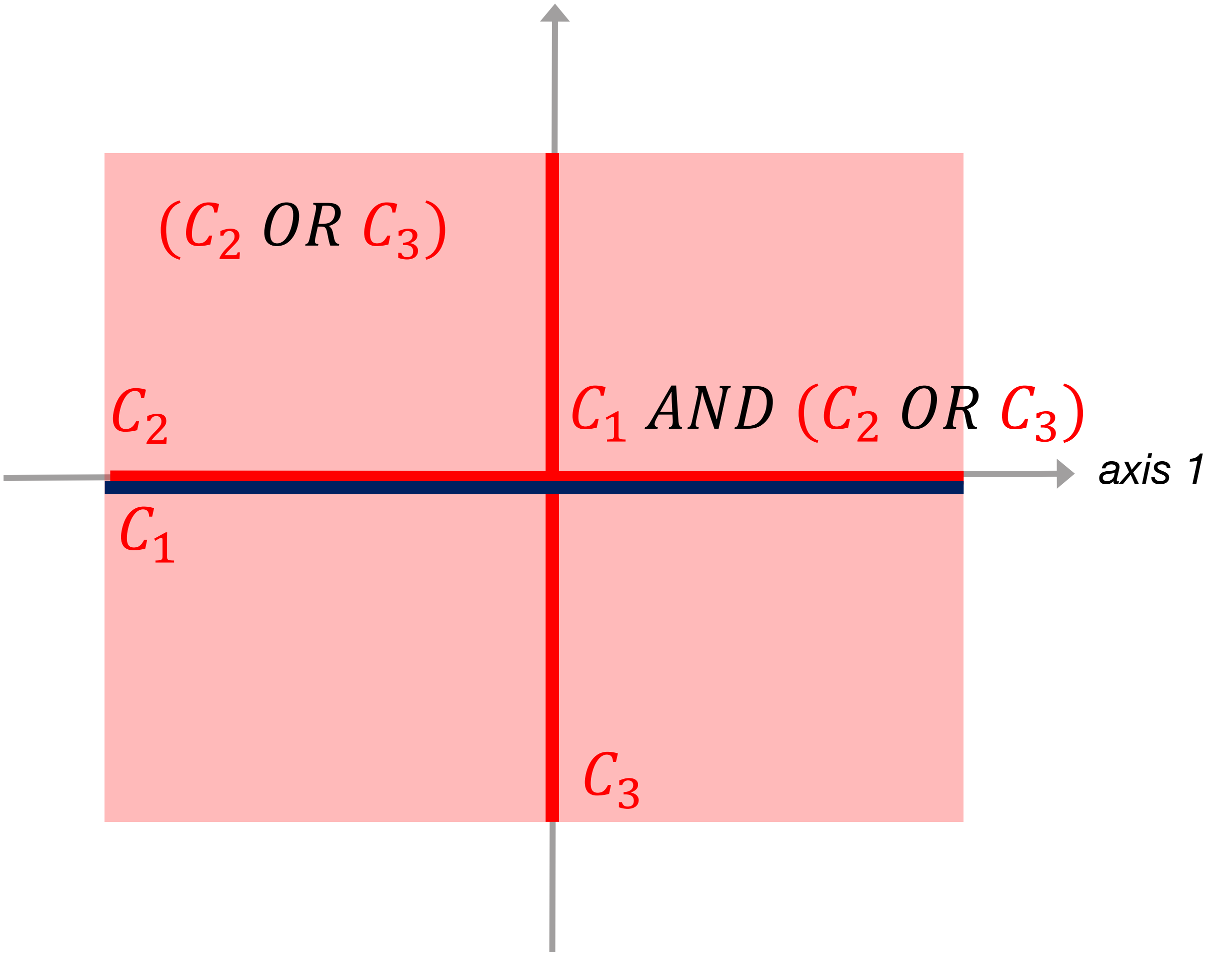
$$\text{LHS} \neq \text{RHS}$$

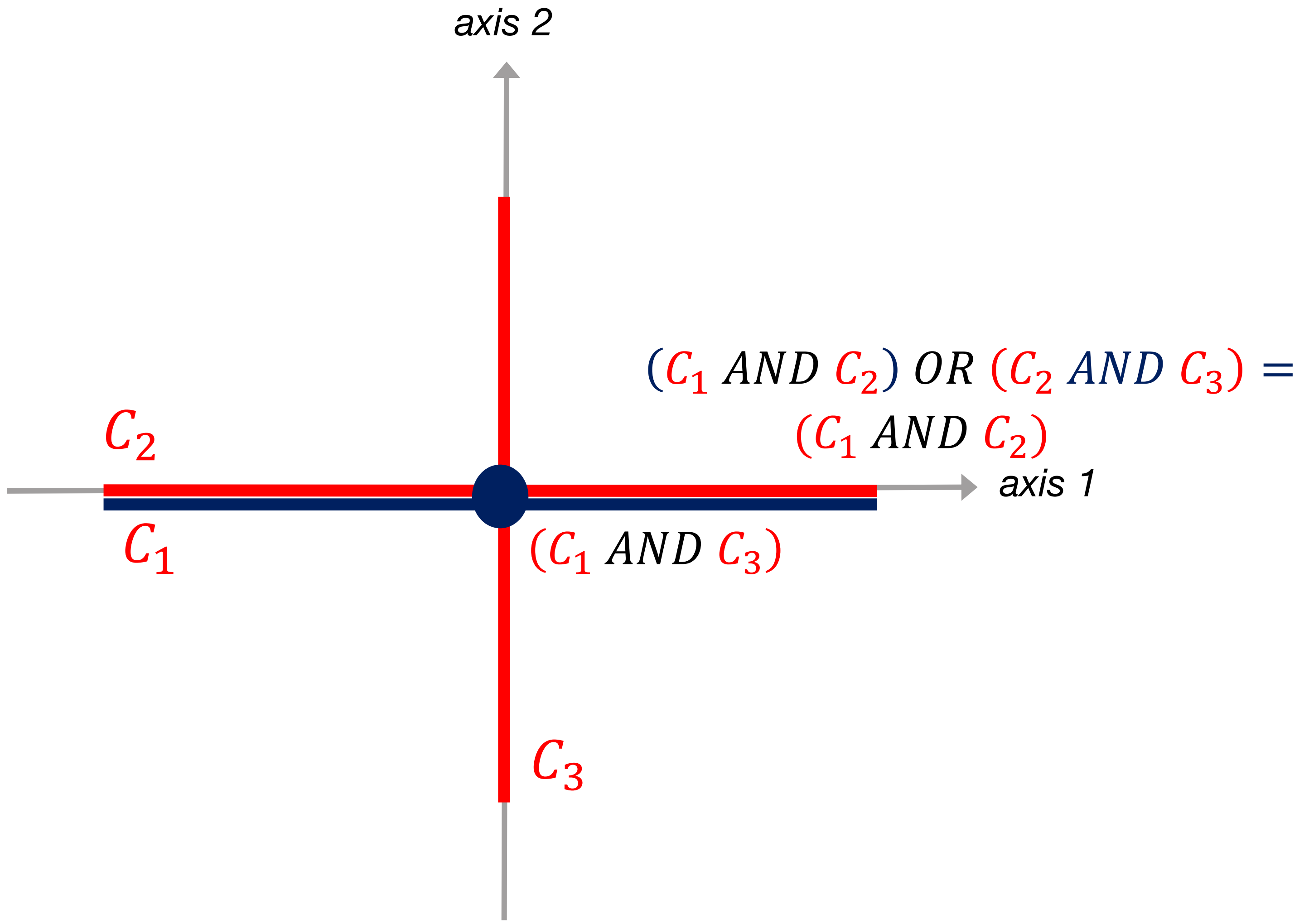

axis 2
axis 1
$(C_2\ OR\ C_3)$
$C_1\ AND\ (C_2\ OR\ C_3)$
$C_2$
$C_1$
$C_3$
axis 2
axis 1
$C_2$
$(C_1\ AND\ C_2) = (C_1\ AND\ C_3)$
$= (C_1\ AND\ C_2)\ OR\ (C_1\ AND\ C_3)$
$C_1$
$C_3$

# The good news is that there is a fix for it

[Theorem] The distributive law will hold true in Quantum Logic if we work with axis-parallel subspaces only.

$$C_1\ AND\ (C_2\ OR\ C_3) \quad = \quad (C_1\ AND\ C_2)\ OR\ (C_1\ AND\ C_3)$$

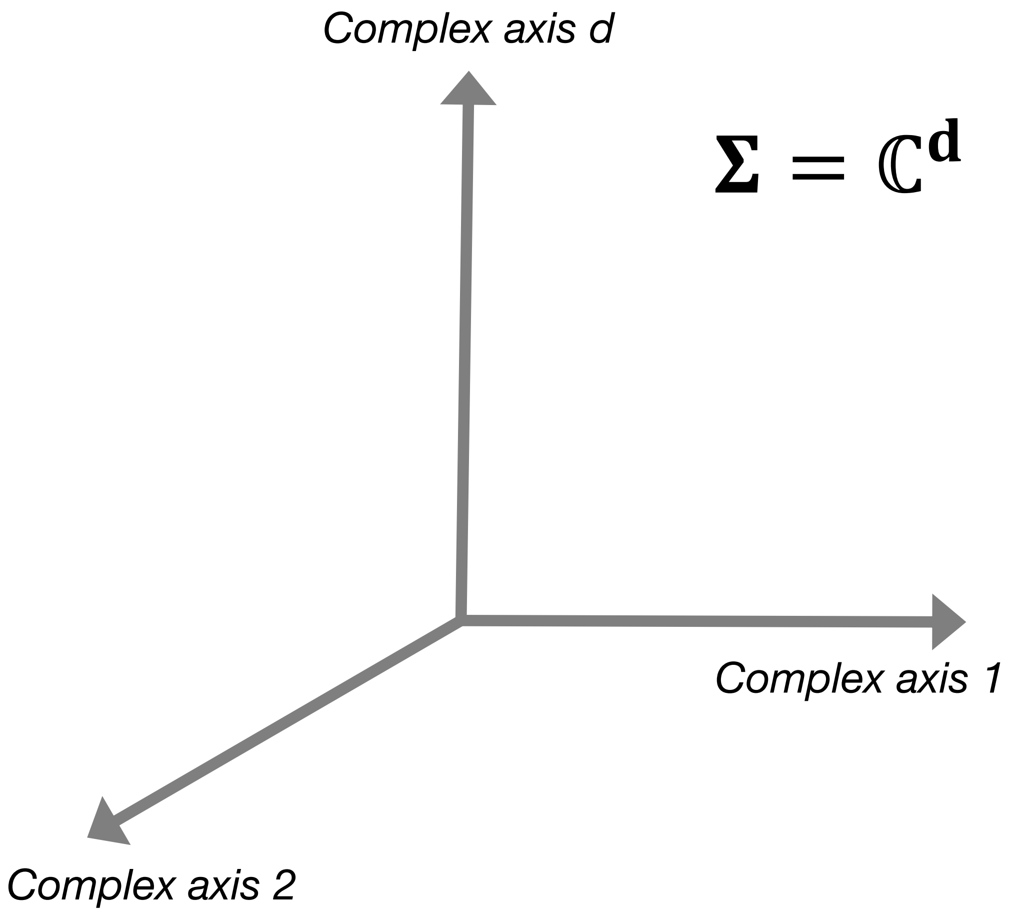




# Learning Axis-Parallel Quantum Embedding

## Setup

| | |
|---|---|
| Embeddings Space | $\mathbb{C}^{\mathbf{d}}$ |
| Underlying Field | $\mathbb{R}$ |
| # axis-parallel subspaces | $2^{2d}$ |
| Number of concepts in input hierarchy | $N_C$ |
| Number of relations in input hierarchy | $N_R$ |
| Lower bound on $d$ | $d > \log_2(\sqrt{N_C + N_R}\,)$ |



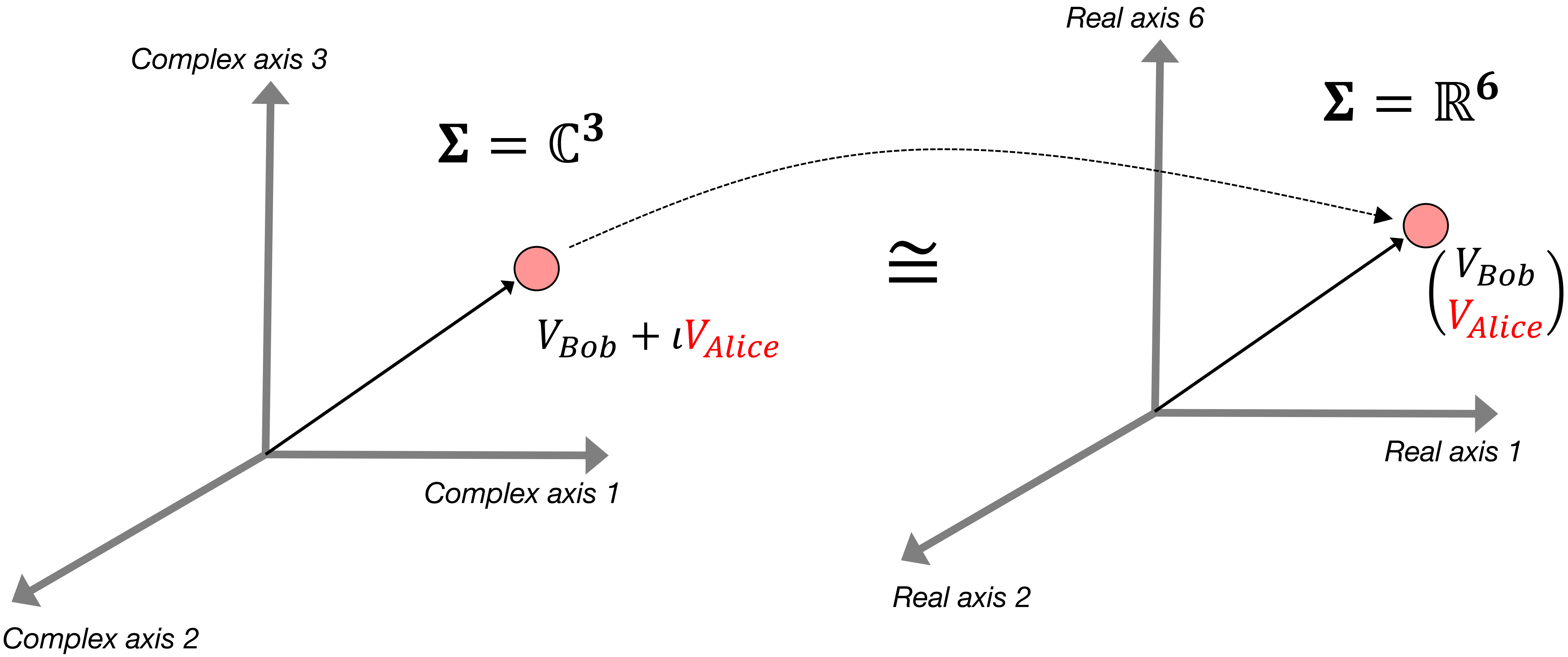
$$\boldsymbol{\Sigma} = \mathbb{C}^{\mathbf{d}}$$

Suppose $d$ = 3 for the sake of illustration. Then, due to the assumption of field being $\mathbb{R}$, we have

$$\text{Set of Standard Basis for } \mathbb{C}^3 = \left(\underbrace{\begin{matrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\end{matrix}}_{\text{Real axes}}\;\;\underbrace{\begin{matrix}1&0&0\\0&1&0\\0&0&0\\\iota&0&0\\0&\iota&0\\0&0&\iota\end{matrix}}_{\text{Imaginary axes}}\right)$$

# Isomorphism between $\mathbb{C}^d$ and $\mathbb{R}^{2d}$

Due to the real field assumption, we have an isomorphism between $\mathbb{C}^d$ and $\mathbb{R}^{2d}$.
For the sake of illustration, suppose $d = 3$, then we can write


Complex axis 3
$\boldsymbol{\Sigma} = \mathbb{C}^{\mathbf{3}}$
$V_{Bob} + \iota V_{Alice}$
Complex axis 1
Complex axis 2
$\cong$
Real axis 6
$\boldsymbol{\Sigma} = \mathbb{R}^{\mathbf{6}}$
$\begin{pmatrix} V_{Bob} \\ V_{Alice} \end{pmatrix}$
Real axis 1
Real axis 2

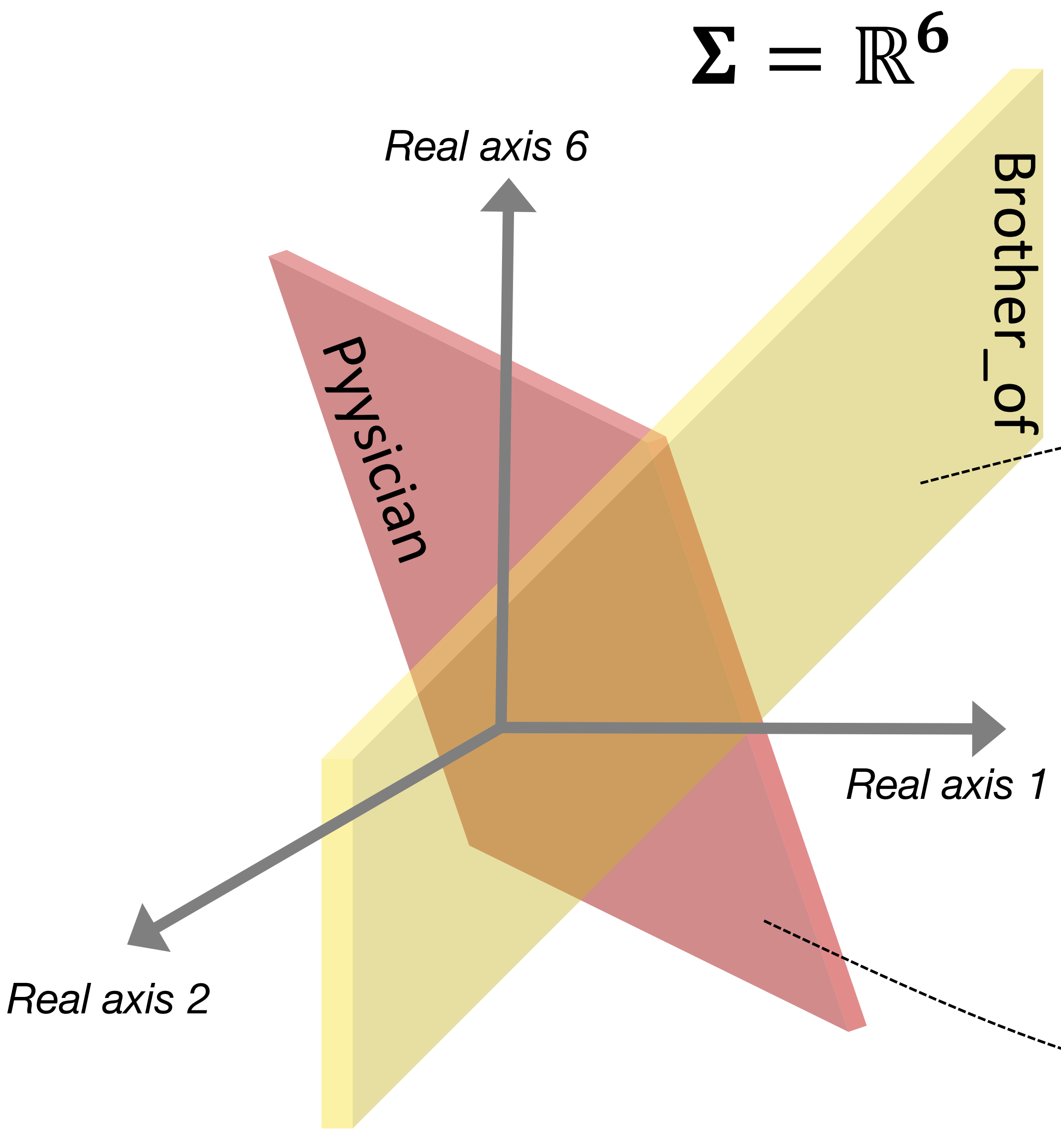
$$\text{Basis for } \mathbb{C}^3 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \iota & 0 & 0 \\ 0 & 0 & 0 & 0 & \iota & 0 \\ 0 & 0 & 0 & 0 & 0 & \iota \end{pmatrix} \cong \text{Basis for } \mathbb{R}^6 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Real axes Imaginary axes

Real axes Imaginary axes

# Learning Axis-Parallel Quantum Embedding

Indicator vectors can be used to represent concept / relation subspaces in $\mathbb{R}^{2d}$ . Suppose $d$ = 3 for the illustration sake. Then,

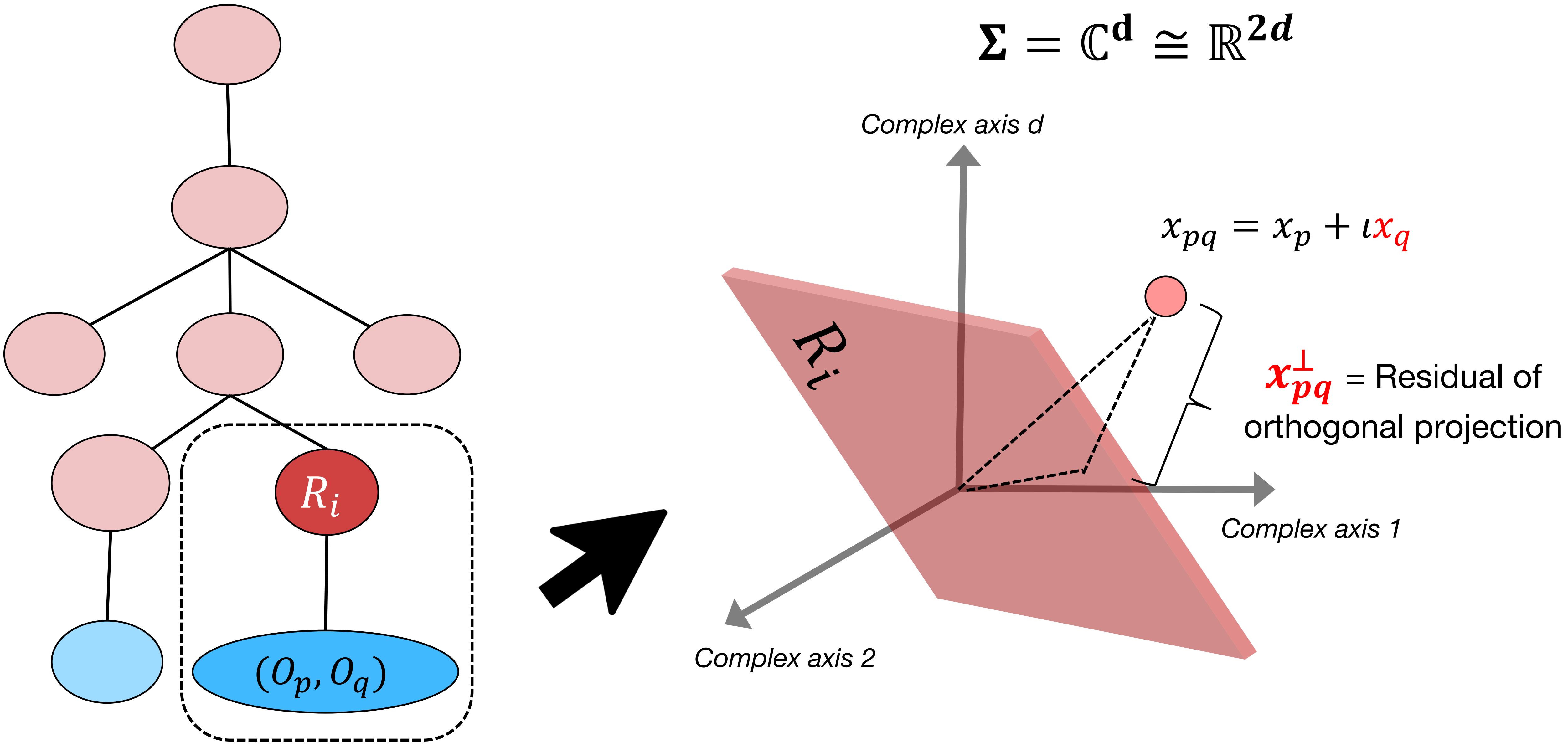


Basis for $\mathbb{R}^6$ = $\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

| | | | | | | |
|---|---|---|---|---|---|---|
| $z_{brother\_of}$ | 0 | 1 | 1 | 0 | 1 | 0 |
| $z_{physician}$ | 0 | 1 | 1 | 0 | 0 | 0 |

Learnable parameters

# Learning Loss for Membership Constraint



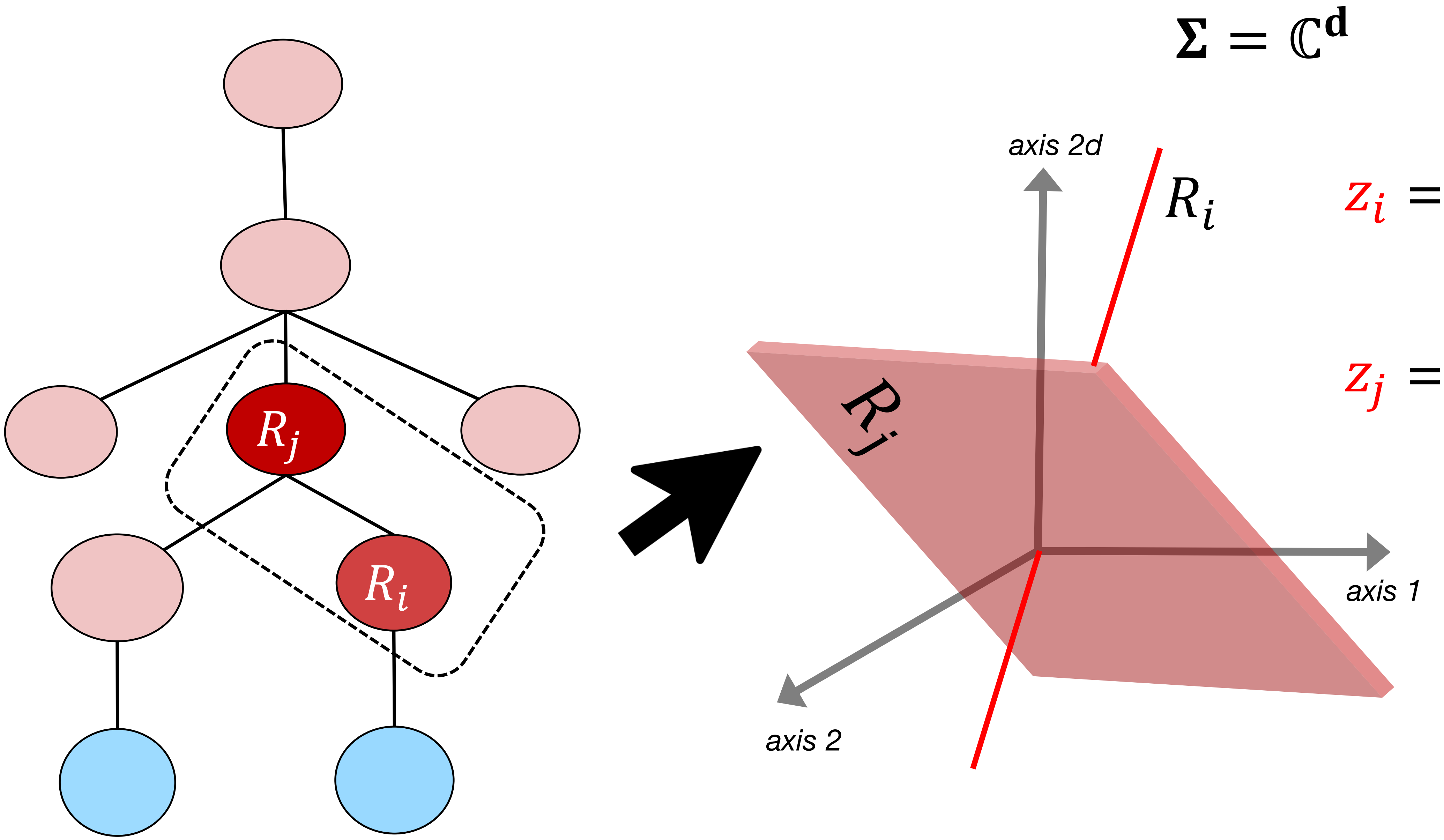
$\Sigma = \mathbb{C}^{\mathbf{d}} \cong \mathbb{R}^{2d}$
Complex axis d
$x_{pq} = x_p + \iota x_q$
$R_i$
$R_i$
$x^{\perp}_{pq}$ = Residual of orthogonal projection
Complex axis 1
Complex axis 2
$(O_p, O_q)$

$$L_{R_i(O_p,O_q)} = \langle x^{\perp}_{pq} , x^{\perp}_{pq} \rangle^2 + \| Re( x_{pq} ) - x_p \|^2 + \| Re( x_{pq} ) - x_q \|^2$$

Because of real field assumption, we use the following valid formula for inner product $\langle x, y \rangle^2 = (y^{*\top} x)^* (y^{*\top} x)$ where $*$ denotes complex conjugate operation.

*For the membership of an entity $O_i$ to a unary predicate $C_j$ , one can write a loss function $L_{(O_i \in C_j)}$ in a similar manner. The last two terms will be missing there but an unitary length constraint would show up there (refer Eq. 4 in the paper).*

# Learning Loss for Implication Constraint



$$\boldsymbol{\Sigma} = \mathbb{C}^{\mathbf{d}}$$

|  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|
| $z_i =$ | 0 | 1 | 1 | 0 | 1 | 0 |
| $z_j =$ |  | 1 | 1 |  | 1 |  |

$z_j$ must have 1 in position where $z_i$ has 1. This can be captured by the loss

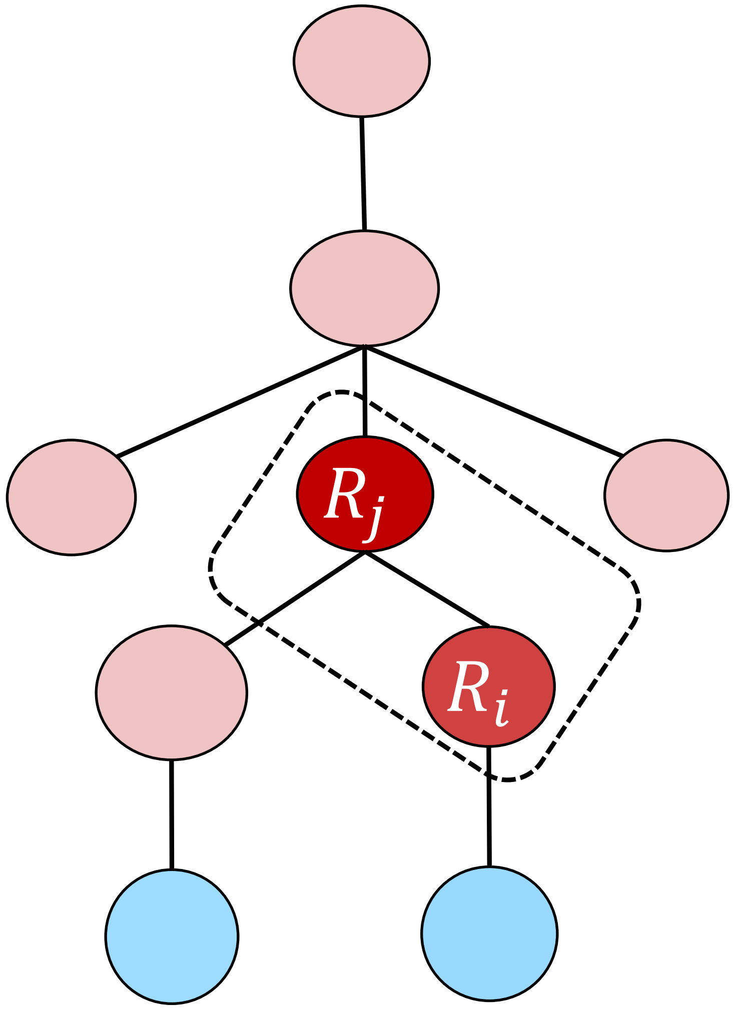
$L_{R_i \sqsubseteq R_j} = \left\| z_i^\top \odot \overline{z_j} \right\|^2$ where, $\overline{z_j} = (1 - z_j)$

*A similar loss function can be written down for the case of unary predicates (refer Eq. 5 in the paper).*

# Regularization Loss

[Observation] A degenerate solution is to map all the predicates to a single subspace and all the entities to a single vector

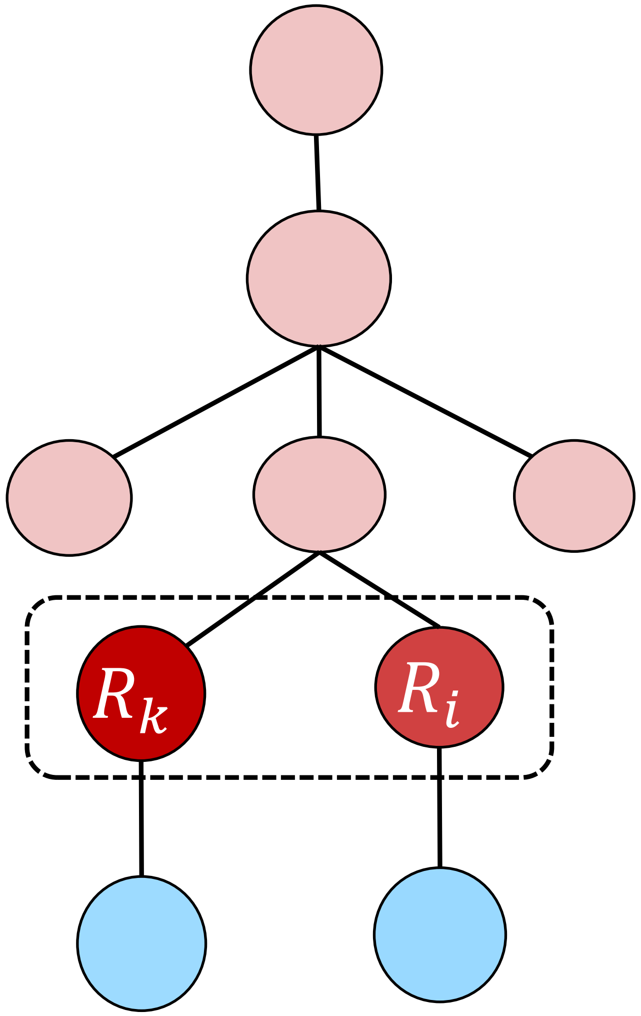
## Regularizing Parental Space



Have minimum difference in the dimensions of the subspaces of $R_i$ and $R_j$

$$\Omega_{\text{parent}} = \left(\sqrt{d} \; - (z_j - z_i)^\top 1\right)^2$$

## Regularizing Sibling Space



Have the subspaces of $R_k$ and $R_i$ orthogonal to each other as much as possible

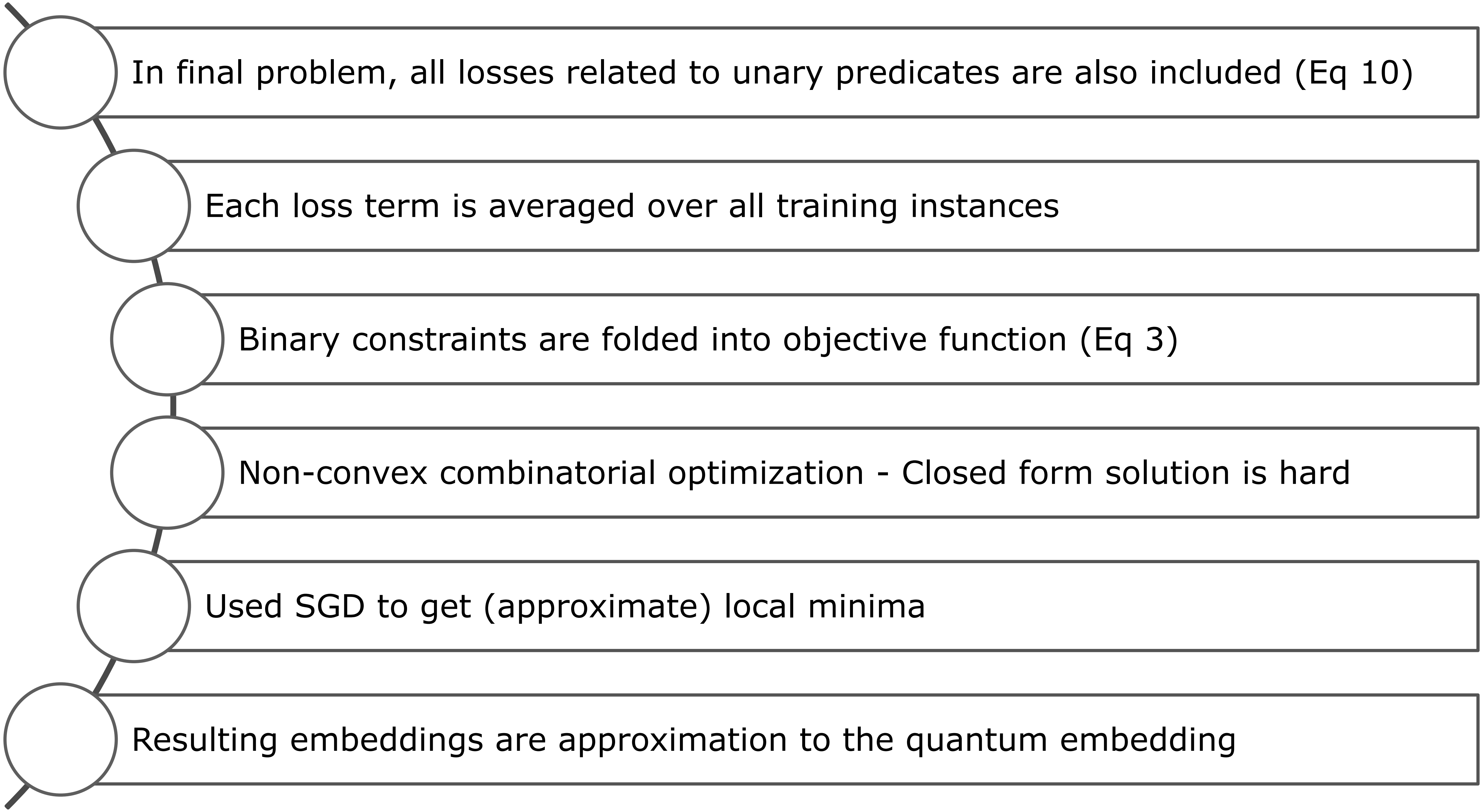
$$\Omega_{\text{sibling}} = \left(z_k^\top z_i\right)^2$$

*A similar regularization loss function can be written down for the case of unary predicates (refer Eq. 11-12 in the paper).*

# Overall Learning Problem

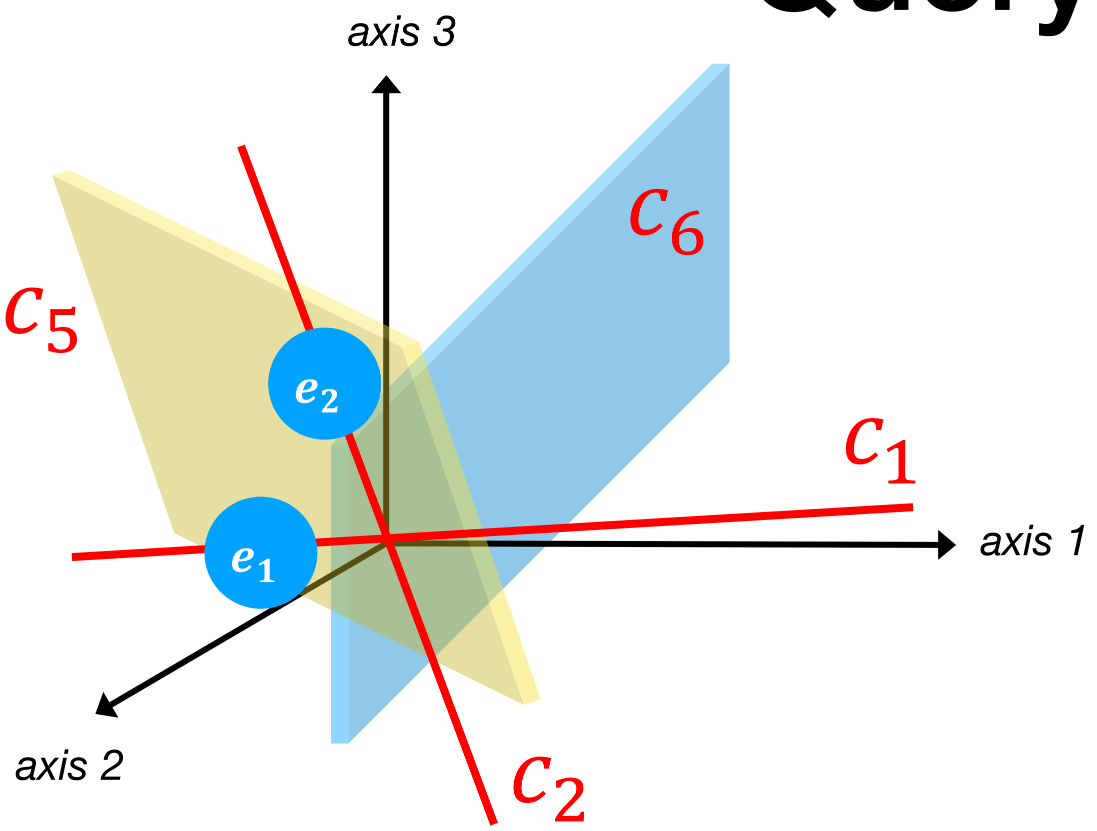
$$\min_{z_i,\, x_e} \left( \bar{L}_{R_i(o_p,o_q)} + \bar{L}_{R_i \sqsubseteq R_j} + \Omega_{parent} + \Omega_{sibling} \right)$$

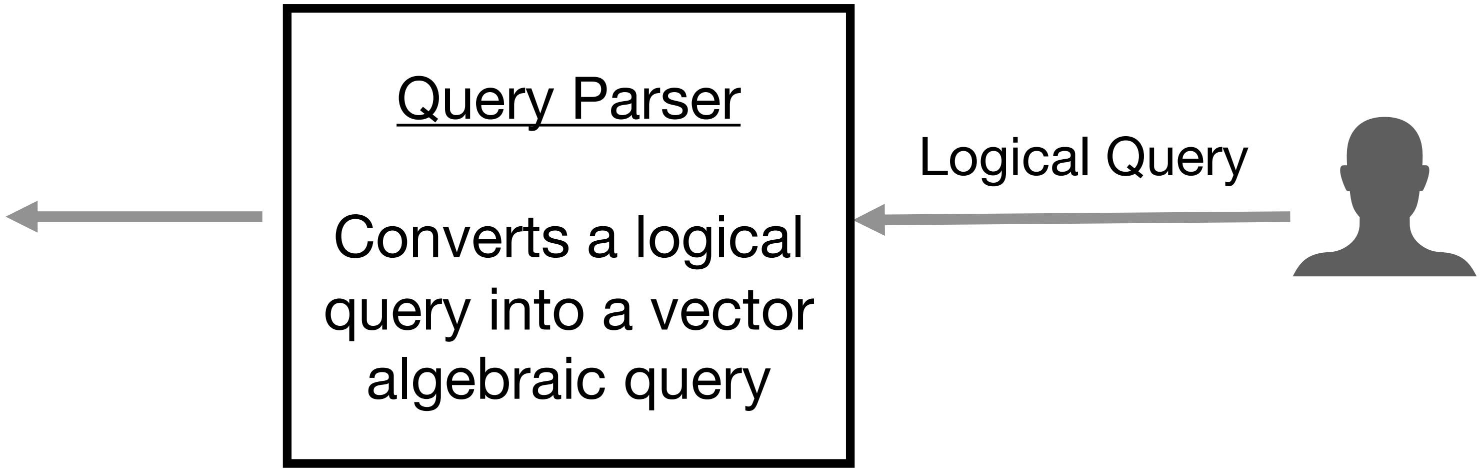
$s.t.$ $z_i$ are binary valued vectors



- In final problem, all losses related to unary predicates are also included (Eq 10)
- Each loss term is averaged over all training instances
- Binary constraints are folded into objective function (Eq 3)
- Non-convex combinatorial optimization - Closed form solution is hard
- Used SGD to get (approximate) local minima
- Resulting embeddings are approximation to the quantum embedding

# Query Processing





Query Parser

Converts a logical query into a vector algebraic query

Logical Query

| Query Type | Example | Transformed Query |
|---|---|---|
| *Membership* | *Find all the entities satisfying the formula* $NOT(C_6)\ AND\ \ (C_1\ OR\ C_2)$ | *Entity that can expressed as a linear combination of* $B_1$ *or* $B_2$ *but not* $B_6$ |
| *Logical Entailment* | *Dose entity* $O_i$ *belong to the concept* $C_j$ | *Check if entity* $x_i$ *can be expressed as a linear combination of the vectors in* $B_j$ |
| *Property Listing* | *List all concepts that entity* $O_i$ *belong to* | *For each concept* $c_j$, *check if the entity* $e_i$ *belongs to the space spanned by basis* $B_j$ |

*In the above table,* $B_i$ *denotes the set of basis vectors for the subspace* $C_i$

# Experiments Plan

| Task | Link Prediction | Finding Members of a Complex Concept |
|---|---|---|
| Type of Reasoning | Predictive | Deductive |
| Dataset | FB15K, WN18 | LUBM1U |
| T-Box | Absent | Present |
| A-Box | Present | Present |
| Predicate Types | [WN18] Only binary Predicates<br>[FB15K] Only binary Predicates | Both Unary and Binary |
| Input KB Size | [A-Box for WN18] 141442 triples<br>[A-Box for FB15K] 483142 triples | [A-Box] 69628 triples<br>[T-Box] 18 triples |
| Test Set Size | [WN18] 5000<br>[FB15K] 59071 | 8 membership queries for non-leaf concepts |
| Performance Metrics | Mean Rank, MRR, Hits@1, Hits@10 | Mean Rank, MRR, Hits@1, Hits@10 |
| Baslines | TransE [1], ComplEx [2] | TransE [1], ComplEx [2] |

[1] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, JasonWeston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In NIPS, pages 2787–2795, 2013.

[2] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In ICML, pages 2071–2080, 2016.

# LUBM1U Dataset Details

- Data is synthetically generated
- Data represent knowledge about various elements of an hypothetical university-setup. For example, *Faculty, Students, Job Roles, Research Interests, Departments*, etc.
- The data has two components:

  1. A-Box – It captures the assertions related to entities and predicates
     * Unary Predicate – Membership of entities to unary predicates (aka concepts)
     * Binary Predicate – Relationship among different pairs of entities
  2. T-Box – It captures information related to the concepts/relations hierarchy
     * Unary Predicate – hierarchy among categories
     * Binary Predicate – hierarchy among relations

- Both A-Box and T-Box are represented in the form of triples

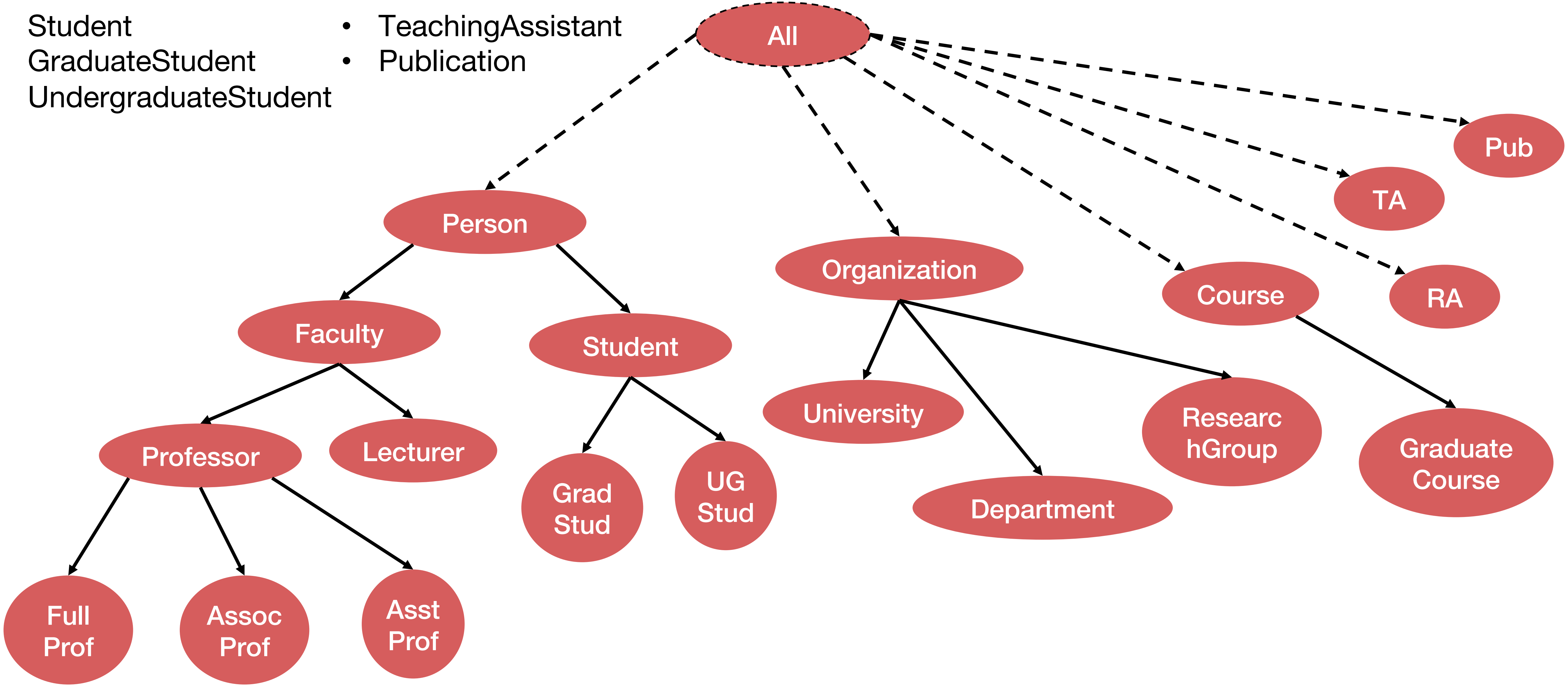
  1. A-Box
     * Unary Predicates – (H, #type, C) – H: head entity, C: concept (aka category)
     * Binary Predicate – (H, R, T) – H: head entity, T: tail entity, R: relation
  2. T-Box
     * unary – (C1, #type, C2) – C1: child category, C2: parent category
     * binary – (R1, #type, R2) – R1: child relation, P2: parent relation

- A-box triples=69628, T-box triples=18, Entities=17174, Categories=19, Relations=13

# LUBM1U – Unary Predicates Hierarchy

**List of Unary Predicates (19 – Concepts)**

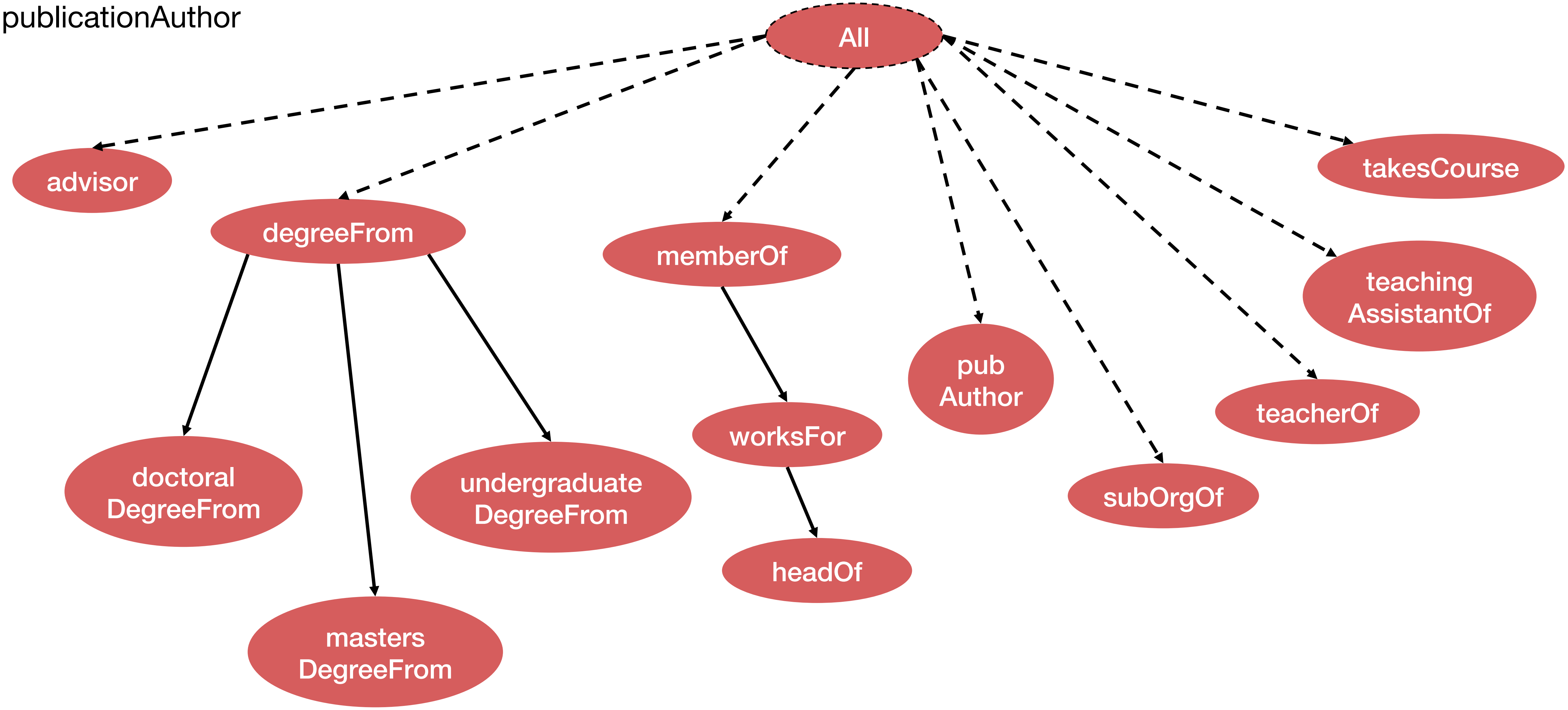
- Person
- Faculty
- Lecturer
- Professor
- FullProfessor
- AssociateProfessor
- AssistantProfessor
- Student
- GraduateStudent
- UndergraduateStudent
- Organization
- University
- Department
- ResearchGroup
- Course
- GraduateCourse
- ResearchAssistant
- TeachingAssistant
- Publication



# LUBM1U – Binary Predicates Hierarchy

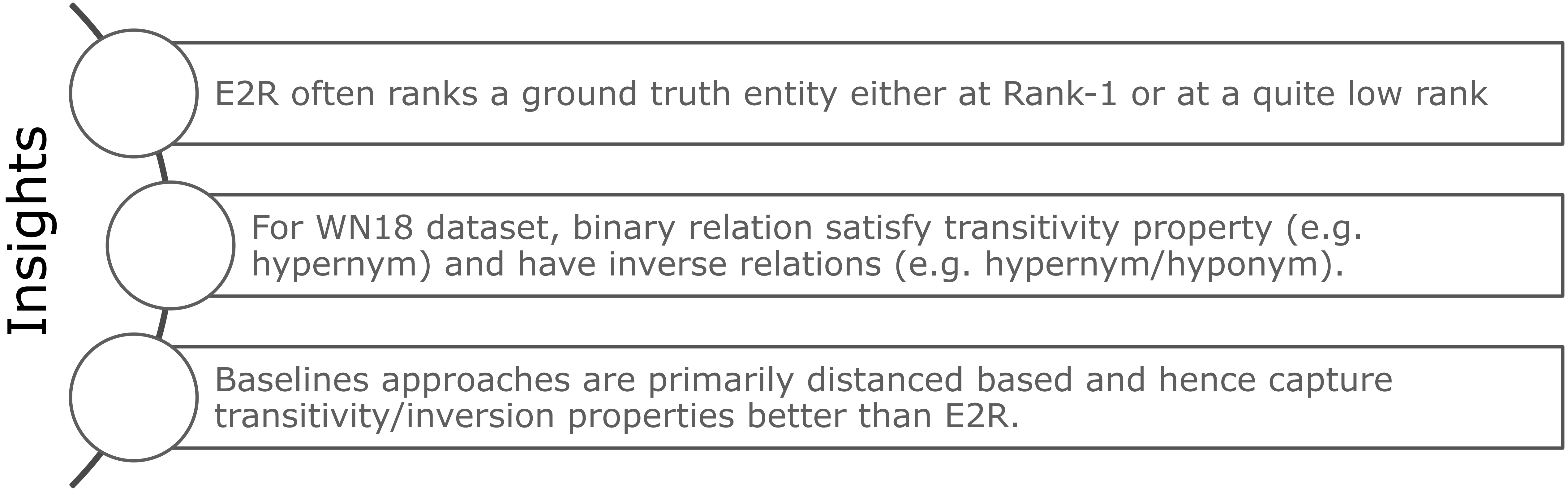
**List of Binary Relations (13 – Relations)**

- advisor
- degreeFrom
- doctoralDegreeFrom
- mastersDegreeFrom
- undergraduateDegreeFrom
- memberOf
- publicationAuthor
- SubOrganizationOf
- teacherOf
- teachingAssistantOf
- takesCourse
- headOf
- worksFor



# Experimental Results

| Data | Mean Rank | | | MRR | | | Hits@1 | | | Hits@10 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | E2R | TE | CE | E2R | TE | CE | E2R | TE | CE | E2R | TE | CE |
| FB15K | 72 | 68 | 114 | **0.96** | 0.49 | 0.61 | **96** | 34 | 49 | **96** | 76 | 81 |
| WN18 | 5780 | 409 | 468 | 0.71 | 0.63 | 0.90 | 71 | 41 | 87 | 71 | 93 | 95 |
| LUBM1U | **220** | 1292 | 5742 | **0.46** | 026 | 0.12 | **45** | 18 | 12 | 45 | 49 | 12 |



Insights

E2R often ranks a ground truth entity either at Rank-1 or at a quite low rank

For WN18 dataset, binary relation satisfy transitivity property (e.g. hypernym) and have inverse relations (e.g. hypernym/hyponym).

Baselines approaches are primarily distanced based and hence capture transitivity/inversion properties better than E2R.

# Conclusions

- The proposed E2R approach can embed a hierarchical KB into a vector space while preserving the structure of all logical propositions.
- Advantage is to be able to answer complex reasoning queries (both deductive and predictive) in accurate manner using distributional representation of the KB.
- In future, we wish to extend this idea for more expressive logic forms as well as natural language KB and also handle noisy inputs.

Thank you