

Embedding-Based Techniques

MATRICES, TENSORS, AND NEURAL NETWORKS

Probabilistic Models: Downsides

Limitation to Logical Relations

- Representation restricted by manual design
 - Clustering? Assymmetric implications?
 - Information flows through these relations
- Difficult to generalize to unseen entities/relations

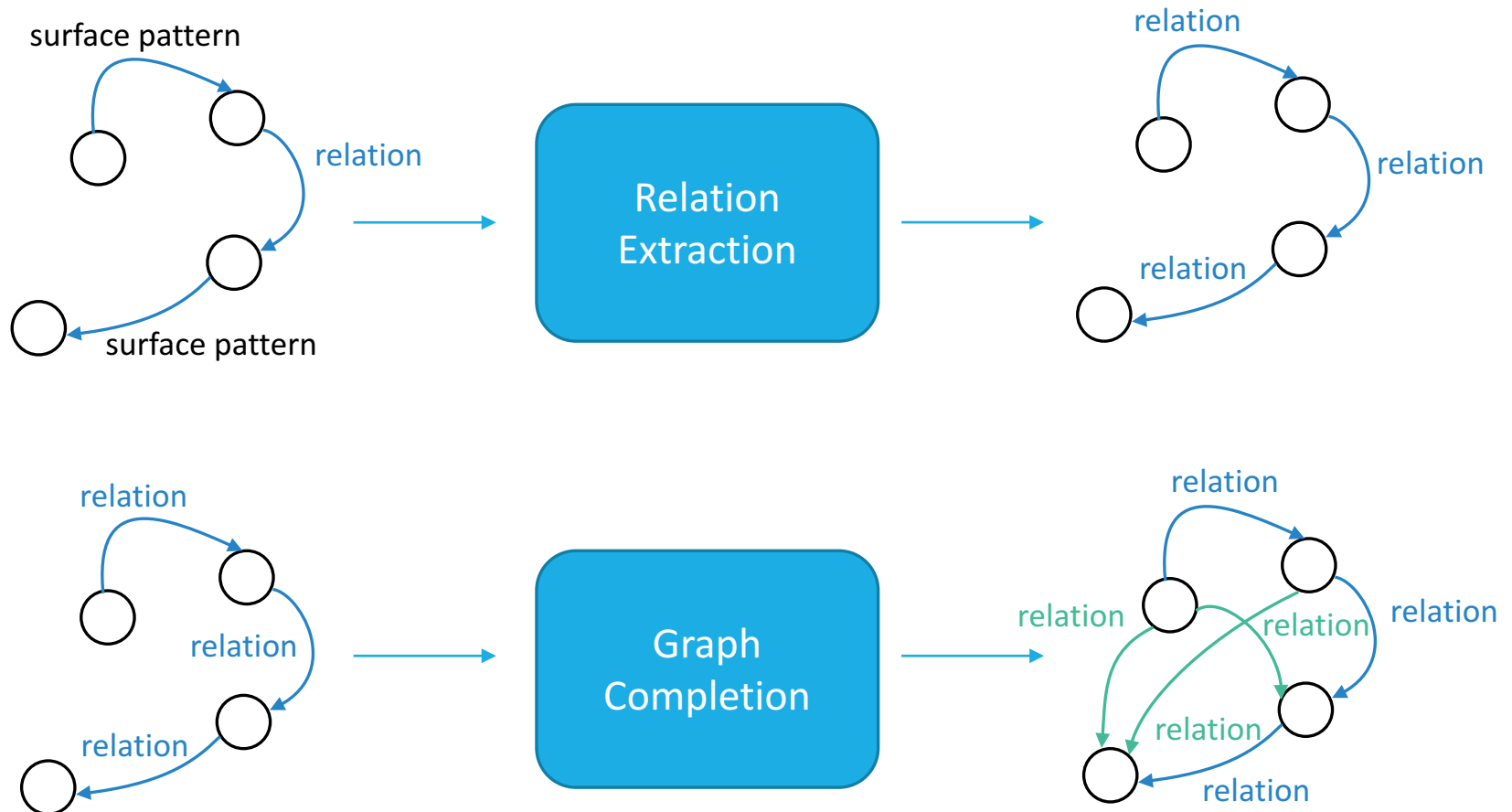
Computational Complexity of Algorithms

- Complexity depends on explicit dimensionality
 - Often NP-Hard, in size of data
 - More rules, more expensive inference
- Query-time inference is sometimes NP-Hard
- Not trivial to parallelize, or use GPUs

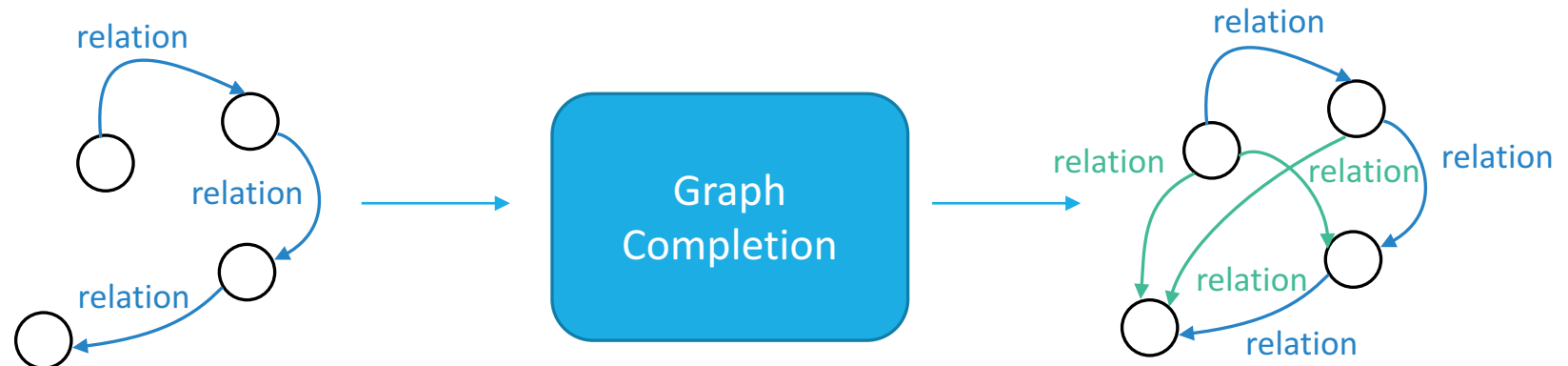
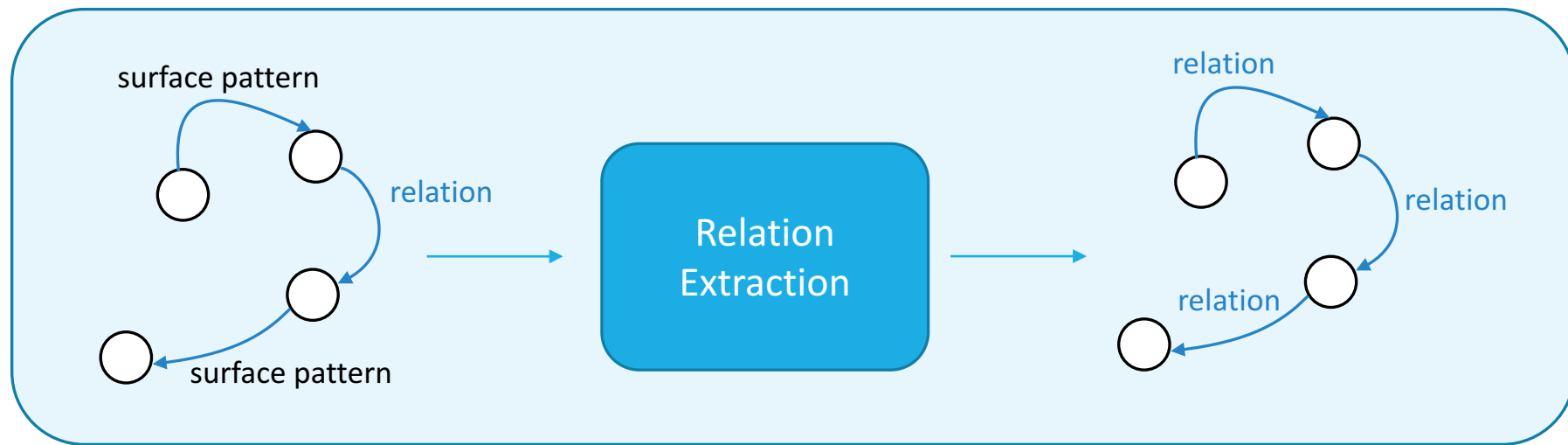
Embeddings

- Everything as dense vectors
- Can capture many relations
- Learned from data
- Complexity depends on latent dimensions
- Learning using stochastic gradient, back-propagation
- Querying is often cheap
- GPU-parallelism friendly

Two Related Tasks

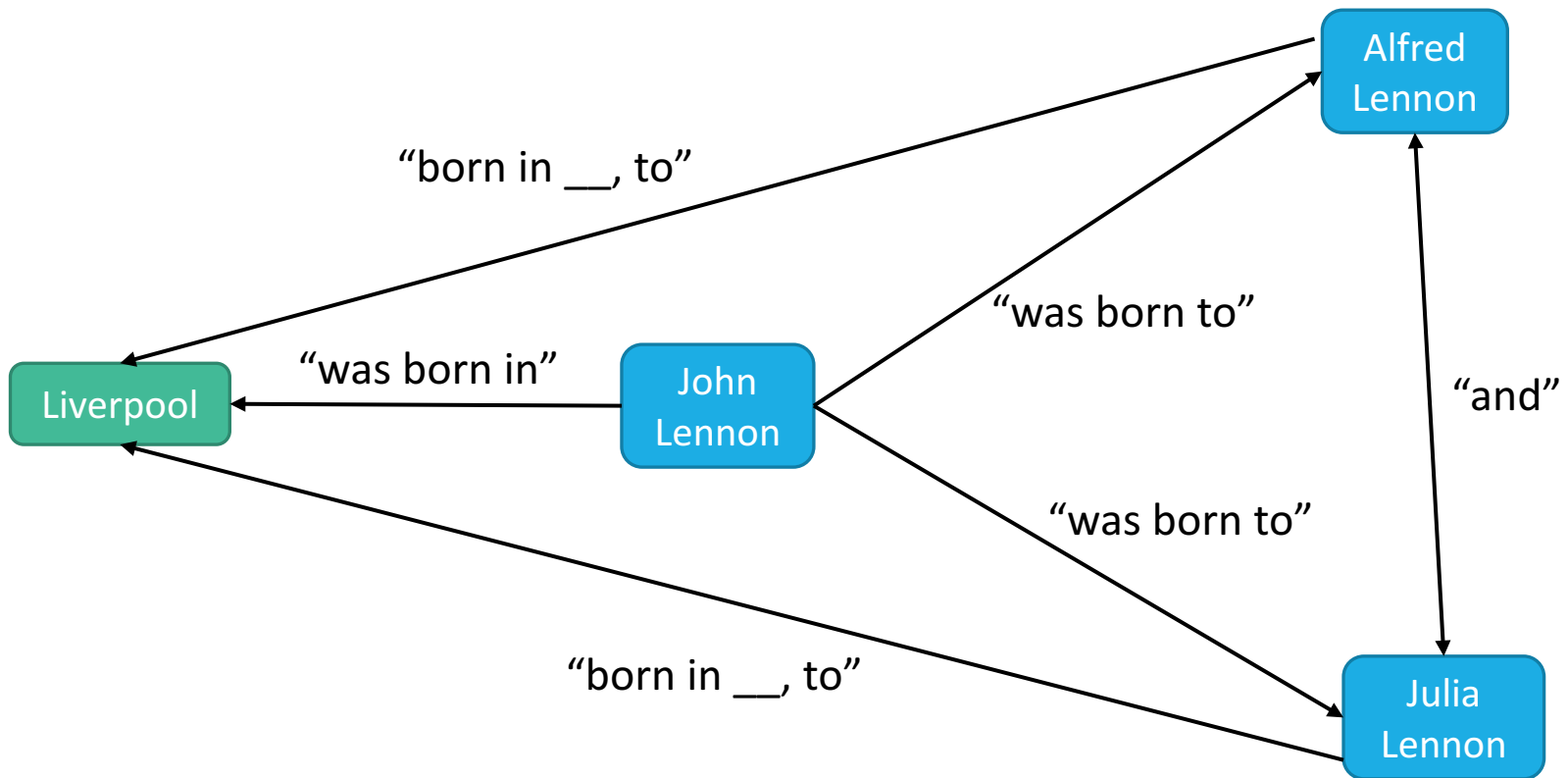


Two Related Tasks



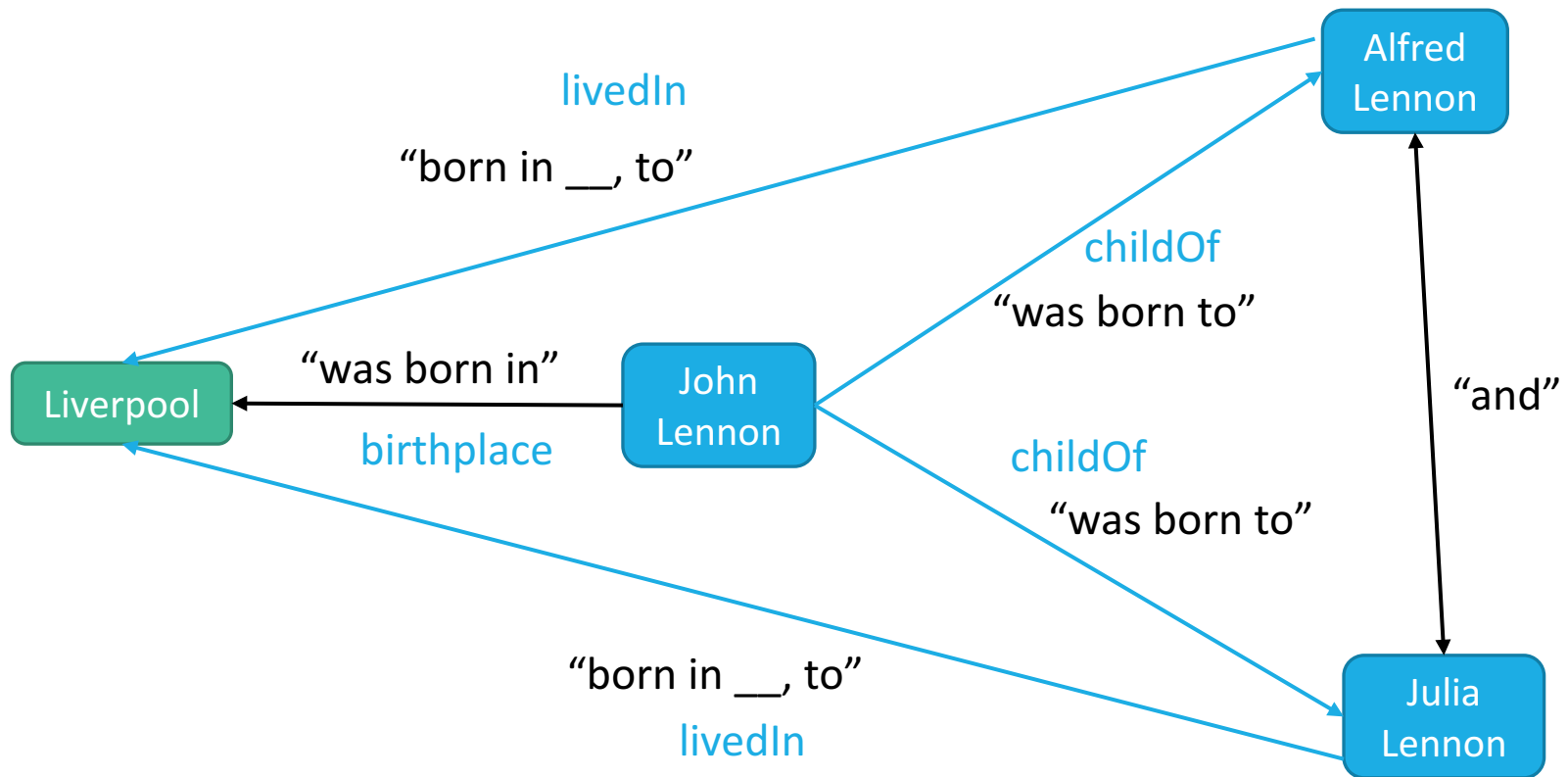
Relation Extraction From Text

John was born in Liverpool, to Julia and Alfred Lennon.

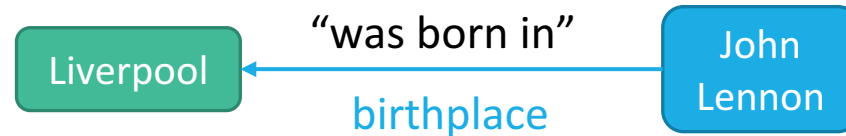


Relation Extraction From Text

John was born in Liverpool, to Julia and Alfred Lennon.



“Distant” Supervision

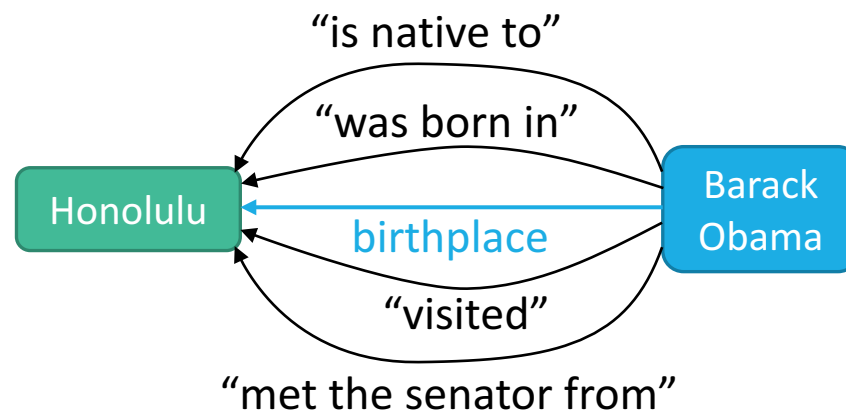


No direct supervision gives us this information.

Supervised: Too expensive to label sentences

Rule-based: Too much variety in language

Both only work for a small set of relations, i.e. 10s, not 100s



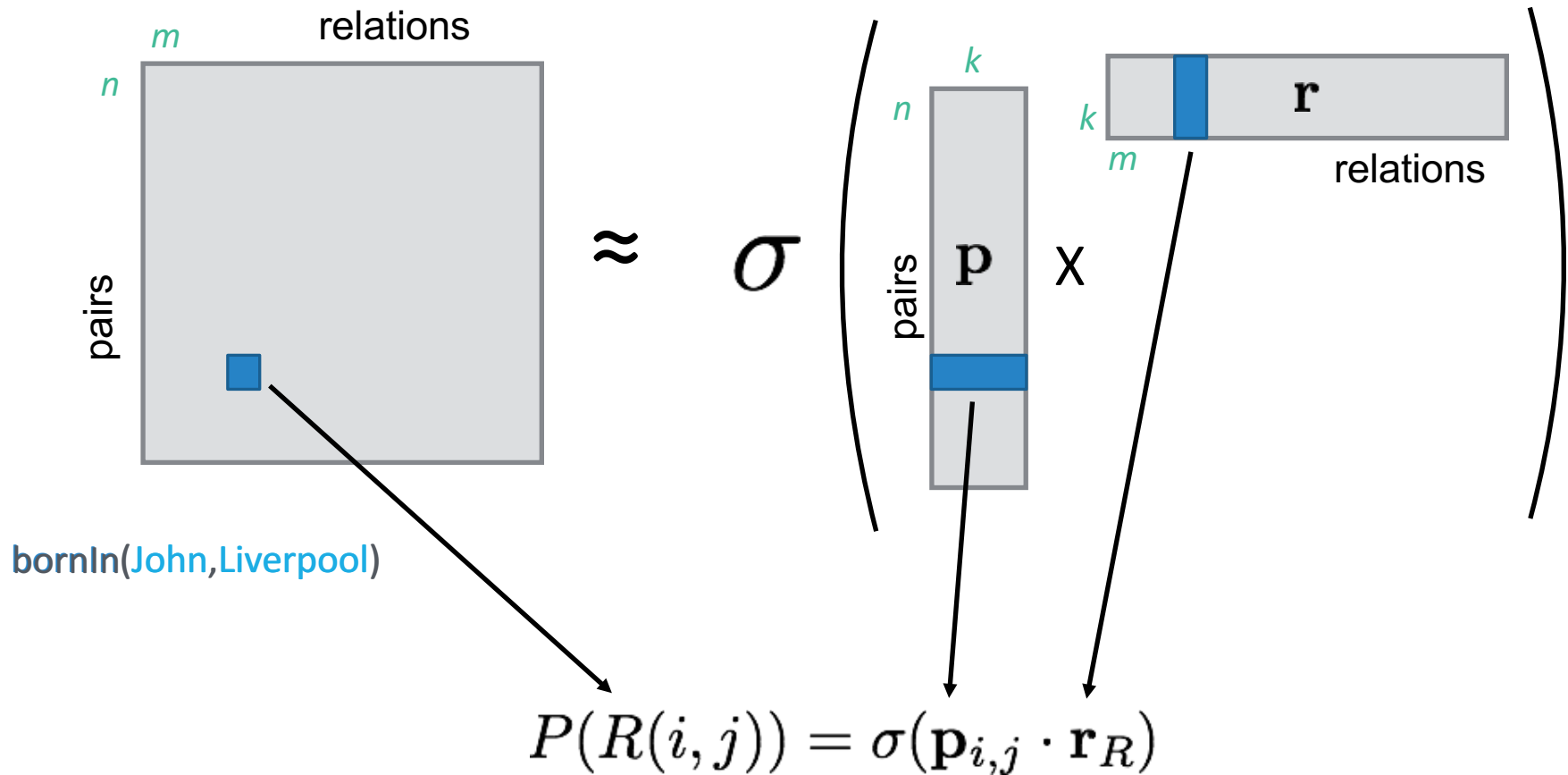
Relation Extraction as a Matrix

John was born in Liverpool, to Julia and Alfred Lennon.

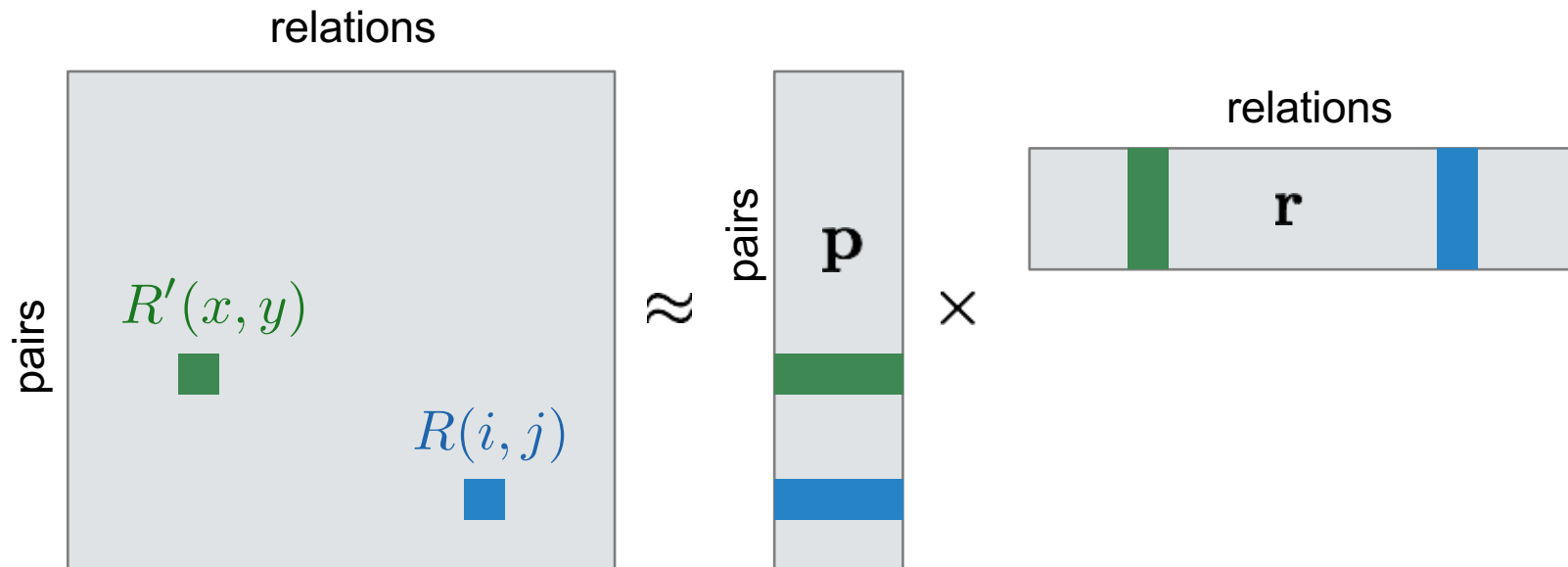
Entity Pairs

| | <i>was born in</i> <small><-nsubjpas-born<-nmod:in-</small> | <i>was born to</i> | <i>and</i> | <i>birthplace(x,y)</i> | <i>spouse(x,y)</i> |
|------------------------------|--|--------------------|------------|------------------------|--------------------|
| John Lennon, Liverpool | 1 | | | ? | |
| John Lennon, Julia Lennon | | 1 | | | |
| John Lennon, Alfred Lennon | | 1 | | | |
| Julia Lennon, Alfred Lennon | | | 1 | | ? |
| Barack Obama, Hawaii | 1 | | | 1 | |
| Barack Obama, Michelle Obama | | | 1 | | 1 |

Matrix Factorization



Training: Stochastic Updates



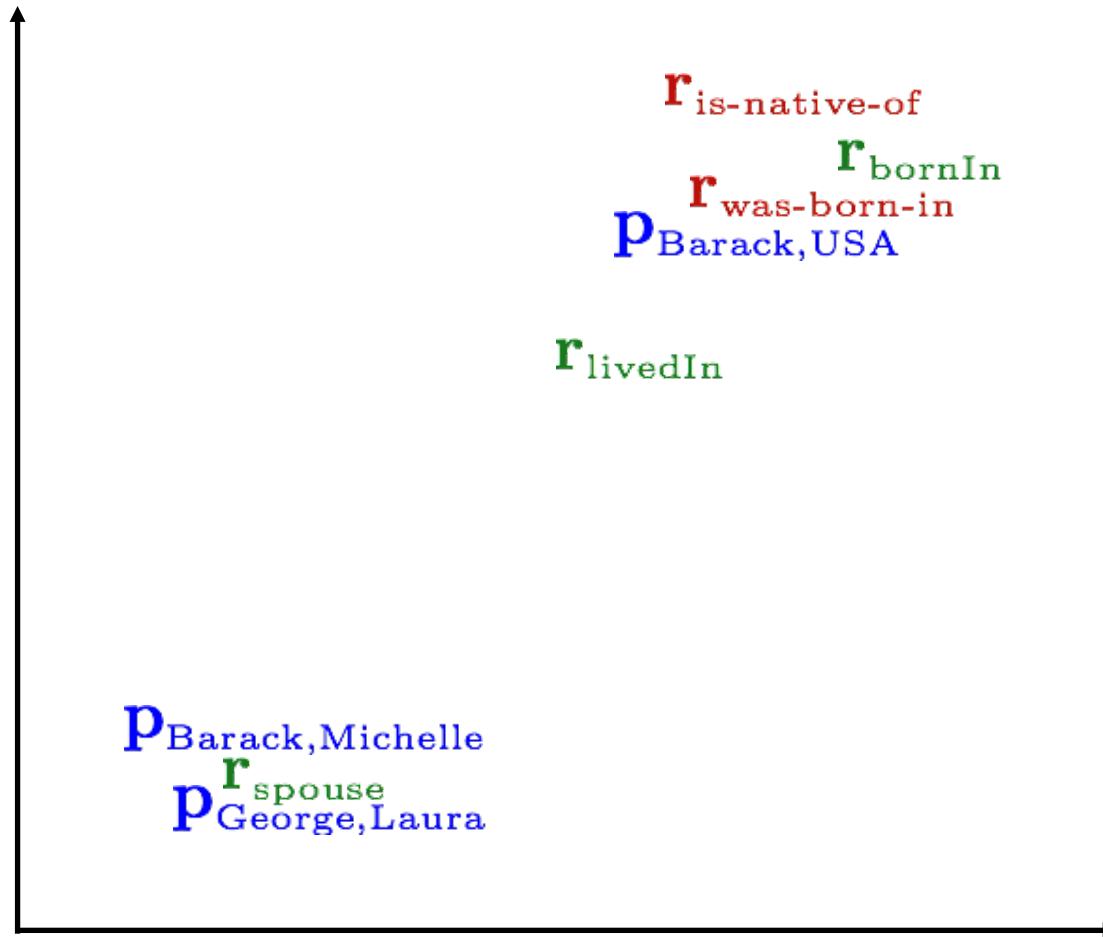
Pick an **observed** cell, $R(i, j)$:

- Update \mathbf{p}_{ij} & \mathbf{r}_R such that $R(i, j)$ is higher

Pick any random cell, assume it is **negative**:

- Update \mathbf{p}_{xy} & $\mathbf{r}_{R'}$ such that $R'(x, y)$ is lower

Relation Embeddings



Embeddings \sim Logical Relations

Relation Embeddings, w

- Similar embedding for 2 relations denote they are paraphrases
 - *is married to*, *spouseOf(X,Y)*, */person/spouse*
- One embedding can be contained by another
 - $w(\text{topEmployeeOf}) \subset w(\text{employeeOf})$
 - $\text{topEmployeeOf}(X,Y) \rightarrow \text{employeeOf}(X,Y)$
- Can capture logical patterns, without needing to specify them!

Entity Pair Embeddings, v

Similar entity pairs denote similar relations between them

Entity pairs may describe multiple “relations”

independent *foundedBy* and *employeeOf* relations

Similar Embeddings

similar underlying embedding

X own percentage of Y X buy stake in Y

similar embedding

| | | |
|-----------------------------------|---|---|
| Time, Inc Amer. Tel. and Comm. | 1 | 1 |
| Volvo Scania A.B. | | 1 |
| Campeau Federated Dept Stores | | |
| Apple HP | | |

Successfully predicts “Volvo owns percentage of Scania A.B.”
from “Volvo bought a stake in Scania A.B.”

Implications

X historian at Y \rightarrow X professor at Y

X professor at Y

X historian at Y

(Freeman, Harvard)
→ (Boyle, OhioState)

Kevin Boyle
Ohio State

R. Freeman
Harvard

| | |
|---|---|
| | 1 |
| 1 | |

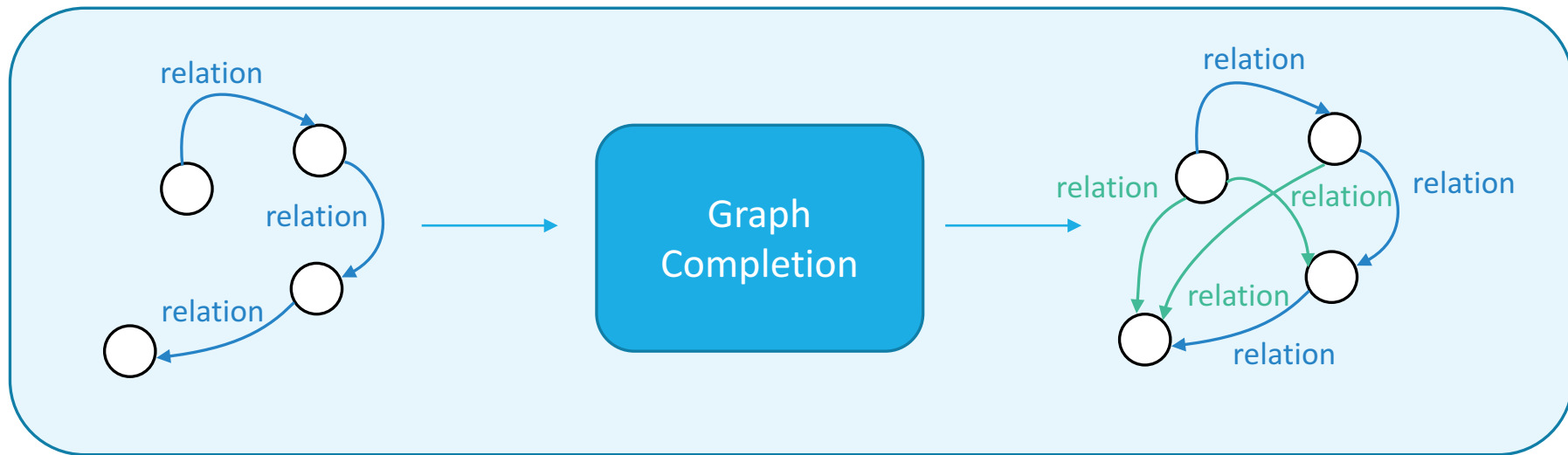
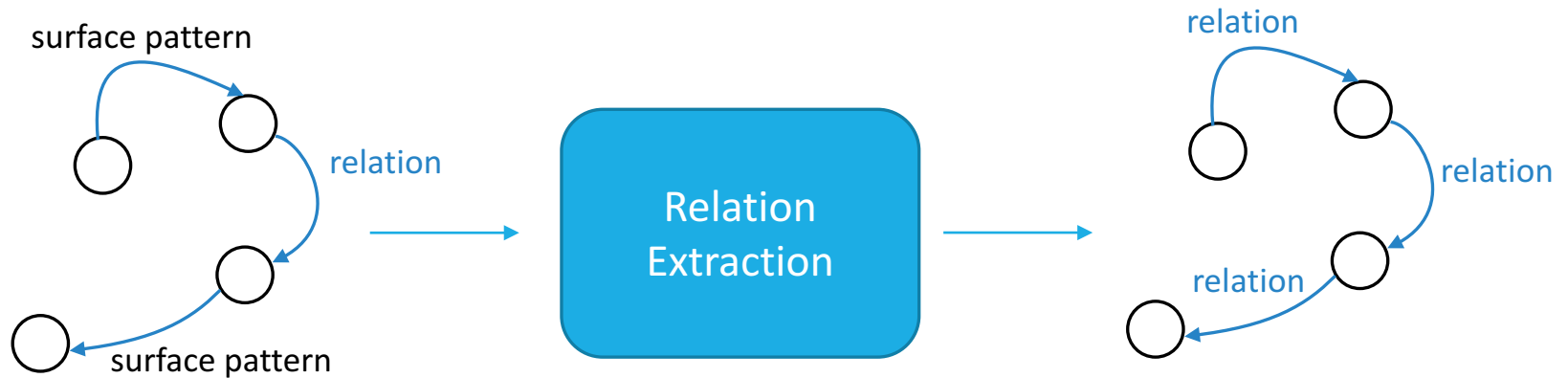
Learns asymmetric entailment:

PER historian at UNIV → PER professor at UNIV

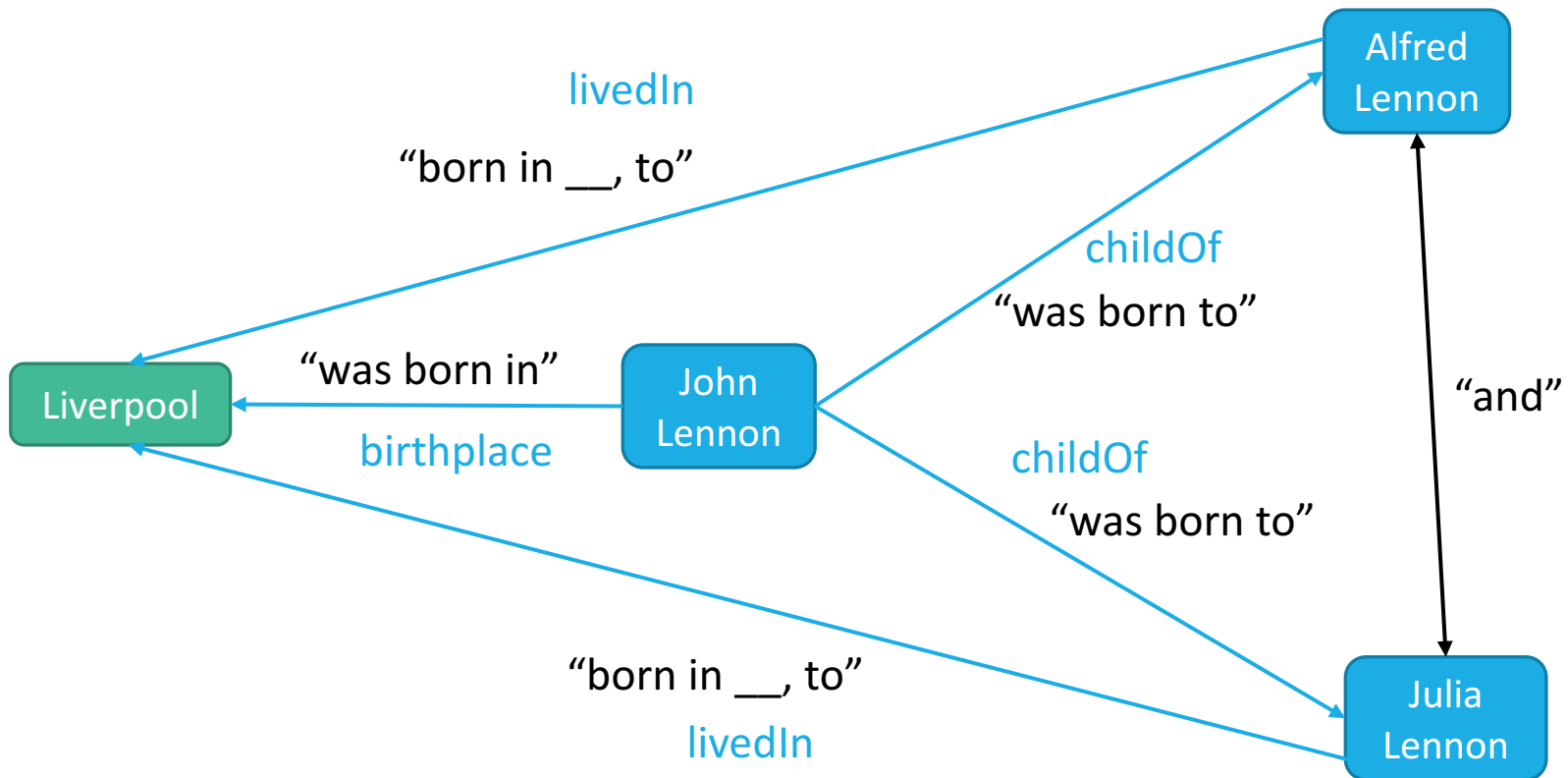
But,

PER professor at UNIV \nrightarrow PER historian at UNIV

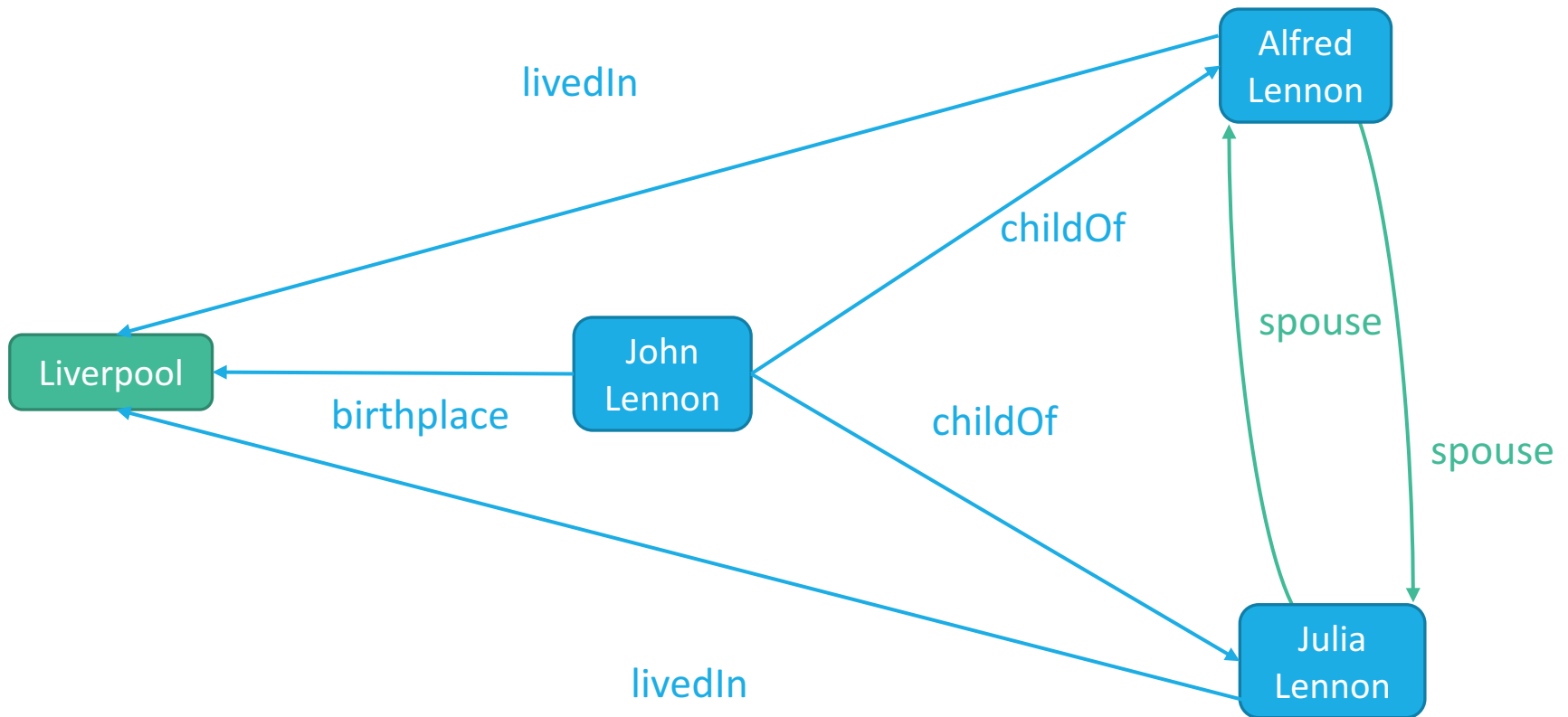
Two Related Tasks



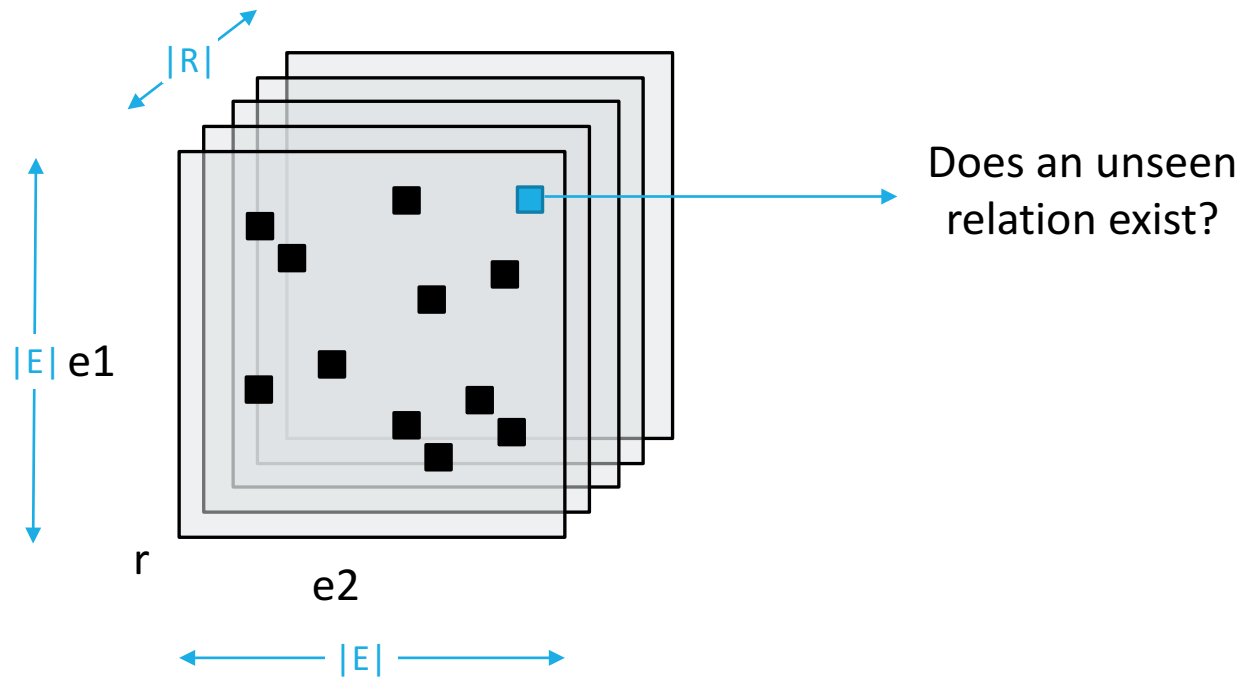
Graph Completion



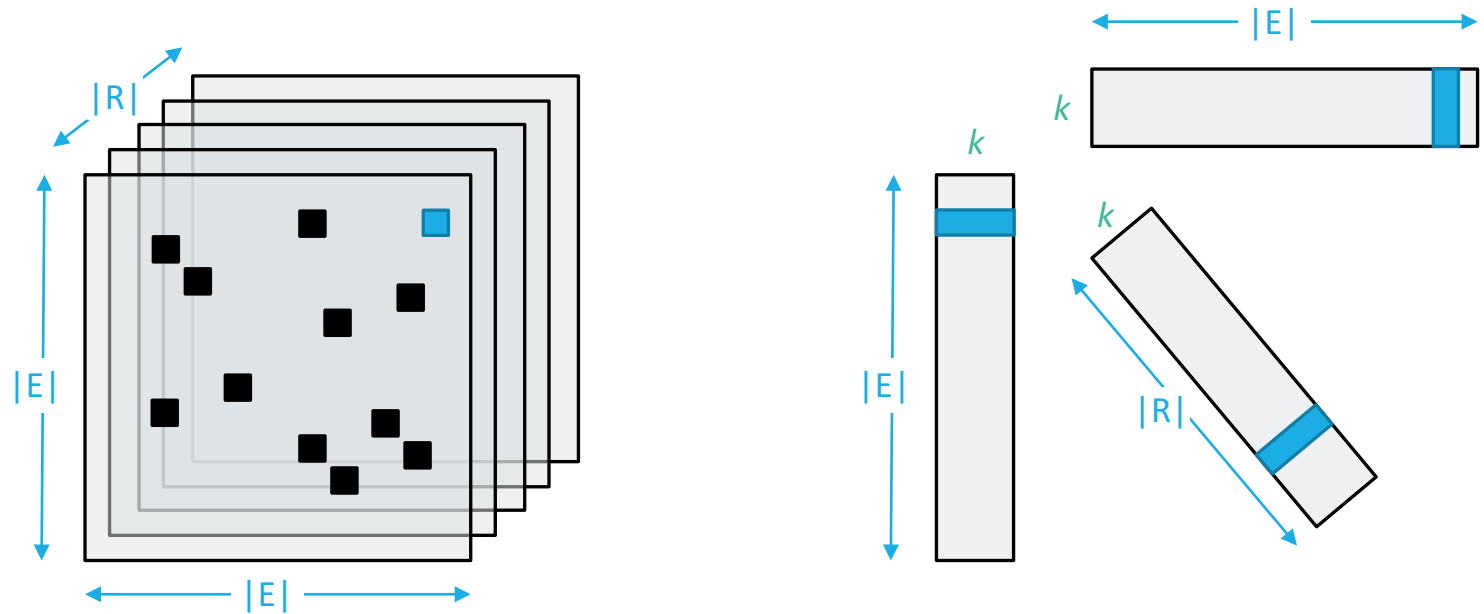
Graph Completion



Tensor Formulation of KG



Factorize that Tensor



$$S(r(a, b)) = f(\mathbf{v}_r, \mathbf{v}_a, \mathbf{v}_b)$$

Many Different Factorizations

CANDECOMP/PARAFAC-Decomposition

$$S(r(a, b)) = \sum_k R_{r,k} \cdot e_{a,k} \cdot e_{b,k}$$

Tucker2 and RESCAL Decompositions

$$S(r(a, b)) = (\mathbf{R}_r \times \mathbf{e}_a) \times \mathbf{e}_b$$

Model E

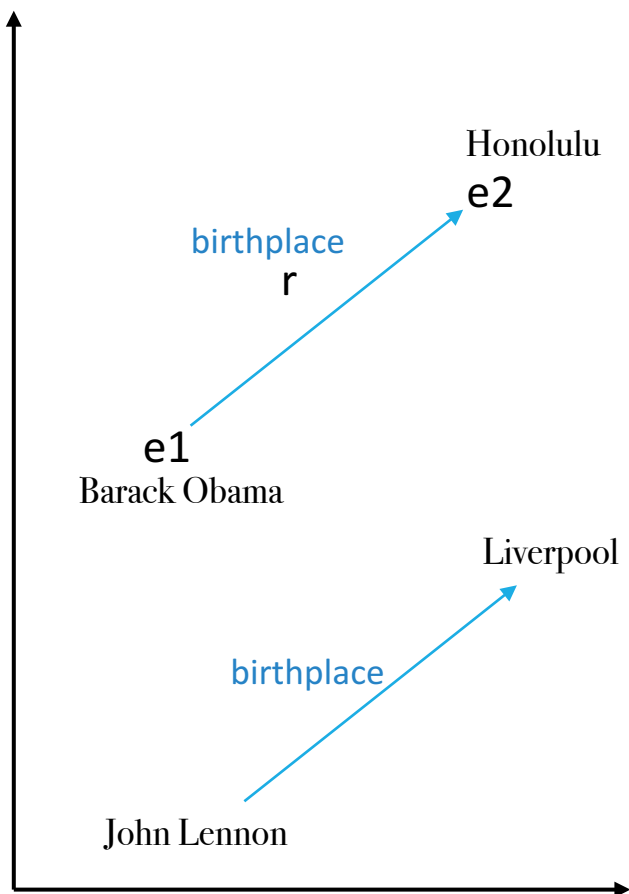
$$S(r(a, b)) = \mathbf{R}_{r,1} \cdot \mathbf{e}_a + \mathbf{R}_{r,2} \cdot \mathbf{e}_b$$

Holographic Embeddings

$$S(r(a, b)) = \mathbf{R}_r \times (\mathbf{e}_a \star \mathbf{e}_b)$$

Not tensor
factorization
(per se)

Translation Embeddings



TransE

$$S(r(a, b)) = -\|\mathbf{e}_a + \mathbf{R}_r - \mathbf{e}_b\|_2^2$$

TransH

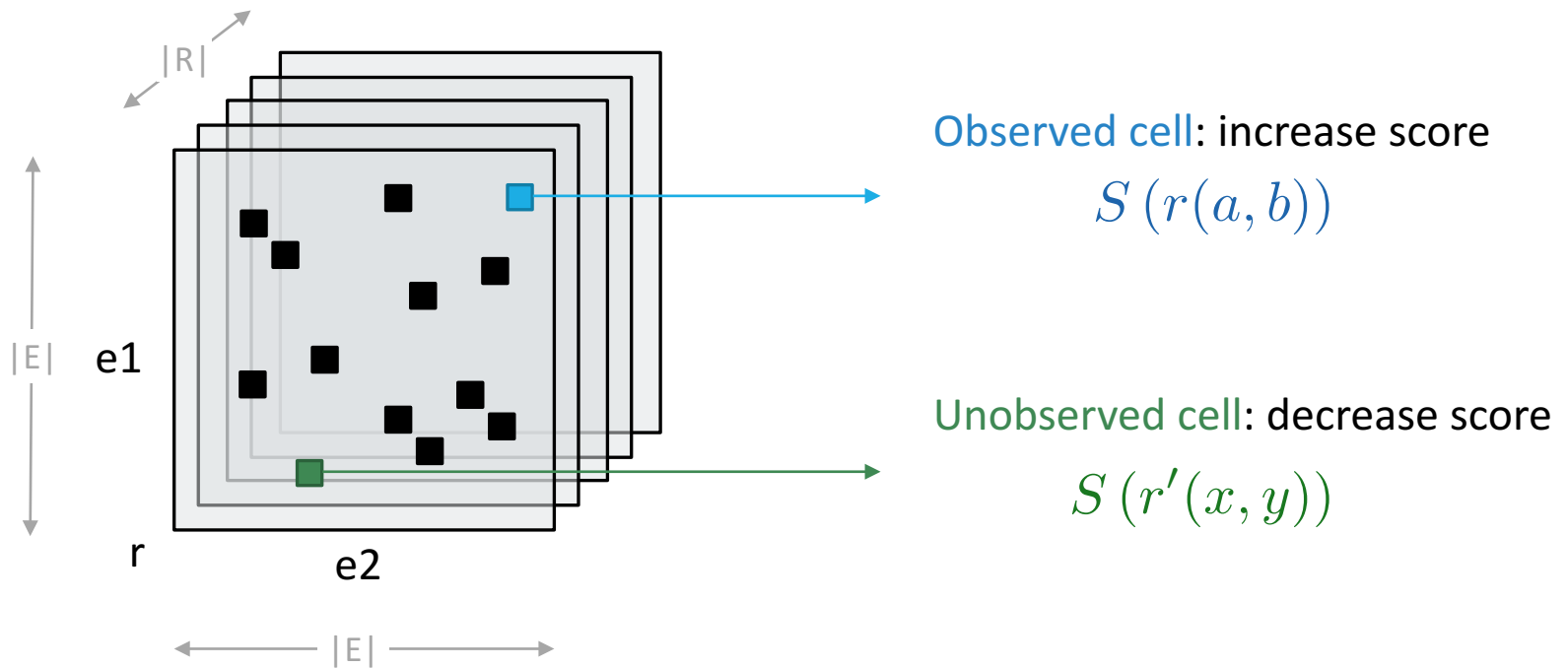
$$S(r(a, b)) = -\|\mathbf{e}_a^\perp + \mathbf{R}_r - \mathbf{e}_b^\perp\|_2^2$$

$$\mathbf{e}_a^\perp = \mathbf{e}_a - \mathbf{w}_r^T \mathbf{e}_a \mathbf{w}_r$$

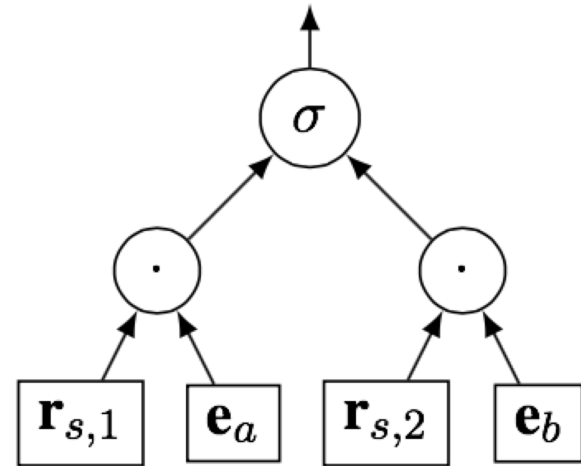
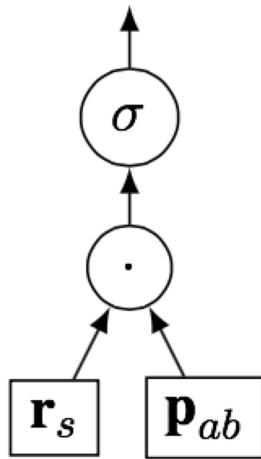
TransR

$$S(r(a, b)) = -\|\mathbf{e}_a \mathbf{M}_r + \mathbf{R}_r - \mathbf{e}_b \mathbf{M}_r\|_2^2$$

Parameter Estimation



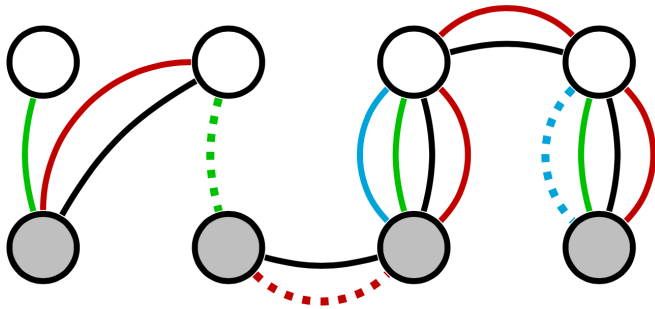
Matrix vs Tensor Factorization



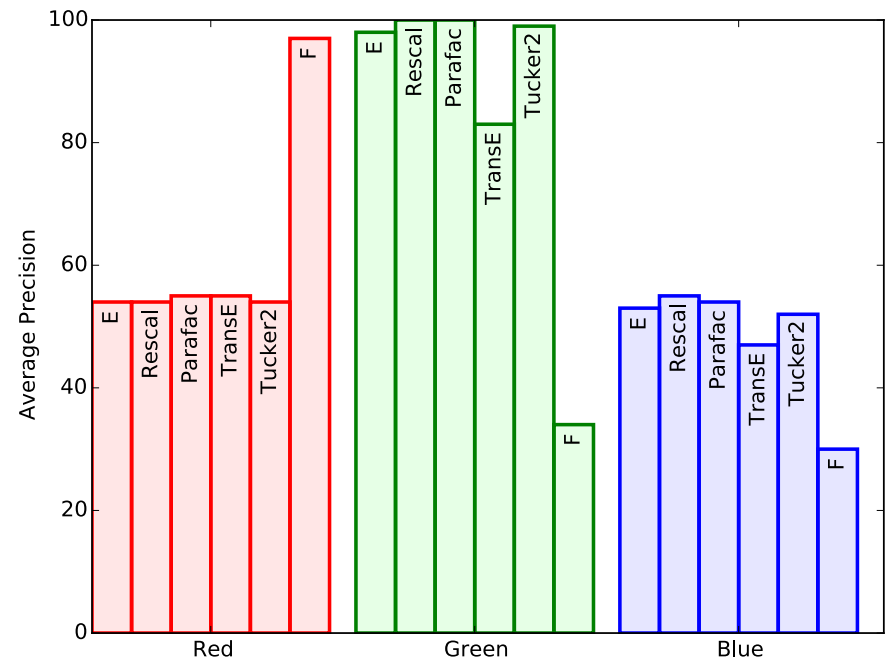
- No encoding of type information
- Can only predict for entity pairs that appear in text together
- Sufficient evidence has to be seen for each entity pair

- Assume low-rank for pairs
- But many relations are not!
- Spouse: you can have only ~ 1
- Cannot learn pair specific information

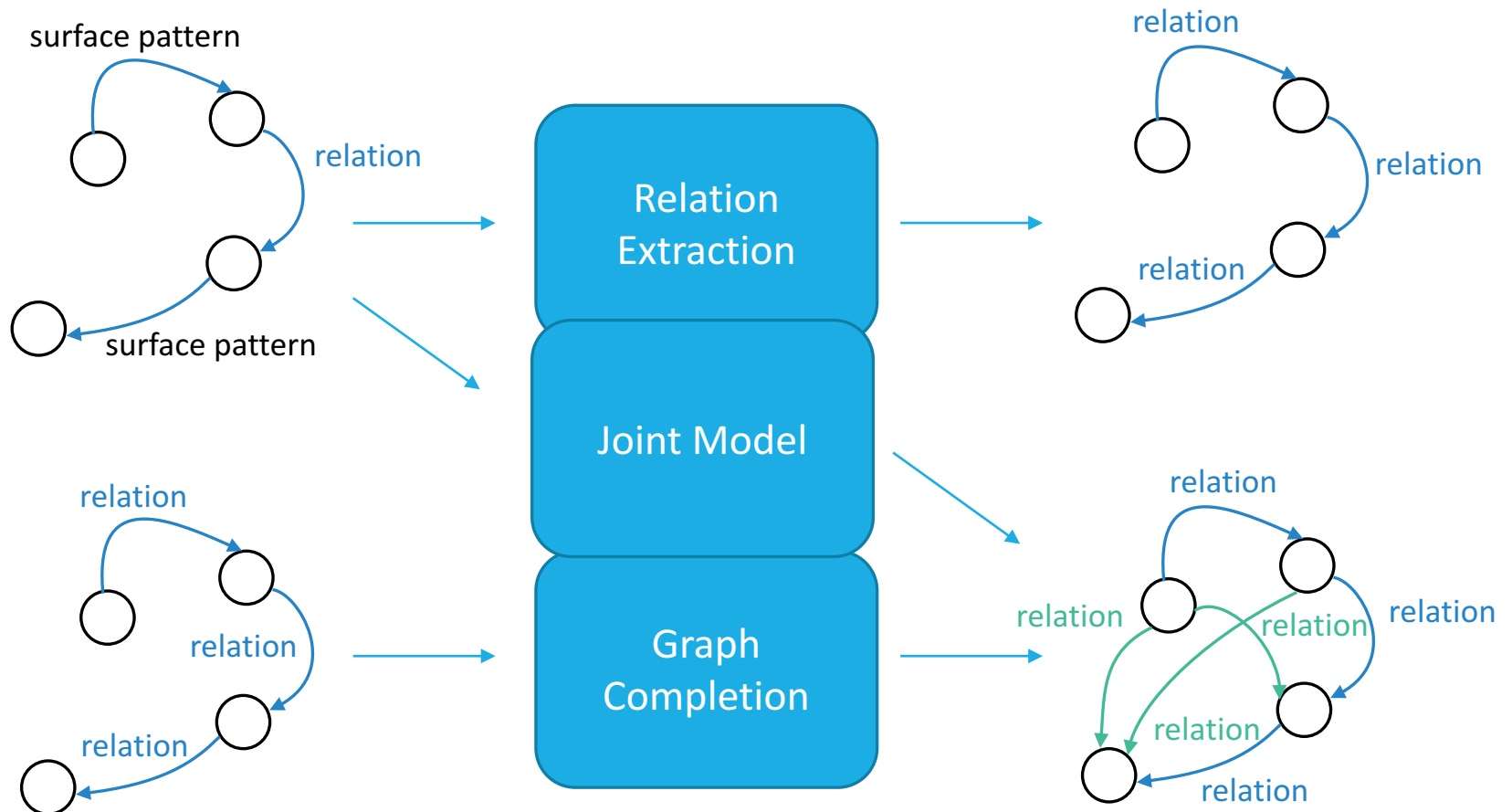
What they can, and can't, do..



- **Red**: deterministically implied by **Black**
 - needs *pair-specific* embedding
 - Only **F** is able to generalize
- **Green**: needs to estimate entity types
 - needs *entity-specific* embedding
 - Tensor factorization generalizes, **F** doesn't
- **Blue**: implied by **Red** and **Green**
 - Nothing works much better than random



Joint Extraction+Completion



Compositional Neural Models

So far, we're learning vectors for each entity/surface pattern/relation..

But learning vectors independently ignores “composition”

Composition in Surface Patterns

- Every surface pattern is not unique
- Synonymy: A is B's spouse.
A is married to B.
- Inverse: X is Y's parent.
Y is one of X's children.
- Can the representation learn this?

Composition in Relation Paths

- Every relation path is not unique
- Explicit: A parent B, B parent C
A grandparent C
- Implicit: X bornInCity Y, Y cityInState Z
X “bornInState” Z
- Can the representation capture this?

Composing Dependency Paths

... was born to ...



... 's parents are ...



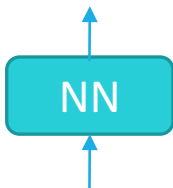
(never appears in
training data)

`\parentsOf`

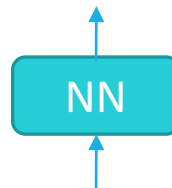


But we don't need linked data to know they mean similar things...

Use neural networks to produce the embeddings from text!



... was born to ...

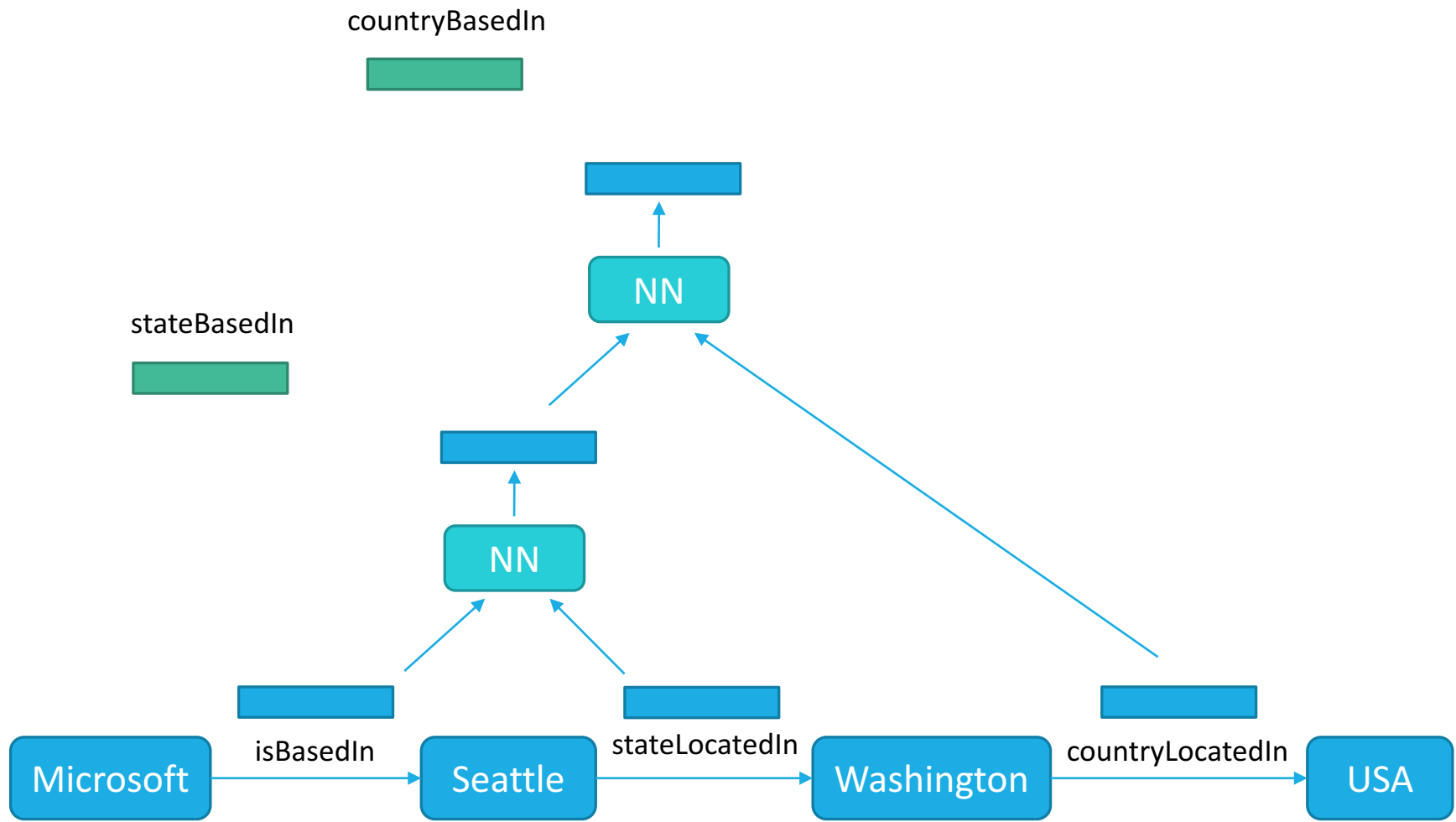


... 's parents are ...



`\parentsOf`

Composing Relational Paths



Review: Embedding Techniques

Two Related Tasks:

- Relation Extraction from Text
- Graph (or Link) Completion

Relation Extraction:

- Matrix Factorization Approaches

Graph Completion:

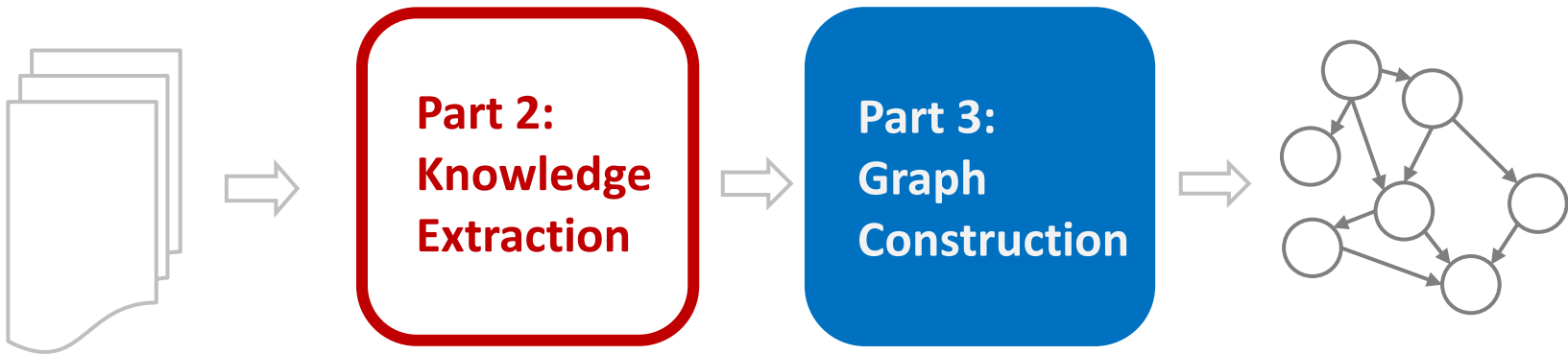
- Tensor Factorization Approaches

Compositional Neural Models

- Compose over dependency paths
- Compose over relation paths

Tutorial Overview

Part 1: Knowledge Graphs



Part 4: Critical Analysis