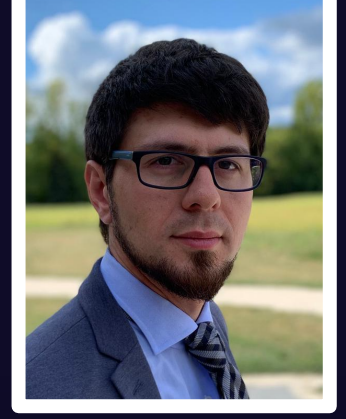


# Graph-Query Suggestions for Knowledge Graph Exploration

**Matteo Lissandrini,** Davide Mottin, Themis Palpanas, Yannis Velegrakis

AALBORG UNIVERSITY DENMARK

Université de Paris

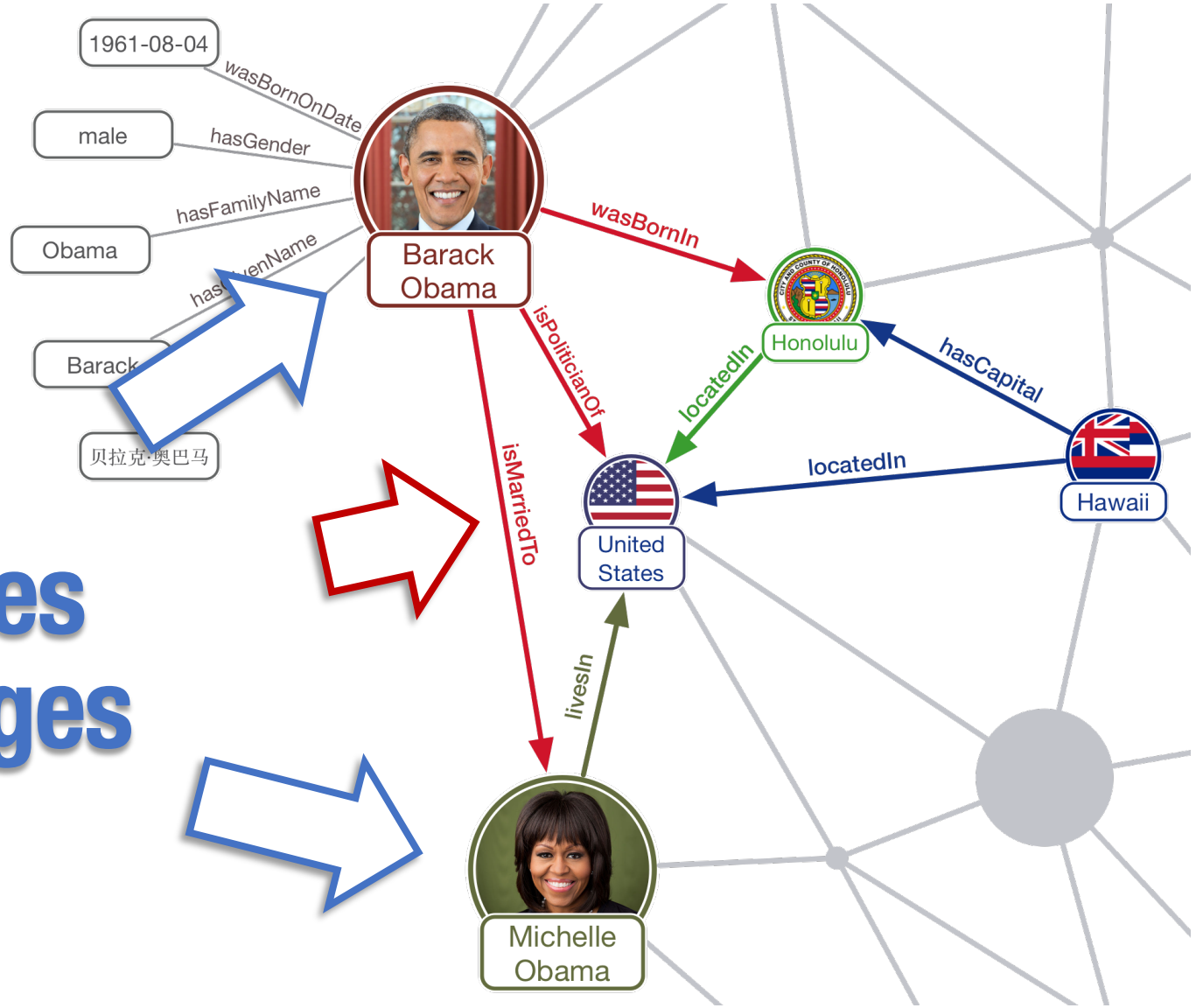
AARHUS UNIVERSITET

Universiteit Utrecht

©Ambiverse GmbH

# Edge-labelled Multigraphs

$\mathbf{G}: \langle V, E, L, \ell \rangle$




## Nodes & Edges





[*] https://queue.acm.org/detail.cfm?id=3332266

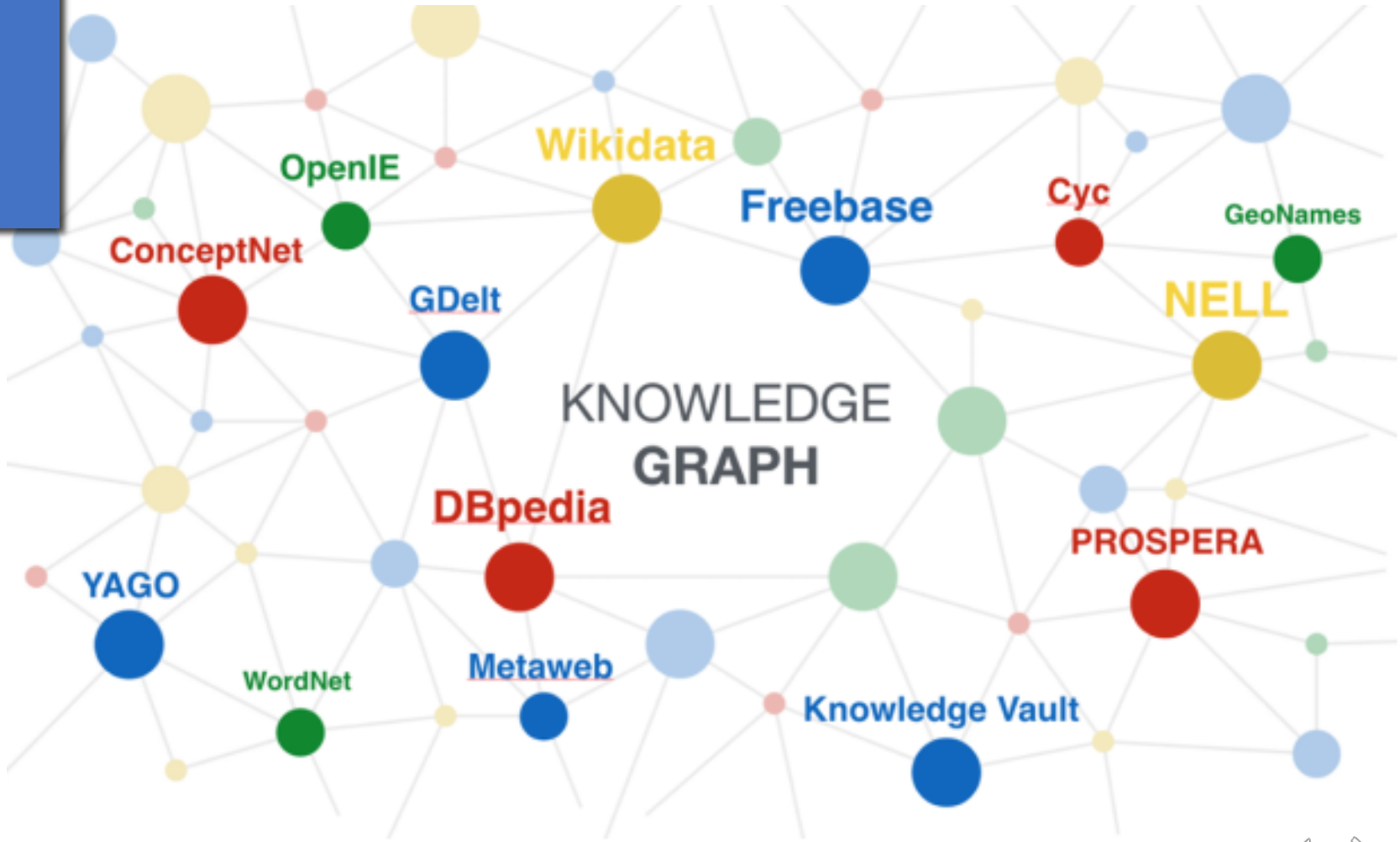
©Ambiverse GmbH

# Edge-labelled Multigraphs

$\mathbf{G}: \langle V, E, L, \ell \rangle$

**In use at:***

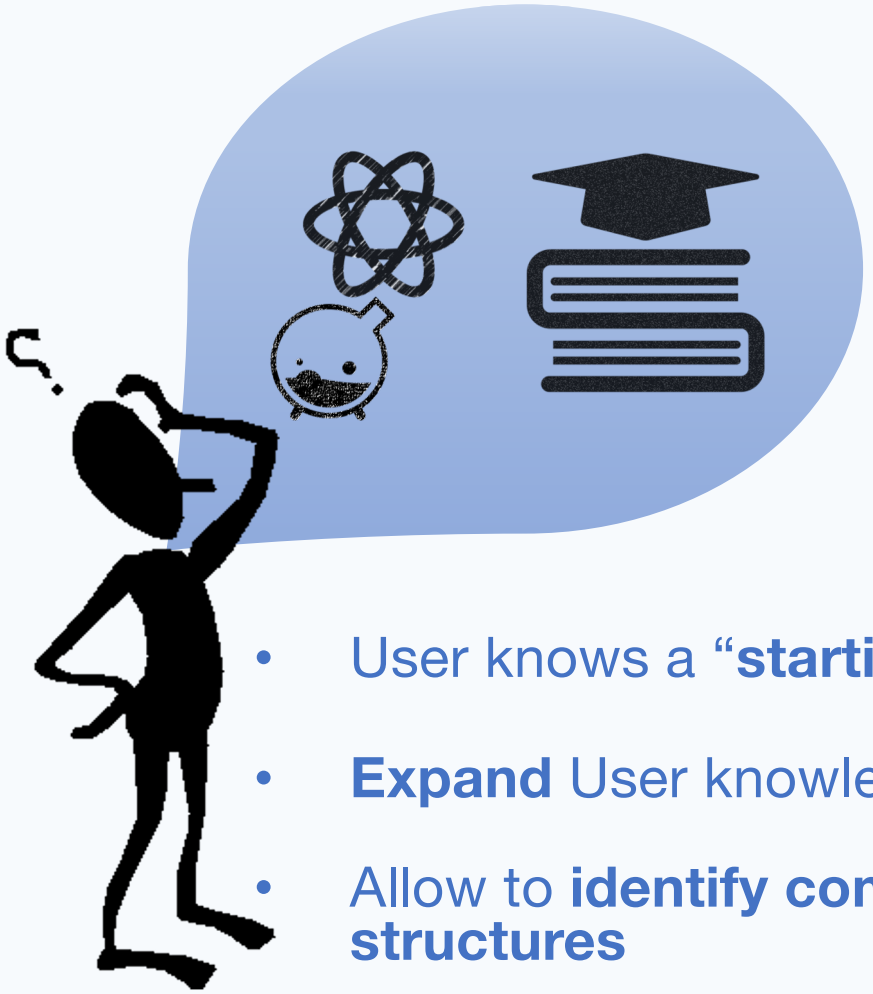
FACEBOOK IBM Microsoft Google ebay



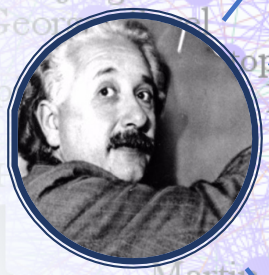



[*] https://queue.acm.org/detail.cfm?id=3332266

# Knowledge Graph Exploration



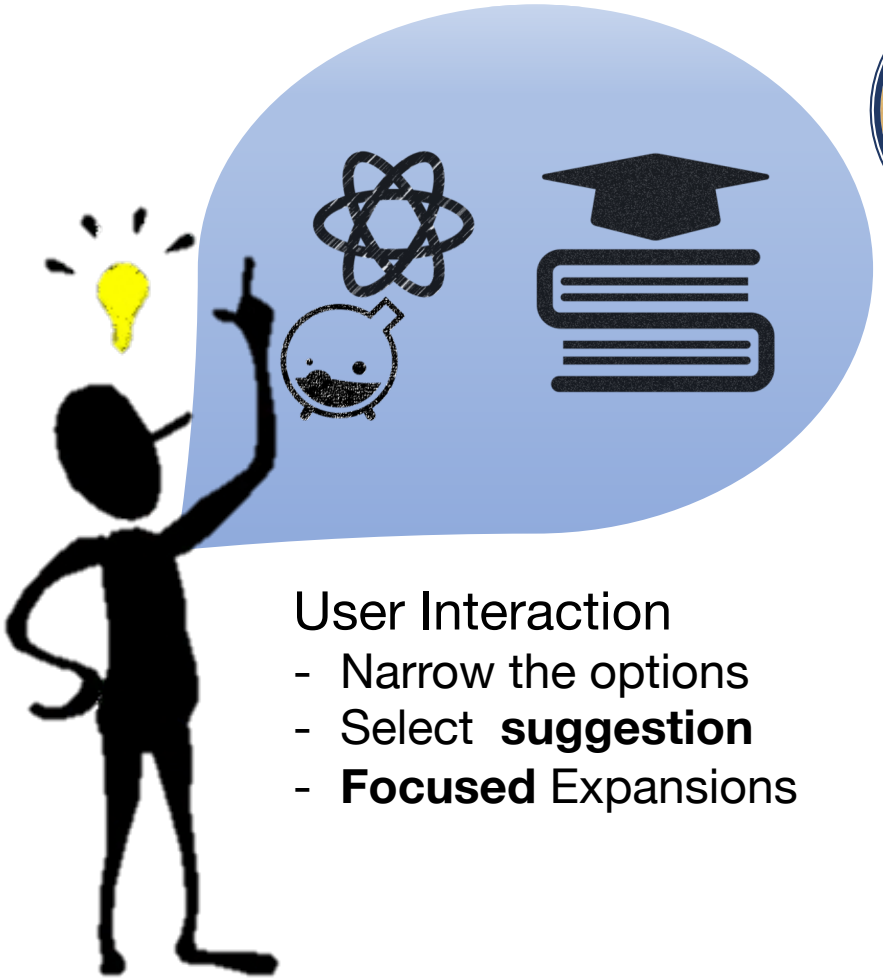
- User knows a “**starting point**”
- **Expand** User knowledge
- Allow to **identify complex structures**



**500+ outgoing relationships**

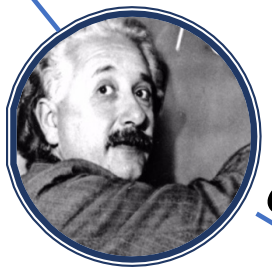
**41 Edge Types**

# Solution: **Interactive Suggestion System**



User Interaction
- Narrow the options
- Select **suggestion**
- **Focused** Expansions

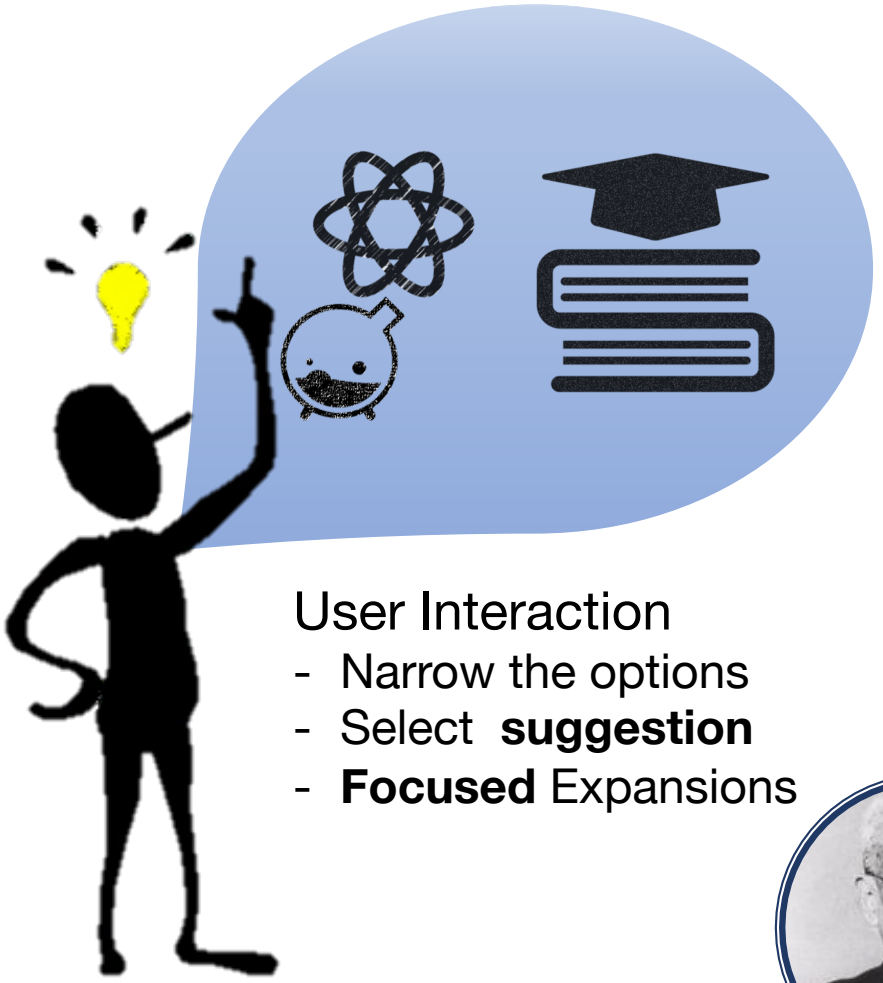
won



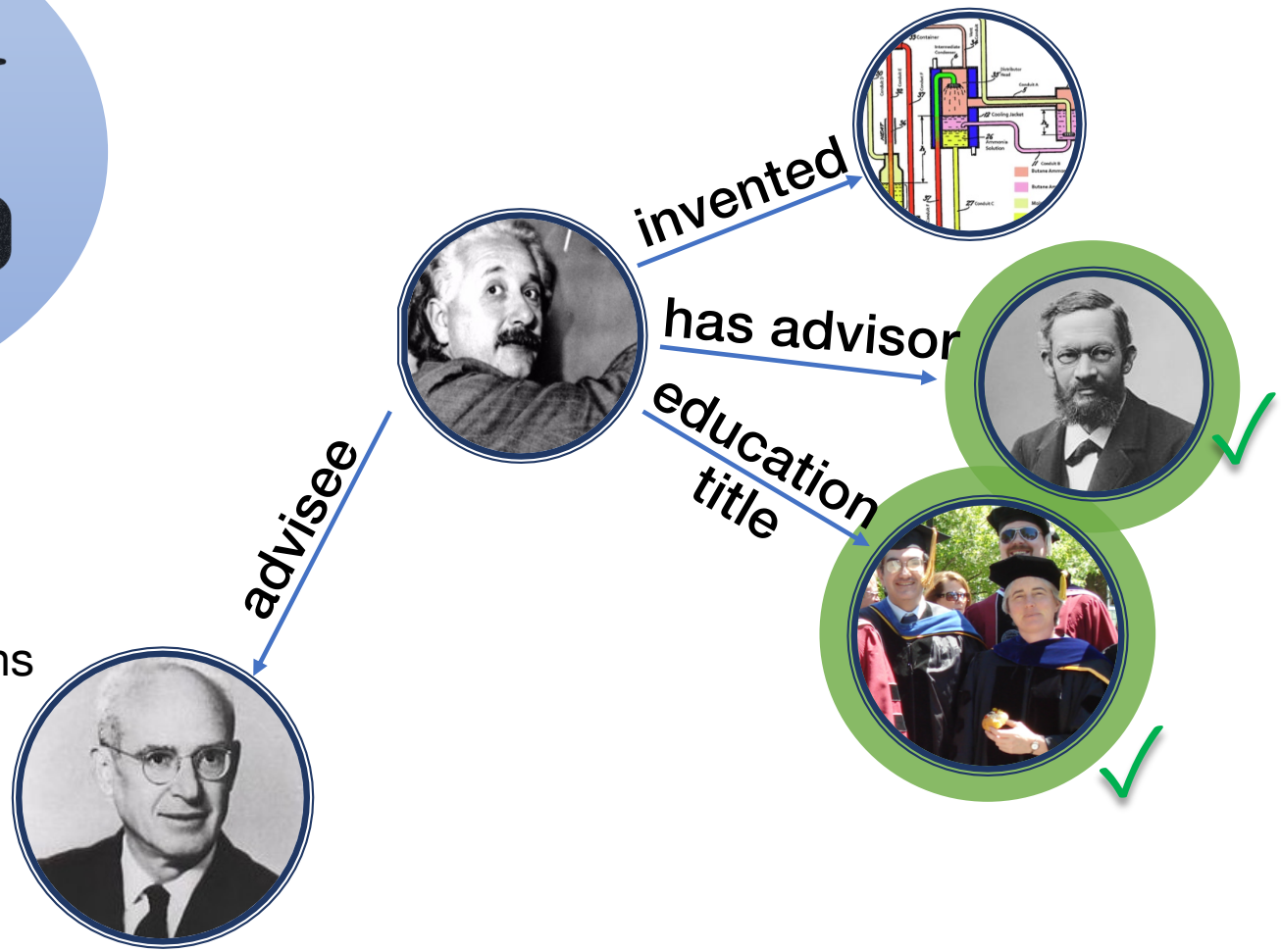
education
title

married

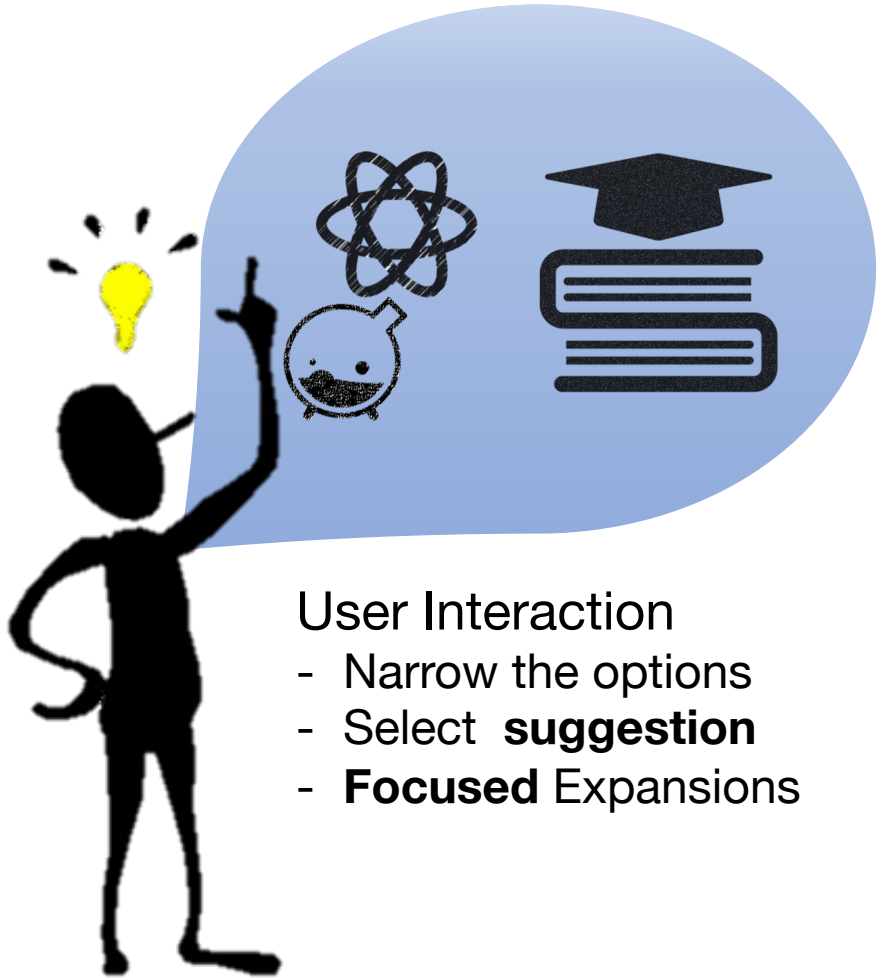
# Solution: **Interactive Suggestion System**



User Interaction
- Narrow the options
- Select **suggestion**
- **Focused** Expansions



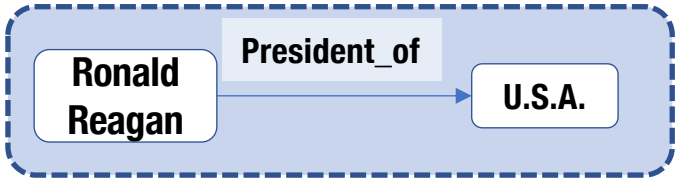
# Solution: **Interactive Suggestion System**



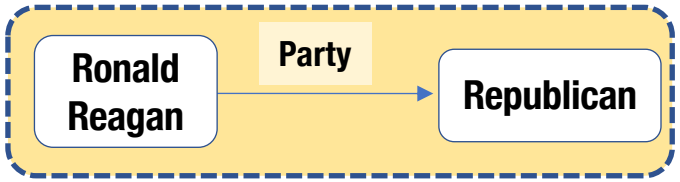
User Interaction
- Narrow the options
- Select **suggestion**
- **Focused** Expansions

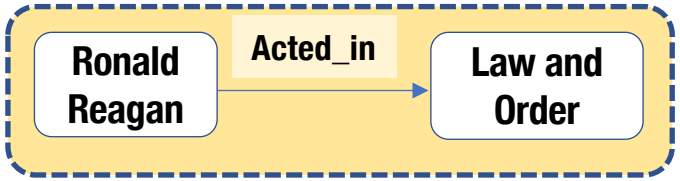
# Suggesting Expansions

## The User Search




## Which Expansion to Suggest?

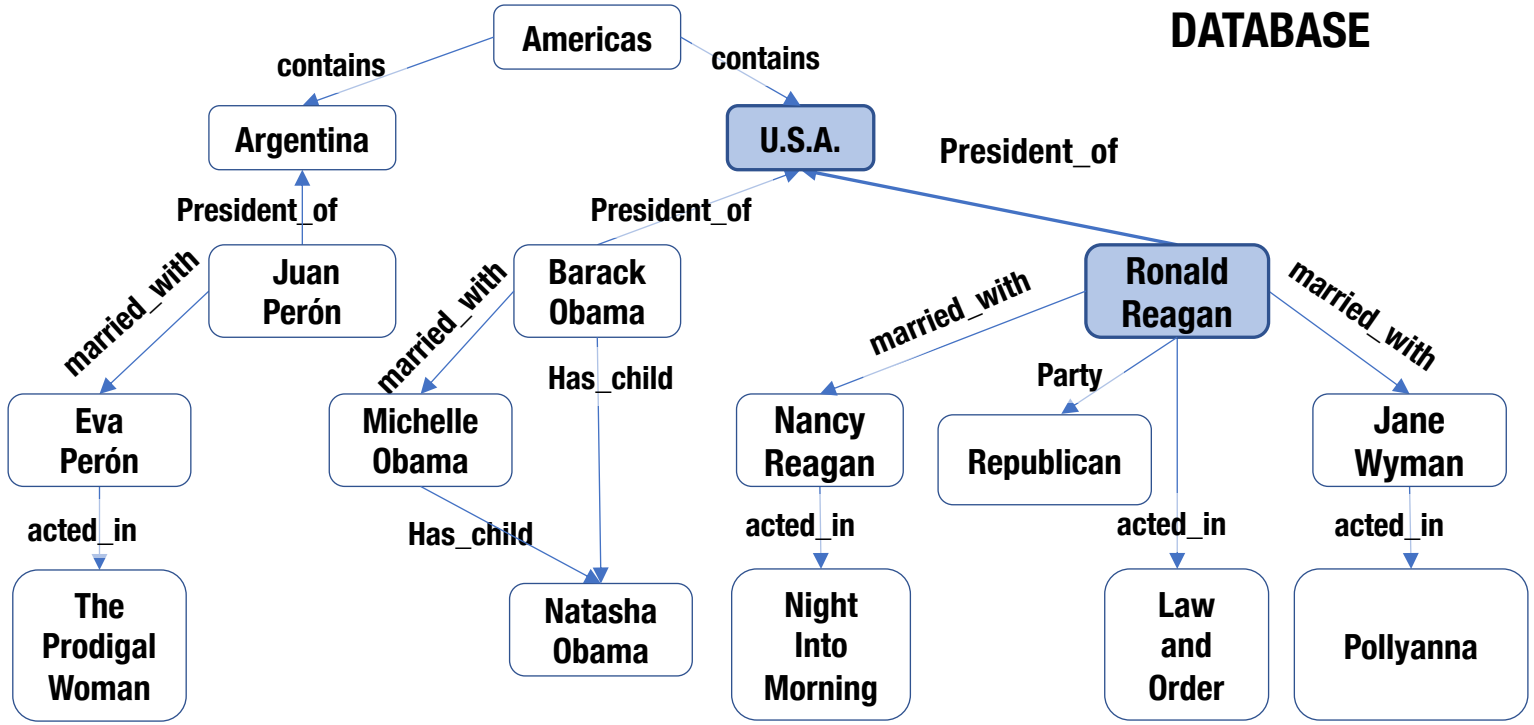


. . .

Rank Expansions

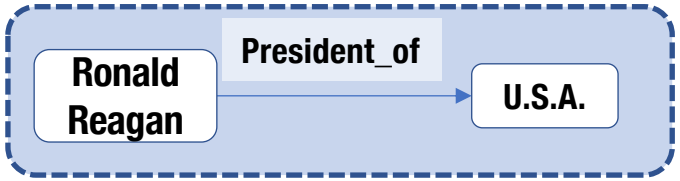
DATABASE



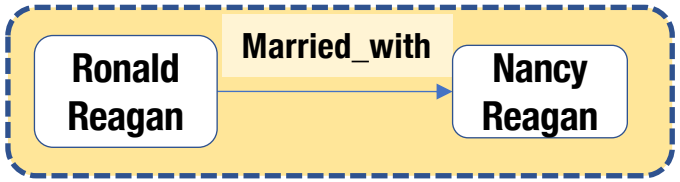

Approach:
**Rank Edges based on Label characteristics**

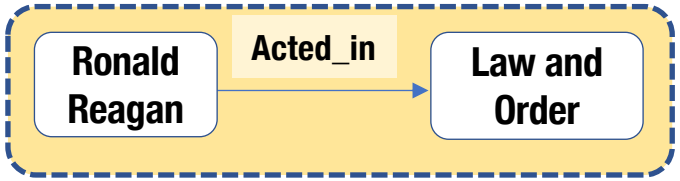
# Suggesting Expansions: **Ranking Labels (I)**

## The User Search




## How to score Expansions?



## Baseline Methods: Language Model

1. **Frequent** in the Graph
2. **Frequent** Around the Query

Probabilisic Model for the Query

$$\hat{p}(l|M_{\bar{Q}})_{MLE} = \frac{|E_Q^l| + \epsilon \hat{p}(l|\mathcal{K})}{|E_Q| + \epsilon}$$

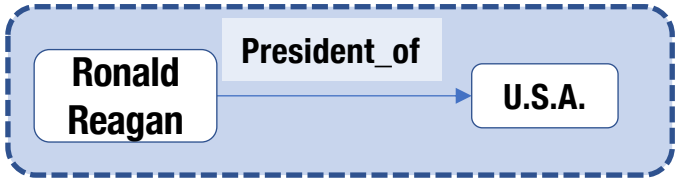
3. **Frequent** Around the Query but **Infrequent in the Graph**

Kullback–Leibler divergence

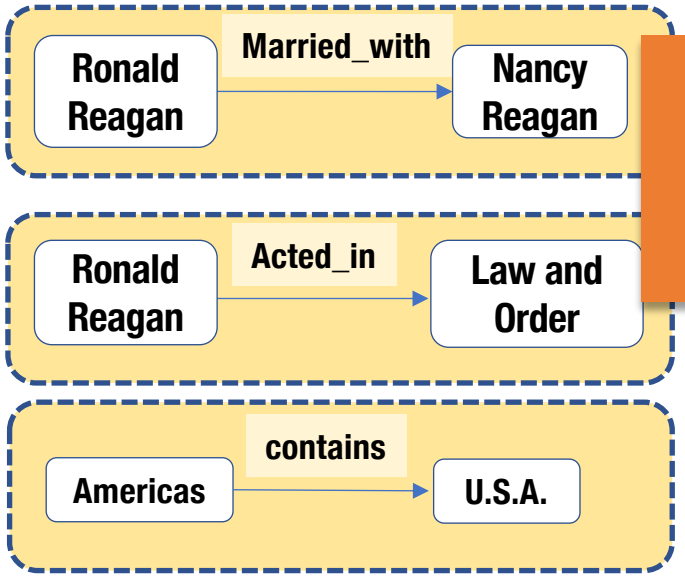
$$\hat{p}(l|M_{\bar{Q}})_{KL} \propto exp\left(\frac{1}{(1-\lambda)} \log\left(\hat{p}(l|M_{\bar{Q}})\right) - \frac{\lambda}{(1-\lambda)} \log\left(\hat{p}(l|\mathcal{K})\right)\right)$$

# Suggesting Expansions: **Ranking Labels (I)**

## The User Search




## How to score Expansions?




## Baseline Methods: Language Model

1. **Frequent** in the Graph
2. **Frequent** Around the Query

Probabilisic Model

$$\frac{|E_Q^l| + \epsilon \hat{p}(l|\mathcal{K})}{|E_Q| + \epsilon}$$

**LIMITED INFORMATION**

Looks only around the query + Global information

...but **Infrequent in the Graph**

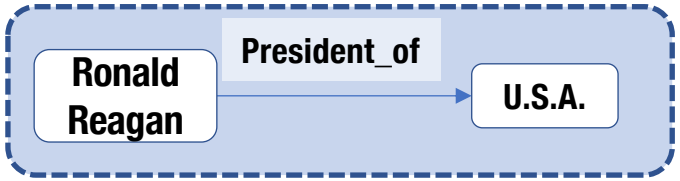
Kullback–Leibler divergence

$$\hat{p}(l|M_{\bar{Q}})_{KL} \propto$$

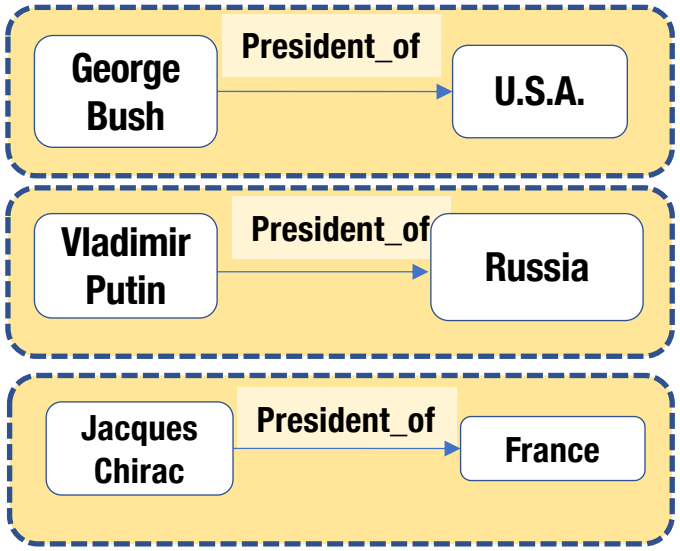
$$exp\left(\frac{1}{(1-\lambda)}\log\left(\hat{p}(l|M_{\bar{Q}})\right) - \frac{\lambda}{(1-\lambda)}\log\left(\hat{p}(l|\mathcal{K})\right)\right)$$

# Suggesting Expansions: **Ranking Labels (II)**

### The User Query Q




### The result-set R




Intuition:
**Exploit Classical IR Model**

## Pseudo Relevance Feedback

1. **Retrieve** "similar" elements **R**,
   resolve user query – e.g., Top-K ExQ Search

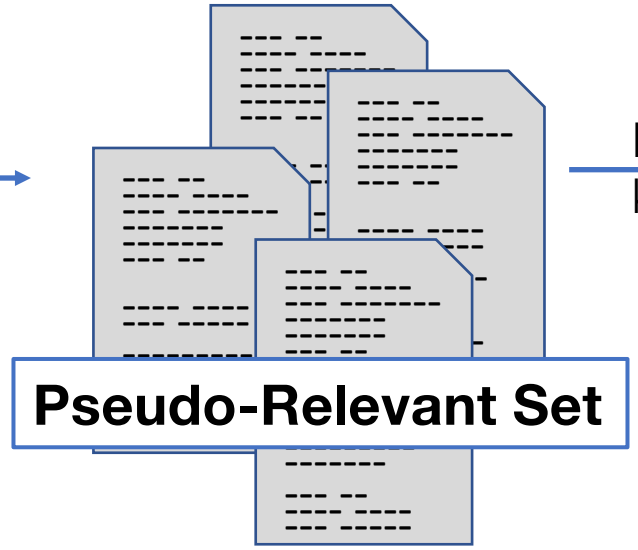
**Pseudo-Relevant Set**

2. **Compute** statistics about **R**
   & use to **score** expansions

# Pseudo Relevance Feedback for Document Search

Keyword Query
**"$w_1$ $w_2$ $w_3$"**

Search "relevant" documents




Extract "new" keywords

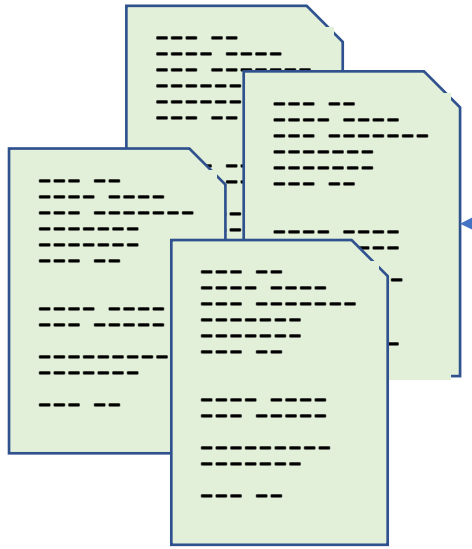
**$w_x$ $w_y$ $w_z$**

Expand query

**"$w_1$ $w_2$ $w_3$ $w_x$ $w_y$ $w_z$"**

**Query** **PRS**

$$\hat{p}(l|M_{\bar{Q}})_{REL} = (1-\lambda)\frac{|E^l_{\bar{Q}}|}{|E_{\bar{Q}}|} + \lambda\hat{p}(l|M_{rel})$$



Search "new" documents

# Pseudo Relevance Feedback Models

## Maximum Likelihood Estimation

$$\hat{p}(l|M_{rel})_{MLE} \approx \sum_{\bar{G} \in \bar{\mathcal{G}}_{rel}} \hat{p}(l|M_{\bar{G}}) \hat{p}(\bar{Q}|M_{\bar{G}})$$

$$\hat{p}(\bar{Q}|M_{\bar{G}}) \propto \prod_{l \in \bar{Q}} \hat{p}(l|M_{\bar{G}})$$

## PRS

$$\hat{p}(l|M_{\bar{Q}})_{REL} = (1-\lambda)\frac{|E^{l}_{\bar{Q}}|}{|E_{\bar{Q}}|} + \lambda \hat{p}(l|M_{rel})$$

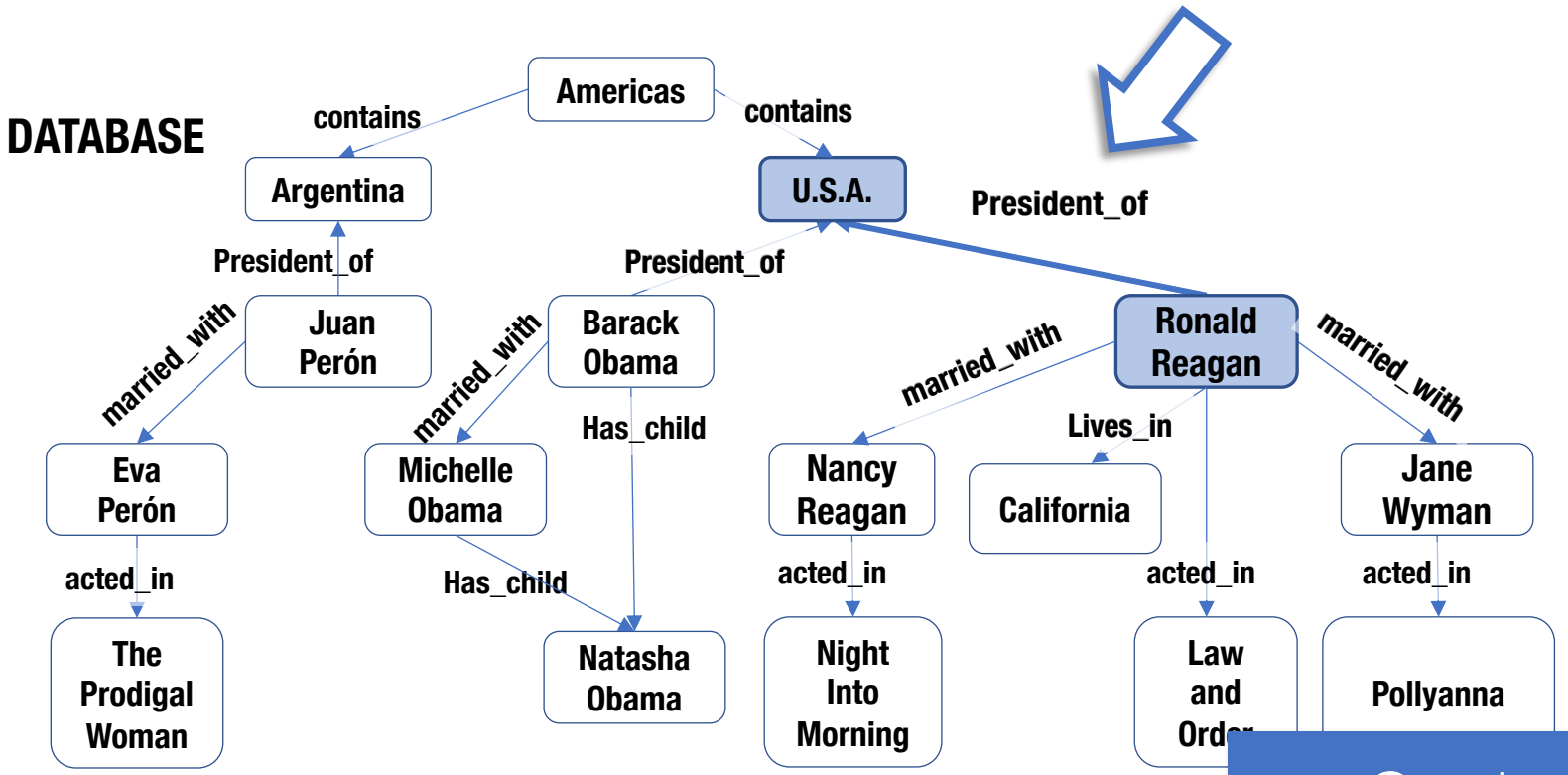
## KL-Divergence

$$\hat{p}(l|M_{rel})_{KL} \propto$$

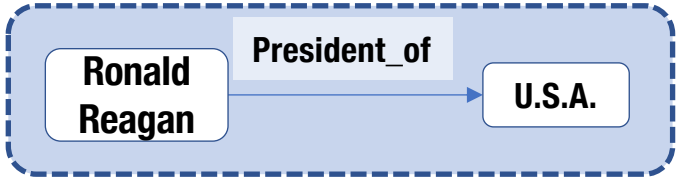
$$exp\left(\frac{1}{(1-\lambda)}\frac{1}{|\bar{\mathcal{G}}_{rel}|}\sum_{\bar{G}}^{\bar{\mathcal{G}}_{rel}} \log\left(\hat{p}(l|M_{\bar{G}})\right) - \frac{\lambda}{(1-\lambda)}\log\left(\hat{p}(l|\mathcal{K})\right)\right)$$

**2 Models of Estimation**
**MLE & KL-Divergence**

# Bag Model for Graphs




## The User Search




How can we convert to the document model?

## The Bag-of-Labels Model

{ President_of, contains, married_with, married_with, acted_in, lives_in }

- Graphs can be modeled as **Bag of Words**
- Describes **MORE than what is in the query**

# Graph Query Suggestion Framework

## Steps:

- Decide the **scoring model** **(**Query & PR-Set**)**
- **Transform** query into Bag of Labels **(**obtain candidate expansions**)**
- Compute **Scoring Model**
- **Score expansions**
- Return top-k expansions

Expansions can be **EDGES** or **LABELS**

**Algorithm 1** Graph-Query Suggestion

**Input:** Knowledge graph $\mathcal{K} : \langle V, E, \ell \rangle$
**Input:** Current query $Q : \langle V_Q, E_Q, \ell \rangle$
**Input:** Current answers $\mathcal{A}_Q$
**Input:** Model $M$ ▷ One defined by Eq 1, 2, 4, 5, or 7
**Input:** Number of expansions $k$
**Output:** Expansions $\langle l_1, l_2, ..., l_k \rangle$

1: $E_\varepsilon \leftarrow \emptyset$
2: **for each** $v_i \in V_Q$ **do**
3: $\quad E_\varepsilon \leftarrow E_\varepsilon \cup$ GETEDGES$(v, E)$
4: $E_\varepsilon \leftarrow E_\varepsilon \setminus E_Q$
5: $L_\varepsilon \leftarrow \{\ell(e) | e \in E_\varepsilon\}$
6: ESTIMATEMODEL$(M, E_\varepsilon, L_\varepsilon, \mathcal{A}_Q)$
7: Scores ←new $Dict()$
8: **for each** $l \in L_\varepsilon$ **do**
9: $\quad$ Scores $\leftarrow \{l : \rho_M(Q, l)\}$
10: Scores $\leftarrow sort(Scores)$
11: **return** $Scores.get(k)$

# Experimental Evaluation (I)

## 5 Ranking Scores:

a) **Based on the Query Alone**
- **Maximum Likelihood Estimation (MLE)**
- **Kullback–Leibler divergence (KL)**

b) **Based on the PR-Set**
- **Maximum Likelihood Estimation (MLE-rel)**
- **Kullback–Leibler divergence (KL-rel)**
- **"Surprise" heuristics (Srp)** [ Sarkas et al.'09 ]

## Baselines:
- **Random**
- **Personalized Page Rank**
- **Distant supervision** [ N.Voskarides, et al.'18 ]

## Real Dataset: Freebase +300M Edges

**Tests (compare NDCG):**
- **65 Queries from QUALD-7 benchmark**
  **Contain 1 entity, 1 edge, 2+ edges**
- **KG-contextualization Dataset** [ N.Voskarides, et al.'18 ]

*Example query: "doctoral supervisor, Albert Einstein"*
- *Mapped to:*
  - *Single Entity: Albert Einstein*
  - *Edge: <Albert Einstein, advisor, A, Kleiner >*
- *Described with general topic*
  - *"Academic Information of Albert Einstein"*

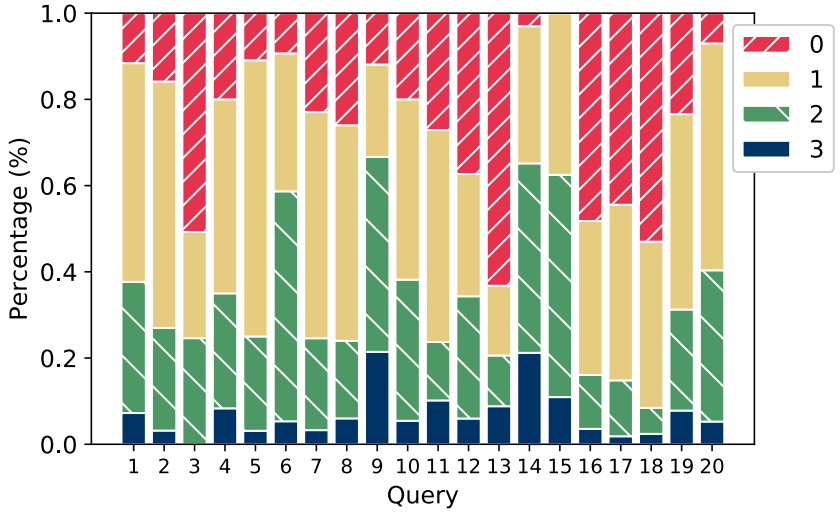
# Experimental Evaluation (II)

**65 Queries from QUALD-7 benchmark** - Contain 1 entity, 1 edge, 2+ edges

For Each Query (graph+description)
& Each Method → Produce 20 Query expansions

Ask 3 Human judges* to evaluate relatedness of each suggestion

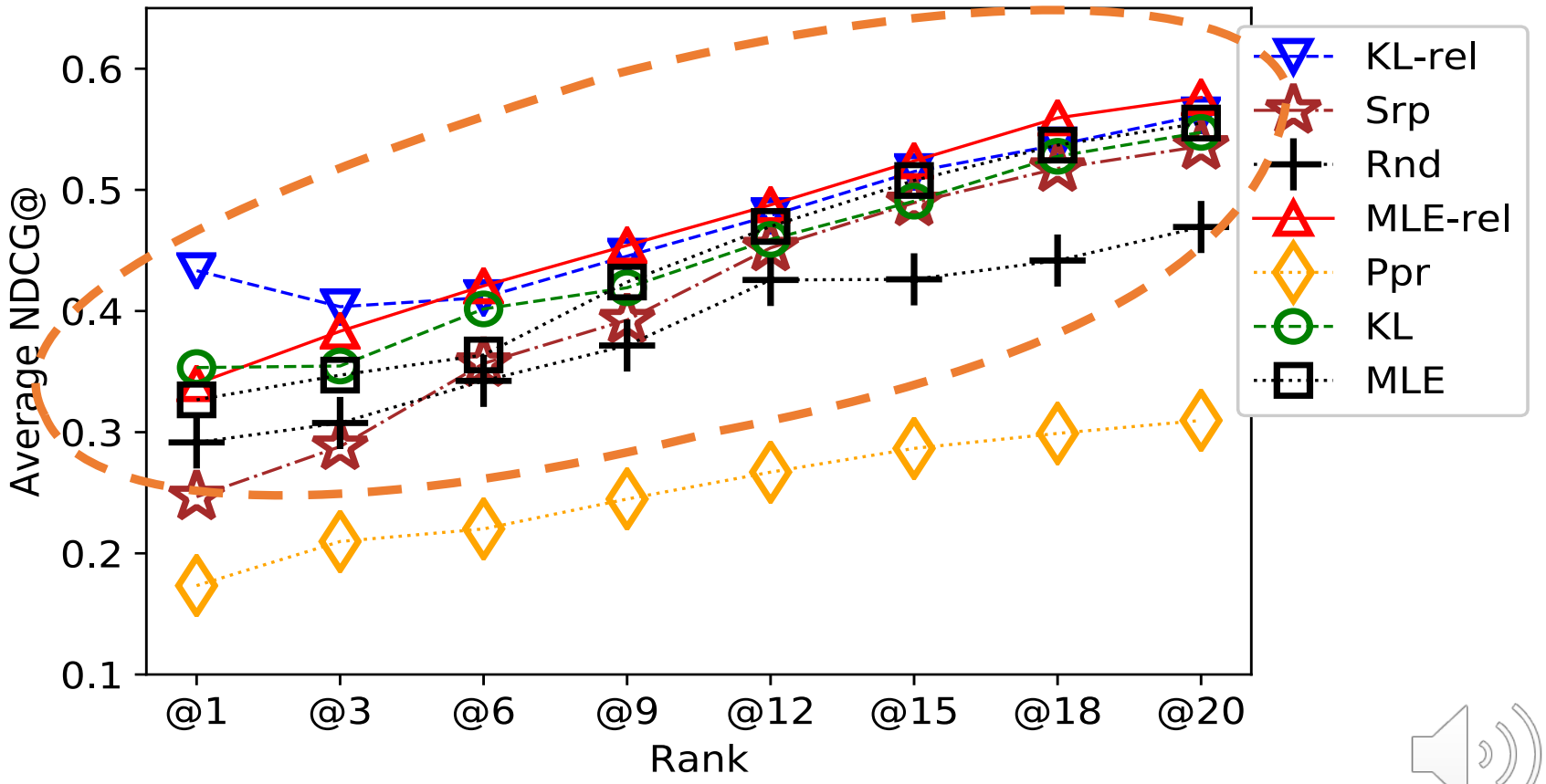
**irrelevant (0), uninteresting (1), fairly interesting (2), really interesting (3)**




Among all rankings
**Few Expansion per Query**
**were considered fairly interesting**

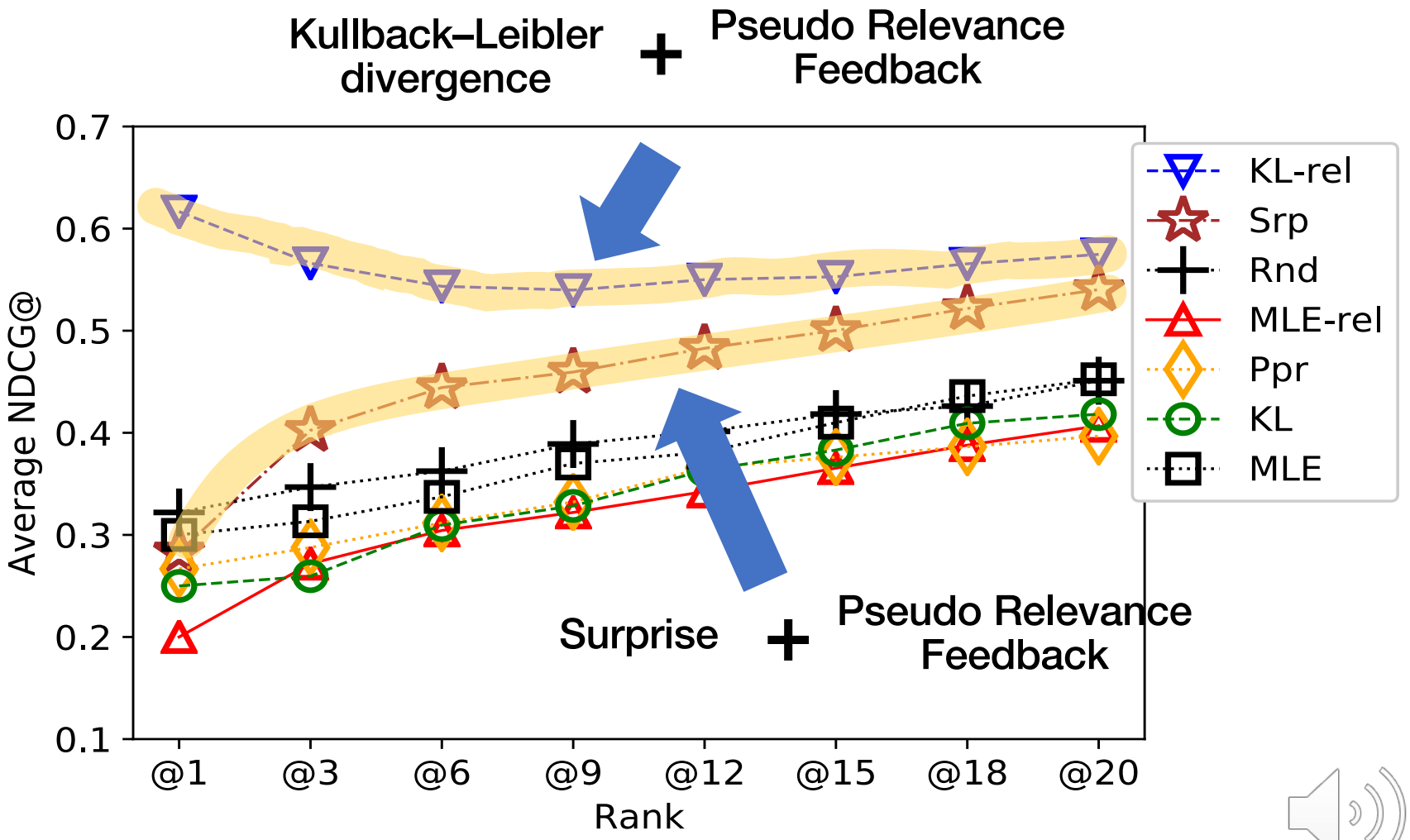
# Evaluation Results: Suggestions with 1 Entity

**With 1 Entity**
**we can only guess**



# Evaluation Results: Suggestions with 1 Edge

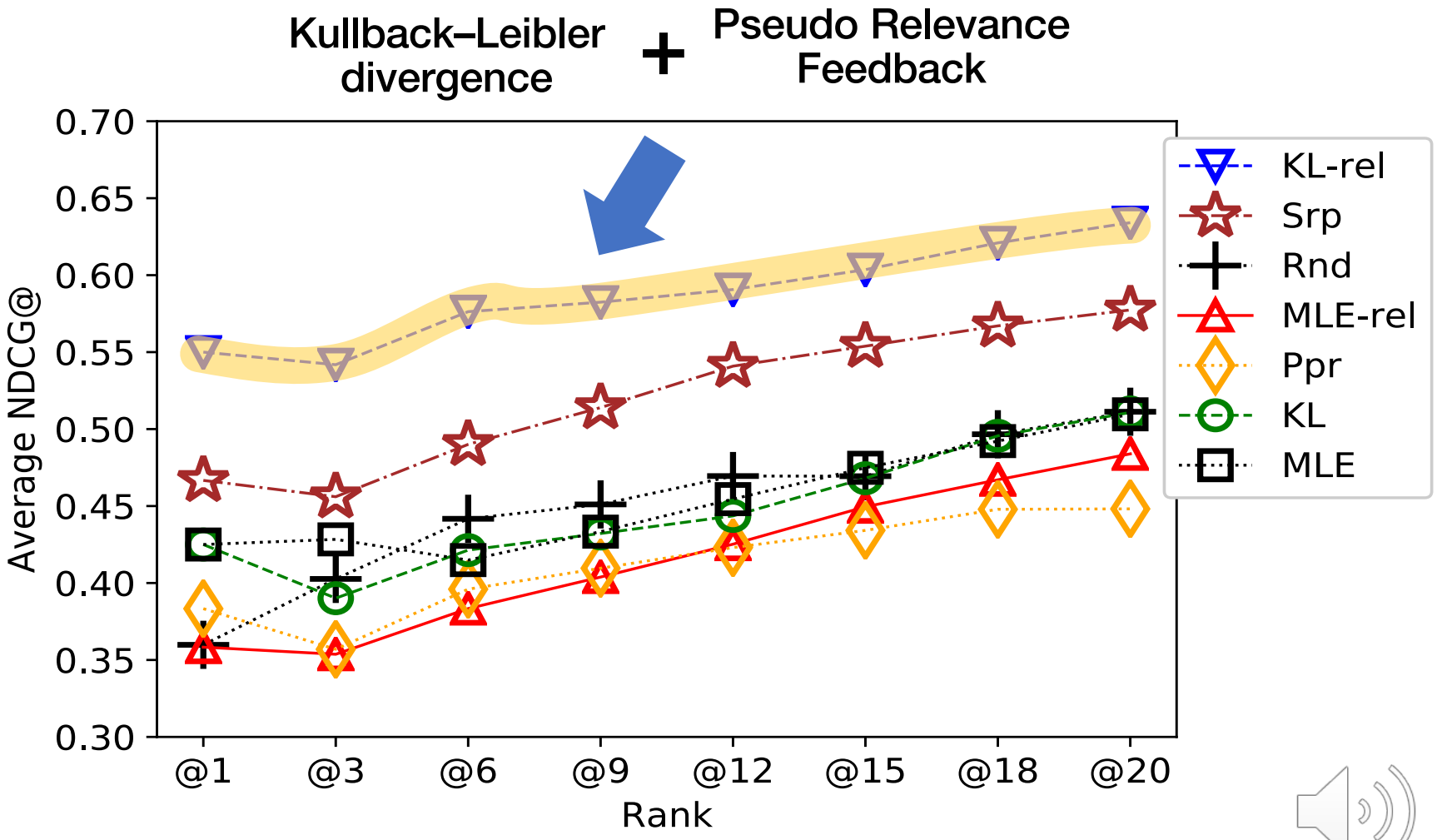
**With 1 Edge**
**Precision**
**Drastically increases**



# Evaluation Results: Suggestions with 2+ Edges

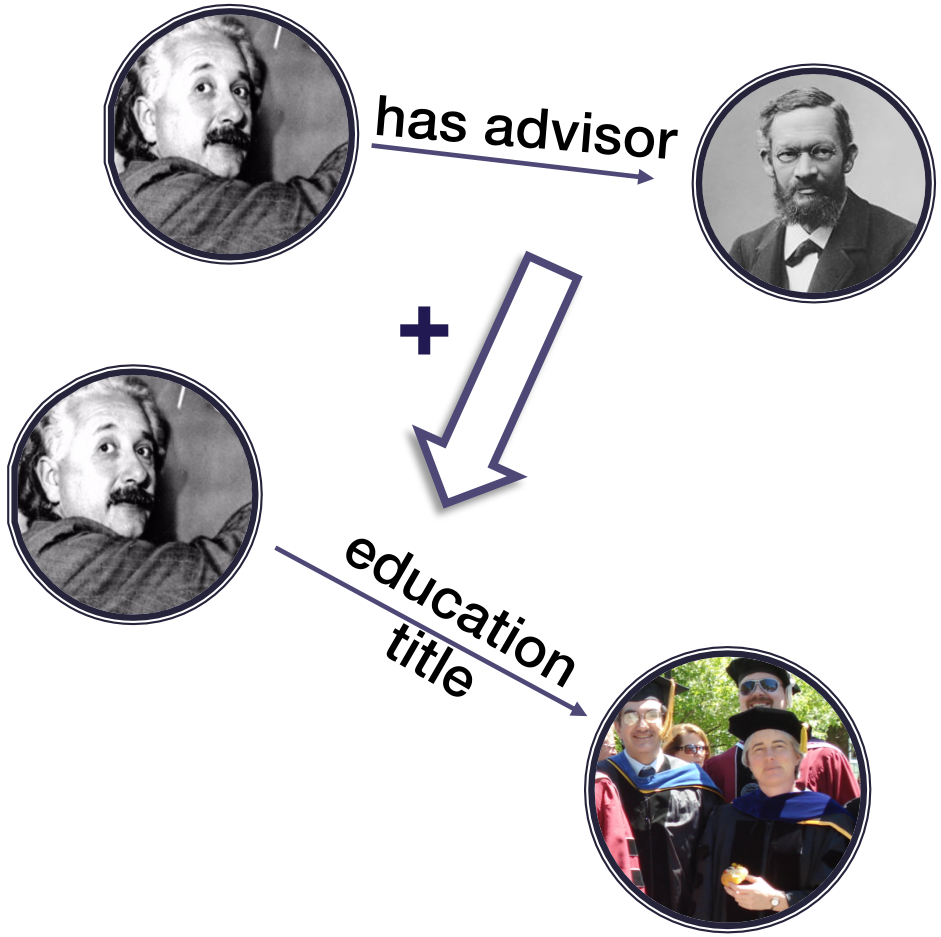
**With 2+ Edges KL-rel does not suffer any loss in precision**

**With more edges in the query the number of alternatives increases**



# KG Query Suggestion

*Given a Graph Query suggest query expansions*




# Summary

**Guide the User in Exploring a KG with complex and rich schema.**

**The Bag-of-Labels model provide and expressive model for Graph Queries on KGs**

1. **Exploit State-of-the-Art IR query suggestion techniques**
2. **Bridge Structural-queries and Semantic of relationship**
3. Provide effective Ranking with just 1 edge
4. **Does not require pre-training of ML models**
5. **Can be expanded with more complex ranking models**

http://j.mp/WebConfKG