RelCrowd: Relational Crowdsourcing for Estimating Knowledge Graph Accuracy under Budget

Prakhar Ojha
prakhar.ojha@csa.iisc.ernet.in

Partha Talukdar
ppt@cds.iisc.ac.in

Department of Computer Science and Automation, Indian Institute of Science, Bangalore India

MOTIVATION FOR RELATIONAL CROWDSOURCING

- Knowledge Graphs (KGs) are multi-relational graphs connecting entities via relations.
- Automatic construction of KGs by reading the web and extracting facts, leads to inaccurate graphs with incorrect facts.

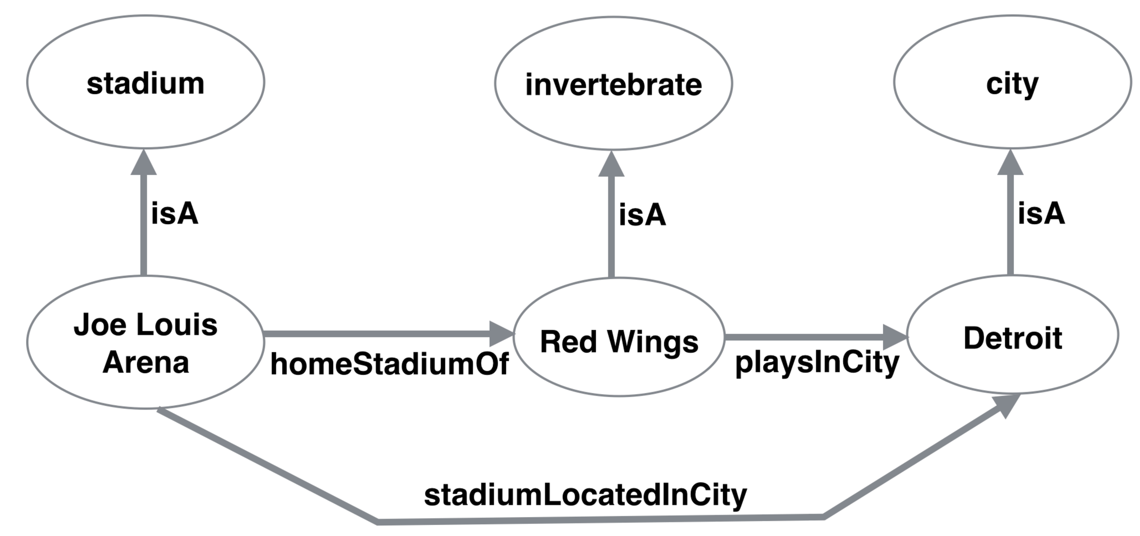
Importance of Accuracy

- Know strengths / weakness for targeted querying.
- Feedback helps in better construction.

How to estimate accuracy? Software Evaluation or Human Judgment

- Budget constraints of real money and human time.
- Generic crowdsourcing ignores structural information.

Idea: Exploit structural information to post fewer tasks and infer the rest.



5 correct
1 incorrect

Evaluate 2 facts

Coupling Constraints

- Type consistency: homeStadiumOf(X,Y) -> stadium(X) ^ sportsTeam(Y)
- Horn-Clause: homeStadiumOf(X,Y) ^ playsInCity(Y,Z) -> stadiumLocatedIn(X,Z)

PROBLEM FORMALIZATION AND APPROACH

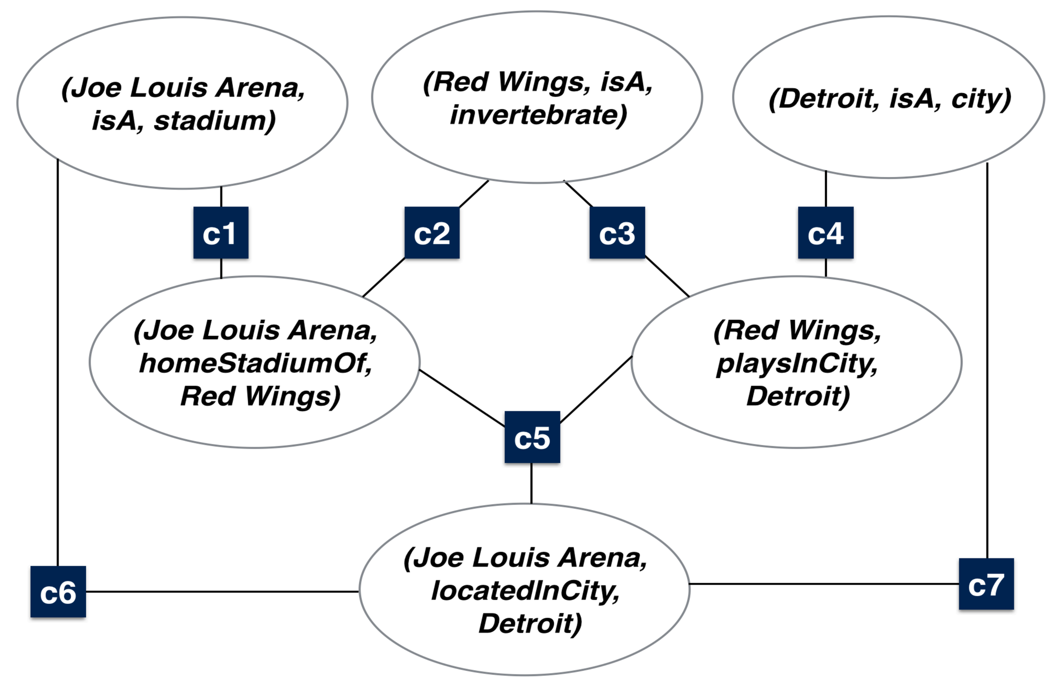
KG with *n* beliefs $\mathcal{H} = \{h_1, \ldots, h_n\}$ and budget $\mathbb{B}$ of to evaluate it.
Coupling constraints $\mathcal{C} = \{(\mathcal{C}_i, \theta_i)\}$ are relationship among tasks.
Inference algorithm uses constraints over already evaluated tasks $\mathcal{Q} \subseteq \mathcal{H}$ and deduces labels for *inferable* set $\mathcal{I}(G, \mathcal{Q}) \subseteq \mathcal{H}$.
$\Phi(\mathcal{Q})$ calculates the average accuracy of evaluated tasks.
Maximize the size of inferable set, taking care of the budget.

$$\arg\max_{\mathcal{Q} \subseteq \mathcal{H}} \ |\mathcal{I}(G, \mathcal{Q})|, \quad \text{s.t.} \sum_{h \in \mathcal{Q}} c(h) \le \mathbb{B} \qquad (1)$$

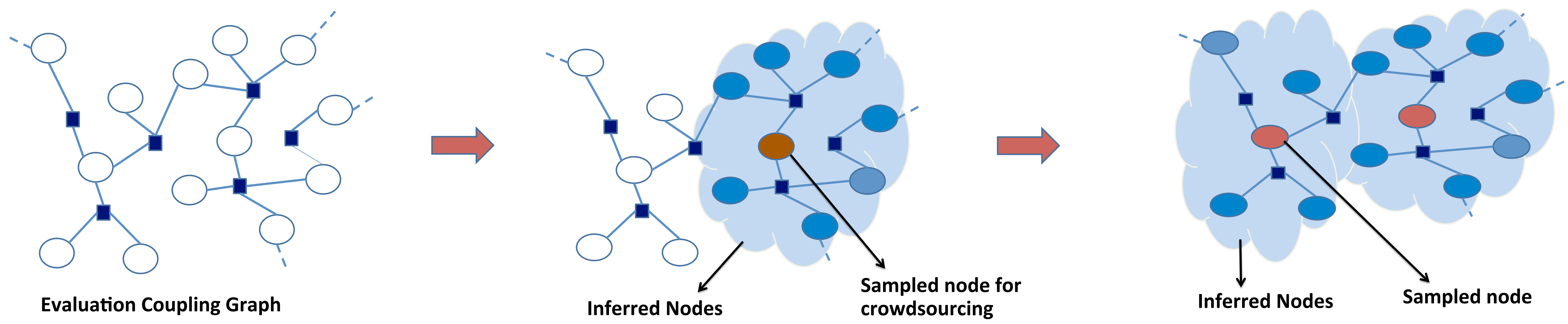
RelCrowd optimization is submodular and subset selection is NP-Hard.

Evaluation Coupling Graph (ECG)

- Bipartite factor graph
- Nodes for each constraint and task
- Edges between constraints and participating tasks



OVERVIEW OF CONTROL AND INFERENCE MECHANISM



METHOD

Inference Mechanism

Probabilistic Soft Logic (PSL) : Distribution over labels given by

$$\mathbb{P}\Big(l(\mathcal{I}(G, \mathcal{Q}))\Big) = \frac{1}{Z} \exp\Big[- \sum_{j=1}^{|\mathcal{C}|} \theta_j \psi_j (\mathcal{I}(G, \mathcal{Q})) \Big]$$

Labels which satisfy more constraints are more probable.

Control Mechanism

Algorithm 1 KGEval: KG-Evaluation

```
G = BuildECG(H, C)
B_r = B
Q_0 = S
while (B_r > 0) do
    h* = arg max_{h∈H} |I(G, Q_{t-1} ∪ {h})|
    CrowdEvaluate(h*)
    Q_t = I(G, Q_{t-1} ∪ {h*})
    B_r = B_r − c(h*)
    Q = Q ∪ Q_t
    if Q ≡ H then
        Exit
    end if
end while
return (1/|Q|) Σ_{h∈Q} l(h)
```

$\mathcal{H}$: set of tasks
$\mathcal{C}$: coupling constraints
G: ECG
$\mathbb{B}$: assigned budget
B_r: residual budget
$\mathcal{Q}$: evaluated tasks
$\mathcal{S}$: seed set
$c(h)$: cost function
Φ: score aggregator

KEY RESULTS

Datasets: NELLsports and Yago2Sample KGs
Crowdsource from Amazon Mechanical Turk

Baselines: Random, Max-Degree, Independent Cascade, and RelCrowd

NELL sports dataset $(\mathcal{H}_N)$

Method	$1-\Delta Acc_{Micro}$	$1-\Delta Acc_{Macro}$	# Queries
Random	0.987	0.9516	623
Max-Degree	0.971	0.9239	1370
Ind-Cascade	0.992	0.9026	232
RelCrowd	**0.995**	**0.9641**	**140**

- RelCrowd estimates are closest to gold accuracy and utilize minimum budget.
- Rate of coverage over Knowledge graph is fastest.

Effectiveness of Coupling Constraints:

More relational Couplings Constraints ➡ Better performance.

Constraint Set	Iterations	$1-\Delta Acc_{Micro}$
$\mathcal{C}$	**87**	**0.993**
$\mathcal{C} - \mathcal{C}_{b3}$	209	0.991
$\mathcal{C} - \mathcal{C}_{b3} - \mathcal{C}_{b2}$	285	0.989

Future directions: Minimize regret incurred in terms of budget spent.
Model aggregation methods for noisier crowd responses.