Computational Fact-Checking through Relational Similarity based Path Mining

Himanshu Ahuja and Alexander Michels

What is a Knowledge Graph?

A knowledge graph *G* is an ordered pair *G = (E, R)* where *E* is a set of entity or concept nodes and *R* is a set of relation or predicate edges.

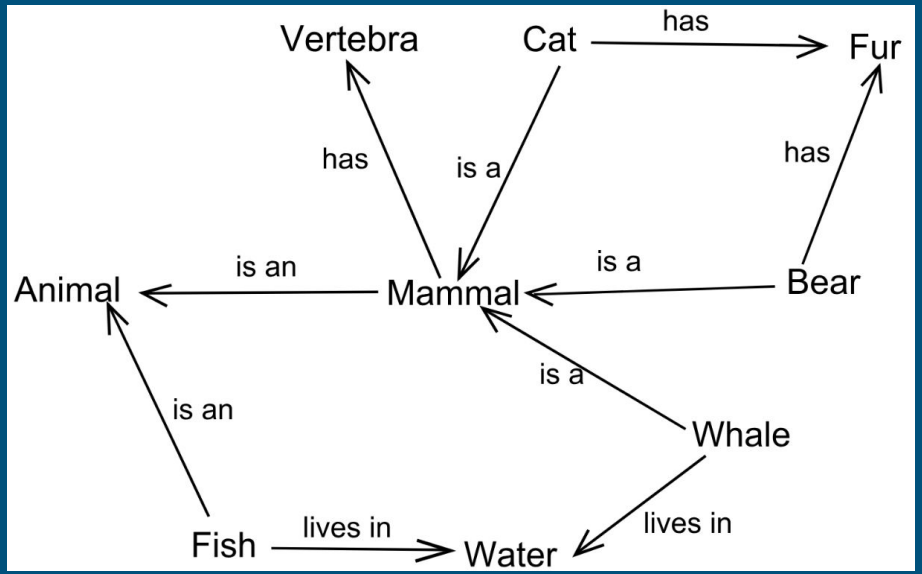


Image: Wikipedia

What is fact-checking?

Fact-checking is the process of:

- putting a claim into context
- gathering relevant information
- conducting thorough analysis
- reporting a conclusion with explanations and evidence.

Image: Politifact

PredPath

Views the problem as a ***link prediction*** one

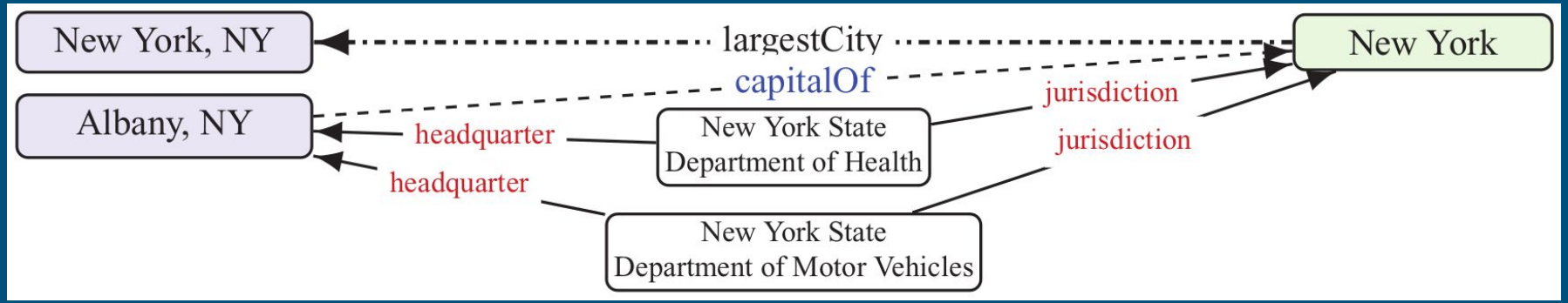
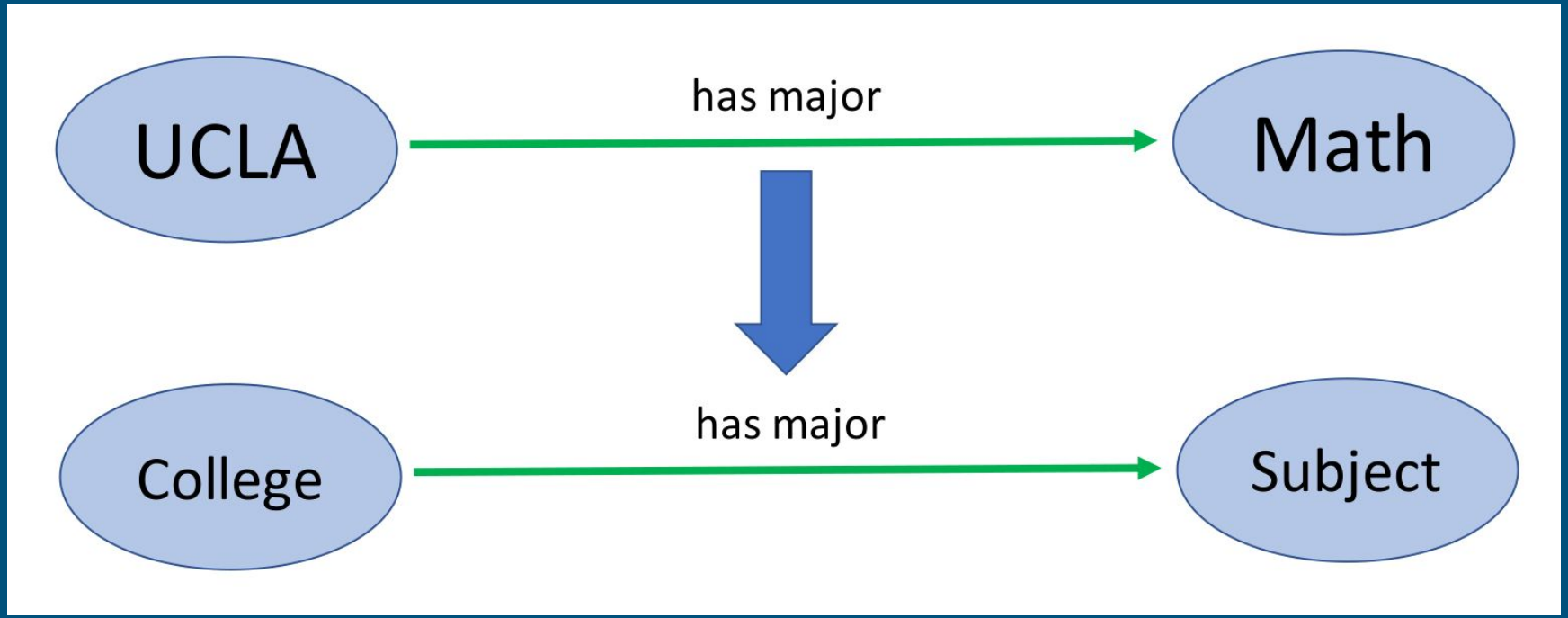
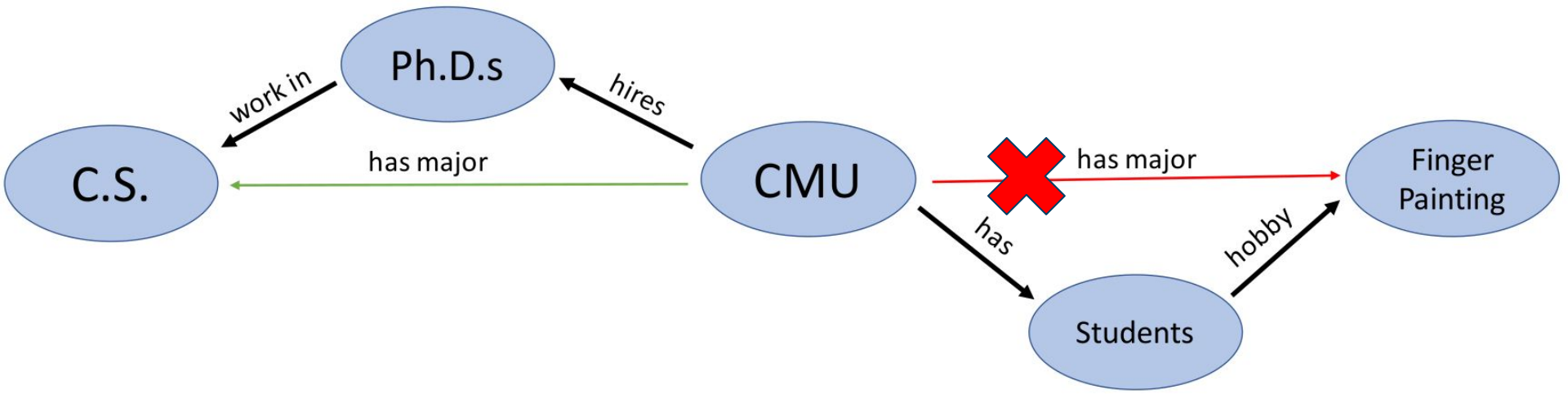


Image: B. Shi and T. Weninger, Fact checking in large knowledge graphs - A discriminative predicate path mining approach, CoRR, abs/1510.05911 (2015).

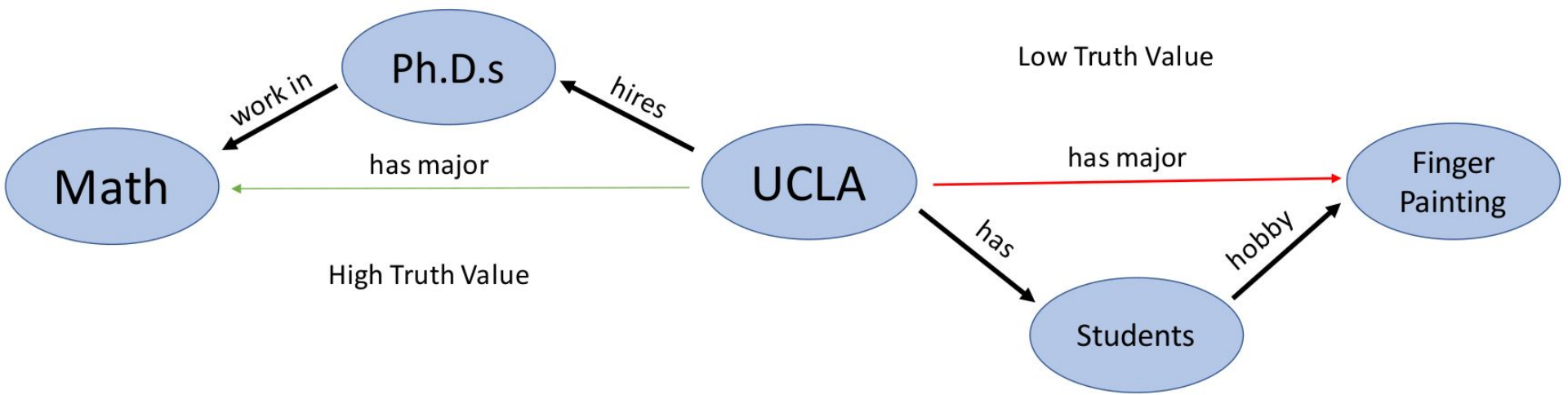
Does UCLA have a Math Major?



What does it mean to “have a major?”

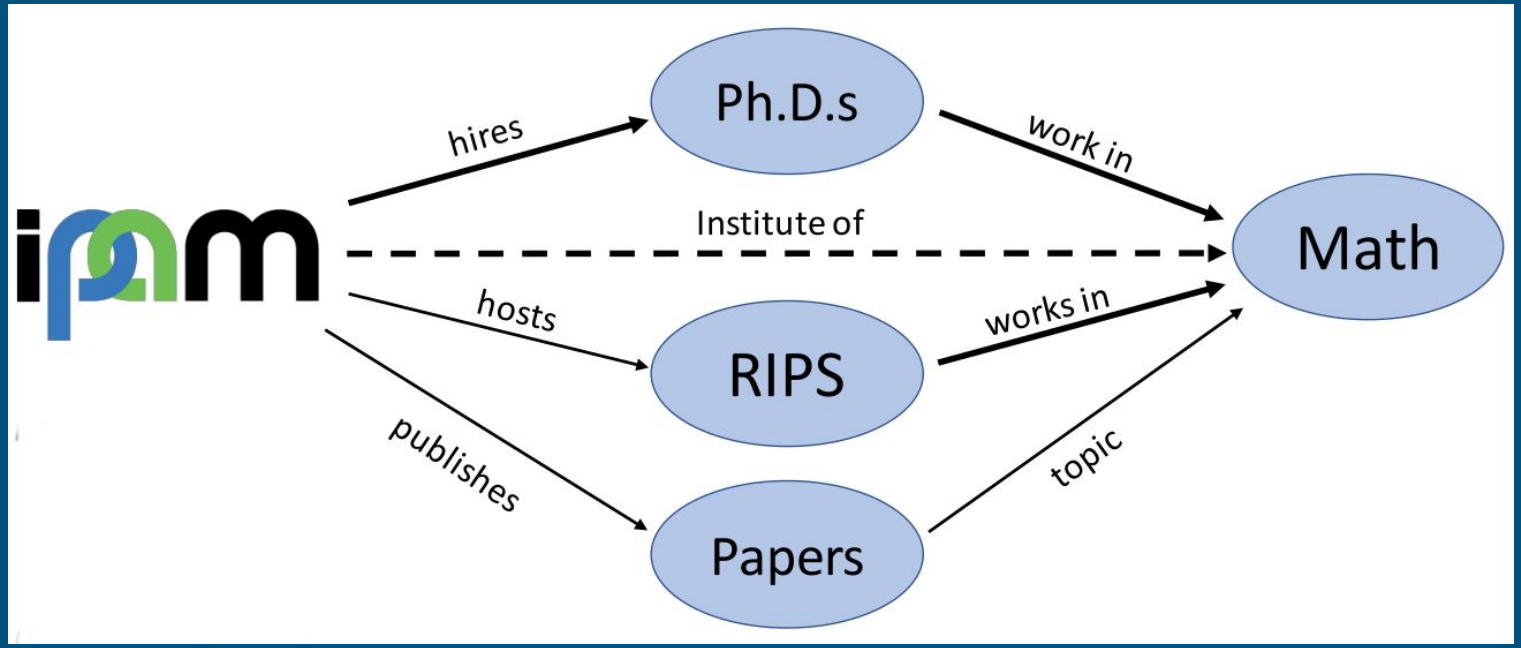


Yes, UCLA has a math major



Knowledge Stream

Views the problem as a ***network flow*** problem.



Path Specificity

Path Specificity = Node Specificity + Edge Similarity

How general the idea of the node is
(measured by log of degree)
Very General: University
Very Specific: This Room, Baltimore

How similar two relations are
e.g.: Mentors
Highly Similar: advises, counsels
Less similar: robs, steals

Path Specificity

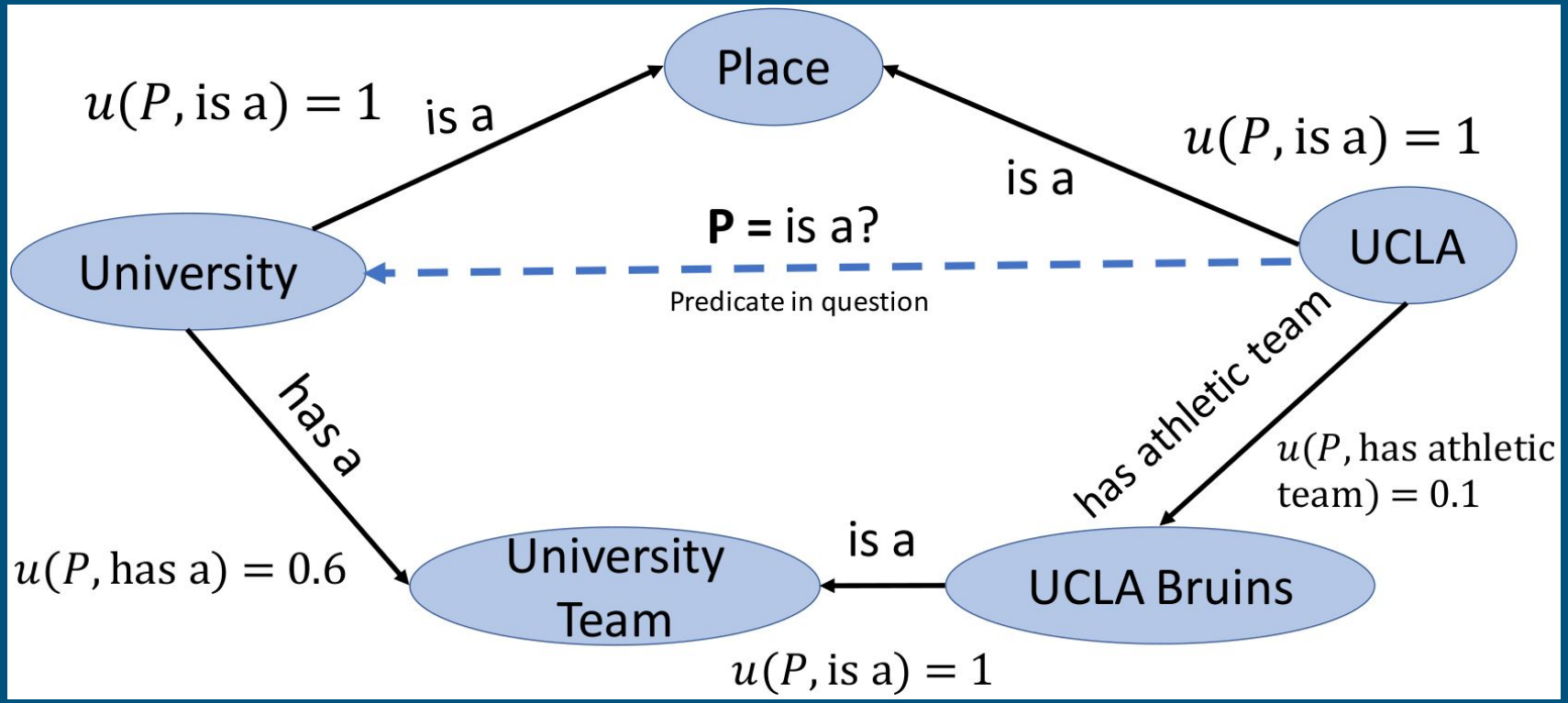
Path Specificity = Node Specificity + Edge Similarity

$$S'(P_{s,p,o}) = \left[\sum_{i=2}^{n-1} \frac{\log k(v_i)}{u(r_{i-1},p)} + \frac{1}{u(r_{n-1},p)} \right]^{-1}$$

Can ***path specificity*** be used to provide a better definition of path length for the *PredPath* algorithm?

Yes, we named it *RelPredPath* (Relational PredPath)

Is UCLA a University?



RelPredPath: Challenges

How do we collect our paths?

- *PredPath* used all paths under length *k*, but with our new definition of path length what do we set *k* to?
- Is the optimal *k* a function of network characteristics (connectivity, sparsity, betweenness centrality, etc.)?

Performance

Dataset	*RelPredPath*	*PredPath*	*KS*
Presidents/First Ladies	1.0000	1.0000	0.9895
Movies/Directors	0.9741	0.9997	0.8500
Nationality	0.8400	0.9520	0.9792
Profession	0.9455	0.9271	0.9866
NBA Player/Team	0.9634	0.9331	0.9996
Company/President	0.7936	0.8867	0.8119
State/Capital	1.0000	0.9968	1.0000
Vice Presidents	0.8537	0.9440	0.7780

Going Further

- Is there a better way to collect the paths?
- How can the path length definition be improved?

Acknowledgements

Susana Serna, Director of RIPS

Stephen Desalvo, Praedicat

Urjit Patel, Praedicat

Questions?