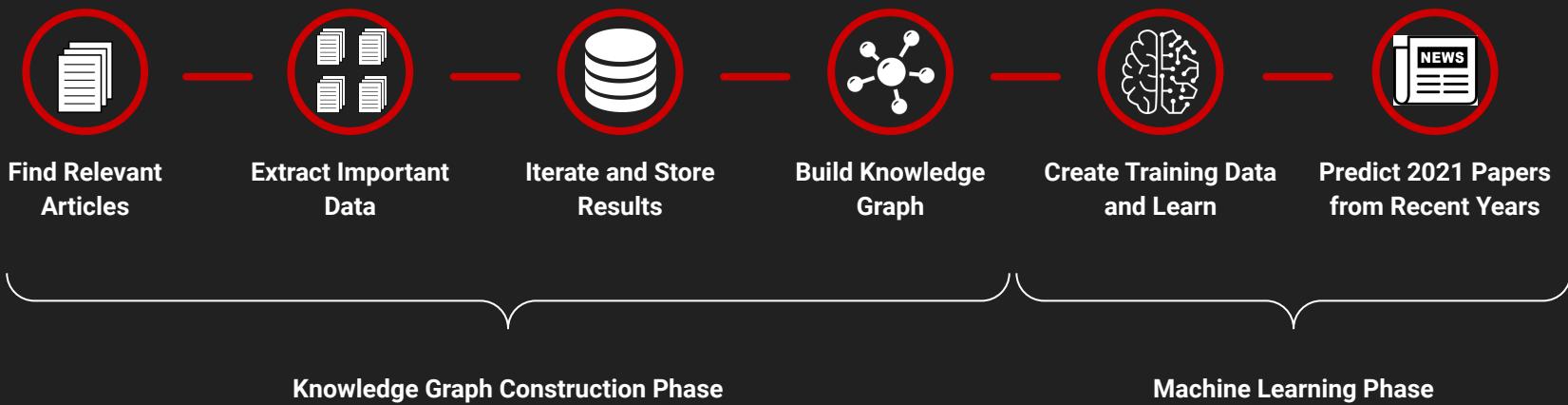


Predicting Future Scientific Research with Knowledge Graphs

Matheus Schmitz
Rehan Ahmed

The Big Picture



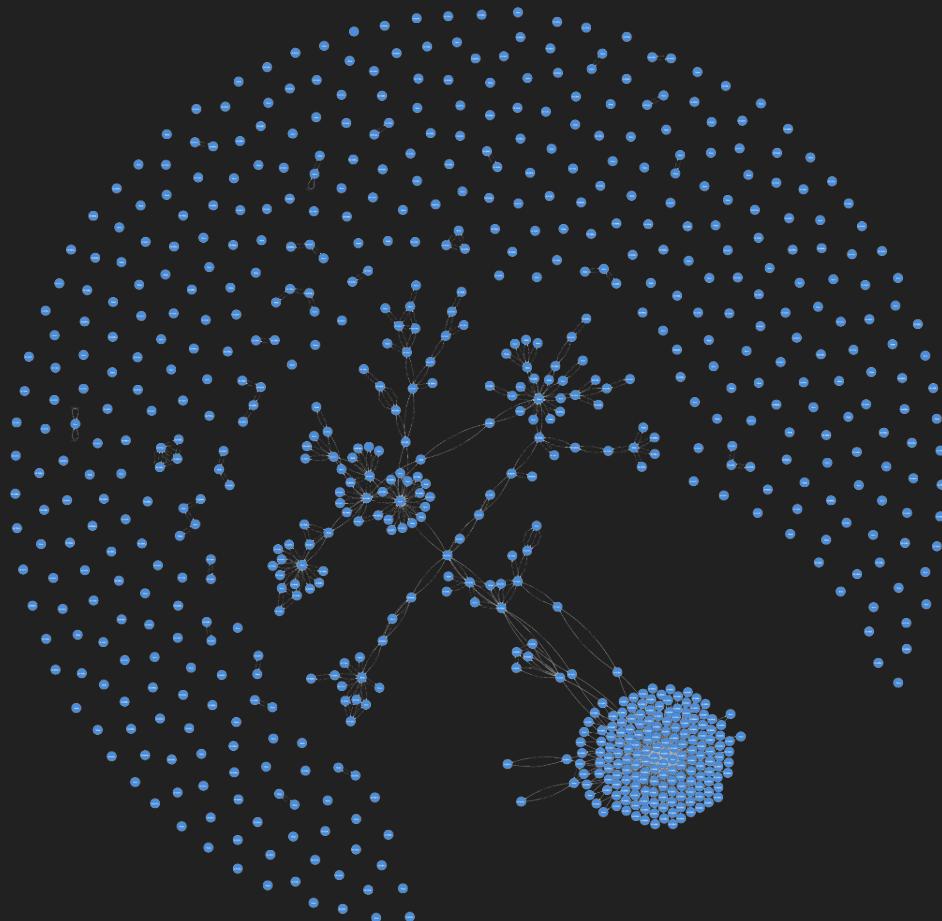
Papers by one author:

Challenges

(1) Building a connected graph:

Too much data out there, over
50k CS papers published per
year on arXiv alone!

(2) Handling papers referenced
in the KG but not crawled



Challenges

(3) Making plausible inferences about the future!

(And not generating predictions that are merely a
reshuffling of their sources' abstracts)



(4) Making the graph useful

(You don't need a KG to look up
papers by author or title)

Data Sources



dblp
computer science bibliography

Extract citations and references



Popular articles



Abstracts for the popular articles



“Seed trail” crawling

Web Scraping



Entity Resolution

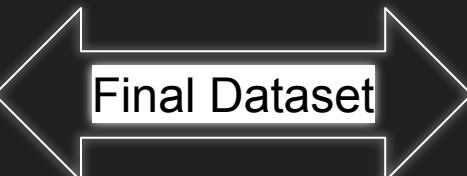


43 Million Pairs
Reduction Ratio: 0.93
3M pairs compared

Result: 20534 records

(1) Link Using

Title
Author
Date



(3) Link Using

Title
Author
DOI



(2) Link Using
DOI



External KG Linkage

Our Knowledge Graph

Author

Info

uri

ns0__url

ns0__birthDate

Geoffrey Hinton

1008713

http://dbpedia.org/resource/Geoffrey_Hinton

1947-12-06

Grammar as a Foreign Language

Regularizing Neural Networks by Penalizing Confident Output Distributions

Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer

Results: 3

2

```
graph LR; GH[Geoffrey Hinton] --> BH[1947-12-06]; BH --> DBpediaPage[DBpedia Entity Page]
```

dbo:academicDiscipline

- dbr:Machine_learning
- dbr:Neural_network
- dbr:Cognitive_science
- dbr:Artificial_intelligence
- dbr:Object_recognition

dbo:almaMater

- dbr:University_of_Cambridge
- dbr:University_of_Edinburgh

dbo:award

- dbr:IEEE_Frank_Rosenblatt_Award
- dbr:Rumelhart_Prize
- dbr:BBVA_Foundation_Frontiers_of_Knowledge_Award
- dbr:IEEE/RSE_James_Clerk_Maxwell_Medal
- dbr:AAAI_Fellow
- dbr:Turing_Award
- dbr:IJCAI_Award_for_Research_Excellence

dbo:birthDate

- 1947-12-06 (xsd:date)

dbo:birthName

- Geoffrey Everest Hinton (en)

dbo:birthPlace

- dbr:Wimbledon,_London

Triple Generation

453k Paper URIs (incl.
references)

51k Author URIs

298 DBpedia Matches

383 Predicted Papers

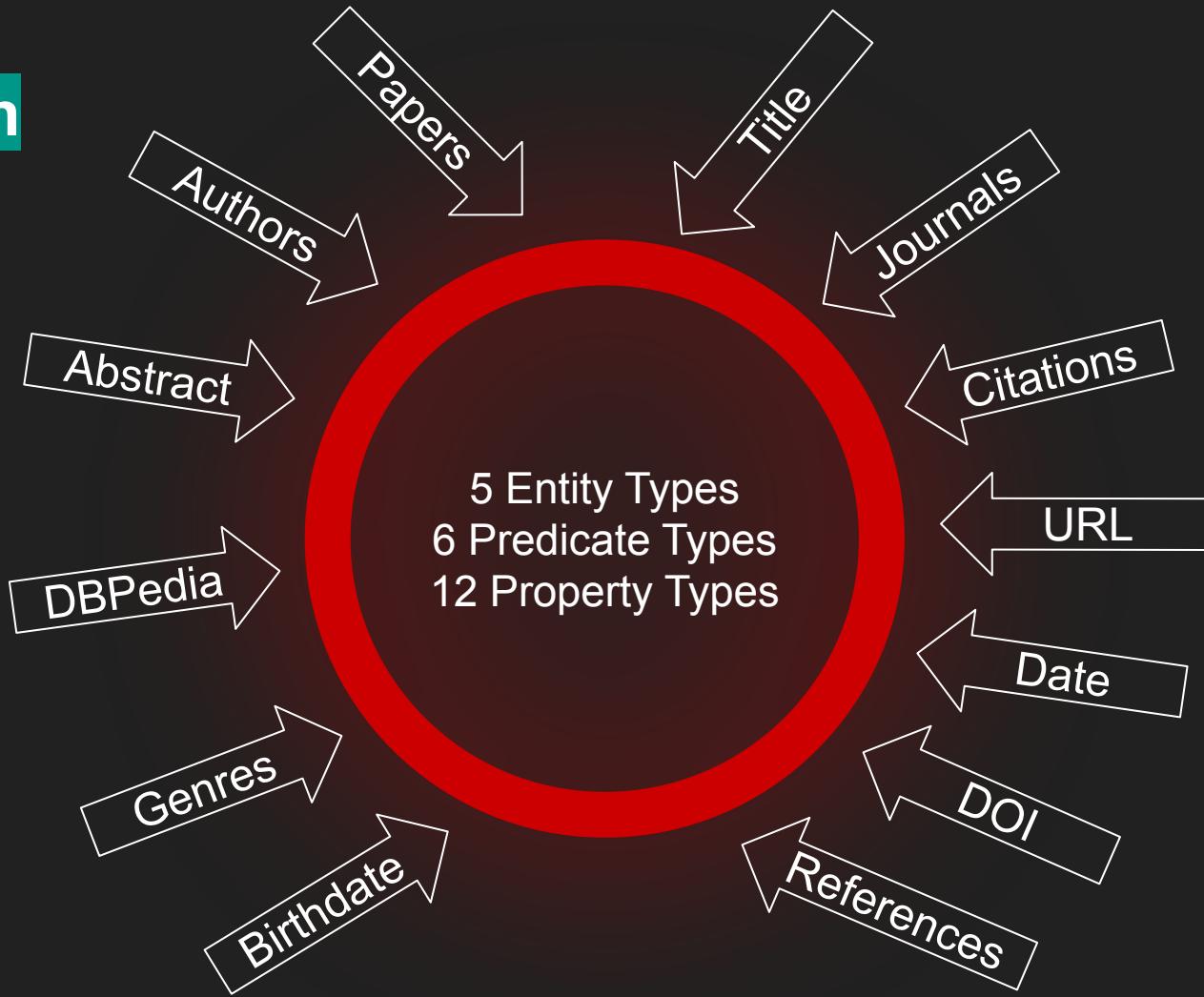
207 Genres

86 Publishers

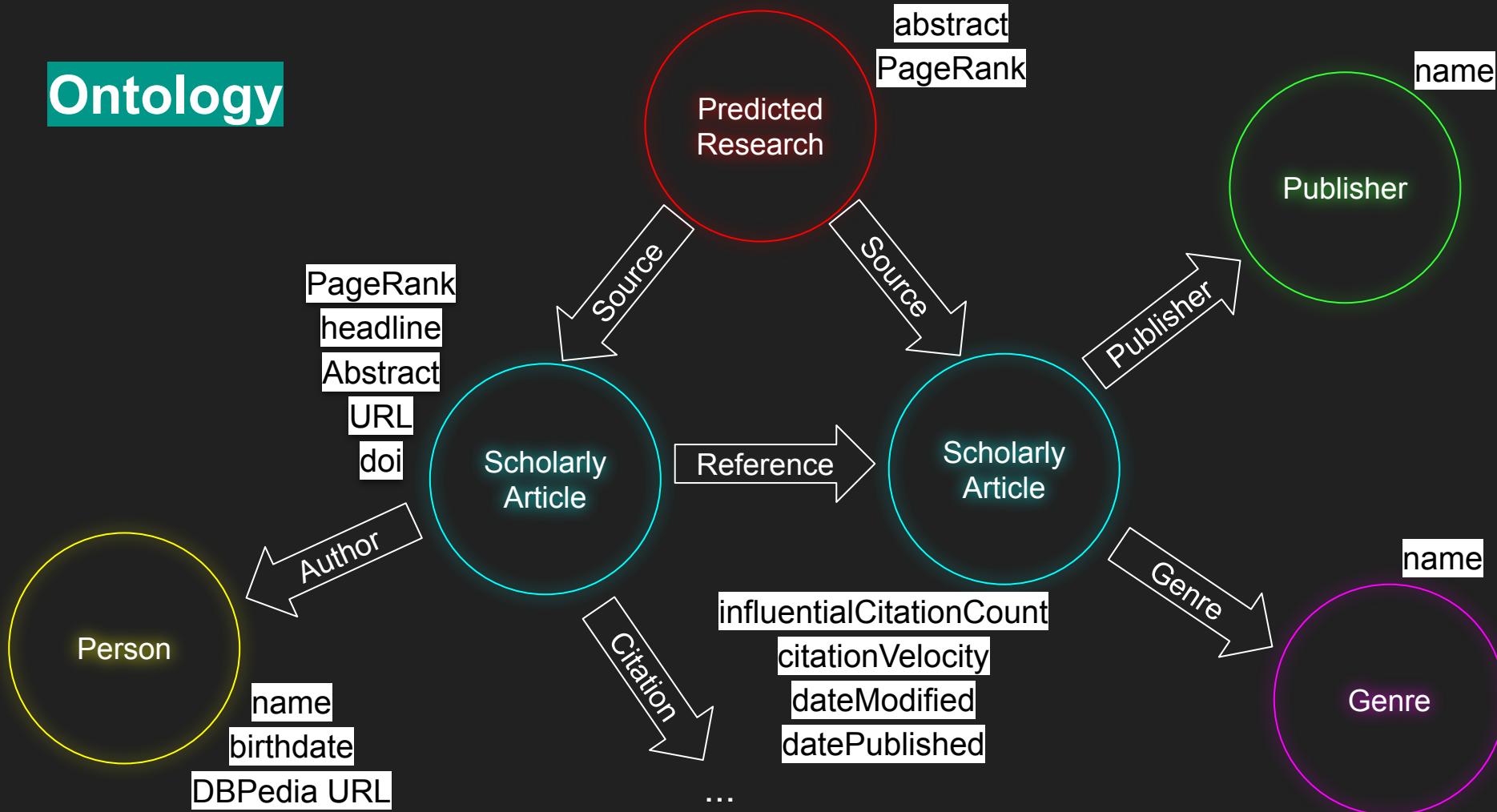
1.5M Relationships

505k Nodes

2.65M Triples!



Ontology



Graph Database



neo4j browser interface showing a query results graph and the underlying Cypher code.

Query:

```
1 MATCH (n) WHERE n.ns0__headline contains "Knowledge Graph" RETURN n
```

The results show a network graph with several clusters of nodes. A large cluster on the right is highlighted in pink, and a smaller one on the left is highlighted in orange. Nodes are represented by circles of varying sizes, and connections between them are shown as lines.

Query:

```
1 MATCH (n) WHERE n.ns0__headline contains "Knowledge Graph" RETURN n
```

```
1 <Record n=<Node id=2891 labels=frozenset({'Resource', 'ns0_ScholarlyArticle'}) properties={'ns0_headline': 'Knowledge Graph', 'ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_title': 'Knowledge Graph', 'ns0_doi': null, 'ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_headline': 'Knowledge Graph', 'ns0_ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_ns0_title': 'Knowledge Graph', 'ns0_ns0_doi': null, 'ns0_ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.'}>
```

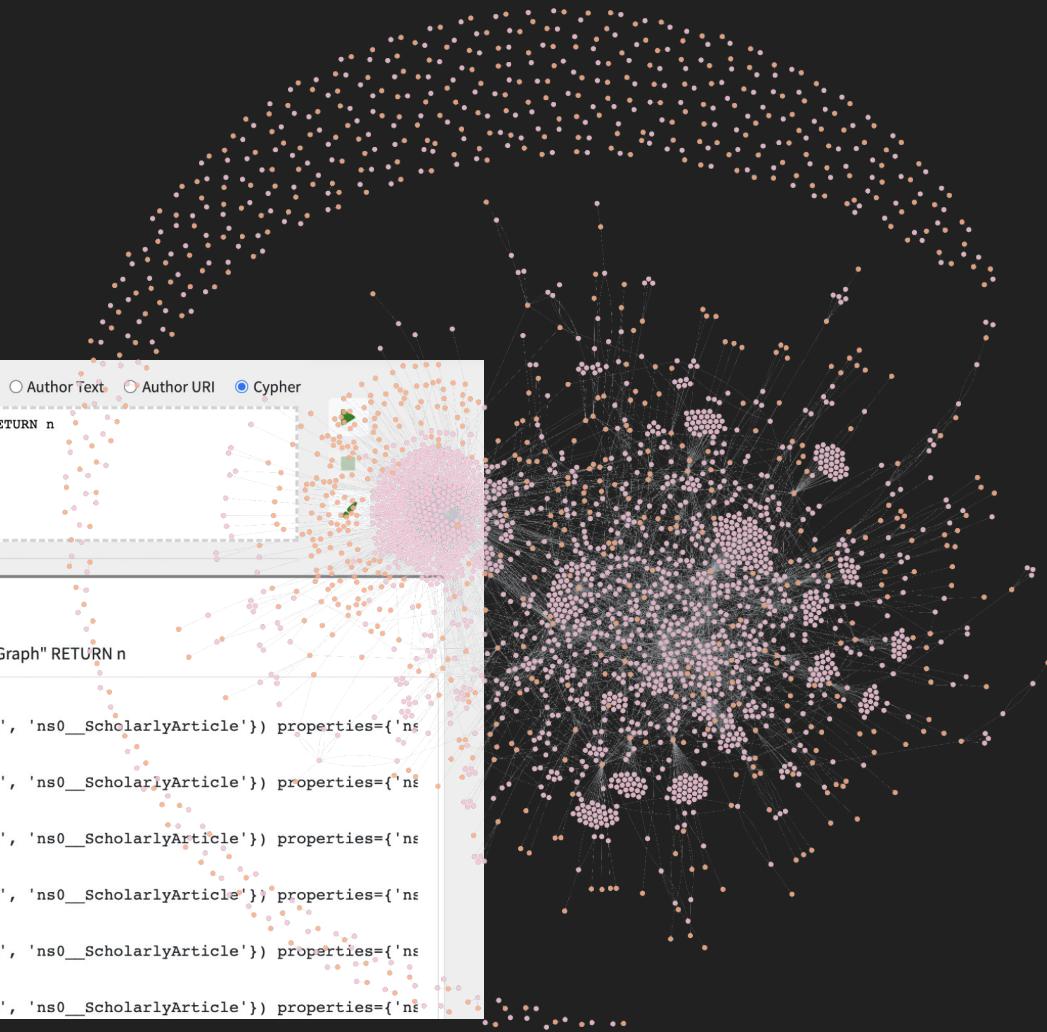
```
2 <Record n=<Node id=4934 labels=frozenset({'Resource', 'ns0_ScholarlyArticle'}) properties={'ns0_headline': 'Knowledge Graph', 'ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_title': 'Knowledge Graph', 'ns0_doi': null, 'ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_headline': 'Knowledge Graph', 'ns0_ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_ns0_title': 'Knowledge Graph', 'ns0_ns0_doi': null, 'ns0_ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.'}>
```

```
3 <Record n=<Node id=5887 labels=frozenset({'Resource', 'ns0_ScholarlyArticle'}) properties={'ns0_headline': 'Knowledge Graph', 'ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_title': 'Knowledge Graph', 'ns0_doi': null, 'ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_headline': 'Knowledge Graph', 'ns0_ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_ns0_title': 'Knowledge Graph', 'ns0_ns0_doi': null, 'ns0_ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.'}>
```

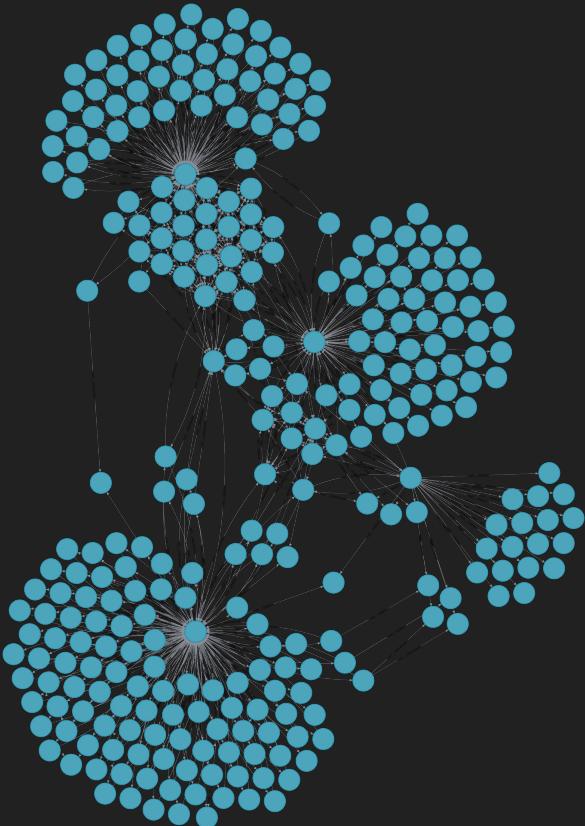
```
4 <Record n=<Node id=6428 labels=frozenset({'Resource', 'ns0_ScholarlyArticle'}) properties={'ns0_headline': 'Knowledge Graph', 'ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_title': 'Knowledge Graph', 'ns0_doi': null, 'ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_headline': 'Knowledge Graph', 'ns0_ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_ns0_title': 'Knowledge Graph', 'ns0_ns0_doi': null, 'ns0_ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.'}>
```

```
5 <Record n=<Node id=6688 labels=frozenset({'Resource', 'ns0_ScholarlyArticle'}) properties={'ns0_headline': 'Knowledge Graph', 'ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_title': 'Knowledge Graph', 'ns0_doi': null, 'ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_headline': 'Knowledge Graph', 'ns0_ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_ns0_title': 'Knowledge Graph', 'ns0_ns0_doi': null, 'ns0_ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.'}>
```

```
6 <Record n=<Node id=6917 labels=frozenset({'Resource', 'ns0_ScholarlyArticle'}) properties={'ns0_headline': 'Knowledge Graph', 'ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_title': 'Knowledge Graph', 'ns0_doi': null, 'ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_headline': 'Knowledge Graph', 'ns0_ns0_abstract': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.', 'ns0_ns0_url': 'https://en.wikipedia.org/w/index.php?title=Knowledge_graph&oldid=1000000000', 'ns0_ns0_title': 'Knowledge Graph', 'ns0_ns0_doi': null, 'ns0_ns0_text': 'A knowledge graph is a graph database specifically designed for storing and querying large amounts of structured data.'}>
```



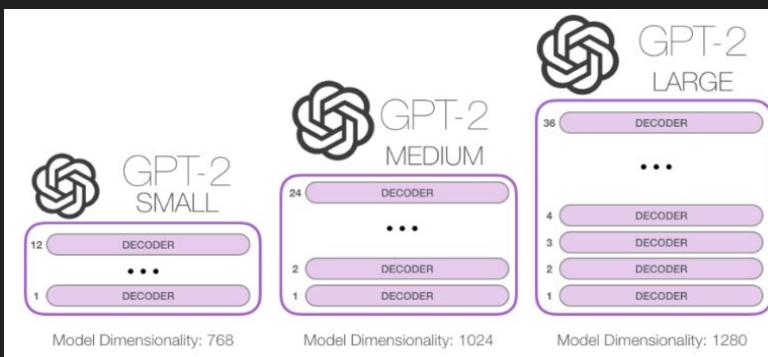
PageRank



Sample Endpoint Query:

Title	Representation Learning: A Review and New Perspectives		
URI	221		
ns1_doi	10.1109/TPAMI.2013.50		
ns0_datePublished	2013		
ns0_url	https://www.semanticscholar.org/paper/184ac0766262312ba76bbdece4e7ffad0aa8180b		
pagerank	21.52842592122827		
ns1_citationVelocity	1384		
ns1_b_influentialCitationCount	372		
onCount			
Authors	Yoshua Bengio	P. Vincent	Aaron C. Courville
Abstract	<p>The success of machine learning algorithms generally depends on data representation, because different representations can entangle and hide more or less the different facets behind the data. Although specific domain knowledge can be used to help define generic priors can also be used, and the quest for AI is motivating the design of algorithms implementing such priors. This paper reviews recent work in the area of deep learning, covering advances in probabilistic models, autoencoders, manifolds, which motivates longer term unanswered questions about the appropriate objectives for computing representations (i.e., inference), and the geometrical connections between estimation, and manifold learning.</p>		

Predicting Future Research



Predicting Future Research

(1) Reading the reference
abstracts (priming the model)



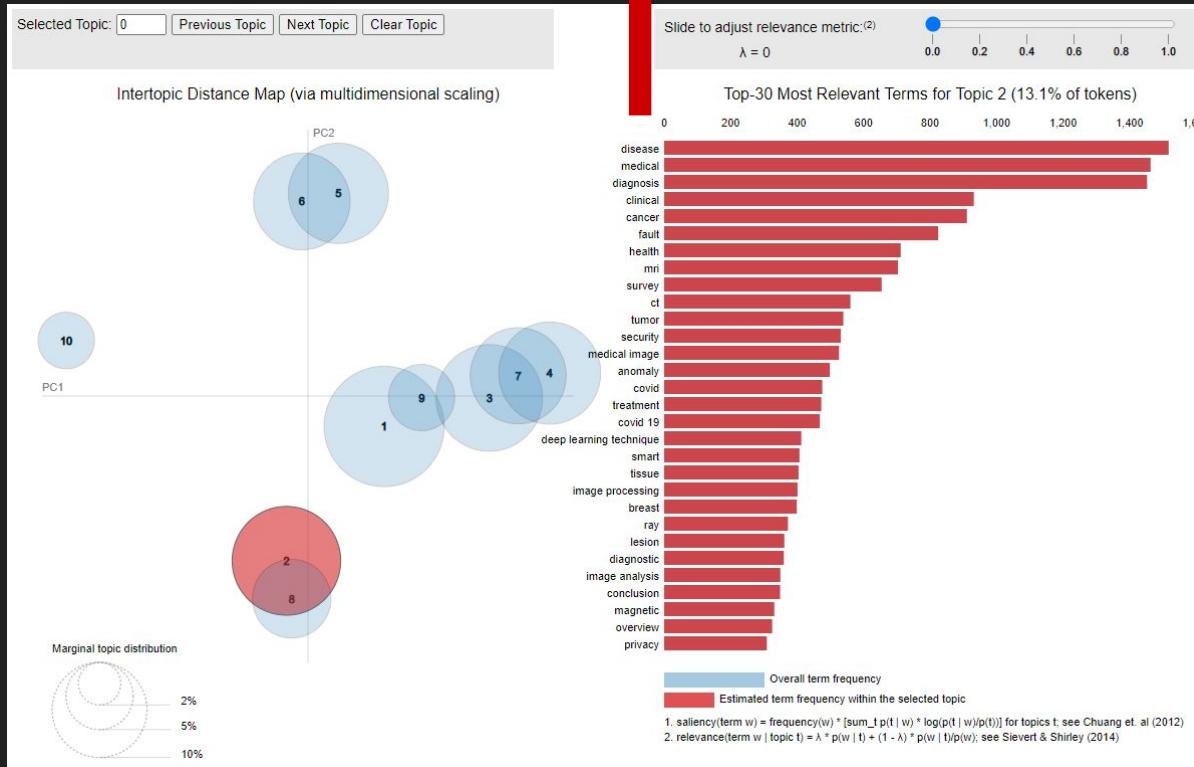
(2) Writing an entire new
paper (text generation)



(3) Writing the paper's
abstract (text summarization)

BONUS!

Topic Modeling



Latent Dirichlet Allocation

BONUS!

Recommender System

```
%time  
  
INPUT_STRING = "Deep Support Vector Quantum Neural Graph Networks"  
  
get_recommendation(INPUT_STRING,  
                   NUM_RECOMMENDATIONS=10,  
                   vectorizer = vectorizer,  
                   LDA_obj = LDA_obj,  
                   LDA_data = LDA_data,  
                   topic_names = topic_names,  
                   papers = papers)
```

Wall time: 1.07 s

ID	Cosine Distance	Title	Modeled Topic	Citations
1	0.011244	Underwater Target Classification Using Deep Ne...	feature; deep; classification; neural; neural ...	0.0
2	0.013166	Deep Learning for Monotonic Support Vector Mac...	feature; deep; classification; neural; neural ...	0.0
3	0.013761	Stacked Sparse Autoencoder (SSAE) based framew...	feature; deep; classification; neural; neural ...	13.0
4	0.020387	DeepGene: an advanced cancer type classifier b...	feature; deep; classification; neural; neural ...	13.0
5	0.021423	Extract Features Using Stacked Denoised Autoen...	feature; deep; classification; neural; neural ...	0.0
6	0.024708	Hybrid shallow and deep learned feature mixtur...	feature; deep; classification; neural; neural ...	0.0
7	0.025183	DeepPPI: Boosting Prediction of Protein-Protei...	feature; deep; classification; neural; neural ...	20.0
8	0.027645	Human emotion recognition using deep belief ne...	feature; deep; classification; neural; neural ...	36.0
9	0.030773	A deep learning method for lincRNA detection u...	feature; deep; classification; neural; neural ...	0.0
10	0.031278	ECG assessment based on neural networks with p...	feature; deep; classification; neural; neural ...	6.0

Trained on the entire graph

Paper recommendations using LDA similarity between input and abstracts

Not reliant on searching keywords in the database

Even hosted on a local machine can return query in 1 second

Validation

Validation Set: 98.4% TPR on
DBpedia authors

Novelty in Generated Abstracts:

Only 7% of synthetic papers find their
sources on the top 5 recommendations

Full dataset: 0.02% self-evident FPs
(impossible 3-way matches)

Validation Set: 99.53% TPR
on record linkage

Reduction Ratio: 0.93

Final Result

Recommender
System

Text Query

Neo4J
Endpoint

Research
Predictions

PageRank

DBpedia

AIKG

- Generated Queries
- Paper with Most In Links
- Highest PageRank Papers
- Paper with Most Out Links
- Most Connected Node with Connections
- All papers about Knowledge Graphs
- Connected papers about Knowledge Graphs
- Most Connected Node
- Papers by both Craig and Jay
- Papers by Wang (popular name)

Recommender System → Text Query → Neo4J Endpoint

+ LDA Topic Modeling

UI Elements:

- Radio buttons: Recommender, Paper Title, Paper URI, Paper DOI, Author Text, Author URI, Cypher (Cypher is selected)

Query:

```
1 MATCH (n)-[r:ns0__creditText]->() RETURN DISTINCT n.uri LIMIT 2
```

Results:

Query	MATCH	(n)-[r:ns0__creditText]->	() RETURN DISTINCT n.uri LIMIT 2
1 <Record n.uri='file:///Users/rehanahmed/Documents/USC/DSCI-558%20Project/notebooks/453709'>			
2 <Record n.uri='file:///Users/rehanahmed/Documents/USC/DSCI-558%20Project/notebooks/453708'>			

Results: 2 Executed at 2021-04-25 18:03:29.201735

Search Title: CNN

Results:

- Towards lightweight convolutional neural networks for object detection
- Fast Eye Detector Using CPU Based Lightweight Convolutional Neural Network

Neural-Net for Traffic Control on Foggy Days

'A sparse deep belief network with efficient fuzzy learning framework. The algorithm is based on a deep neural network that is able to learn to distinguish between different types of blurred scenes'. The algorithm can be applied to any scene, including static and dynamic ones, and it can also be used to improve the performance of existing algorithms. The authors demonstrate the use of deep learning in traffic prediction by using a network trained on real traffic data to train a model that could predict how fast a particular vehicle would be traveling at any given time of day, based on its speed in relation to other vehicles, such as cars, trucks, buses, motorcycles, bicycles, etc. In this way, the researchers demonstrate how deep networks can be used in a wide range of problems, including traffic management, vehicle safety, traffic congestion, urban planning, transportation planning and transportation planning. They hope that this work will inspire other researchers to explore the potential of this approach in other areas of this type of engineering. The work was supported by the U.S. National Science Foundation (NSF)'

Training a Facial Recognition Neural-Net from Brain Activity

"The reconstructions capture individual-specific information, such as the location of the eyes, nose, mouth, and other facial features, as well as their shape and size. This information can be used to improve the model's performance in a number of ways. For example, it is possible to use the reconstructed images to train a model to recognize faces. In this way, we are able to capture the individual features of a face and use them to predict the face's identity. We also use this information to enhance the performance of our model in other ways, for instance, by using it to detect faces that are not in the image. Our method is based on a combination of two approaches: (1) the use of a network-based approach and (2) a neural network based approach. Our model is able to extract features from the neural activity of a human subject, including the shape and shape of their dendritic trees. We show that our method can also be used as a tool for reconstruct"



typo

Thank You!

Questions?

