Knowledge Graph Fact Prediction via Knowledge-Enriched Tensor Factorization

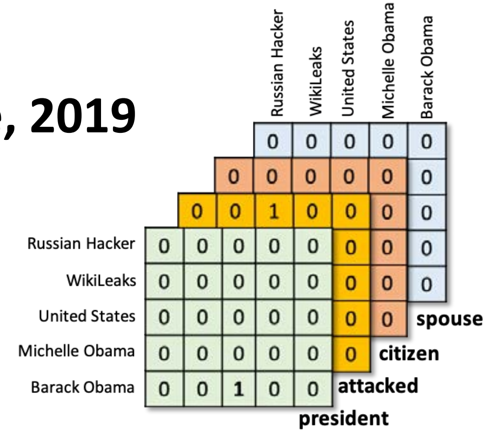
Ankur Padia, Konstantinos Kalpakis, Francis Ferraro and Tim Finin

University of Maryland, Baltimore County

Baltimore, MD, USA

International Semantic Web Conference, 2019

UMBC



Knowledge graphs of one kind or another have been used for more than 60 years for AI tasks, especially those involving language understanding

Today it’s important to combine knowledge graphs and machine learning to make both better

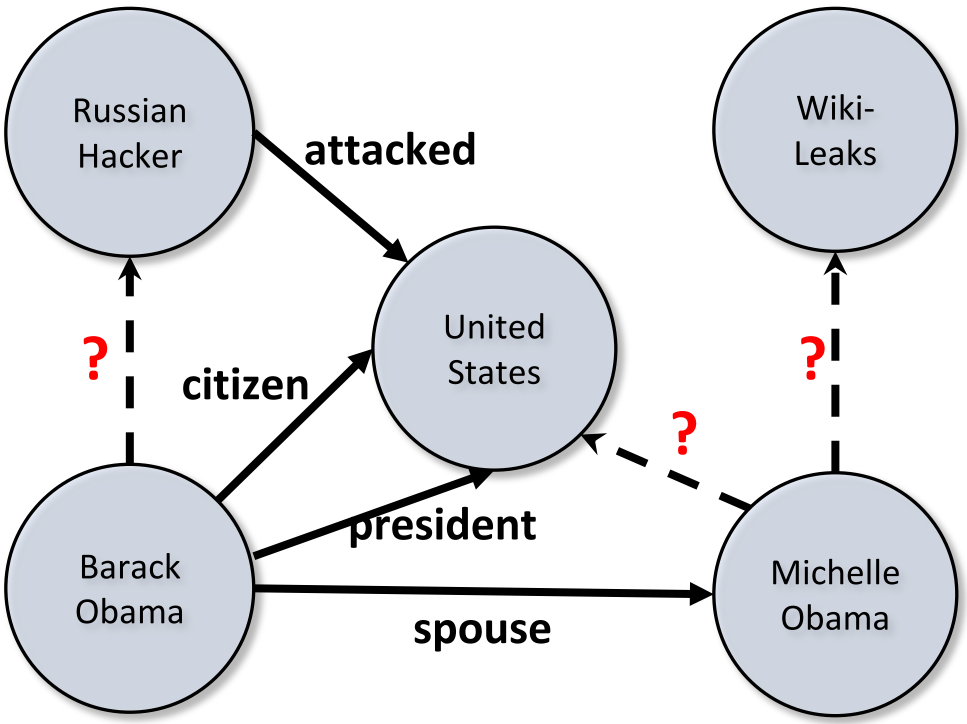
Knowledge Graphs ⇔ Machine Learning

- Both important in intelligent computing systems
  - **KG⇨ML**: Use existing KGs to support ML applications
  - **KG⇦ML**: Use ML to enrich and improve KGs
- **KGFP** has elements of both
  - Uses simple meta-knowledge extracted from an existing KG as data to ...
  - Improve a machine learning system's ability to predict relations that should be in the graph

KGFP Contributions

- **KGFP predicts likely relations** in a KG and outperforms RESCAL and similar systems
- Identifies relations *believed to hold* (**link verification**) rather than a *ranked list* of possible relations **(link ranking)**
- Does not require KG schemas, but computes a **relational similarity matrix** from an existing graph
- **Evaluated** on eight existing KG datasets, including ones based on DBpedia and Freebase

Knowledge Graph



Belief 1: Russian_Hacker attacked United_States .

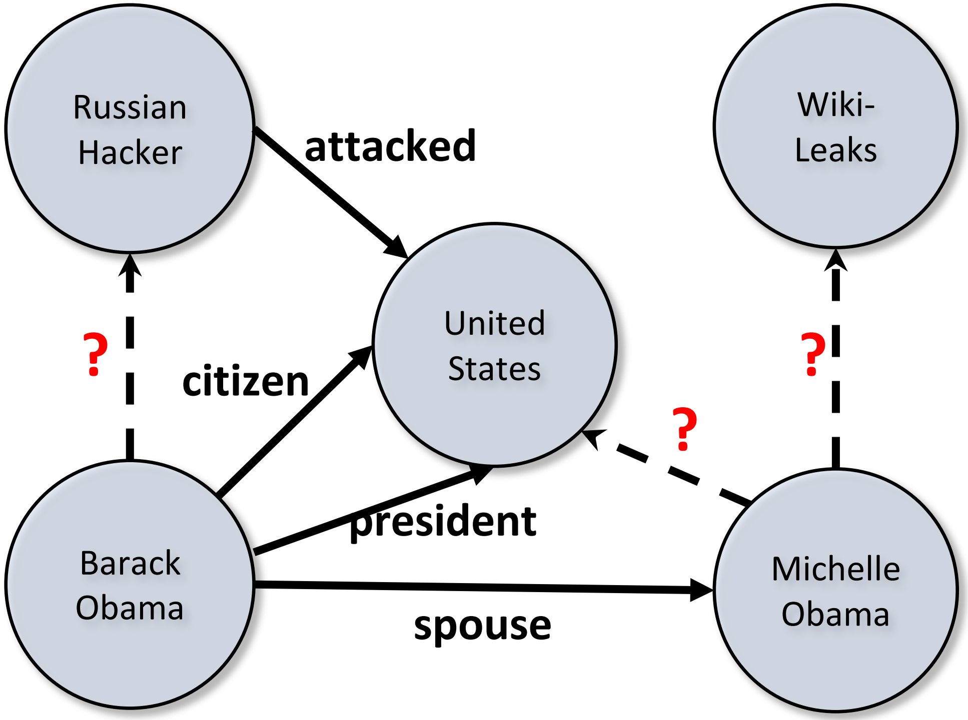
Belief 2: Barack_Obama Spouse Michelle_Obama .

Belief 3: Barack_Obama President United_States .

Belief 4: Barack_Obama citizen United_States .

Belief 5: Wikileaks isa non-profit_organization .

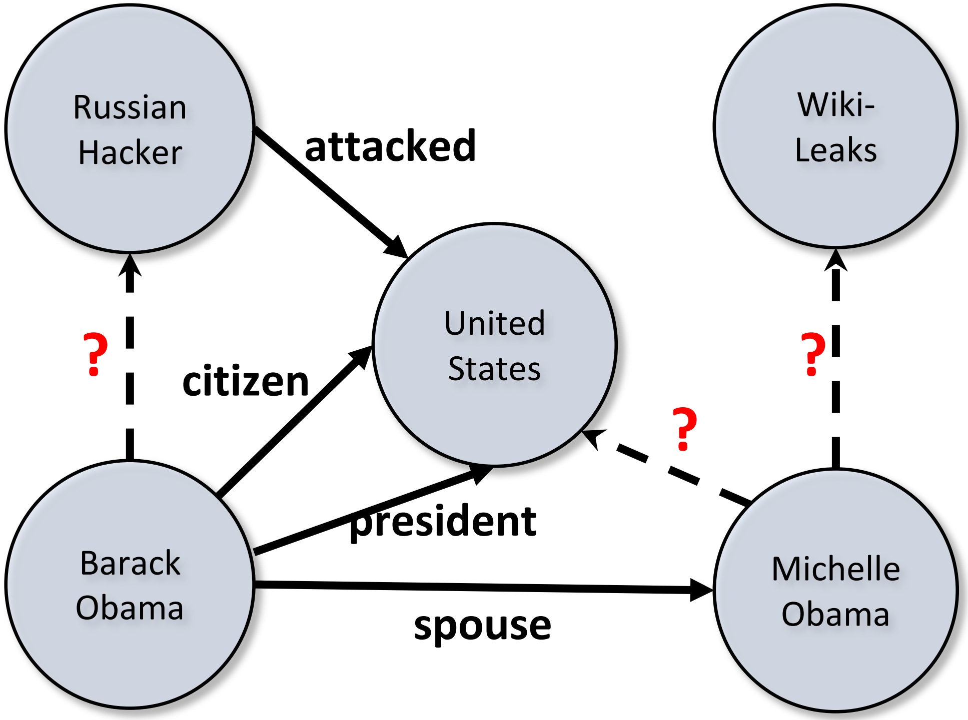
Knowledge Graph as Tensor



	Russian Hacker	WikiLeaks	United States	Michelle Obama	Barack Obama
Russian Hacker	0	0	1	0	0
WikiLeaks	0	0	0	0	0
United States	0	0	0	0	0
Michelle Obama	0	0	0	0	0
Barack Obama	0	0	0	0	0

attacked

Knowledge Graph as Tensor



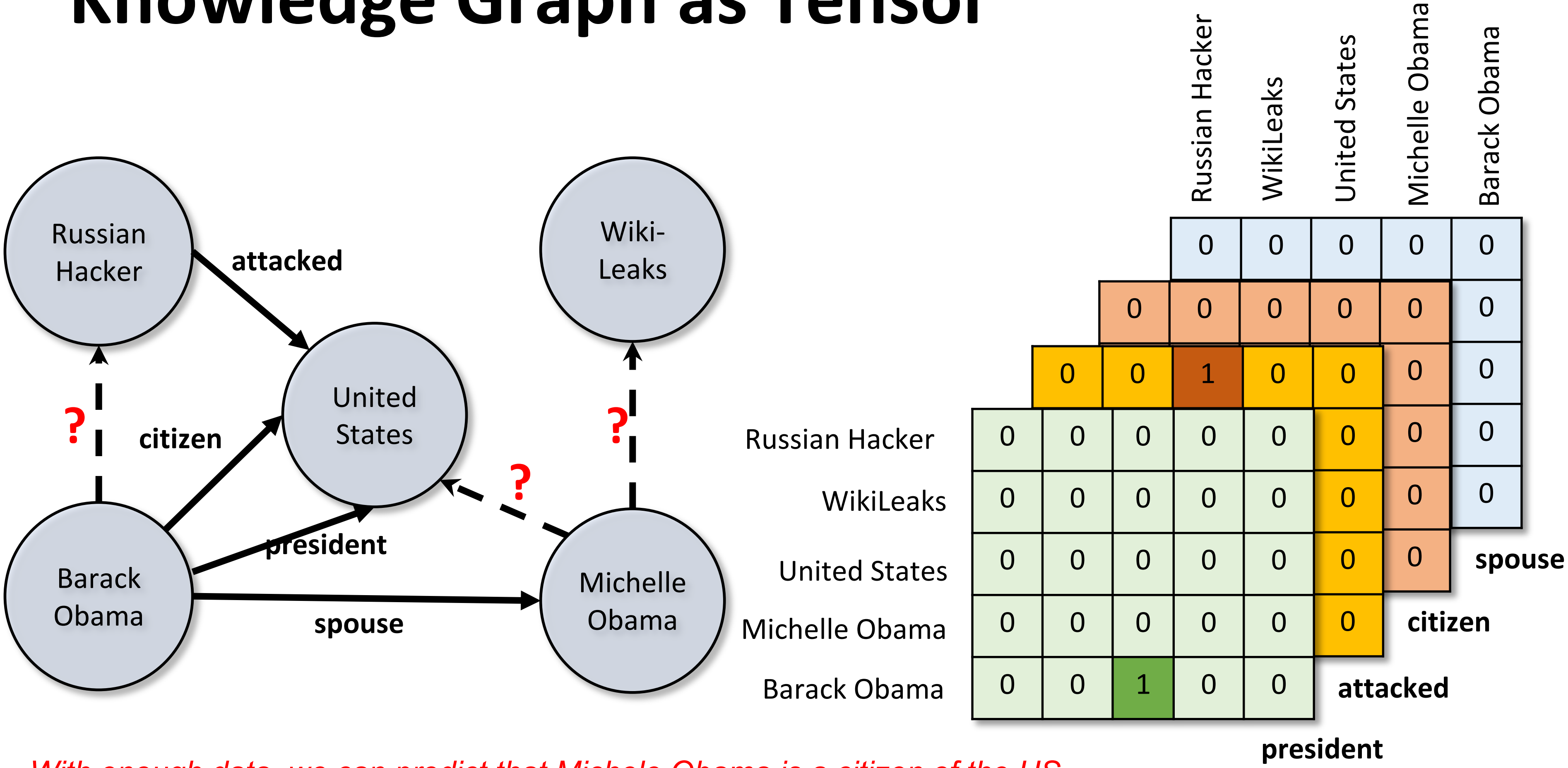
	Russian Hacker	WikiLeaks	United States	Michelle Obama	Barack Obama
Russian Hacker	0	0	1	0	0
WikiLeaks	0	0	0	0	0
United States	0	0	0	0	0
Michelle Obama	0	0	0	0	0
Barack Obama	0	0	0	0	0

attacked

	Russian Hacker	WikiLeaks	United States	Michelle Obama	Barack Obama
Russian Hacker	0	0	0	0	0
WikiLeaks	0	0	0	0	0
United States	0	0	0	0	0
Michelle Obama	0	0	0	0	0
Barack Obama	0	0	1	0	0

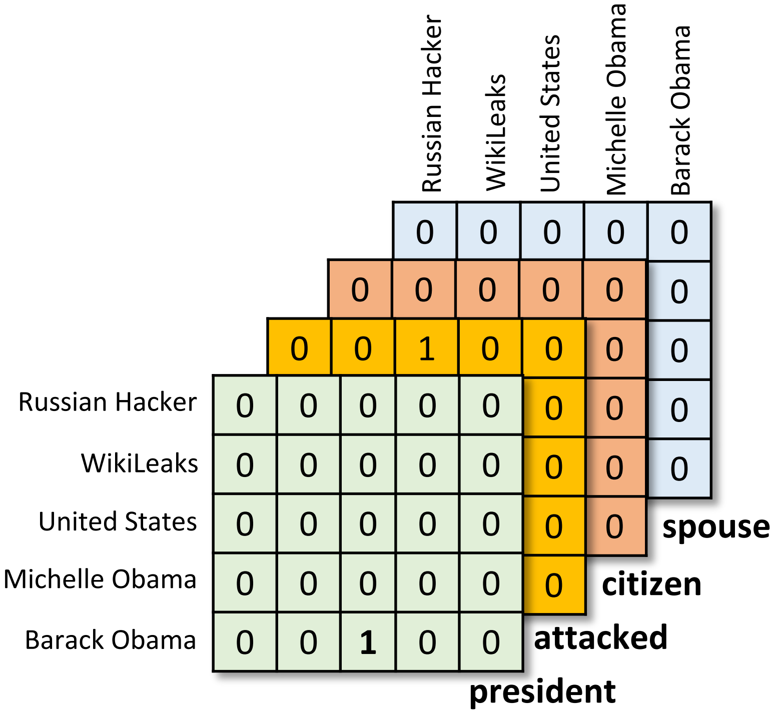
president

Knowledge Graph as Tensor



With enough data, we can predict that Michele Obama is a citizen of the US

Learning Embeddings for Entities and Relations



$$\chi$$

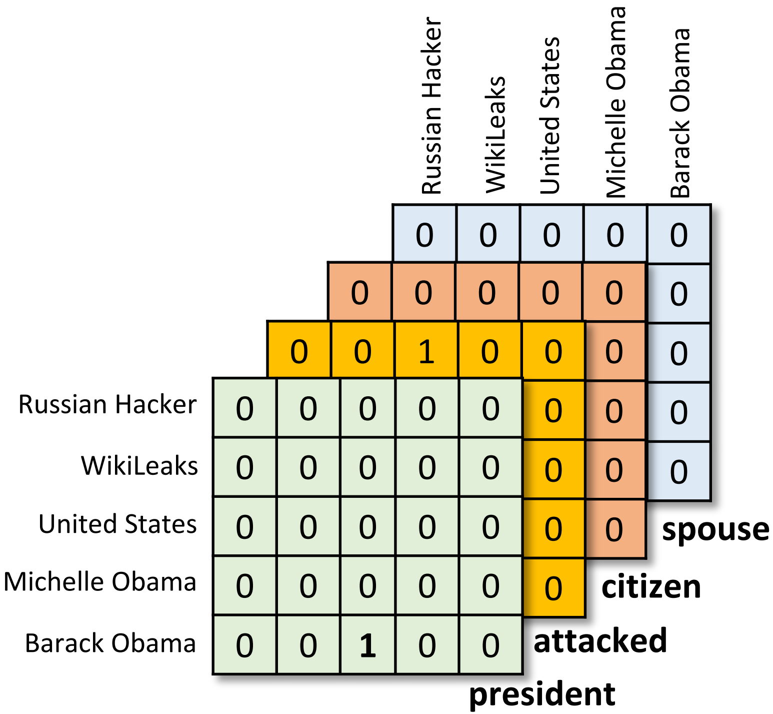
χ: a data tensor of size: e x e x k

e = number of entities

k = number of relations

Learning Embeddings for Entities and Relations

Jointly learn Entity (**E**) and Relation (**R**) embeddings



$$\chi \approx E R E^T$$

χ: a data tensor of size: e x e x k

e = number of entities

k = number of relations

p = latent dimension

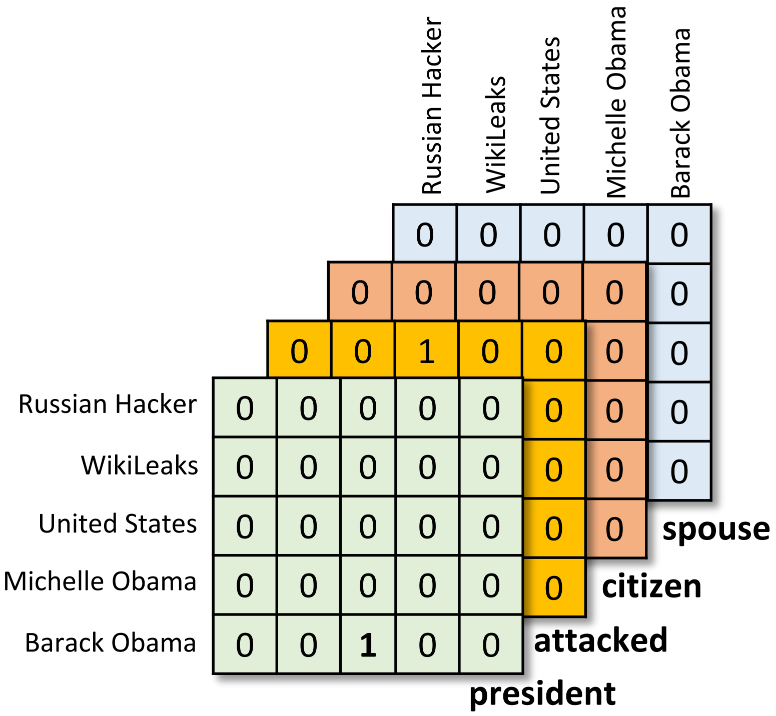
E =shared entity matrix of size: e x p

R = compact relation tensor of size: p x p x k

Learning Embeddings for Entities and Relations

Jointly learn Entity (**E**) and Relation (**R**) embeddings

$$\chi \approx ERE^T$$



We learn parameters to minimize error in reconstructing tensor from entity and relation embeddings

- Since
  Obama president US = 1
- We want
  E(Obama) • R(president) • E(US)^T → 1

Objective Function to Jointly Learn Embeddings

We modify RESCAL's approach, adding a new component with prior knowledge derived from the graph

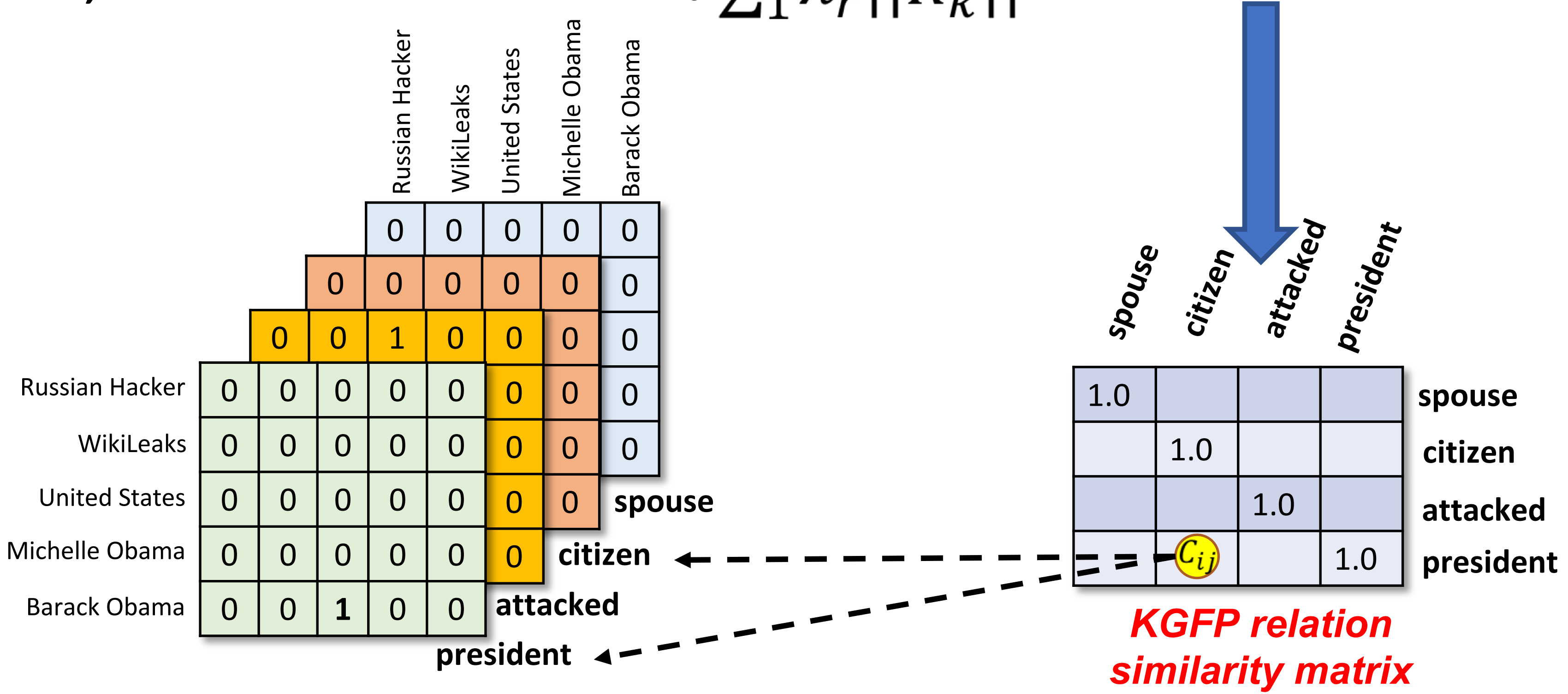
$$\underbrace{\underset{E,\,R}{\text{minimize}}\ \underbrace{||\chi - ERE^T||^2}_{\textit{reconstruction error}} + \underbrace{\lambda_a||E||^2 + \sum_1^k \lambda_r ||R_k||^2}_{\textbf{Avoid Overfitting}}}_{\textit{Base model – RESCAL (Nickel et al. ICML 2011)}} + \underbrace{C_{ij}||R_i - R_j||^2}_{\textbf{Use prior information}}$$

C_{ij} similarity matrix based on one of four similarity models

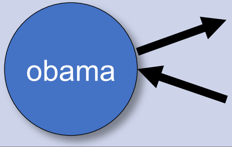
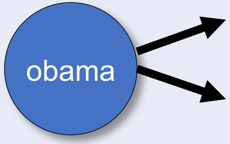
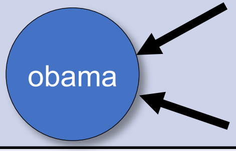
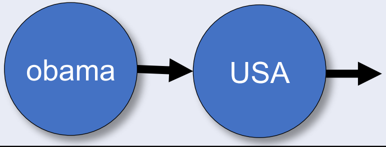
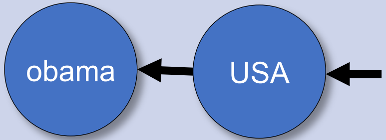
KGFP relation similarity matrix

Relation Similarity as Prior Information

$$\underset{E,\,R}{\text{minimize}} \; \|\chi - ERE^T\|^2 + \lambda_a\|E\|^2 + \sum_1^k \lambda_r\|R_k\|^2 + C_{ij}\|R_i - R_j\|^2$$



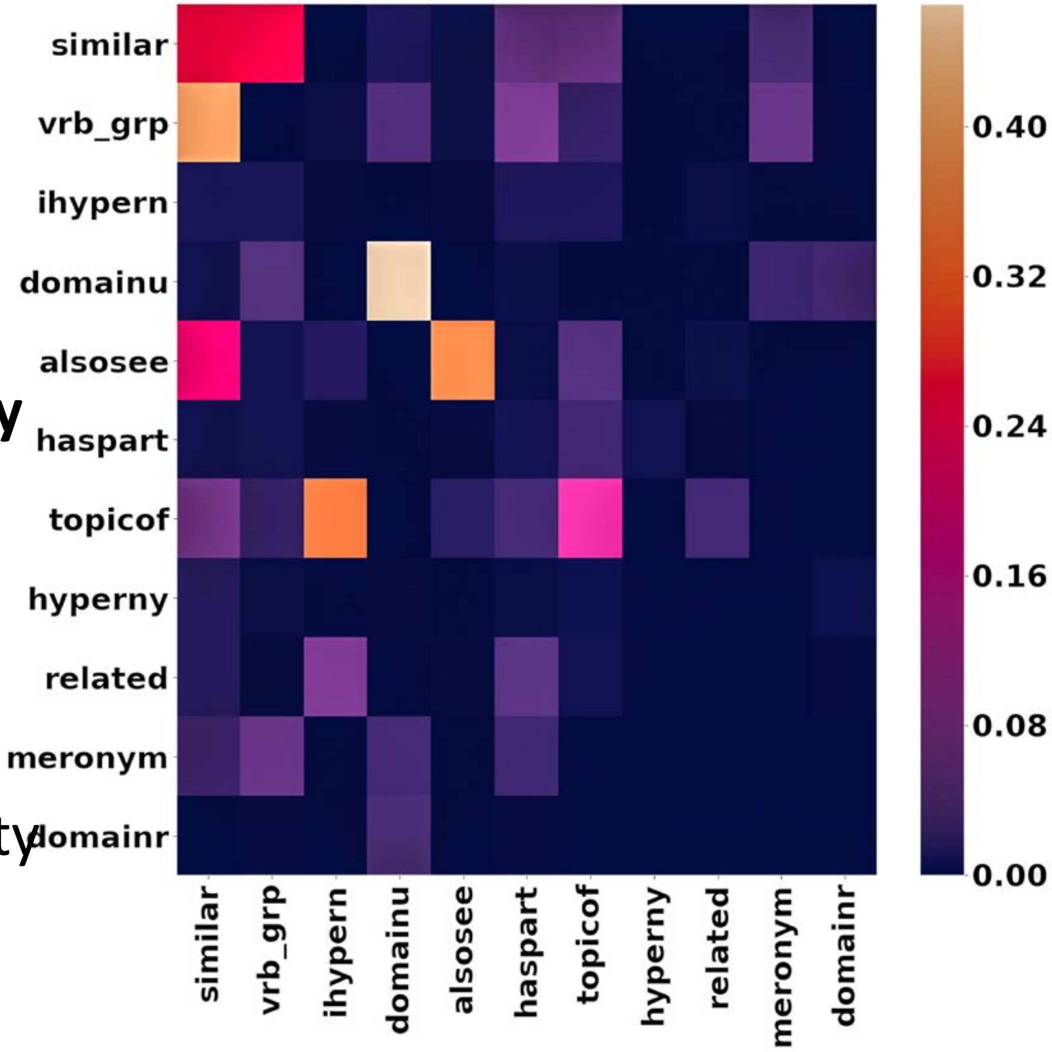
Relation Similarity

symmetry	How often do Ri & Rj share subject or object	
agency	How often do Ri & Rj share same subject	
patient	How often do Ri & Rj share same object	
transitivity	How often is Ri's object Rj's subject	
reverse transitivity	How often is Ri's subject Rj's object	

We evaluated five simple, easy to compute relation similarity metrics for every pair of relations in initial KG

Transitivity Best

- Which similarity metric is best depends on the KG
- Experiments showed **transitivity** gave **best** performance overall
- Especially for graphs derived from DBpedia and Freebase
- Heatmap for WIN18RR relation similarity matrix using transitivity shown



Evaluation on eight datasets

Dataset	Domain	Entities	Relations	Facts	Avg. Deg	Graph Density
Kinship	Social	104	26	10.7K	102.75	0.98798
UMLS	Medical	135	49	6.8K	50.01	0.37048
FB15-237	General	14.5K	237	310.1K	21.32	0.00147
DB10k	General	4.3K	140	10.0K	2.27	0.00052
FrameNet	Language	22.3K	16	62.3K	2.79	0.00013
WN18	Language	40.9K	18	151.4K	3.7	0.00009
FB13	General	81.1K	13	360.5K	4.45	0.00005
WN18RR	Language	40.9K	11	93.0K	2.27	0.00005

high

low

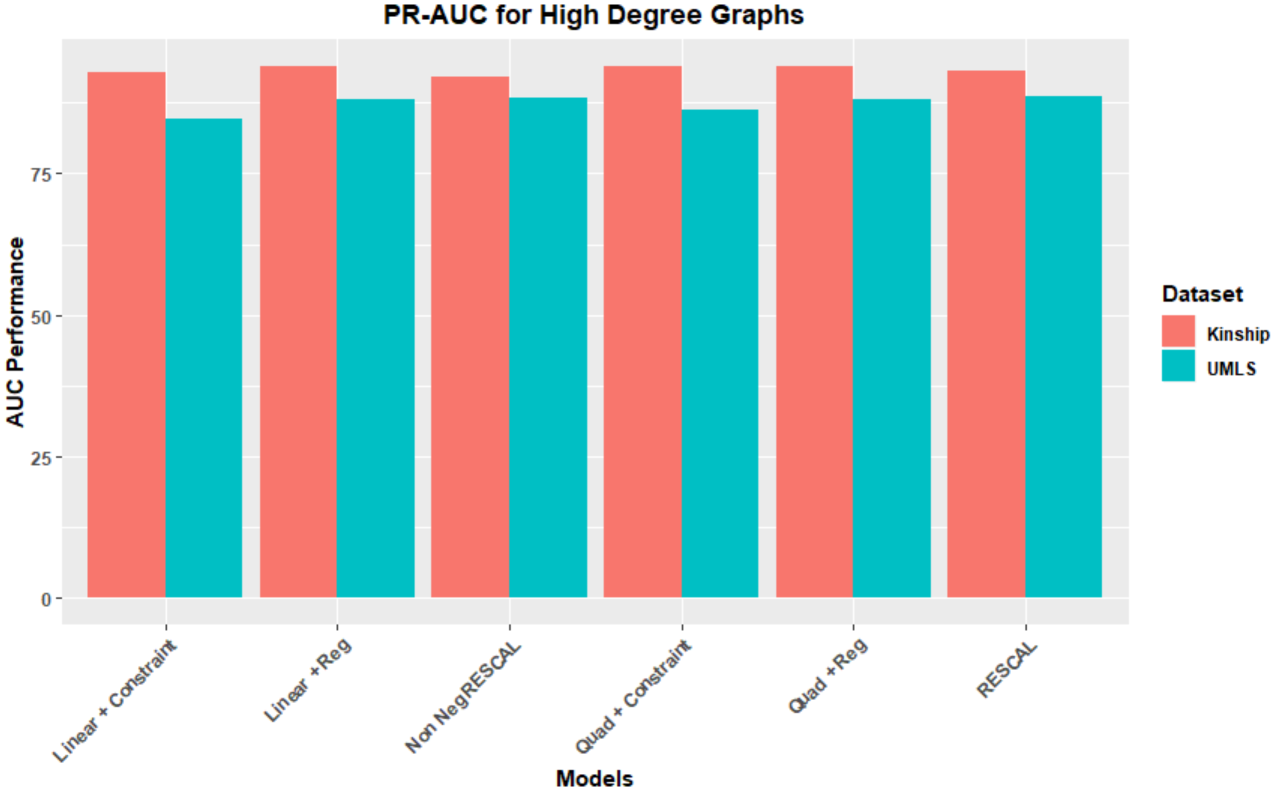
- Comparison with SOTA tensor factorization & translation-based models
- Used Precision-Recall AUC evaluation metric
- Note that Kinship and UMLS are outliers w.r.t. graph density

Experimental results

Model Name	Kinship	UMLS	WN18	FB13	DB10	Framenet	WN18RR	FB15-237
Previous tensor factorization models								
RESCAL	93.24	**88.53**	62.13	65.37	61.27	82.54	66.63	92.56
NN-RESCAL	92.19	88.37	83.93	79.13	**81.72**	82.6	68.49	93.03
Linear/Quadratic Regularized/Constrained tensor factorization models								
LR	**93.99**	88.22	81.86	80.07	80.79	78.11	69.15	90.00
QR	93.89	88.11	84.41	79.12	80.47	82.34	66.73	**93.07**
LC	92.87	84.71	80.18	75.79	80.67	73.64	66.46	81.88
★ QC	93.84	86.17	**91.07**	**85.15**	81.69	**86.24**	**72.62**	86.47

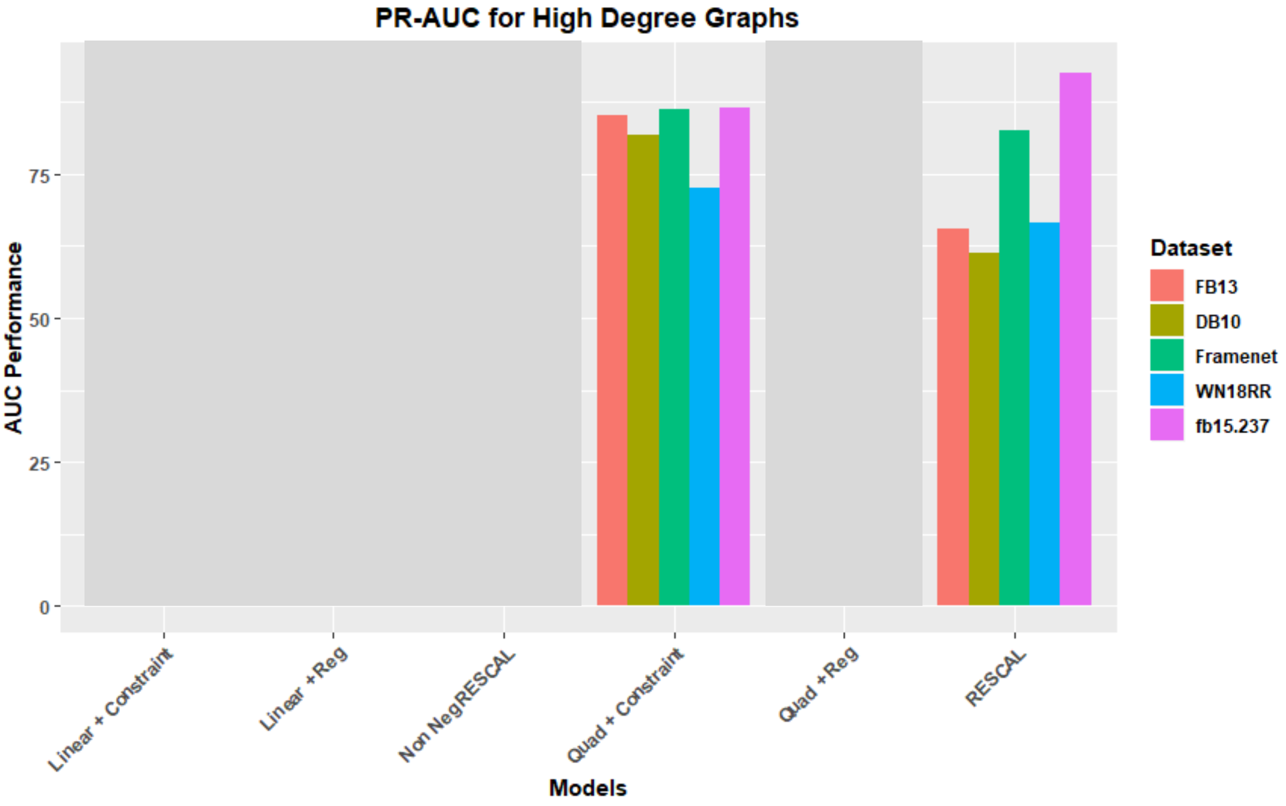
Fact prediction AUC performance for all models. The **quadratic-constrained** model is best overall for graphs with low density, which include the those derived from DBpedia and Freebase.

Results: High Density KGs



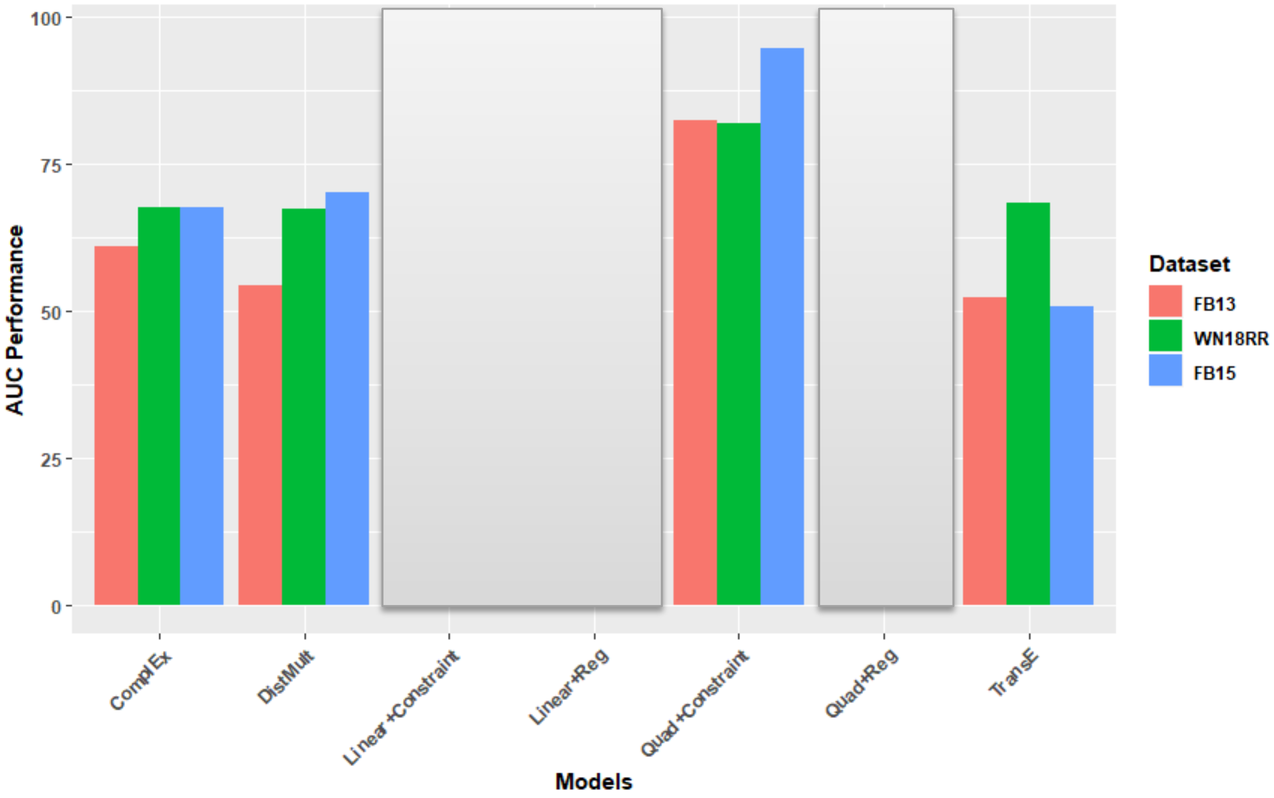
When graph density high, regularization or constraining does not help much and results are similar to RESCAL

Results: Low Density KGs



When graph density low, the quadratic + constraint model is significantly better

Results: Compared to Neural Systems

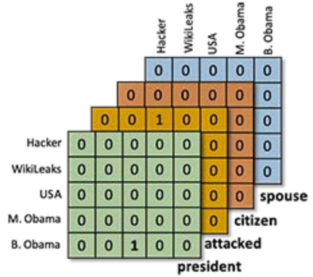


KGFP performed better than TransE, ComplEx & DistMult for our experiments on low density graphs

Future Work

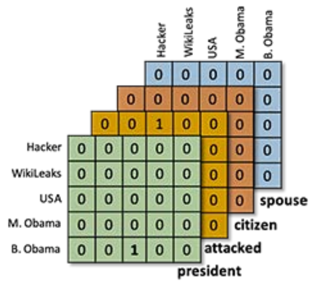
- Exploring additional **relation similarity models**
- Using KGFP to help identify **NLP information extraction system errors**
- Using KGFP as a module in a system to **clean noisy knowledge graphs** by identifying unlikely relations
- Apply KGFP to **cybersecurity knowledge graphs** created from STIX cyberthreat intelligence feeds, NVD and other semi-structured cybersecurity data resources

Conclusions



- **KGFP** is a novel **tensor factorization** approach for KG **fact prediction**, giving SOTA results on many graphs
- Exploits **relation similarity** extracted from existing graphs, with the **transitivity prior** generally best
- Performs well on relatively **sparse graphs** like DBpedia and Freebase

More information



- Code & datasets: https://github.com/Ebiquity/KGFP
- Preprint of JWS paper
- ISWC journal track paper
- For more information, contact Dr. Ankur Padia, Philips Research Americas, pankur1@umbc.edu