Knowledge Graphs in Recommending and CBR

Derek Bridge
University College Cork

Keynote invited talk, ICCBR'23

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

Insight
SFI RESEARCH CENTRE FOR DATA ANALYTICS

HOST INSTITUTIONS

PARTNER INSTITUTIONS

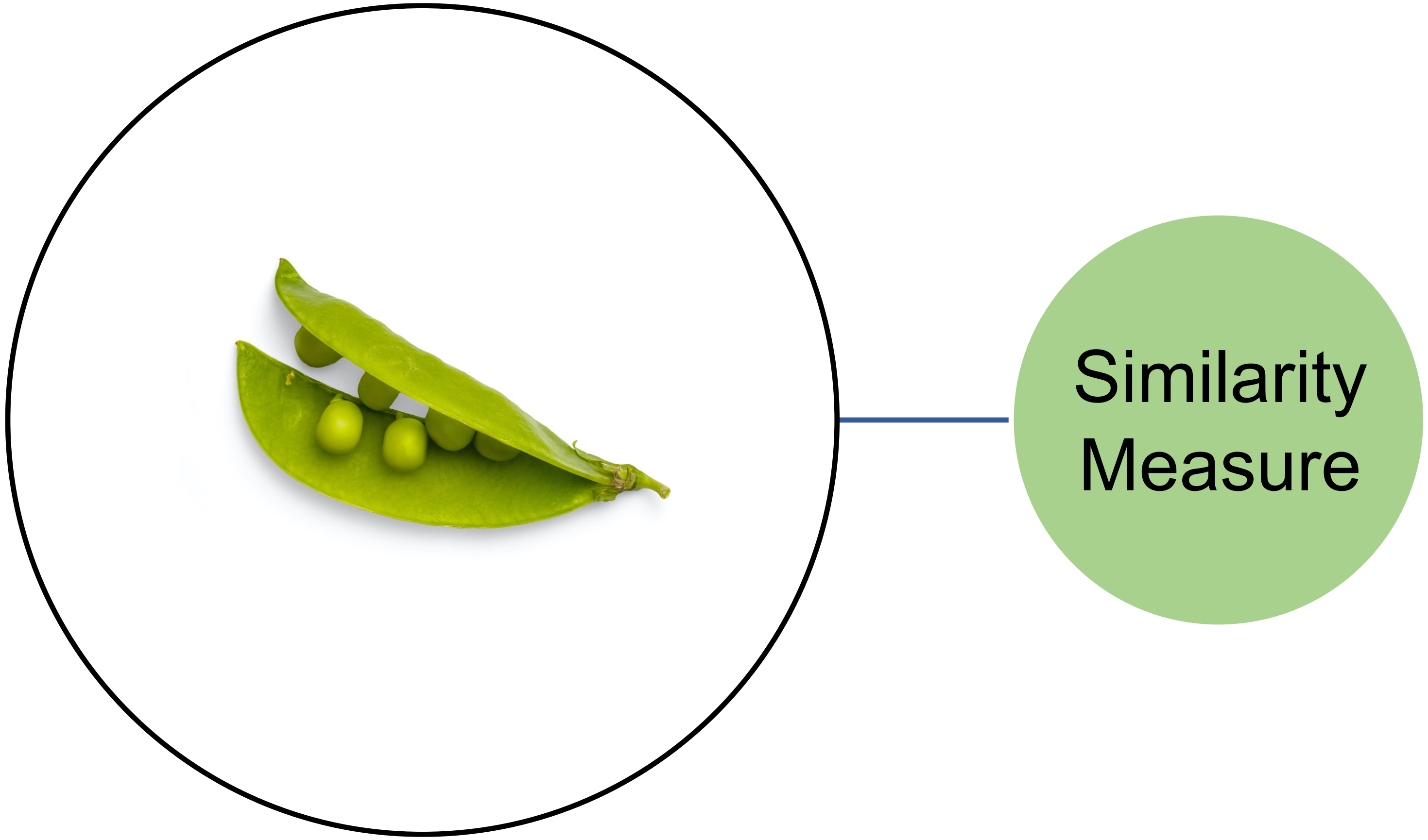
FUNDED BY:

Derek Bridge

- Senior Lecturer,
  School of Computer Science & IT,
  University College Cork
- Principal Investigator,
  Insight, SFI Centre for Data Analytics
- Co-director,
  SFI Centre for Research Training in AI
- Principal Investigator,
  iSee Project

Overview

- Introduction
- Applications
  - Interesting paths
  - Neighbourhood classification
  - Heuristic embeddings
  - Learned embeddings
- Conclusions


Similarity
Measure

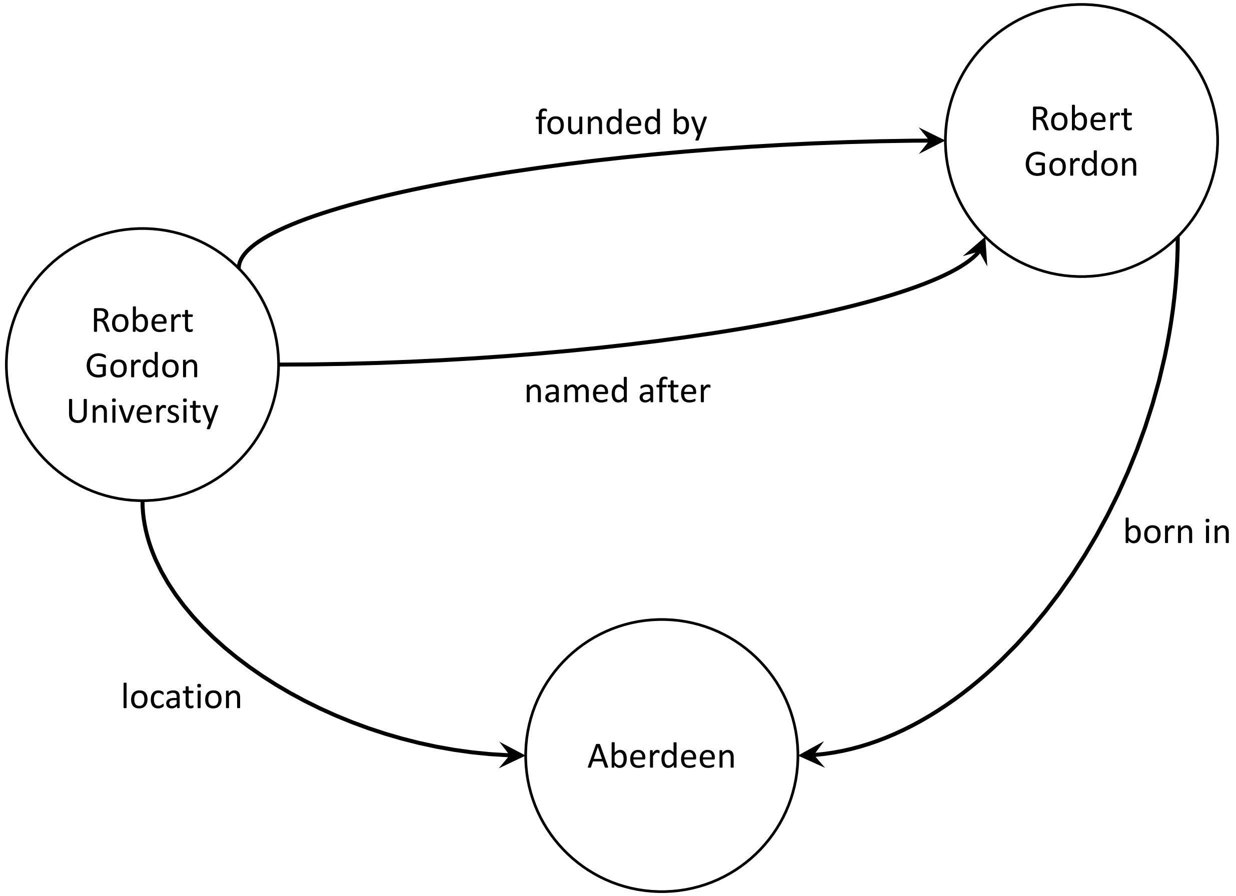
Knowledge Graphs

Introduction

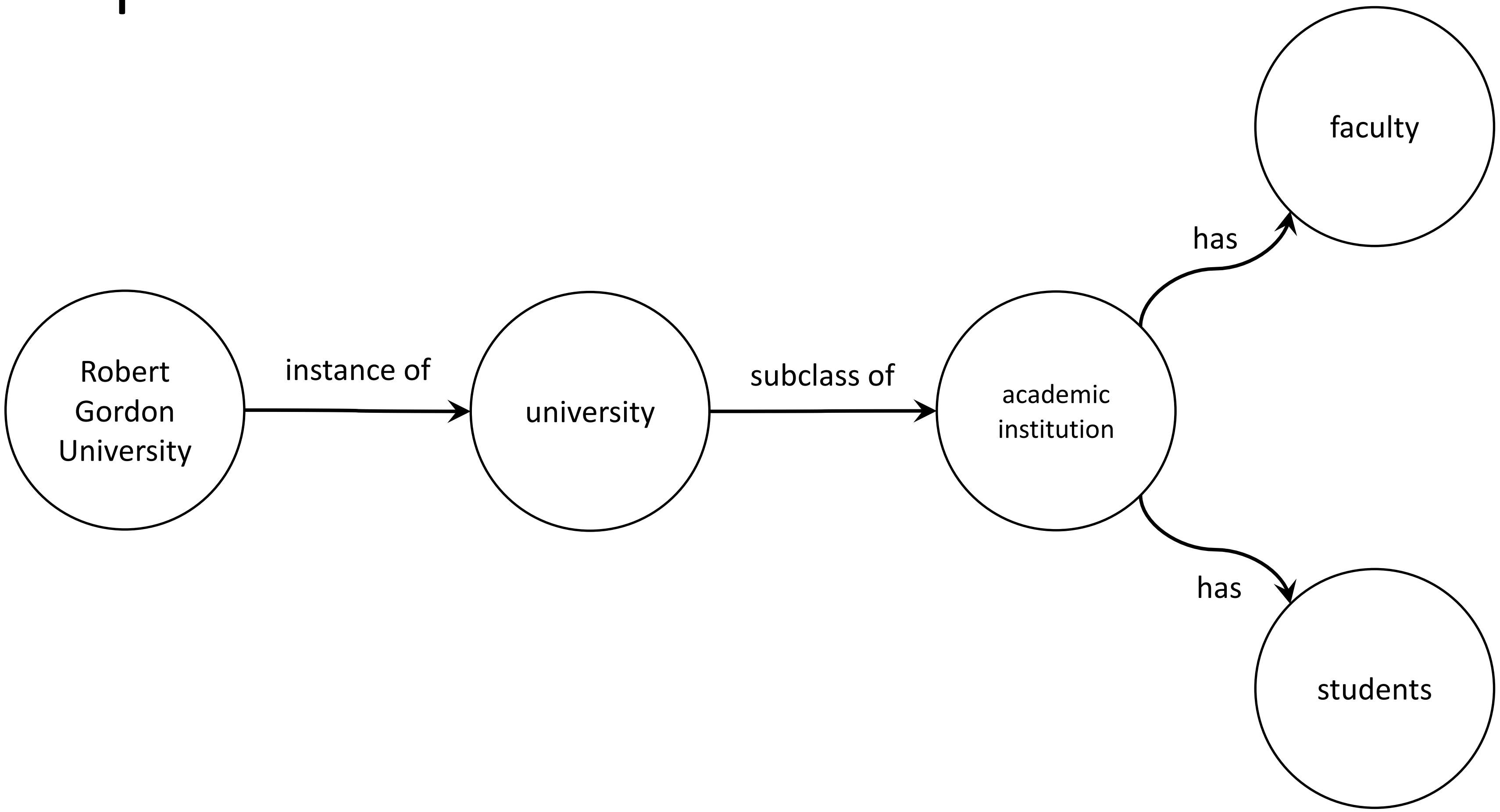
Facts: entities and relations

SOURCE ENTITY	RELATION	TARGET ENTITY
Robert Gordon University	founded by	Robert Gordon
Robert Gordon University	named after	Robert Gordon
Robert Gordon University	location	Aberdeen
Robert Gordon	born in	Aberdeen

Knowledge graph



Concepts



Robert Gordon University
instance of
university
subclass of
academic institution
has
faculty
has
students

Scope

Enterprise-specific

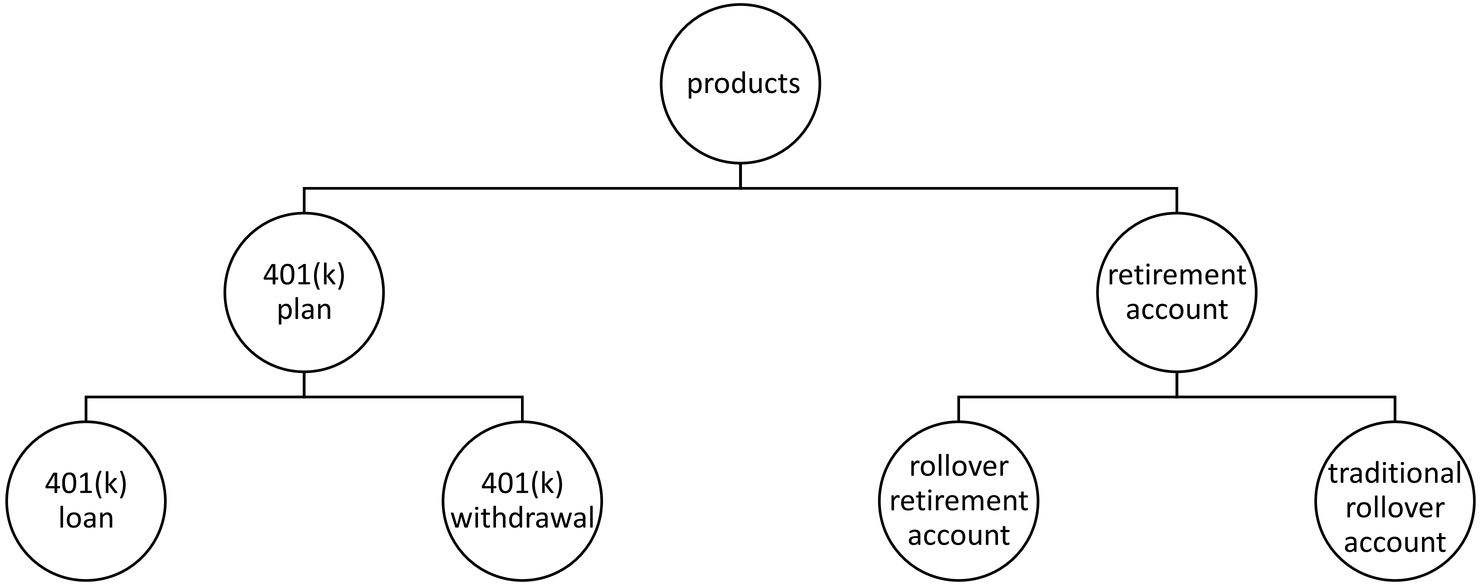
Domain-specific

Encyclopedic

Enterprise-specific

Taxonomy Extraction for Customer Service Knowledge Base Construction

Bianca Pereira[1(✉)], Cecile Robin[1], Tobias Daudert[1], John P. McCrae[1], Pranab Mohanty[2], and Paul Buitelaar[1]

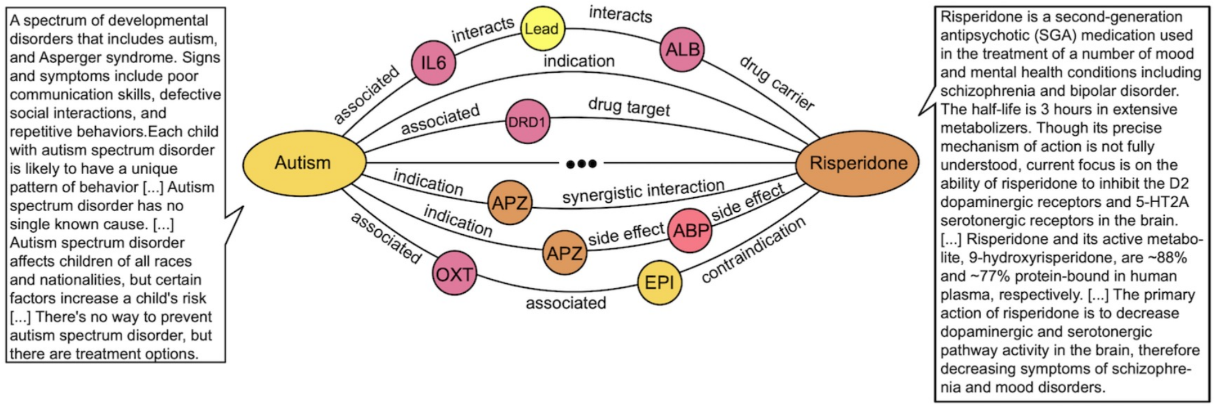


Saffron
Knowledge Extraction Framework

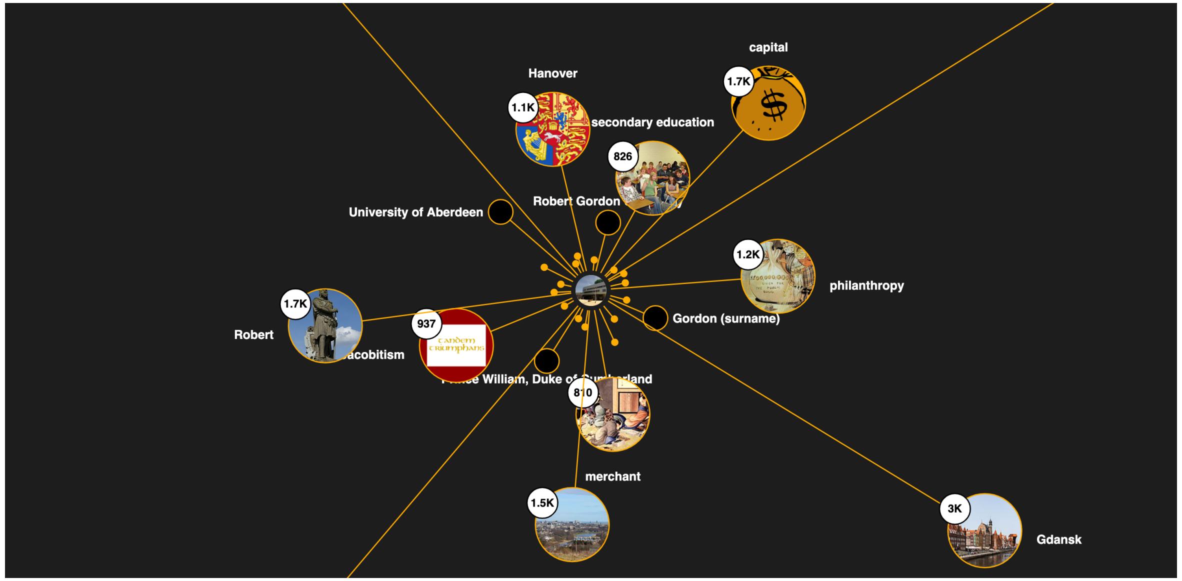
Domain-specific

MusicBrainz

PrimeKG

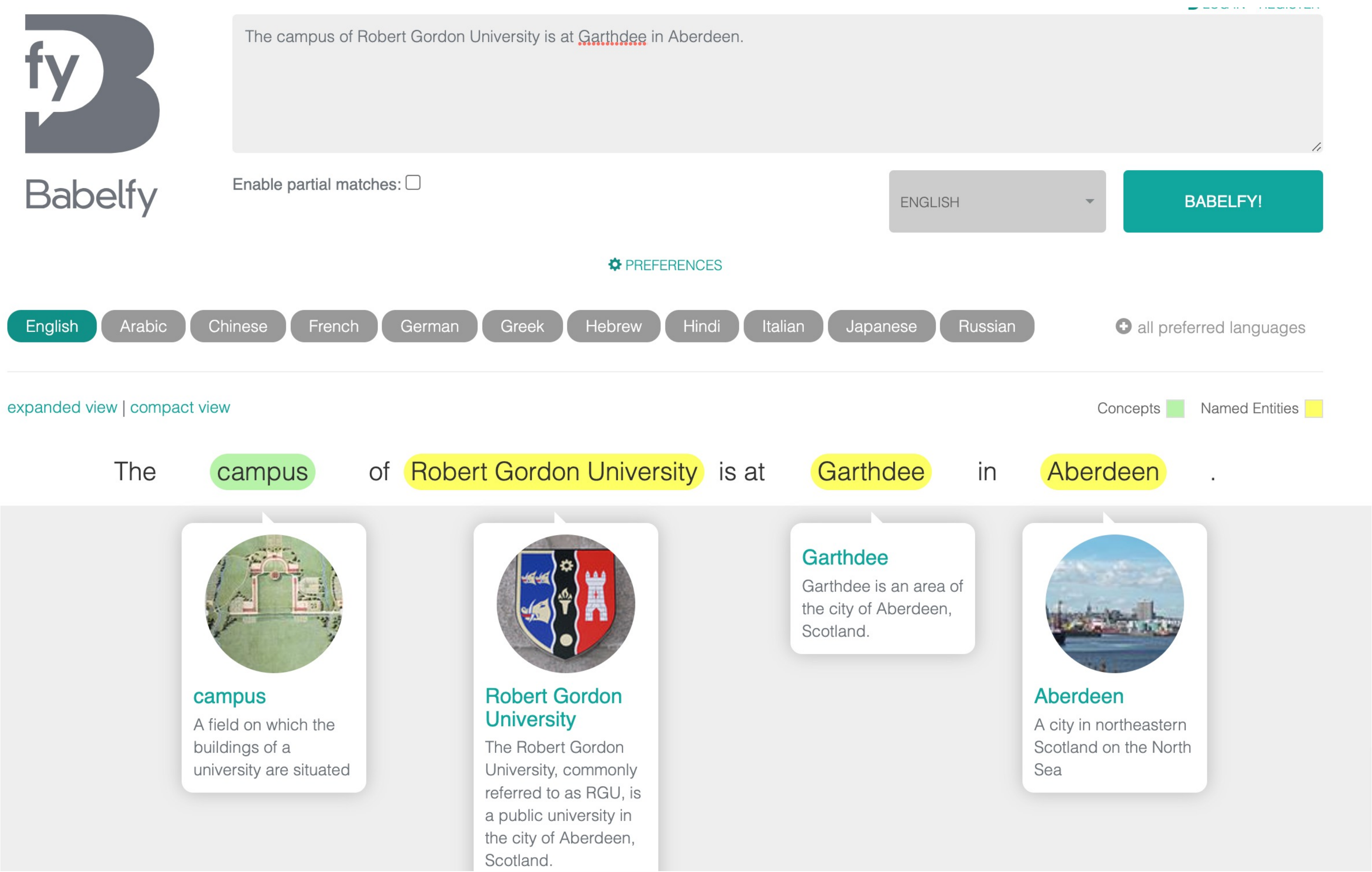


Encyclopedic

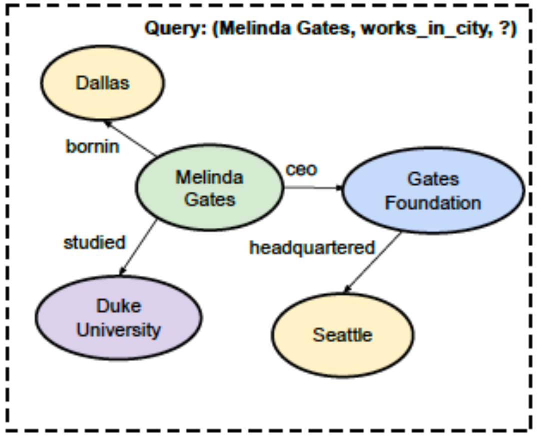


- a multilingual encyclopedic dictionary
- covers 520 languages
- obtained from the automatic integration of
  - Wikipedia
  - WikiData
  - WordNet
  - GeoNames
  - …

Entity Linking/Word Sense Disambiguation



Link prediction



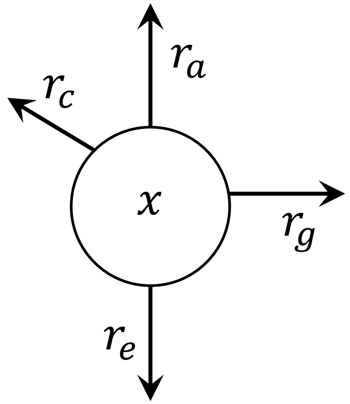
A Simple Approach to Case-Based Reasoning in Knowledge Bases

Rajarshi Das[1] RAJARSHI@CS.UMASS.EDU
Ameya Godbole[1] AGODBOLE@CS.UMASS.EDU
Shehzaad Dhuliawala[2] SHEHZAAD.DHULIAWALA@MICROSOFT.COM
Manzil Zaheer[3] MANZILZAHEER@GOOGLE.COM
Andrew McCallum[1] MCCALLUM@CS.UMASS.EDU

Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion

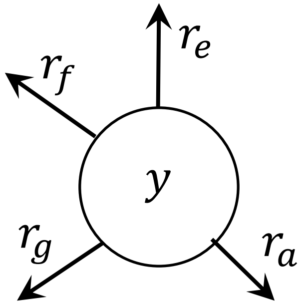
Rajarshi Das, Ameya Godbole, Nicholas Monath, Manzil Zaheer, Andrew McCallum

Node similarity



r_a	r_b	r_c	r_d	r_e	r_f	r_g	r_h
1	**0**	**1**	**0**	**1**	**0**	**1**	**0**

inner product



1	**0**	**0**	**0**	**1**	**1**	**1**	**0**

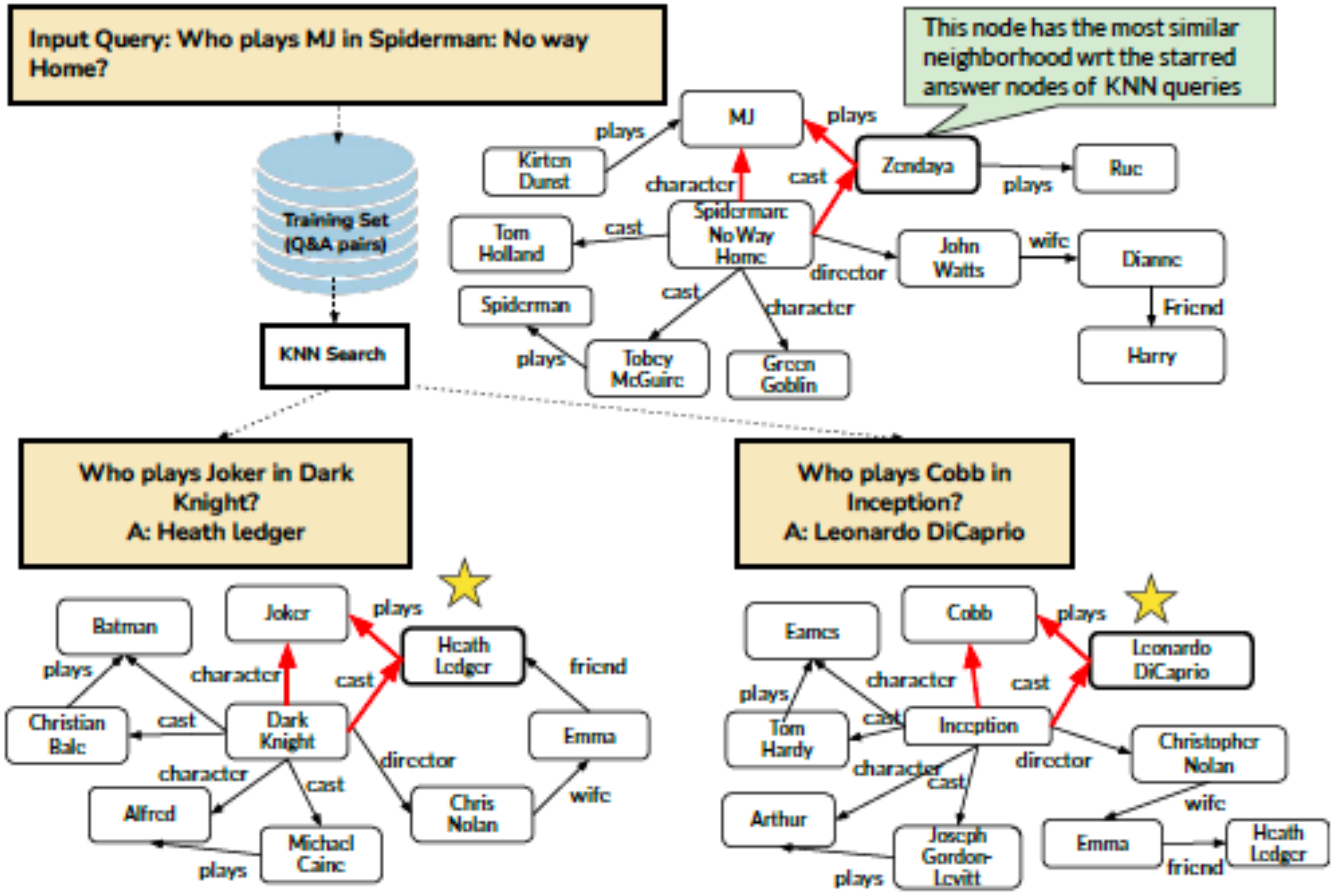
Knowledge Graphs

Applications

Question-answering

Knowledge Base Question Answering by Case-based Reasoning over Subgraphs

Rajarshi Das *[1] Ameya Godbole *[2] Ankita Naik [1] Elliot Tower [1] Robin Jia [2]
Manzil Zaheer [3] Hannaneh Hajishirzi [4] Andrew McCallum [1]



Overview

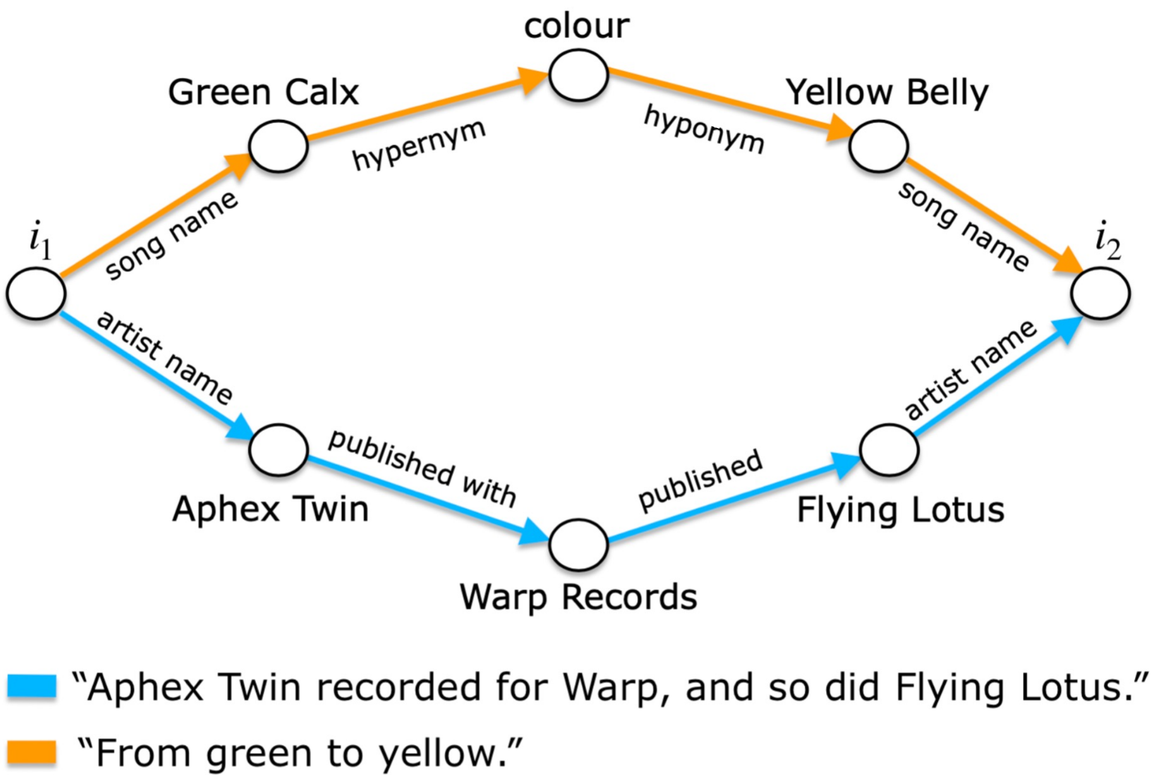
- Introduction
- Applications
  - Interesting paths
  - Neighbourhood classification
  - Heuristic embeddings
  - Learned embeddings
- Conclusions

Segues between songs

first song	segue	second song
Green Calx by *Aphex Twin*	Aphex Twin recorded for Warp and so did Flying Lotus...	*Yellow Belly* by *Flying Lotus*

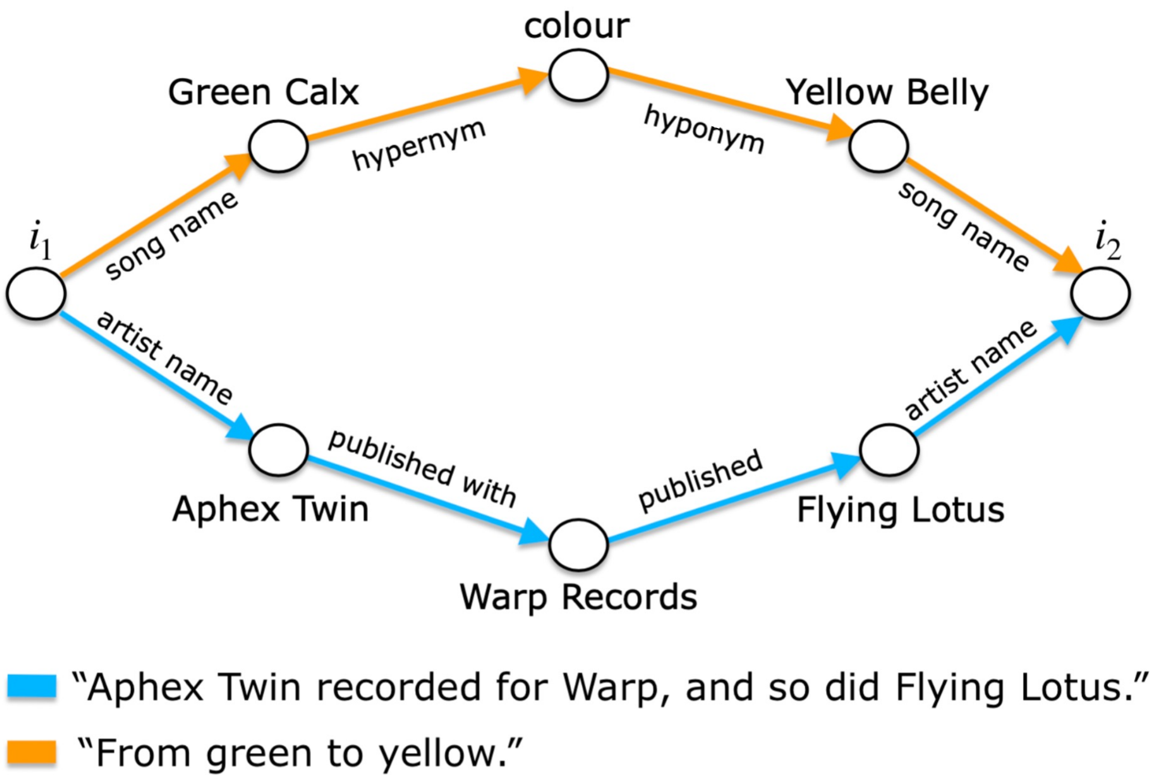
Dave: a method to generate segues

- Find **paths** in a knowledge graph that connect the songs
- Score the paths for **interestingness**
- Convert the most interesting path into **natural language**



Dave’s theory of interestingness

- **Infrequency** adds to interestingness
  - infrequent path type
  - infrequent entities/concepts
- **Conciseness** adds to interestingness



We are so fragile

Tubeway Army

Gary Numan was a member of Tubeway Army

I am dust

Gary Numan

Whine and grine

Prince Buster

From prince to princess…

Princess Olivia

Al Stewart

Dave user trial (*n* = 158)

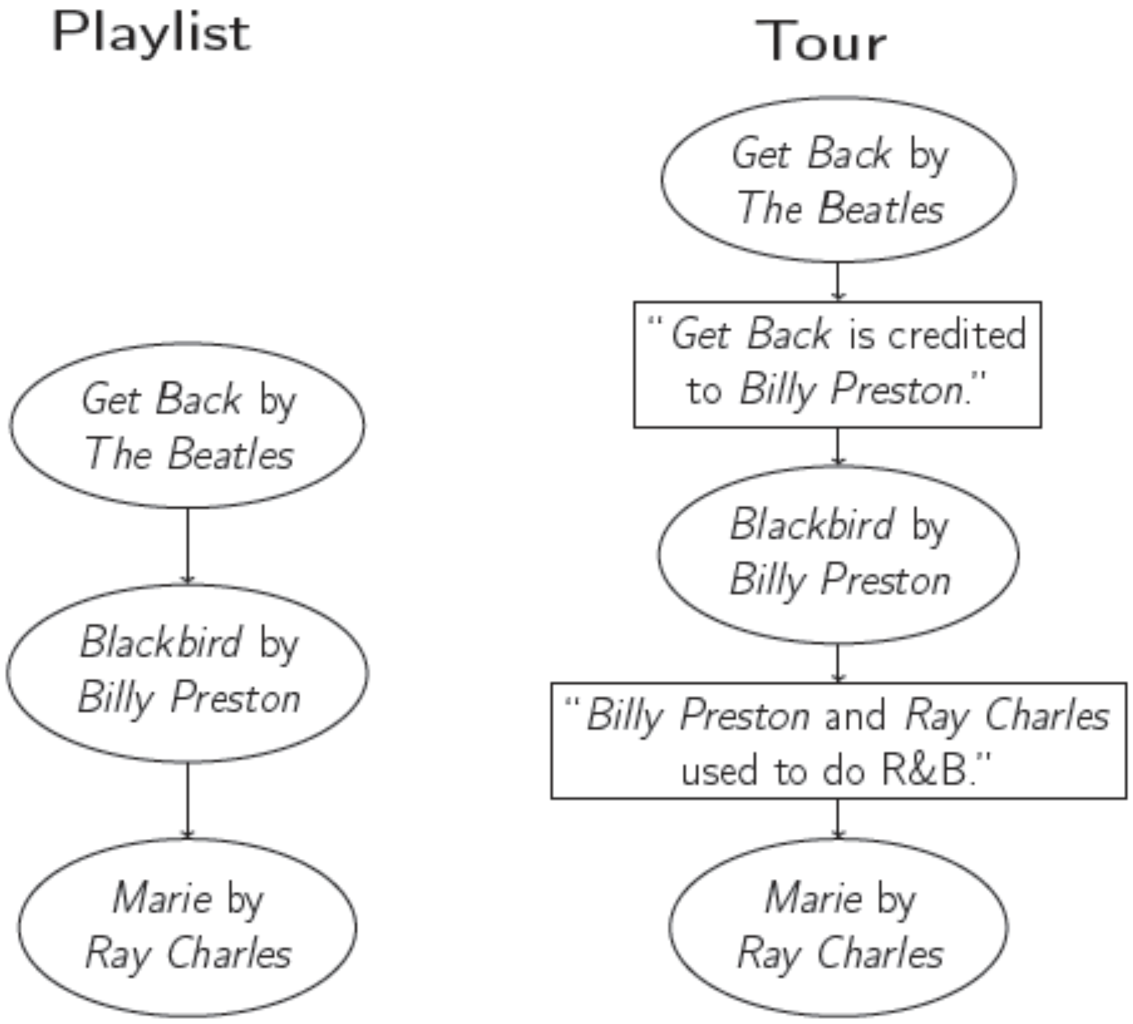
- Dave's **factual** segues are *as good as* manually-created, curated ones

	likeable	*high-quality*	*sparked-interest*	*funny*	*informative*	*creative*	*understandable*	*well-written*
DAVE	3.26	3.22	3.06	2.39	3.73**	3.33	3.84	3.46
THE CHAIN	3.20	3.20	3.13	2.64*	3.43	3.59*	3.69	3.33

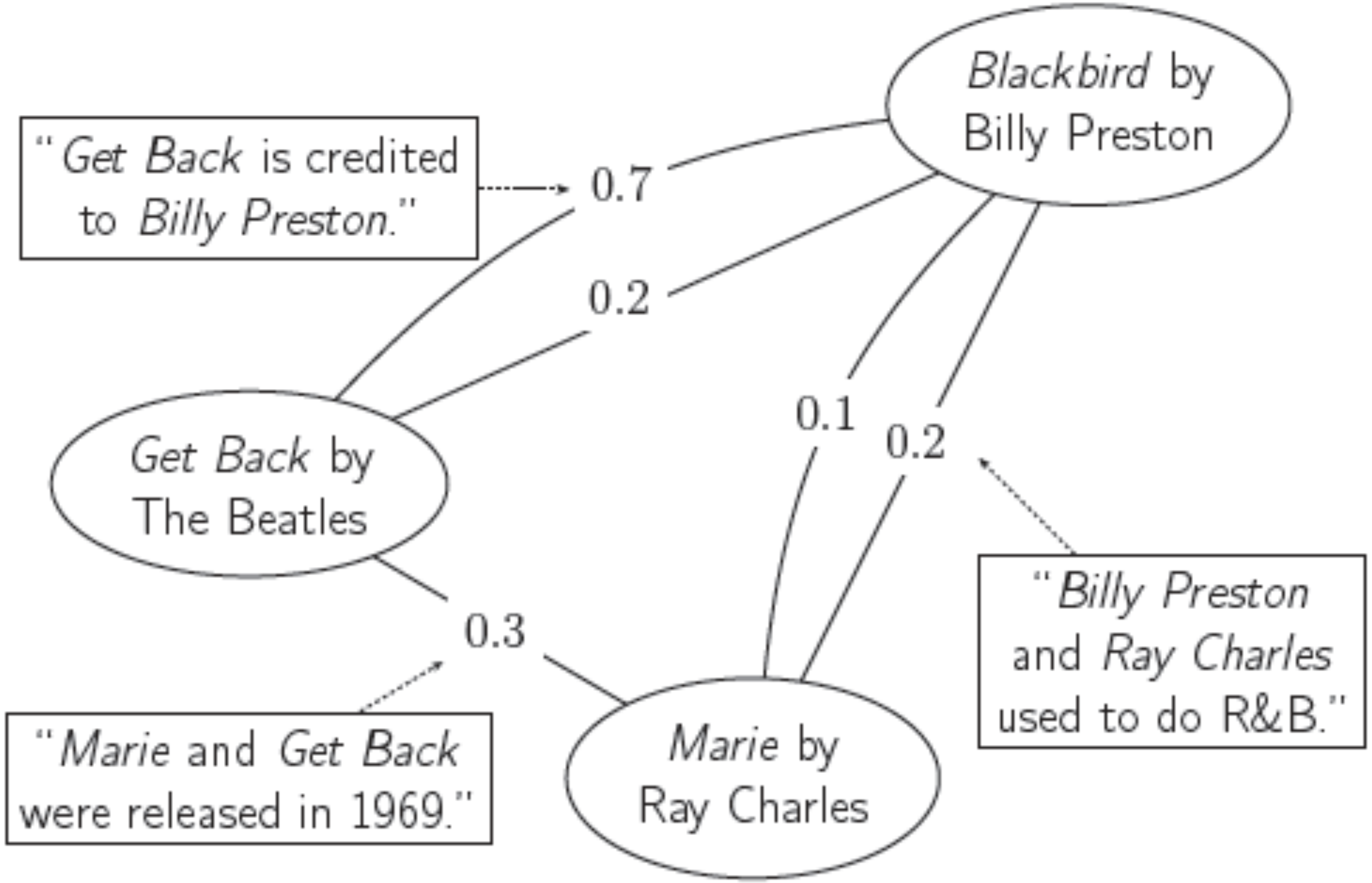
- Dave's **humorous** segues are *worse than* manually-created, curated ones

	likeable	*high-quality*	*sparked-interest*	*funny*	*informative*	*creative*	*understandable*	*well-written*
DAVE	2.94	2.77	2.76	2.71	2.66	3.22	3.58	3.13
THE CHAIN	3.28***	3.14***	2.99*	2.89*	2.90**	3.57***	3.78*	3.35*

Tours: playlists augmented with segues



Sam: a method to generate tours

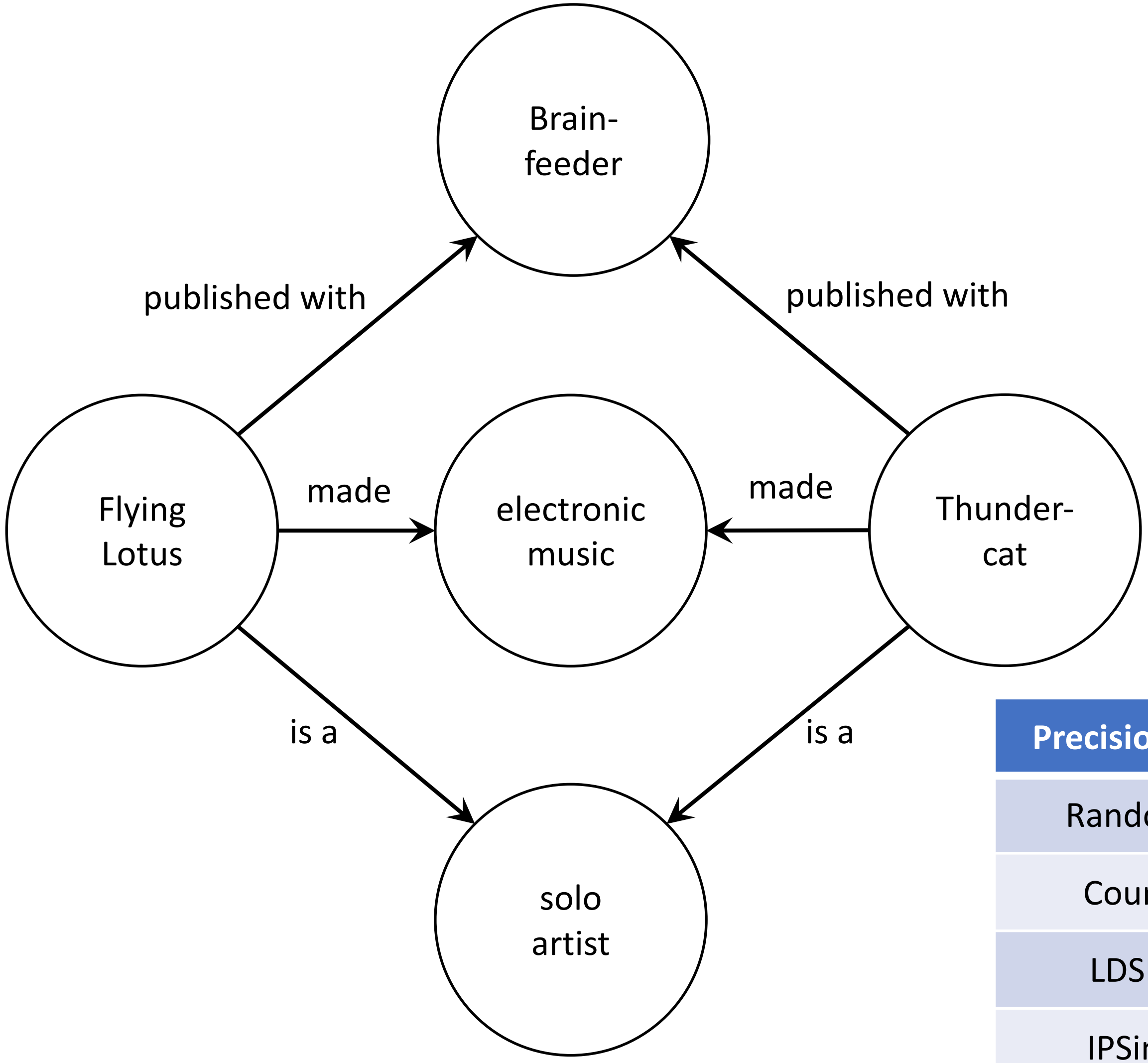


Sam user study (n = 16)

Tours functionality would be welcome for active listening	"I really like it! It is nice to see how all songs are linked together. It is really cool!" "I'd certainly use this … if it was implemented in a music streaming app." "A lot of times you listen to music as a background, whereas I think [for a tour you need] designated hours in your day, like I'm going to sit down and listen to this."
Segue diversity is important	"If I was getting the same information … that would be becoming boring after a while, I … want diversity in information."
Personalisation is needed to avoid familiar segues	"One of the segues was that U2 and Fionn Regan are Irish and … I know that because I'm Irish! … That is just tedious and boring." "[In good tours] you'd have information that you didn't know and you were just learning. Something interesting, like one of those moments \`Oh my God! I didn't know that these two bands were connected'."

Spotify’s My DJ

Node similarity: PathSim & IPSim

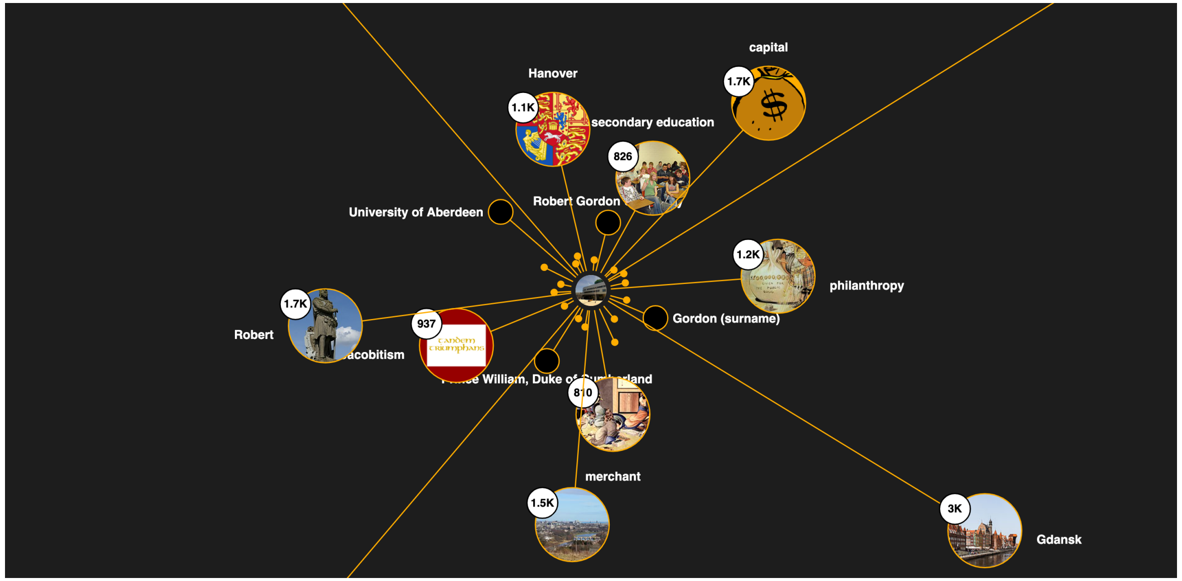


Precision@5	MREX	LastFM(a)	LastFM(b)	Facebook
Random	2.6%	0.2%	0.1%	0.1%
Count	6.2%	3.3%	1.4%	1.6%
LDSD	9.6%	9.5%	**2.2%**	1.9%
IPSim	**13.0%**	**10.3%**	1.7%	**2.3%**

Overview

- Introduction
- Applications
    - Interesting paths
    - Neighbourhood classification
    - Heuristic embeddings
    - Learned embeddings
- Conclusions

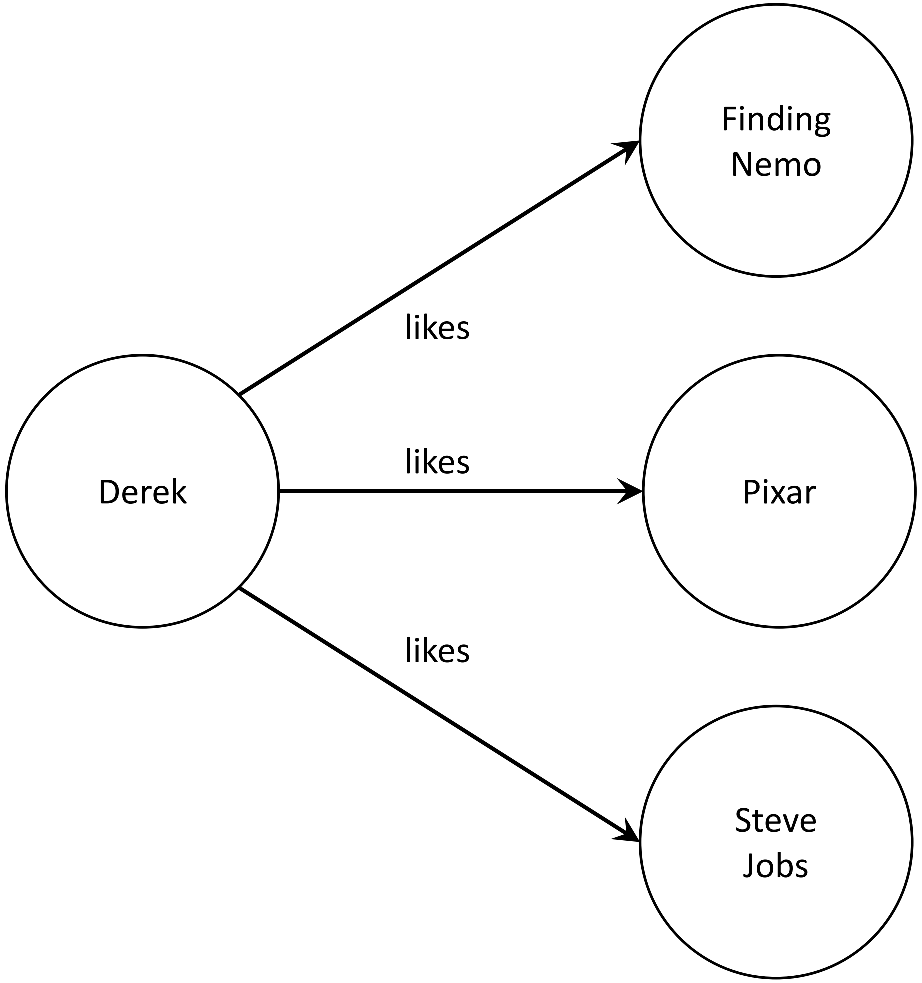
Knowledge graph browser



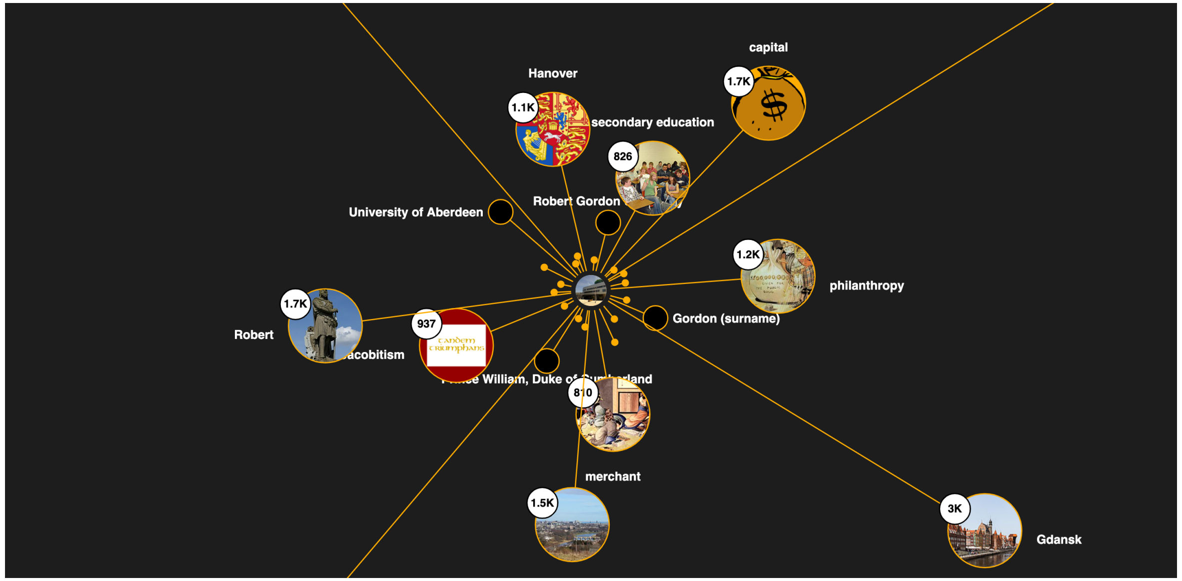
- Browsers support knowledge discovery
- But high connectivity brings choice overload
- We propose to incorporate recommendation techniques

User profiles

- We extend the graph to include user profile facts
- As the user browses, they can give a ‘thumbs-up’
  - inserts a new fact
- From a RecSys point-of-view, these are
  - explicit, unary ratings

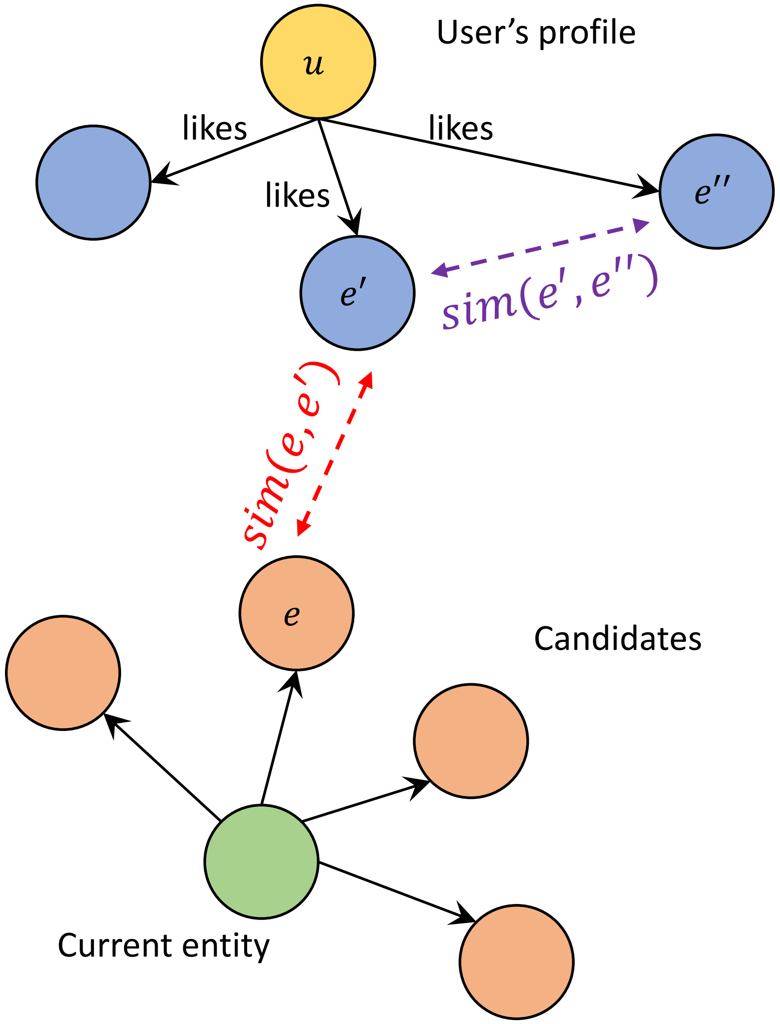


Recommendation as classification



- Candidates for recommendation are entities in the neighbourhood of the current entity
- We must predict which candidates the user will like – this is *classification*

One-class classification



- Let
  - e be a candidate
  - e' be the nearest-neighbour of e that is in the user profile
  - e'' be the nearest-neighbour of e' that is in the user profile

- Predict u likes e if $sim(e, e') > sim(e', e'')$

Case-Based Collective Classification

Luke K. McDowell[1], Kalyan Moy Gupta[2], and David W. Aha[3]

Collective classification

Traditional classification

- Entities are classified in isolation
- However, accuracy can sometimes be improved by taking into account the related entities
    - e.g. faculty web pages often link to postgraduate student pages

Collective classification

- Classify sets of entities together
- When classifying an entity, the classifier can also use the *predicted class* of related entities

Iterative classification algorithm: LDRec

- Repeatedly:
  - Classify the candidates using the one-class classifier
  - For each candidate that the classifier predicts the user will like
    - Temporarily insert a new fact into the user profile

until classification stabilizes

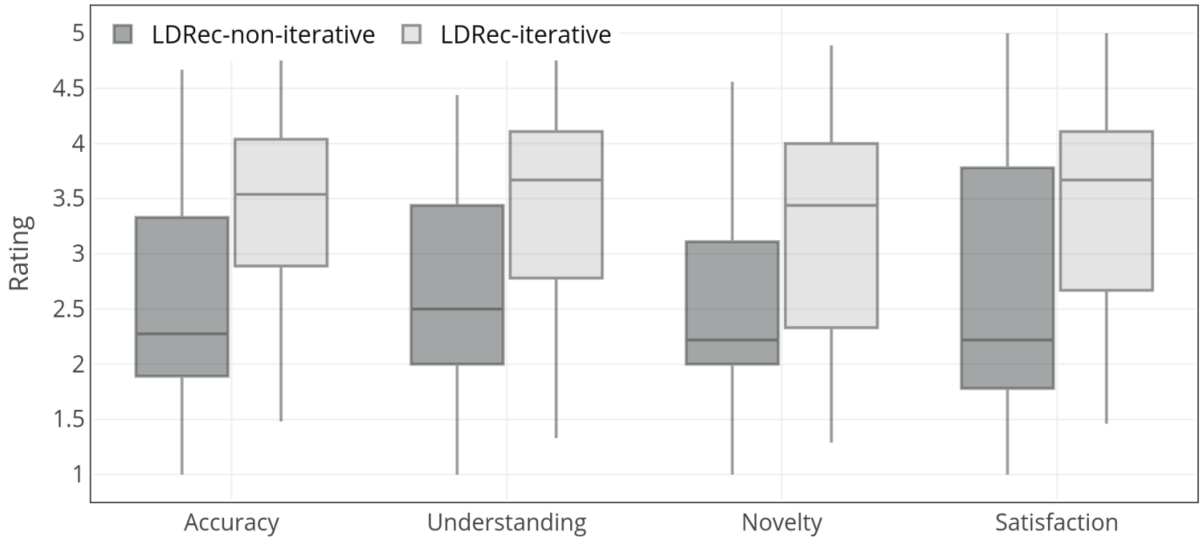
LDRec offline evaluation

- Knowledge graph:
  - DBPedia
- User profiles:
  - Facebook data from the 2015 Linked Open Data-enabled Recommender Systems Challenge
- Method:
  - Koren’s one-plus-random on 100 randomly-chosen users

Precision@N	N=1	N=3	N=5	N=10
LDRec-non-iterative	11	17	25	34
LDRec-iterative	14	31	43	57

LDRec user trial (n = 100)

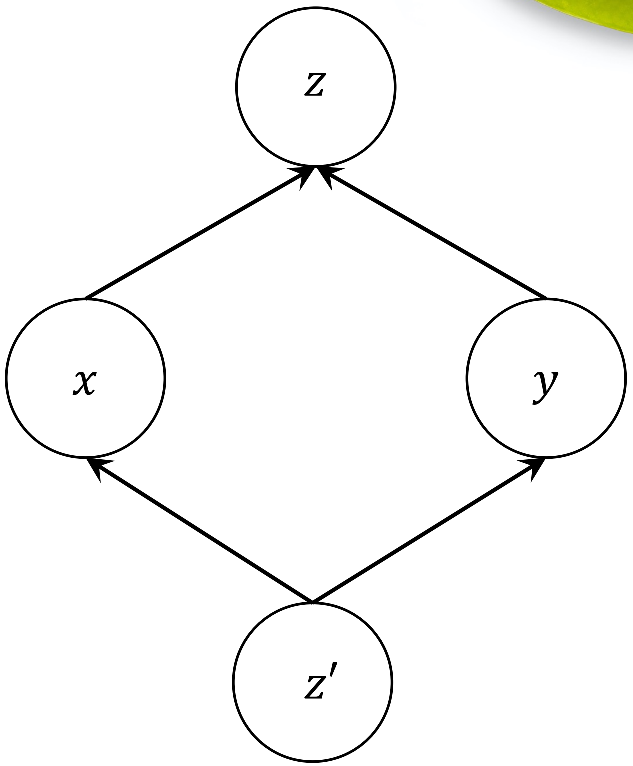
- Knowledge graph:
  - DBPedia
- User profiles:
  - Each user creates a profile of 20 movies, books, musicians
- Method: repeat three times
  - Users choose a seed
  - They receive three recommendations
  - They answer questions

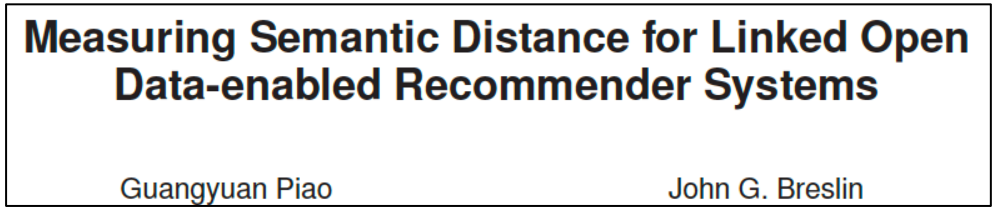


	LDRec-non-iterative	LDRec-iterative
total recs given thumbs-up	133	234
total recs chosen as next seed	52	70

Node similarity: LDSD

- $sim(x, y)$ is an aggregate of
  - number of times x and y both link to the same third entity z
  - number of times a third entity z' links to both x and y





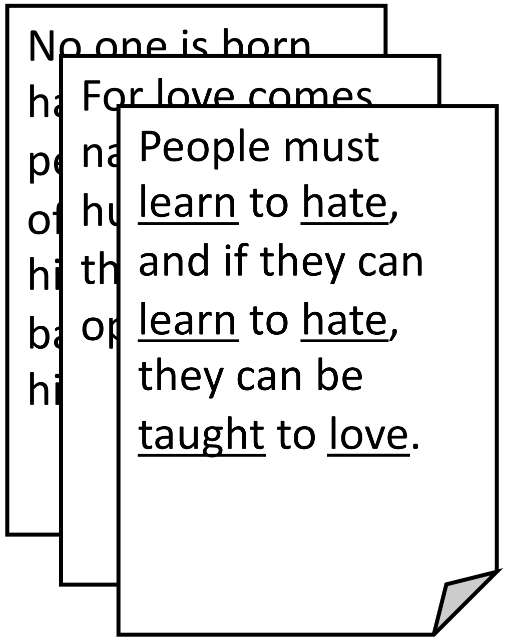
Measuring Semantic Distance for Linked Open Data-enabled Recommender Systems

Guangyuan Piao John G. Breslin

Overview

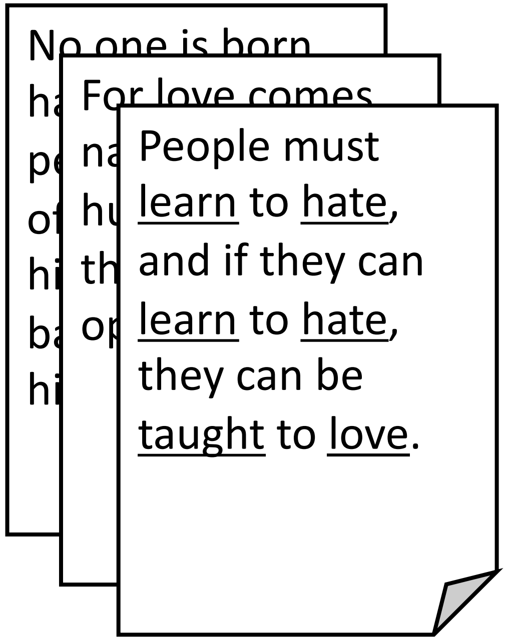
- Introduction
- Applications
    - Interesting paths
    - Neighbourhood classification
    - Heuristic embeddings
    - Learned embeddings
- Conclusions

Vector-space model for documents



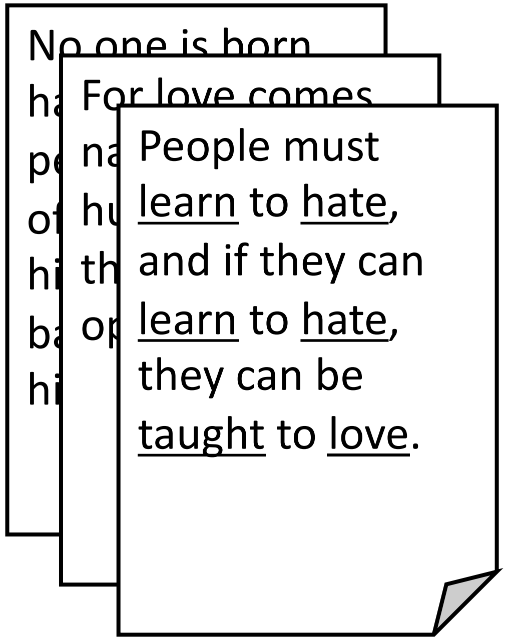
born	hate	learn	love	taught
0	1	1	1	1
0	0	0	1	0
1	1	0	0	0

Vector-space model for documents



born	hate	learn	love	taught
0	2	2	1	1
0	0	0	1	0
1	1	0	0	0

Vector-space model for documents

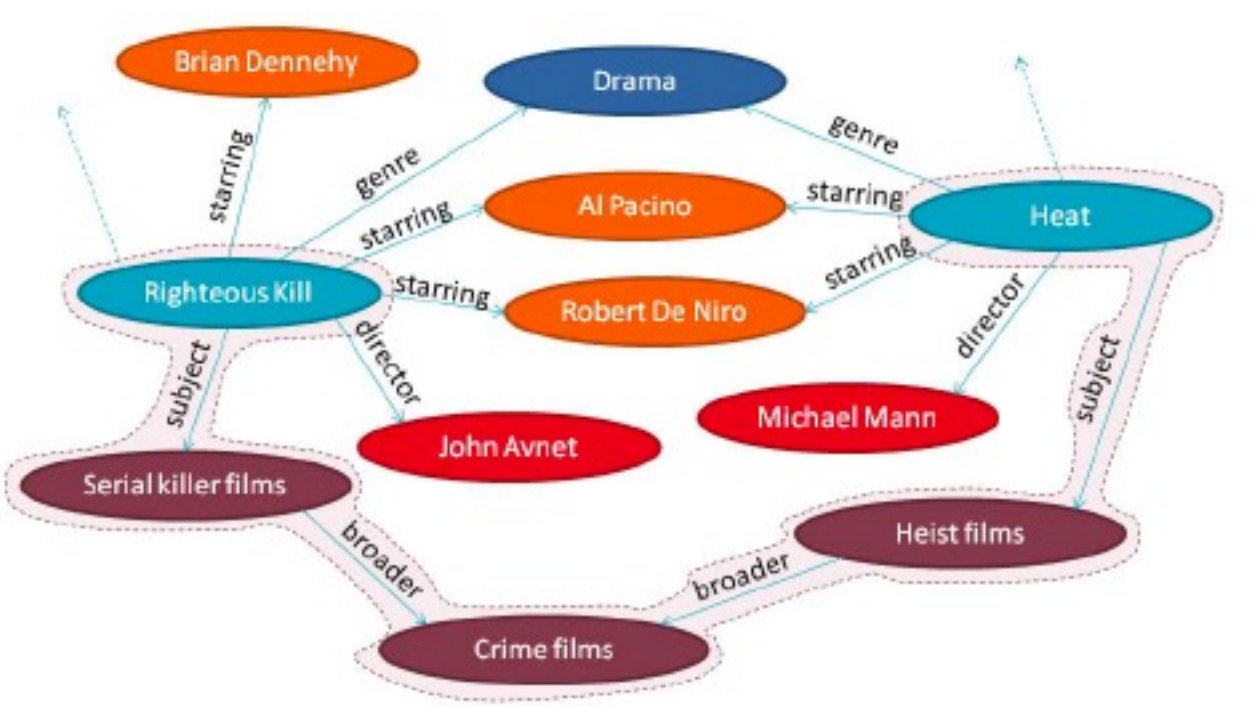


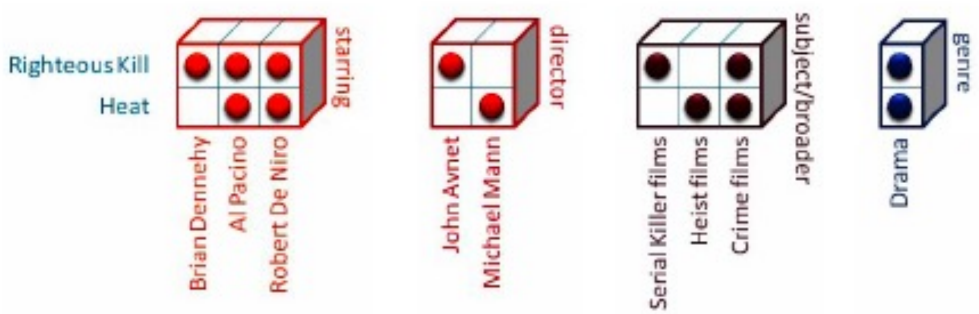
born	hate	learn	love	taught
0	0.6	0.4	0.3	0.4
0	0	0	0.3	0
0.4	0.3	0	0	0

Vector-space model for knowledge graphs

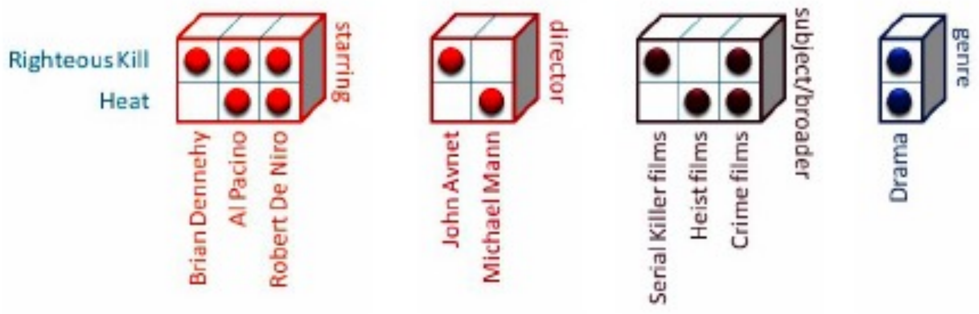
Linked Open Data to support Content-based Recommender Systems

Tommaso Di Noia[1], Roberto Mirizzi[1], Vito Claudio Ostuni[1], Davide Romito[1], Markus Zanker[2*]





Node similarity

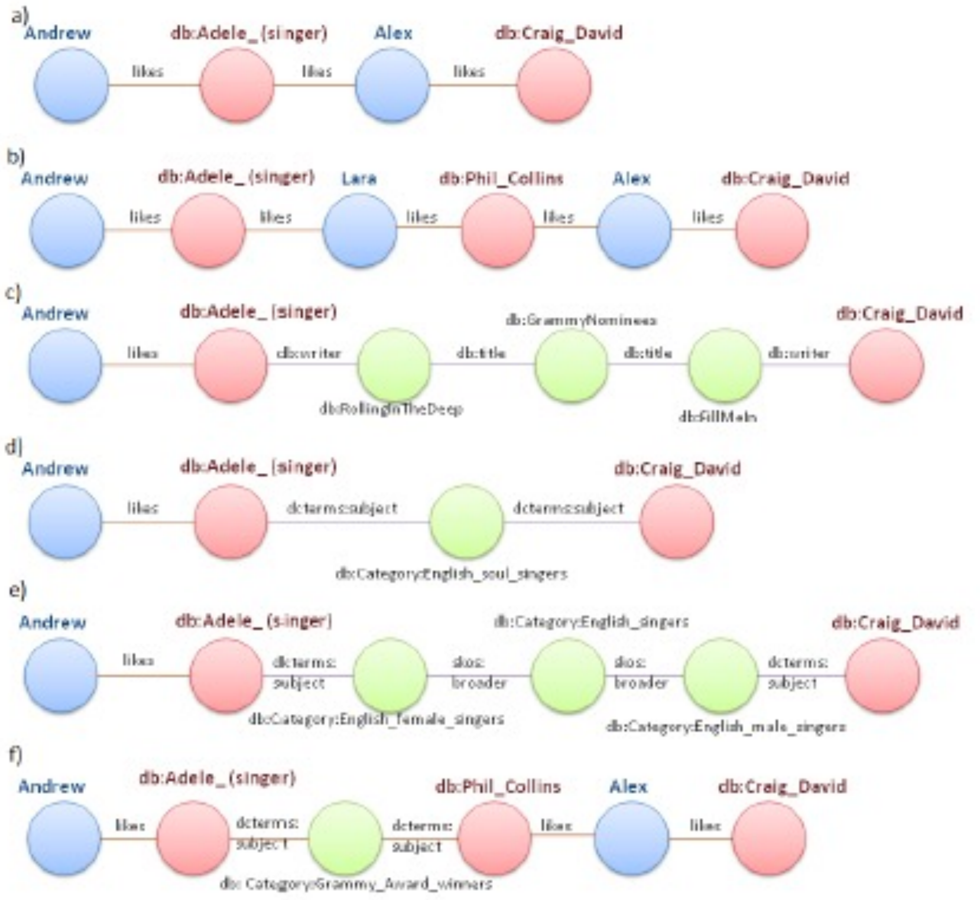


- Local similarity w.r.t. relation *r*
  - cosine
- Global similarity
  - weighted sum

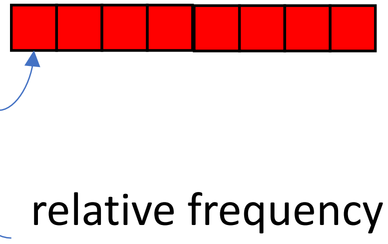
Top-N Recommendations from Implicit Feedback leveraging Linked Open Data

Vito Claudio Ostuni, Tommaso Di Noia, Eugenio Di Sciascio, Roberto Mirizzi

Another vector-space model for knowledge graphs



$x_{\langle Andrew, Craig\ David \rangle}$



relative frequency of paths of a given type

cosine similarity

Top-N Recommendations from Implicit Feedback leveraging Linked Open Data

Vito Claudio Ostuni, Tommaso Di Noia, Eugenio Di Sciascio, Roberto Mirizzi

ExpLOD: a Framework for Explaining Recommendations based on the Linked Open Data Cloud

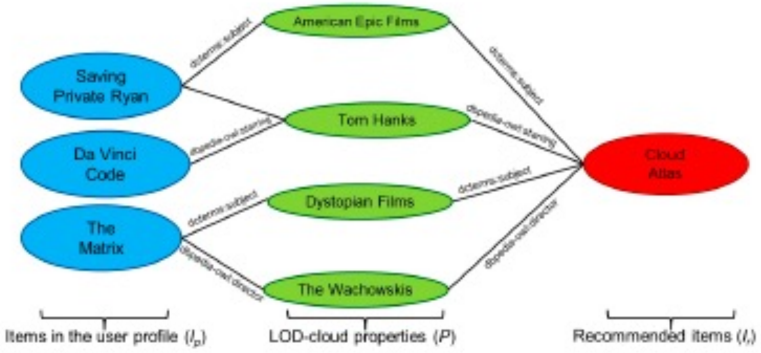
Cataldo Musto Fedelucio Narducci Pasquale Lops
Marco De Gemmis Giovanni Semeraro

Recommender systems

- Accurate hybrid recommenders based on a unified representation of user data, item data and user-item data

	MovieLens			LastFM		
Alg.	r@5	r@10	r@20	r@5	r@10	r@20
given 5						
SPrank	**0.420**	**0.578**	**0.745**	**0.349**	**0.457**	**0.551**
BPRMF	0.353	0.502	0.672	0.213	0.308	0.413
SLIM	0.218	0.363	0.552	0.077	0.152	0.287
BPRLin	0.218	0.314	0.442	0.289	0.381	0.440
SMRMF	0.216	0.354	0.526	0.111	0.181	0.280

- Justifications of recommendation based on connectivity

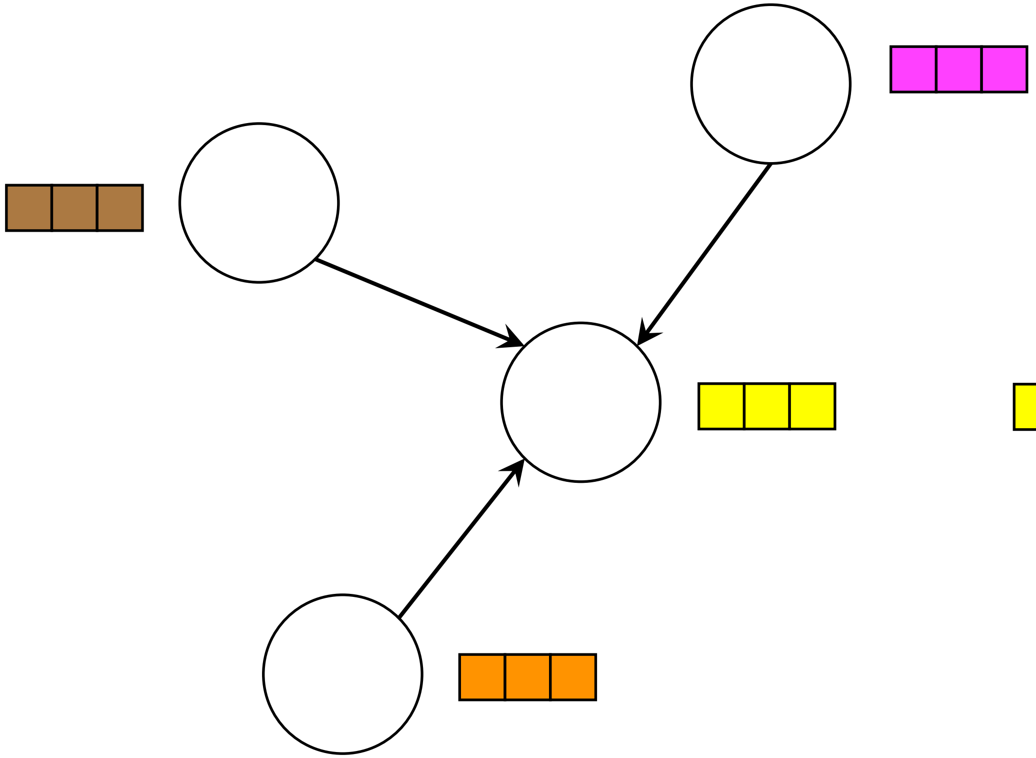


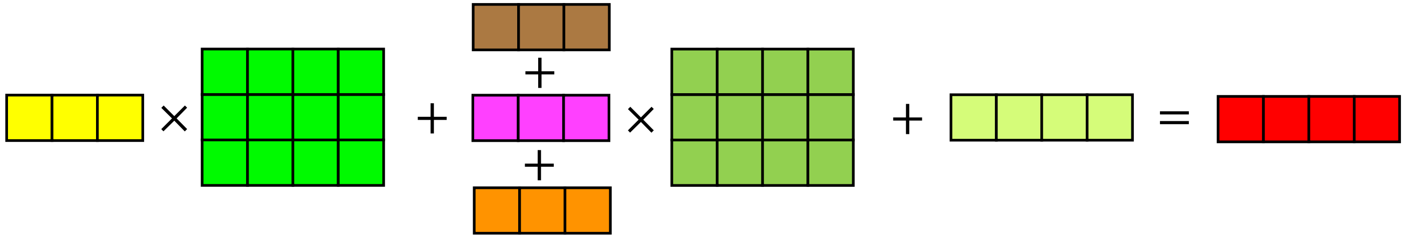
I recommend Cloud Atlas since you often like movies starring Tom Hanks, such as The Da Vinci Code and Saving Private Ryan. Moreover, I recommend it because you sometimes like Dystopian Movies, such as The Matrix, and American Epic Films, such as Saving Private Ryan.

Overview

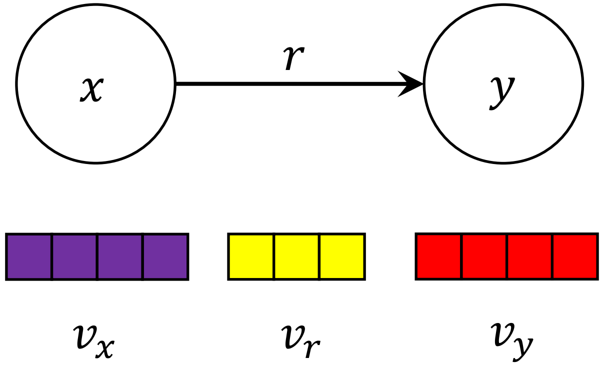
- Introduction
- Applications
  - Interesting paths
  - Neighbourhood classification
  - Heuristic embeddings
  - Learned embeddings
- Conclusions

Graph convolutions




$$\times \quad + \quad \times \quad + \quad =$$

Graph embeddings



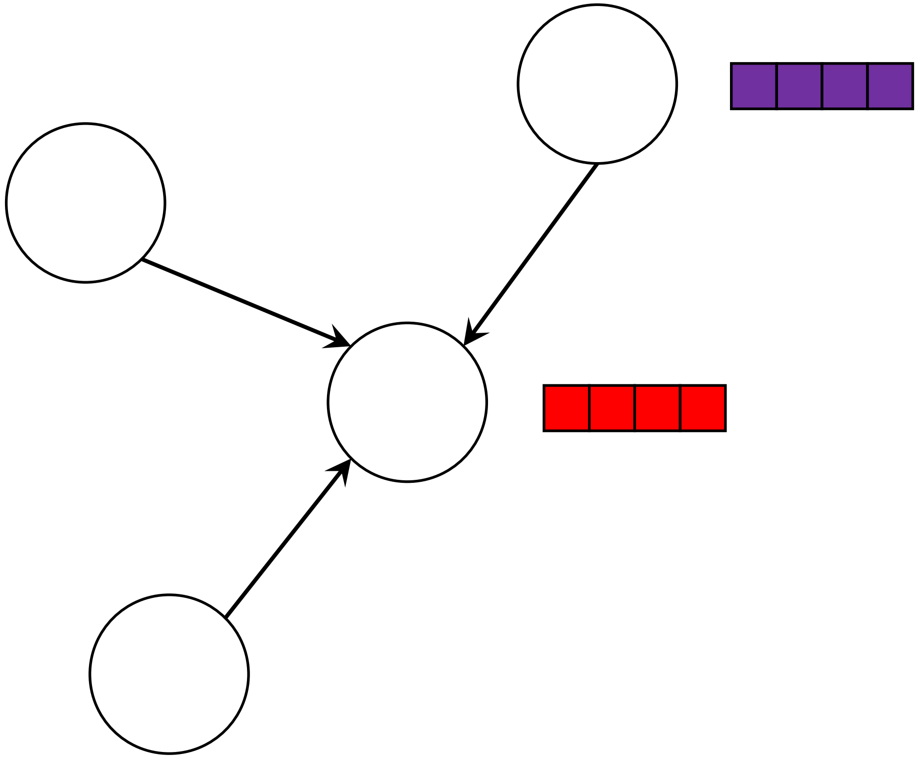
- Let
  - x, r, y be a fact
  - x', r', y' be a non-fact

- We want

$$f\left(v_x, v_r, v_y\right) > f(v_{x'}, v_{r'}, v_{y'})$$

- Different embedding methods (TransE, TransR, TransD,…) differ in their definitions of f.

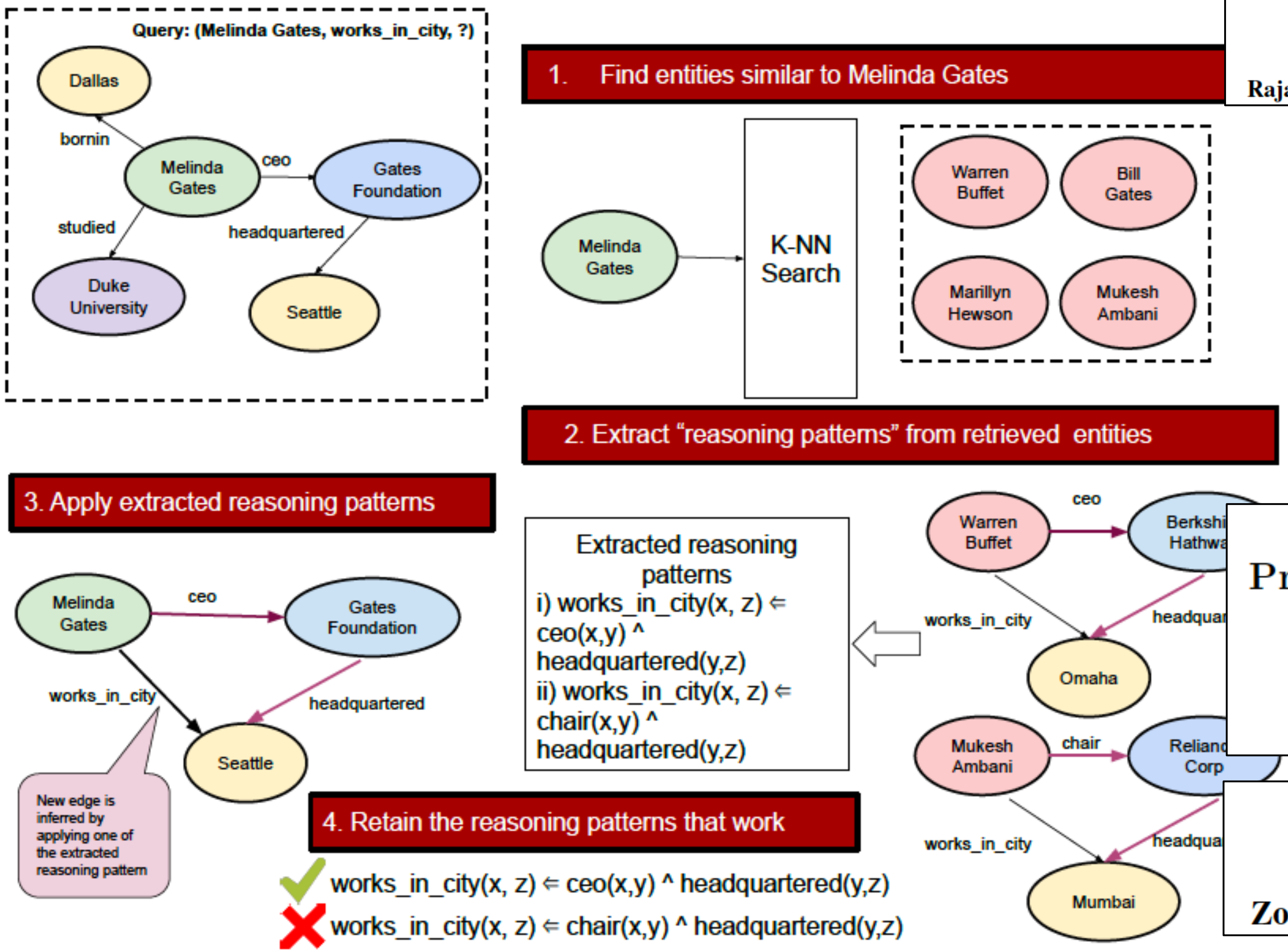
Node similarity



cosine

(or learn the embeddings but regularized by a conventional similarity measure such as PathSim)

Link prediction



A Simple Approach to Case-Based Reasoning in Knowledge Bases

Rajarshi Das[1] RAJARSHI@CS.UMASS.EDU
Ameya Godbole[1] AGODBOLE@CS.UMASS.EDU
Shehzaad Dhuliawala[2] SHEHZAAD.DHULIAWALA@MICROSOFT.COM
Manzil Zaheer[3] MANZILZAHEER@GOOGLE.COM
Andrew McCallum[1] MCCALLUM@CS.UMASS.EDU

Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion

Rajarshi Das, Ameya Godbole, Nicholas Monath, Manzil Zaheer, Andrew McCallum

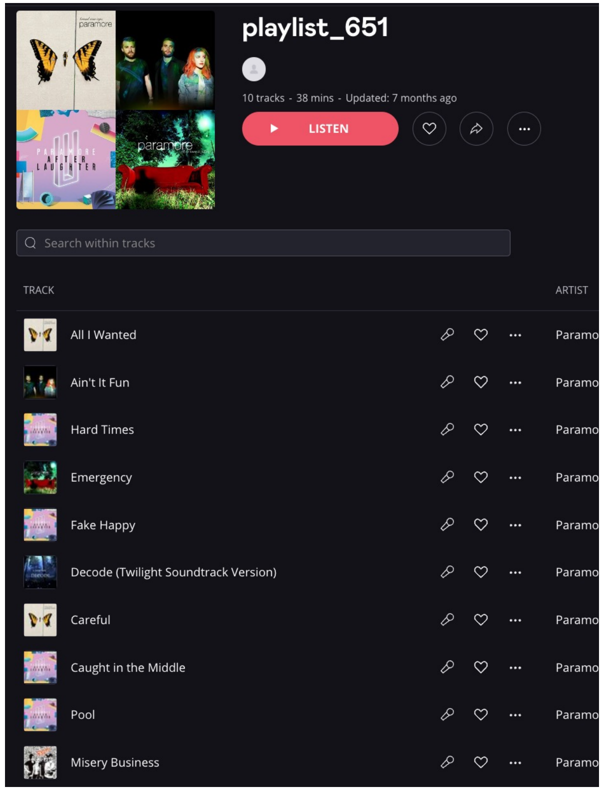
A Contextual Information-augmented Probabilistic Case-based Reasoning Model for Knowledge Graph Reasoning

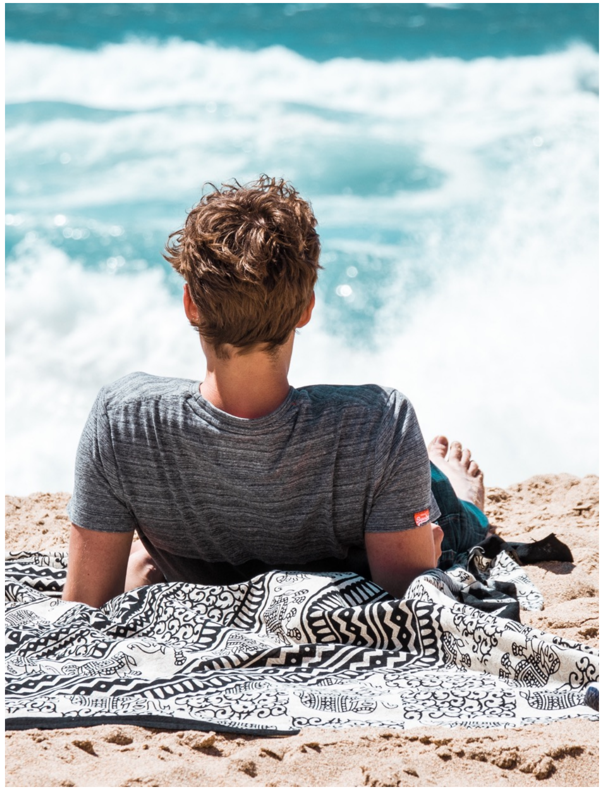
Yuejia Wu and Jian-tao Zhou ✉

End-to-end Case-Based Reasoning for Commonsense Knowledge Base Completion

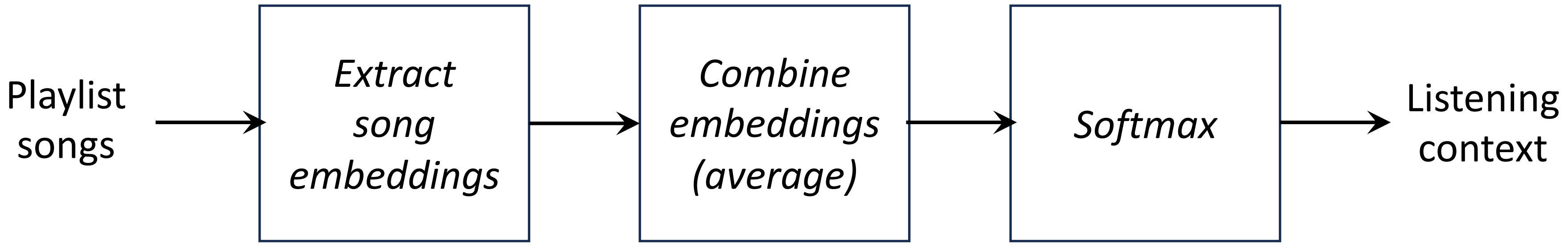
Zonglin Yang♣* Xinya Du♠* Erik Cambria♣ Claire Cardie♦

Predicting playlist listening contexts




Neural network playlist classifier



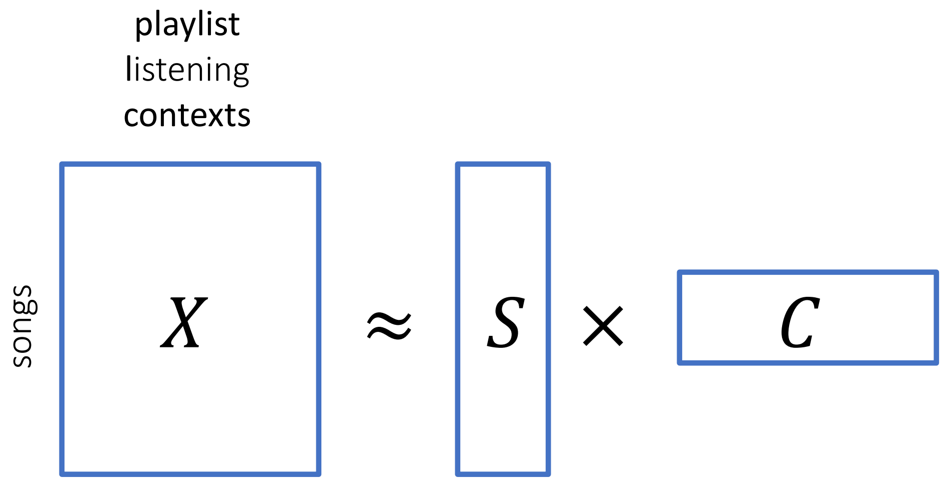
Prediction of User Listening Contexts for Music Playlists

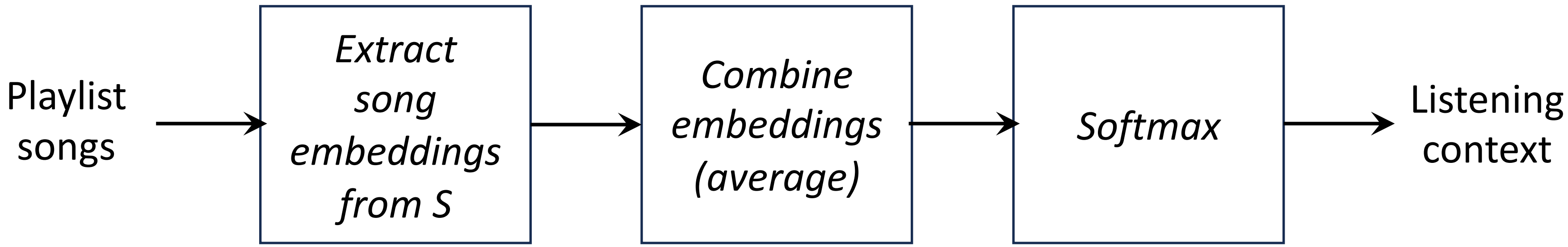
Jeong Choi *
Knowledge AI Lab.
NCSOFT
jchoi@ncsoft.com

Anis Khlif
Deezer Research
research@deezer.com

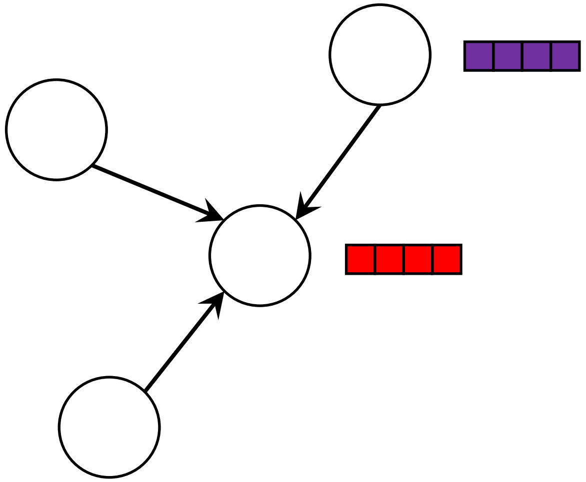
Elena V. Epure
Deezer Research
research@deezer.com

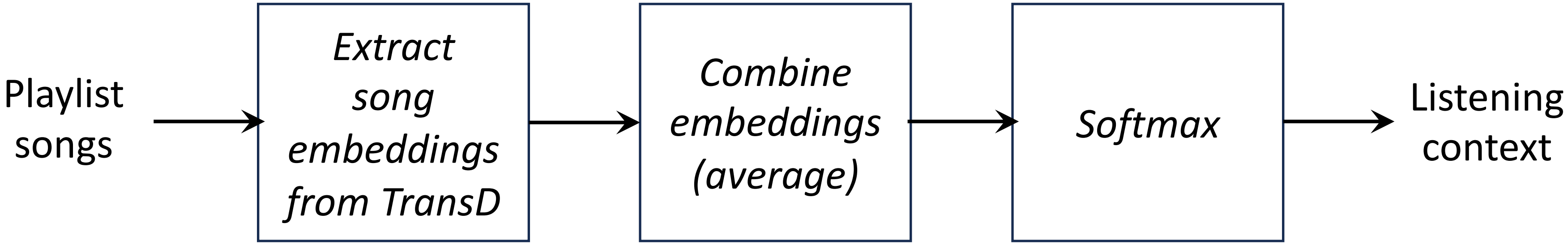
Matrix factorization



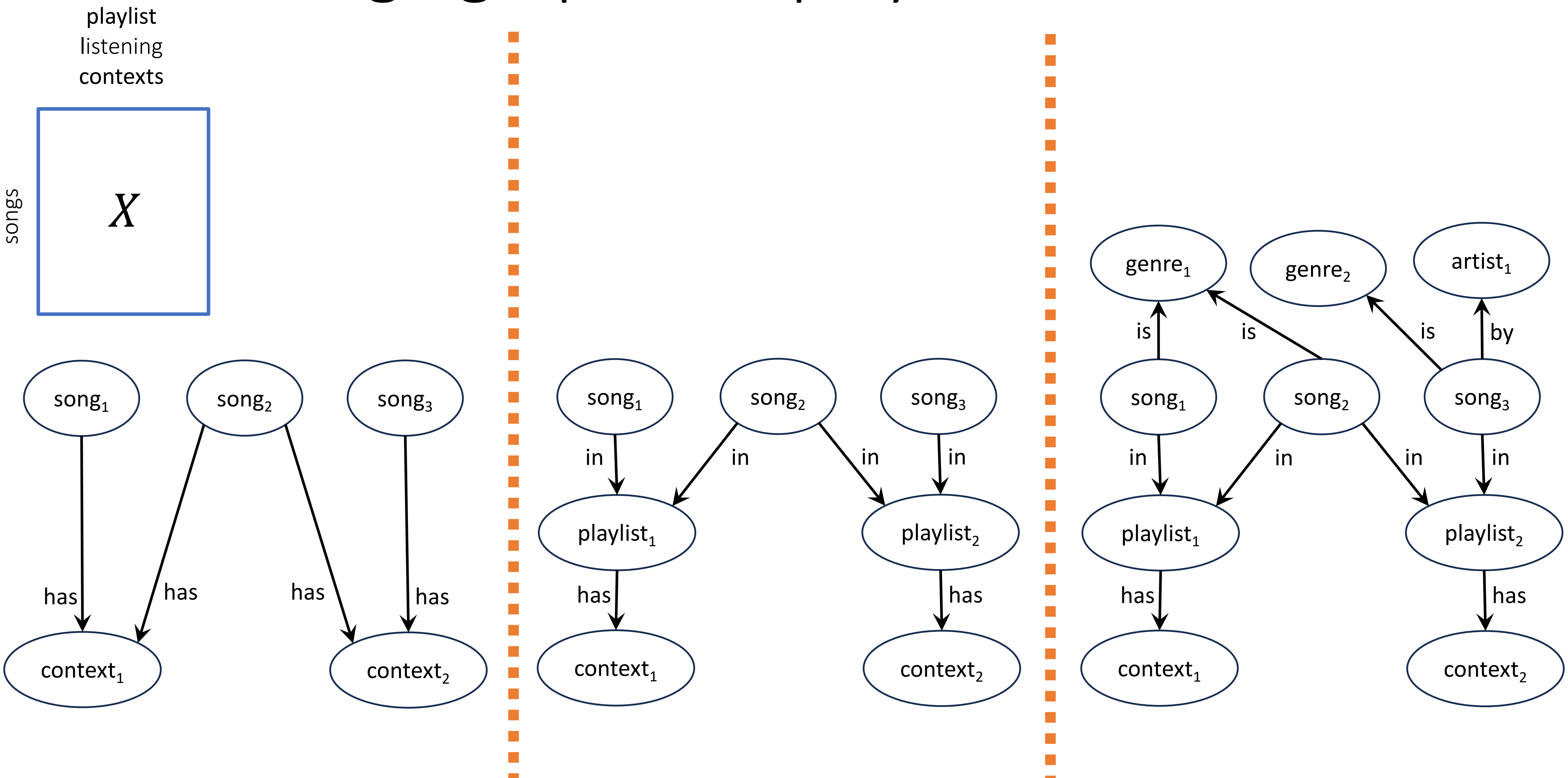


Graph embeddings

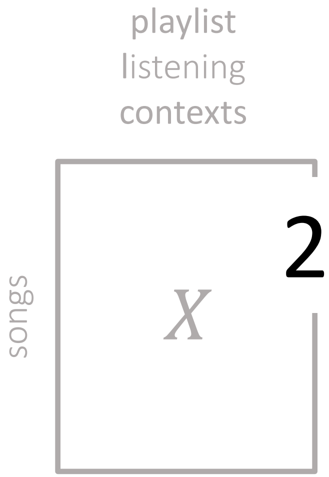




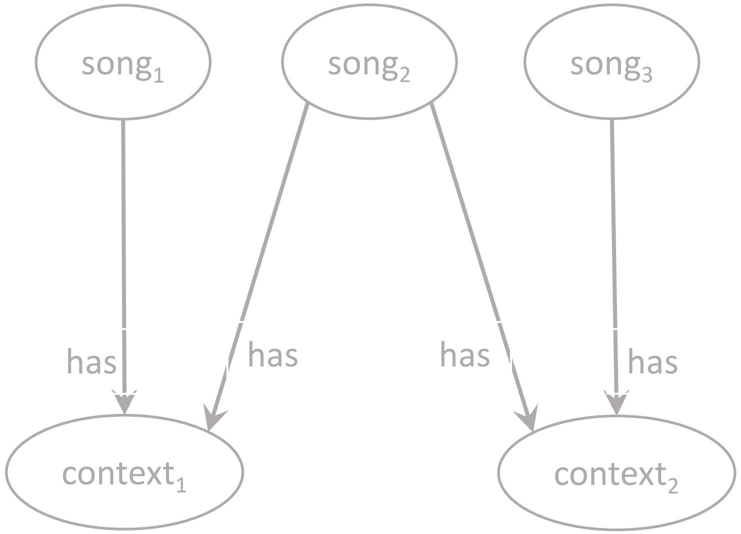
Knowledge graphs for playlists



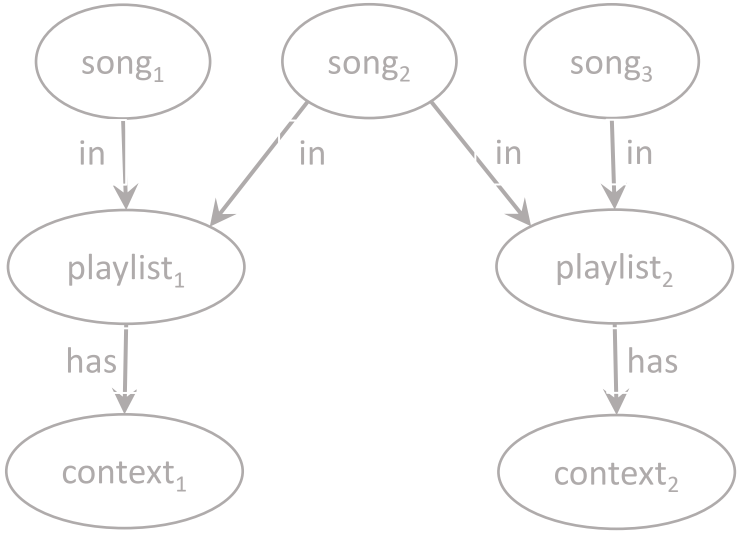
Classification accuracy



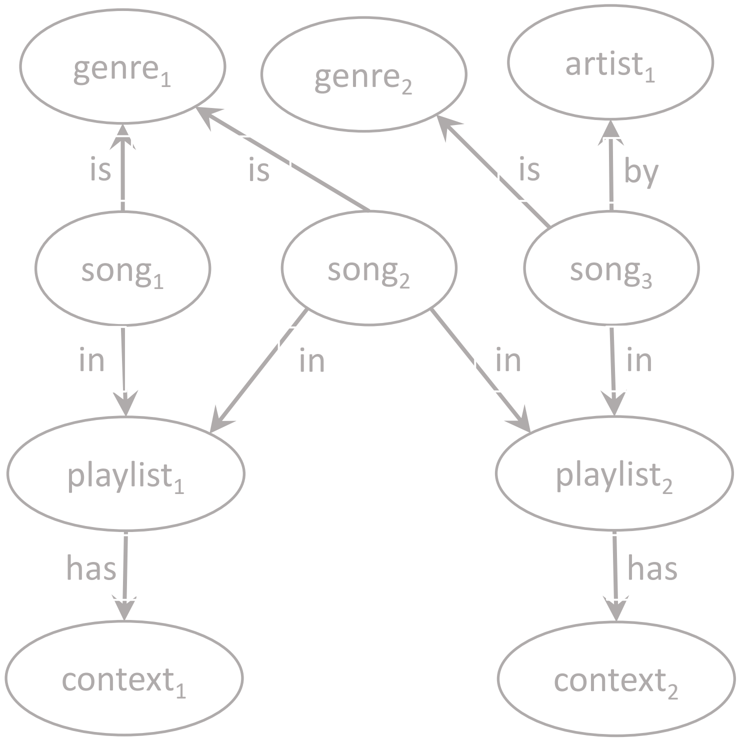
29.9%



37.5%



38.8%



Classification accuracy

Statistic	Value
Number of playlists	114,689
Average playlist length	62.6
Number of unique songs	418,767
Number of unique listening contexts	102

Training 60% - Validation 20% - Testing 20%

Audio	29.1%
Matrix Factorization	29.9%
KG without metadata	37.5%
KG with metadata	38.8%
Hybrid audio + KG	39.5%

Knowledge Graphs

Conclusions

Summary

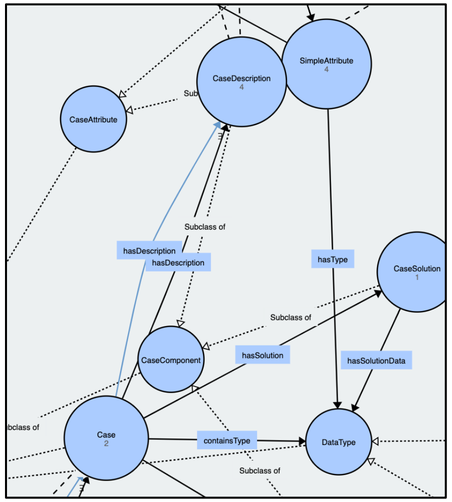
- Introduction
- Applications
    - Interesting paths
    - Neighbourhood classification
    - Heuristic embeddings
    - Learned embeddings

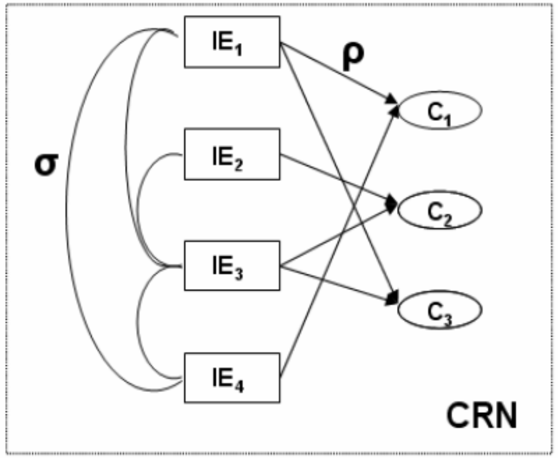
Using Graph Embedding Techniques in Process-Oriented Case-Based Reasoning

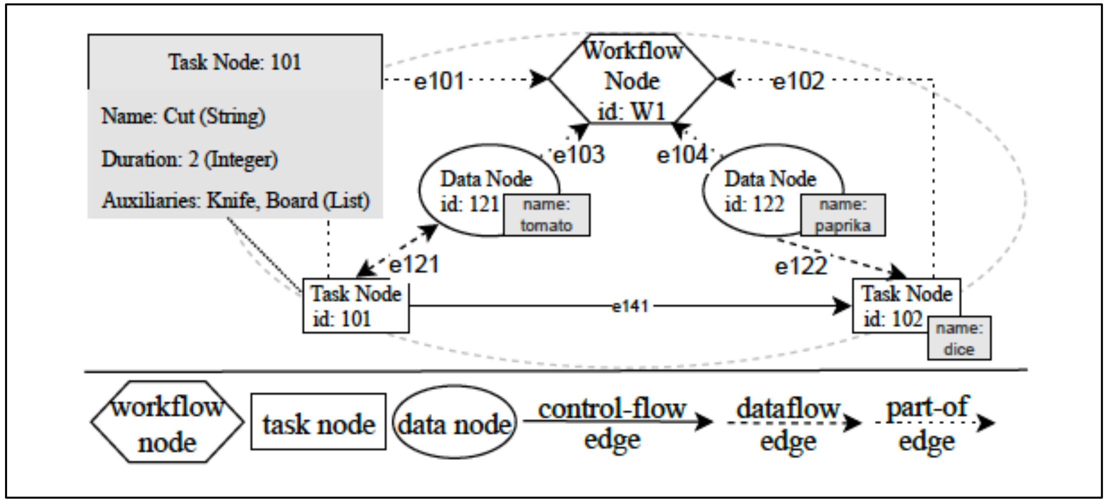
Maximilian Hoffmann [1,2,*] and Ralph Bergmann [1,2]

Knowledge graphs for CBR

- The value of ontologies in CBR is well-recognized
- The value of representing cases as graphs is also well-recognised







- Even graph embeddings have been tried

Knowledge graphs for CBR

- However, the value of knowledge graphs in CBR is under-explored

- Reasons why they may be worth exploring:

  - Knowledge graphs are now readily available and there are tools for constructing new graphs

  - They allow cases to be situated within a wider body of background knowledge

  - As we have seen, they offer many ways to compute similarity

People

Liam de la Cour

Current student @ UCC

Fred Durao

Former post-doc @ UCC

Current lecturer @ Universidad Federal da Bahia, Brazil

Giovanni Gabbolini

Former student @ UCC

Current research scientist @ Apple Music, UK

References

- Saffron https://saffron.insight-centre.org/
  - Paper: Pereira, B., Robin, C., Daudert, T., McCrae, J.P., Mohanty, P., Buitelaar, P. (2019). Taxonomy Extraction for Customer Service Knowledge Base Construction. In: Acosta, M., Cudré-Mauroux, P., Maleshkova, M., Pellegrini, T., Sack, H., Sure-Vetter, Y. (eds) Semantic Systems. The Power of AI and Knowledge Graphs. SEMANTiCS 2019. Lecture Notes in Computer Science(), vol 11702. Springer. https://doi.org/10.1007/978-3-030-33220-4_13
  - Taxonomy comes from this presentation: https://2020-eu.semantics.cc/sites/2020-eu.semantics.cc/files/SEMANTICS2019_presentation_20190911%20copy.pdf
- MusicBrainz https://musicbrainz.org/
- PrimeKG https://zitniklab.hms.harvard.edu/projects/PrimeKG/
  - Paper: Chandak, P., Huang, K. & Zitnik, M. Building a knowledge graph to enable precision medicine. *Sci Data* **10**, 67 (2023). https://doi.org/10.1038/s41597-023-01960-3
  - Code: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IXA7BM
- BabelNet https://babelnet.org/
- Babelfy http://babelfy.org/
  - Paper: Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics*, 2:231–244
- Link prediction using CBR
  - Rajarshi Das, Ameya Godbole, Shehzaad Dhuliawala, Manzil Zaheer and Andrew McCallum: A Simple Approach to Case-Based Reasoning in Knowledge Bases, Automated Knowledge Base Construction, 2020, https://openreview.net/forum?id=AEY9tRqlU7
  - Rajarshi Das, Ameya Godbole, Nicholas Monath, Manzil Zaheer & Andrew McCallum: Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion, Findings of EMNLP, 2020
- Question-answering using CBR
  - Rajarshi Das, Manzil Zaheer, Dung Thai, Ameya Godbole, Ethan Perez, Jay-Yoon Lee, Lizhen Tan, Lazaros Polymenakos & Andrew McCallum: Case-based Reasoning for Natural Language Queries over Knowledge Bases, Procs of EMNLP, 2021

References, cont’d

- Interesting paths
  - Giovanni Gabbolini and Derek Bridge: Generating Interesting Song-to-Song Segues With Dave, Procs. the 29th ACM Conference on User Modeling, Adaptation and Personalization, ACM Press, pp.98-107, 2021
  - Giovanni Gabbolini and Derek Bridge: Play It Again, Sam! Recommending Familiar Music in Fresh Ways, Procs. of the 15th ACM Conference on Recommender Systems, ACM Press, pp.697–-701, 2021
  - Giovanni Gabbolini and Derek Bridge: A User-Centered Investigation of Personal Music Tours, Procs. of the 16th ACM Conference on Recommender Systems, ACM Press, pp.25–-34, 2022
  - Giovanni Gabbolini and Derek Bridge: An interpretable music similarity measure based on path interestingness, Procs. of the 22nd International Society for Music Information Retrieval Conference, pp.213-219, 2021
- Neighbourhod classification
  - Luke K. McDowell||Kalyan Moy Gupta||David W. Aha Case-Based Collective Classification Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2007) (2007)
  - Frederico Araújo Durão and Derek Bridge: A Linked Data Browser with Recommendations, Procs. of the Thirtieth IEEE International Conference on Tools with Artificial Intelligence, 2018
  - Guangyuan Piao and John G. Breslin. 2016. Measuring semantic distance for linked open data-enabled recommender systems. In Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16). Association for Computing Machinery, New York, NY, USA, 315–320. https://doi.org/10.1145/2851613.2851839
- Heuristic embeddings
  - Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, and Markus Zanker. 2012. Linked open data to support content-based recommender systems. In Proceedings of the 8th International Conference on Semantic Systems (I-SEMANTICS '12). Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/2362499.2362501
  - Vito Claudio Ostuni, Tommaso Di Noia, Eugenio Di Sciascio, and Roberto Mirizzi. 2013. Top-N recommendations from implicit feedback leveraging linked open data. In Proceedings of the 7th ACM conference on Recommender systems (RecSys '13). Association for Computing Machinery, New York, NY, USA, 85–92. https://doi.org/10.1145/2507157.2507172

References, cont’d

- Heuristic embeddings, cont’d
  - Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2016. ExpLOD: A Framework for Explaining Recommendations based on the Linked Open Data Cloud. In Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16). Association for Computing Machinery, New York, NY, USA, 151–154. https://doi.org/10.1145/2959100.2959173
- Learned embeddings
  - Image of graph convolution inspired by Figure 18.9 in Sebastian Raschka: Machine Learning with PyTorch and Scikit-Learn, Packt Publishing, 2022
  - Yuejia Wu and Jian-Tao Zhou: A Contextual Information-augmented Probabilistic Case-based Reasoning Model for Knowledge Graph Reasoning, Procs of ICCBR, 2023
  - Jeong Choi, Anis Khlif, and Elena Epure. 2020. Prediction of user listening contexts for music playlists. In Proceedings of the 1st Workshop on NLP for Music and Audio (NLP4MusA), pages 23–27, Online. Association for Computational Linguistics
  - Giovanni Gabbolini and Derek Bridge: Predicting the Listening Contexts of Music Playlists Using Knowledge Graphs, Procs. of the European Conference on Information Retrieval, pp.330-345, 2023
- Conclusions
  - Image of CBROnto taken from https://gaia.fdi.ucm.es/ontologies/
  - Image of case retrieval net from Chakraborti, S., Lothian, R., Wiratunga, N., Orecchioni, A., Watt, S.: Fast Case Retrieval Nets for Textual Data. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds) Advances in Case-Based Reasoning. ECCBR 2006. Lecture Notes in Computer Science(), vol 4106. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11805816_30

References, cont'd

- Conclusions, cont'd
  - Image of NEST workflow from Alexander Schultheis, Maximilian Hoffmann, Lukas Malburg and Ralph Bergmann: Explanation of Similarities in Process-Oriented Case-Based Reasoning by Visualization, Procs of ICCBR, 2023
  - Hoffmann M, Bergmann R. Using Graph Embedding Techniques in Process-Oriented Case-Based Reasoning. Algorithms. 2022; 15(2):27. https://doi.org/10.3390/a15020027
- Peas in a pod
  - Photo by Mockup Graphics on Unsplash

Summary

- Introduction
- Applications
  - Interesting paths
  - Neighbourhood classification
  - Heuristic embeddings
  - Learned embeddings