From Knowledge Graphs to Semantic Spaces

Knowledge Transfer Seminar

Agenda

Knowledge Graphs and Semantic Spaces

A primer to embeddings

Graph Neural Network experiments using PyTorch Geometric

A primer to the SentenceTransformers framework

SentenceTransformers experiments

Problem definition

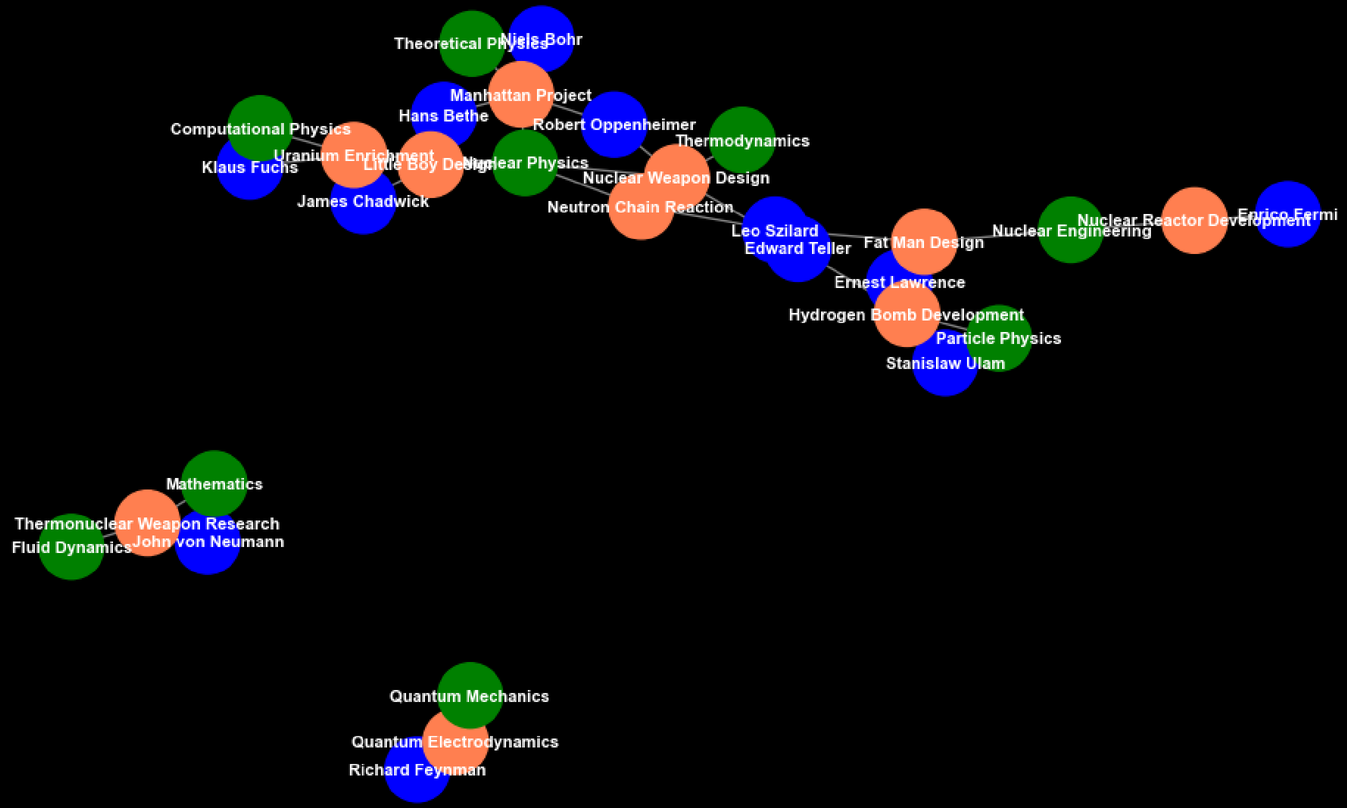
Case: „Subjects verbing objects” //
„Dolgok izélnek bigyókat”

Q: How to relate different entities in such cases? "A good sci-fi director"

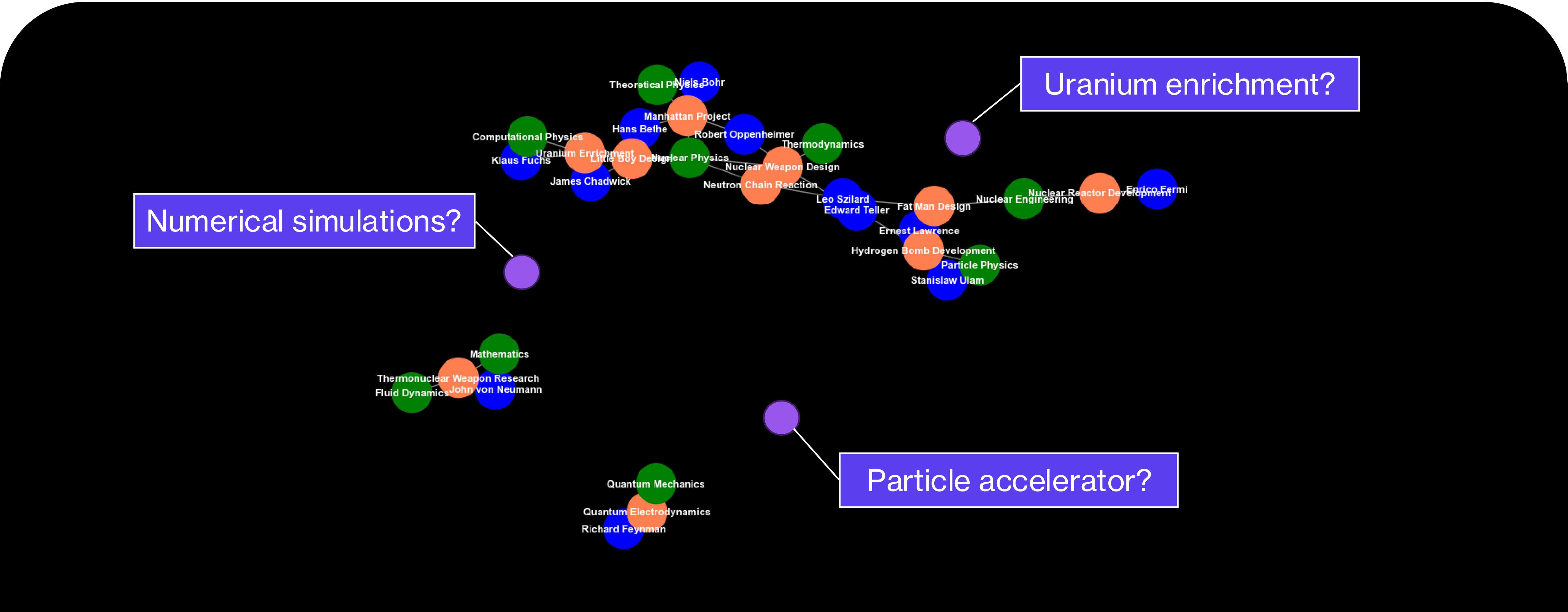
Example

+ Explicit: LinkedIn: People – Skill direct assignment
+ Implicit: University Activities: People – Project + Role description

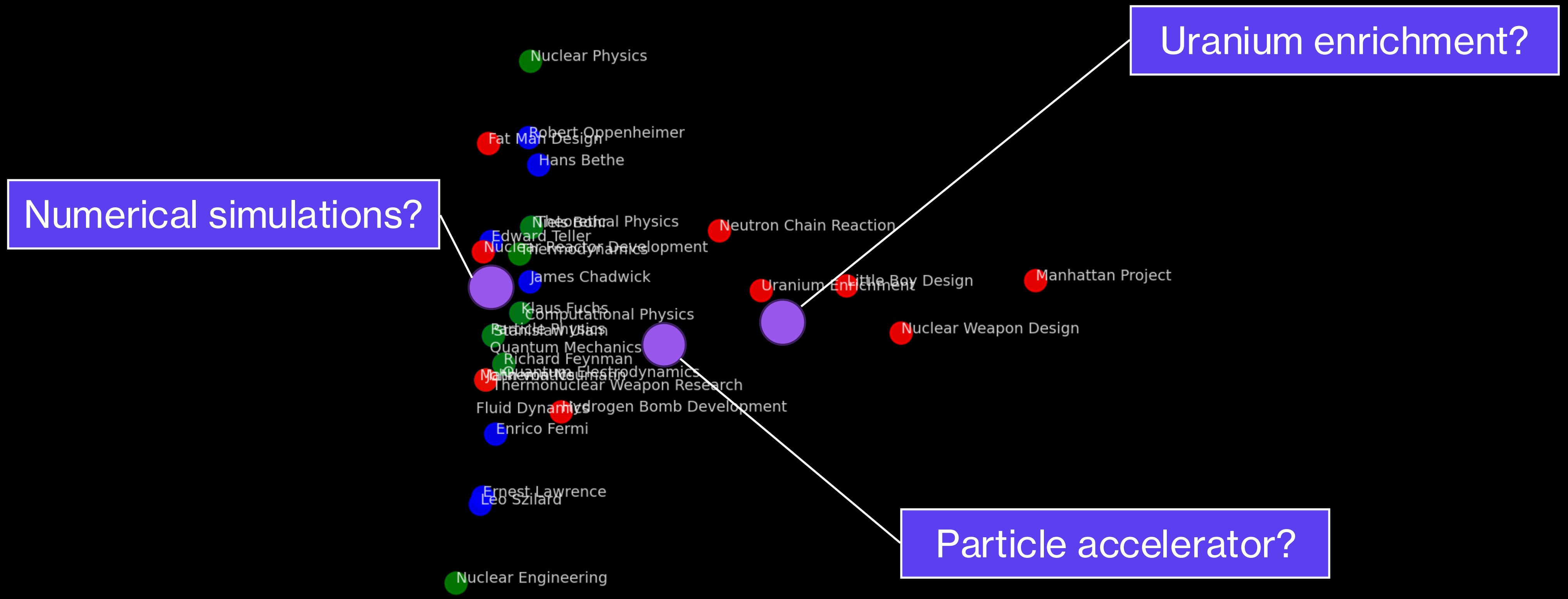
What we can easily have: A vast knowledge graph



What we actually want: A searchable database

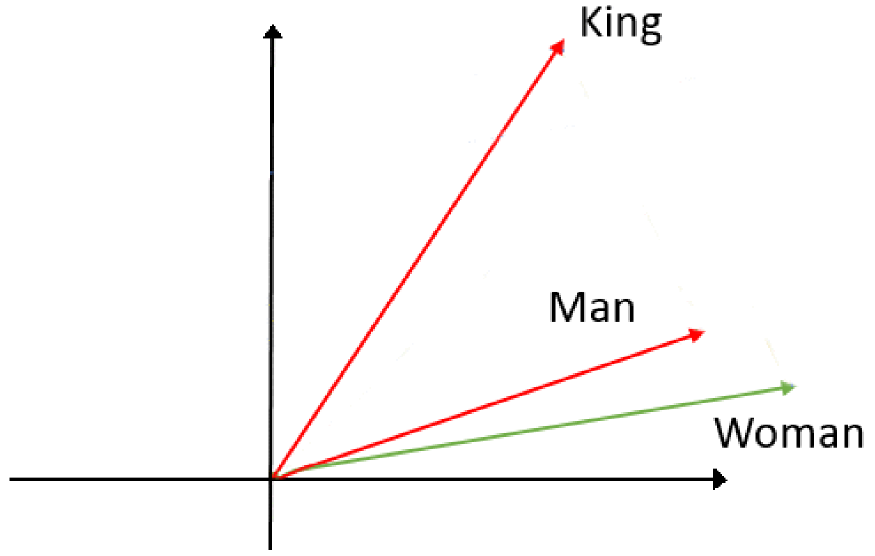


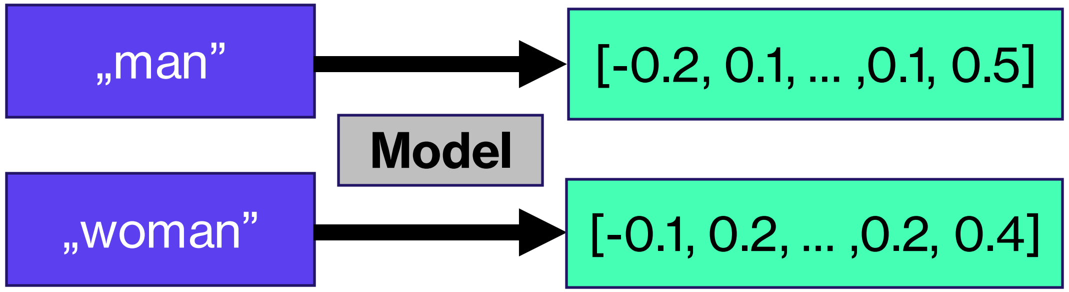
What we possibly need:
An embedded semantic space



A world of words: Primer to embeddings

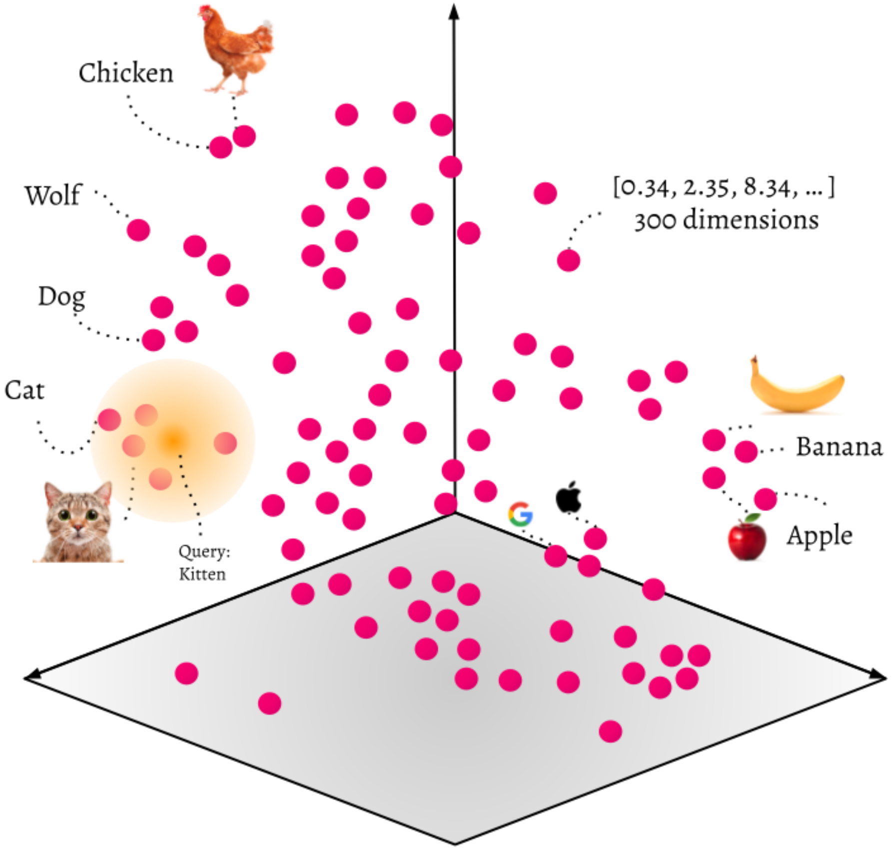
- Text-to-numerical data (n-dimension)
- Fixed-size vector representations of „words”
- Continuus space
- Captures (encodes) semantics: meanings, relationships, context
- Classification, search, translation etc.





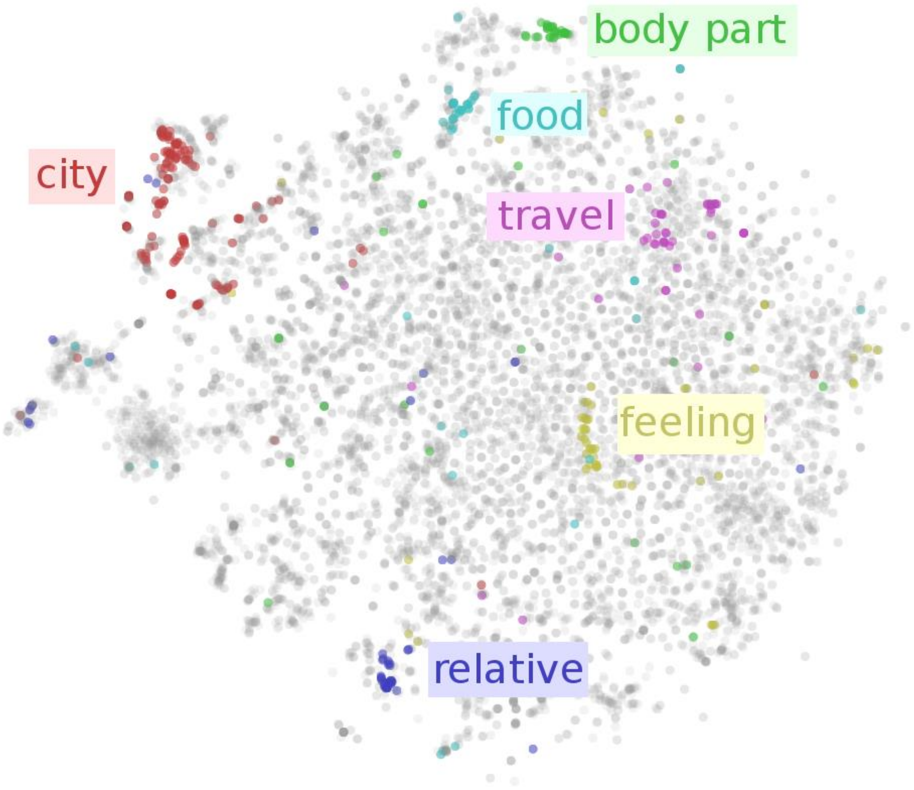
A world of words: Primer to embeddings

- Representation of statistical relationships between **tokens**
- Similarity: „distance” (cosine, dot product)
- Arbitrary dimensions

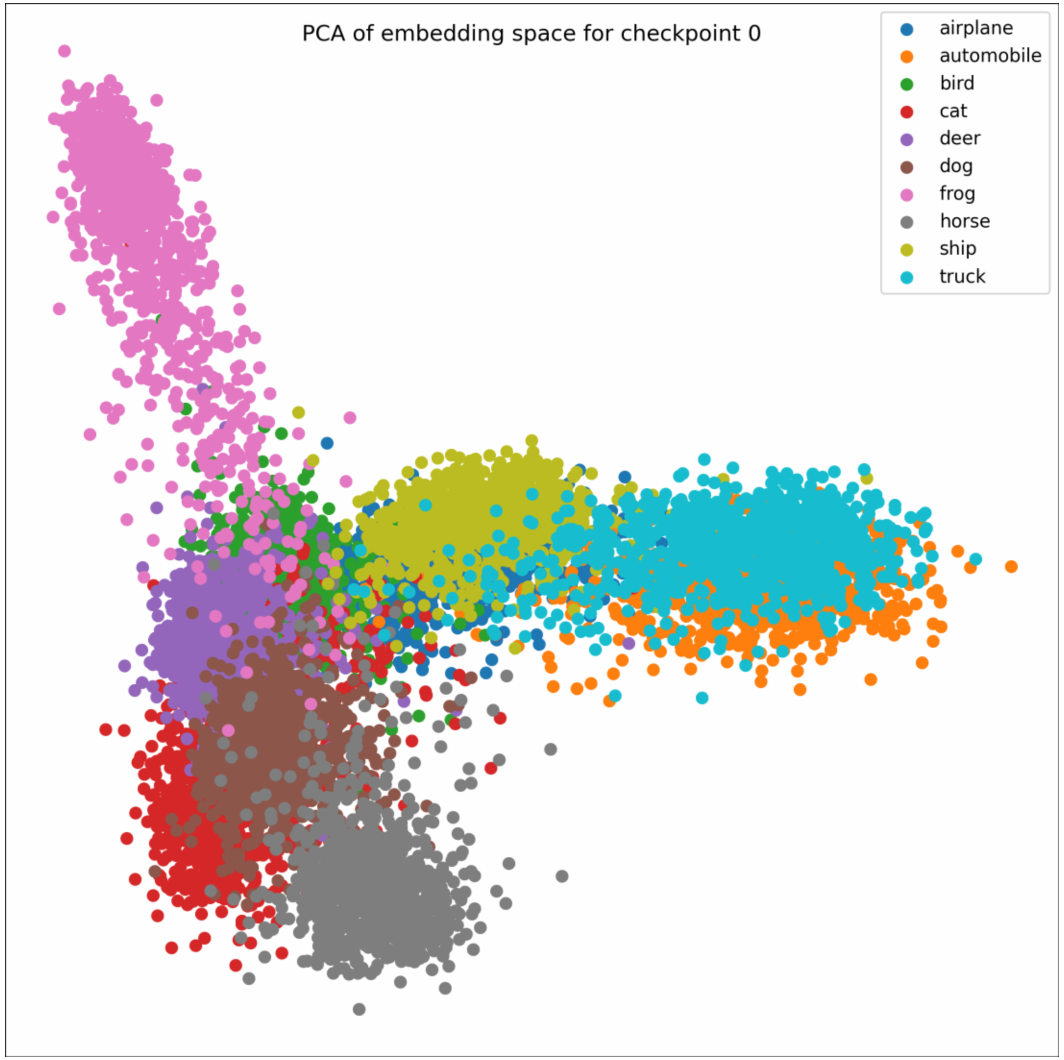


A world of words*: Primer to embeddings

- *Token representation
- Co-occurence based
  + Direct (location)
  + Contextual (meaning)
- Based on/Biased by training data
  + Written human language(s)
  + Biological data
  + Images
  + Sounds
  + Multi-modal



A world of words: Primer to embeddings



Finite embeddings

Finite embeddings

Count-based

- TF-IDF
  - + Term Frequency-Inverse Document Frequency
- GloVe
  - + Word co-occurence ~ Word features

Prediction-based

- Word2Vec
  - + CBOW: surrounding → word
  - + Skip-gram: word → surrounding
- FastText
  - + n-grams*

PRO: simple to implement, captures global statistics, lower-dimensional representation
CON: requires full re-training, limited context-awareness, limited vocabulary (OOV)*

*Infinite embeddings

English

Tokens 105 **Characters** 468

One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed into a horrible vermin. He lay on his armour-like back, and if he lifted his head a little he could see his brown belly, slightly domed and divided by arches into stiff sections. The bedding was hardly able to cover it and seemed ready to slide off any moment. His many legs, pitifully thin compared with the size of the rest of him, waved about helplessly as he looked.

Norwegian

Tokens 157 **Characters** 465

En morgen, da Gregor Samsa våknet fra urolige drømmer, fant han seg forvandlet i sengen sin til et fryktelig skadedyr. Han lå på sin panserlignende rygg, og hvis han løftet litt på hodet kunne han se den brune magen hans, lett hvelvet og delt av buer i stive partier. Sengetøyet klarte så vidt å dekke ham og virket klar til å gli av når som helst. De mange bena hans, ynkelig tynne sammenlignet med størrelsen på resten av ham, vinket hjelpeløst mens han så ut.

Arabic

Tokens 286 **Characters** 402

في صباح أحد الأيام، عندما استيقظ جريجور سامسا
من أحلام مزعجة، وجد نفسه قد تحول في سريره إلى
حشرة رهيبة. كان يرقد على ��هره الذي يشبه الد
رع، وإذا رفع رأسه قليلا�� كان بإمكانه ر��ية بطنه
البني، المقبب قليلا�� والمقسم بواسطة أقواس إ
لى أقسام صلبة. كان الفراش بالكاد يغطيه، وبدا
أنه مستعد للانزلاق في أي لحظة. كانت أرجله الع
ديدة، الرفيعة بشكل مثير للشفقة مقارنة بحجم
بقية جسده، تلوح بلا حول ولا قوة وهو ينظر.

TEXT TOKEN IDS

*Infinite embeddings

*Infinite embeddings

Character-level

- FastText
- CharCNN/RNN

Transformer-based

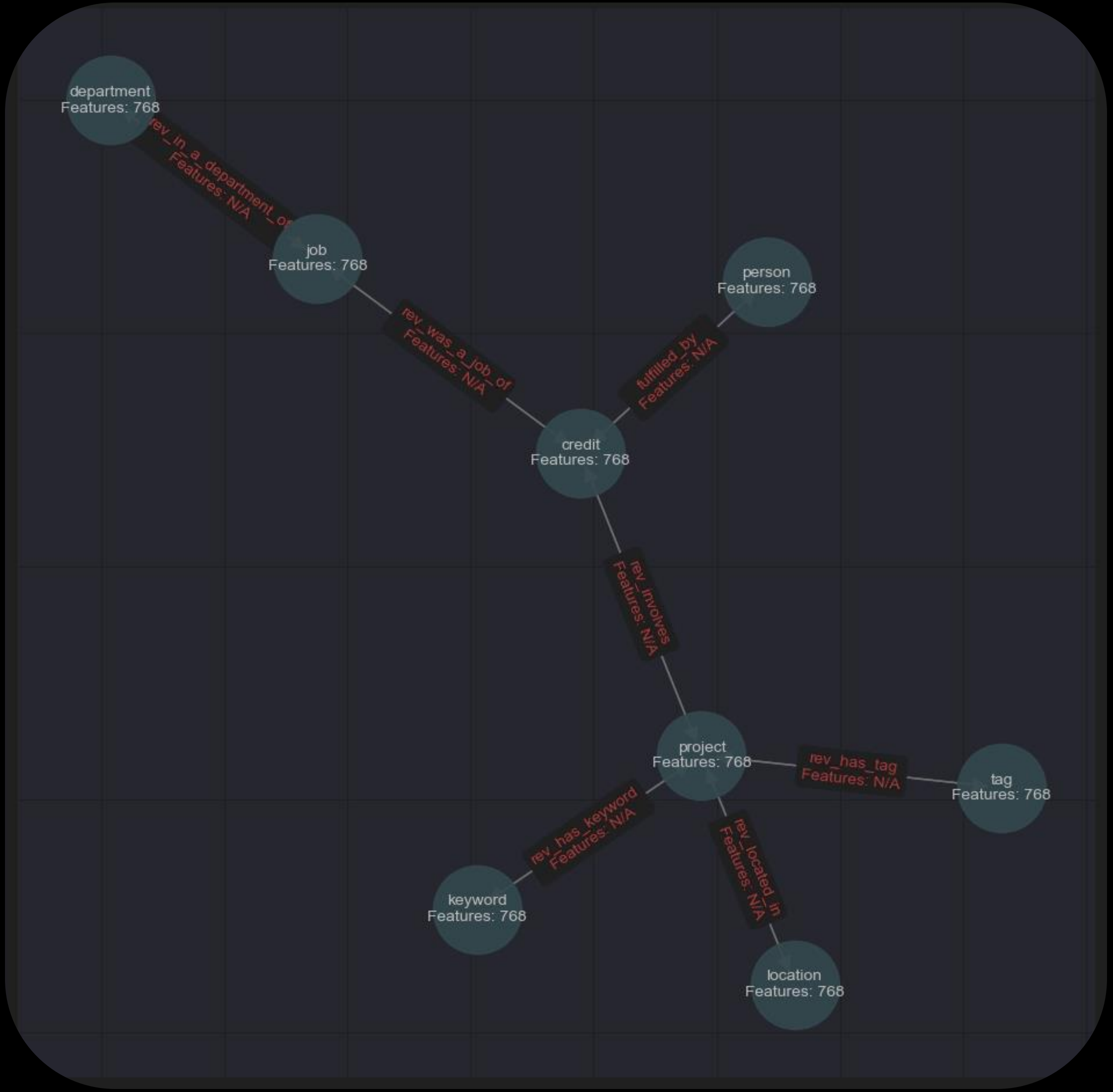
Dynamic, contextual representation

- ELMo
- BERT-generation (all-mpnet-base-v2; all-MiniLM-L6-v2)
- GPTs

PRO: robust to vocabulary (OOV, misspelling), context-sensitive, dynamic embeddings (utterance, sentences)
CON: GPU goes BRRRRR!!!!

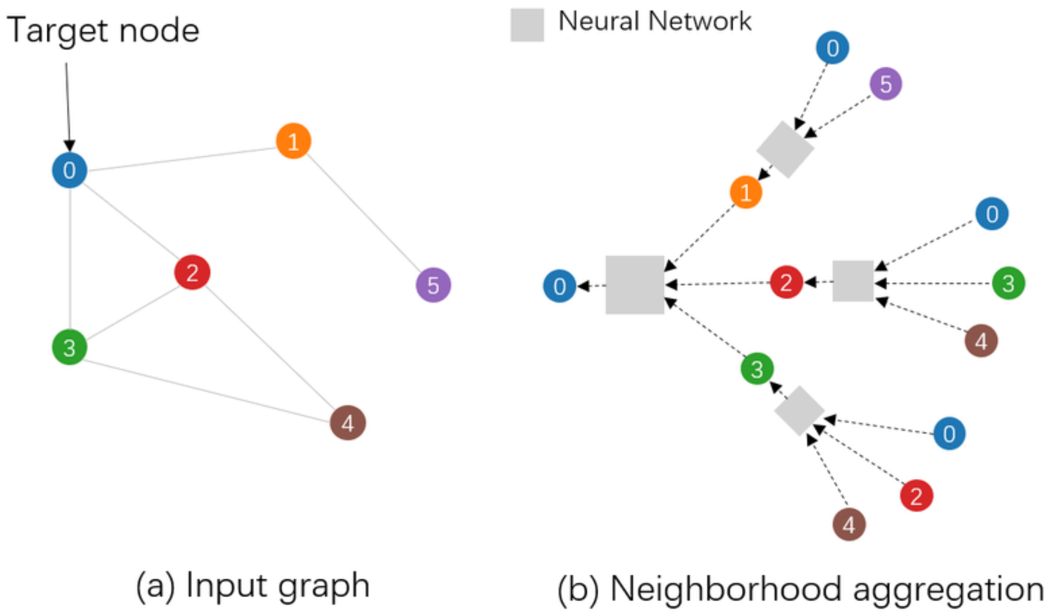
Sounds great! What's next?

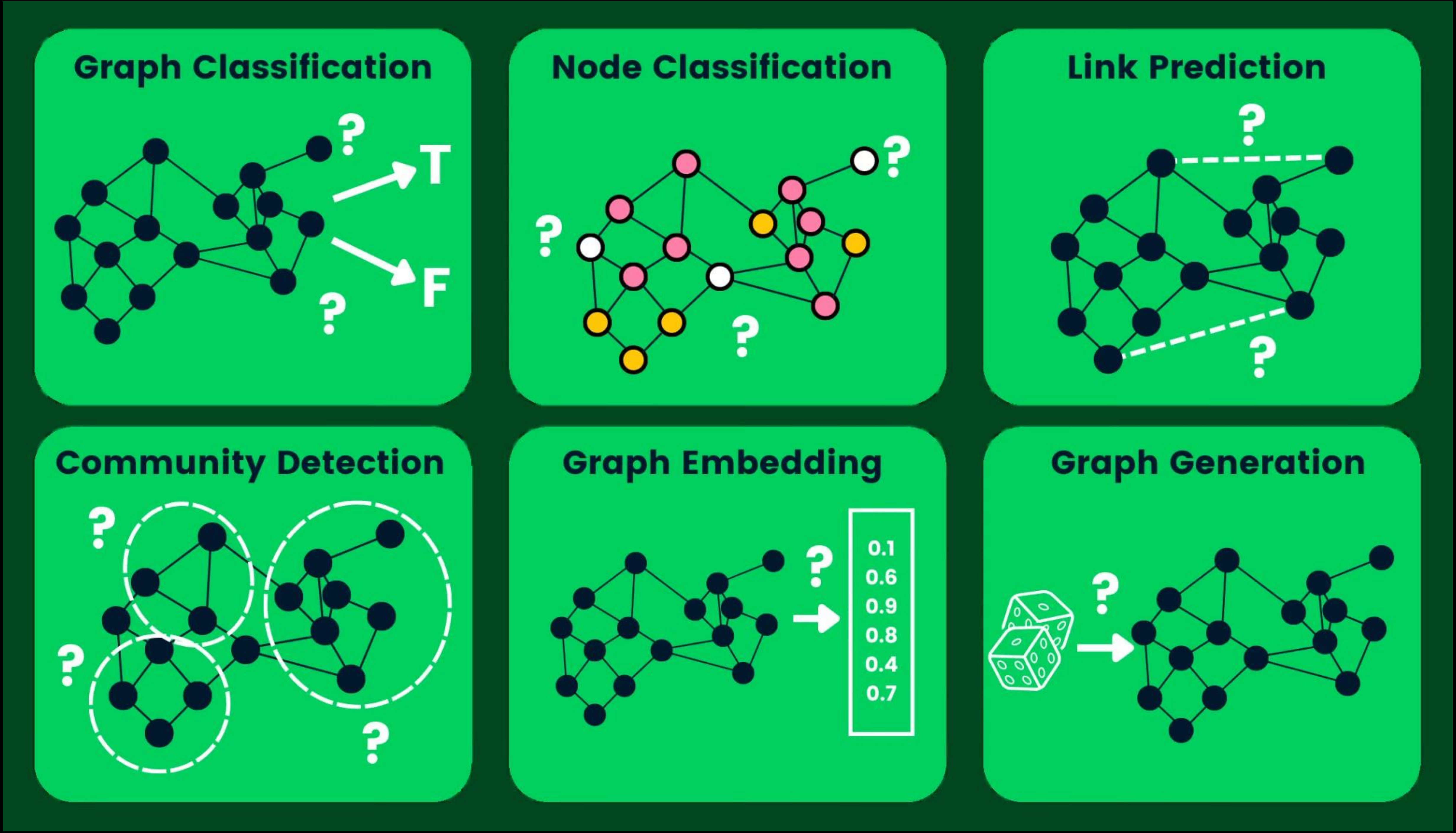
- Activities – rough, mixed data (Hun&Eng)
- Alternative datasets w/ similar metadata
  + IMDB
  + arxiv
- Graph Neural Networks
  + GraphSAGE without embeddings
  + GraphSAGE with starter embeddings
- SentenceTransformers
  + Fine-tuning an existing semantic space



Graph Neural Networks (GNN)

- Reasoning: We have knowledge graphs... why not try out GNNs?
- Data: nodes connected by edges
- Mechanism: nodes share information with neighbours
- Result: learn representations capturing content and connections



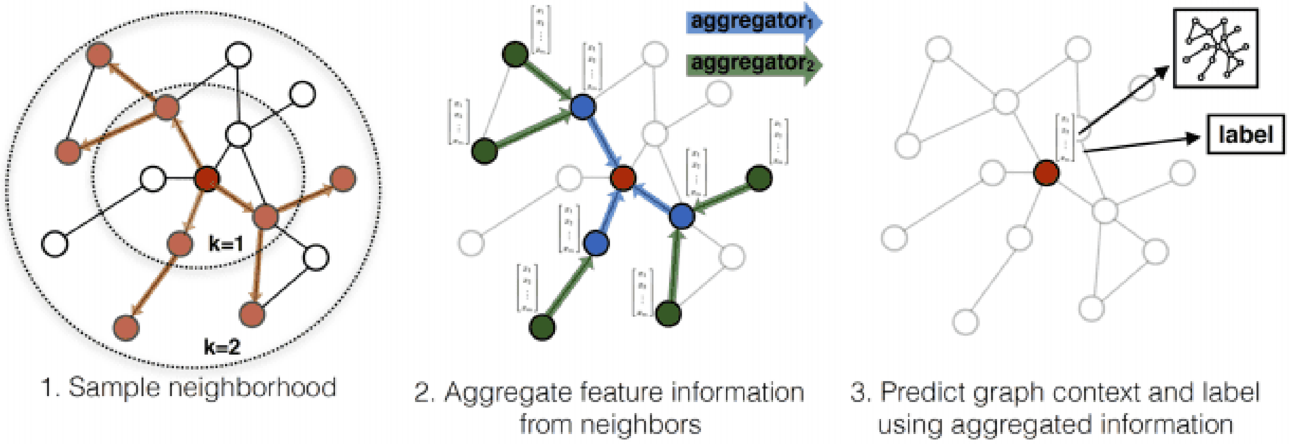

Graph Classification
?
T
F
?
Node Classification
?
?
?
Link Prediction
?
?
Community Detection
?
?
?
Graph Embedding
?
0.1
0.6
0.9
0.8
0.4
0.7
Graph Generation
?

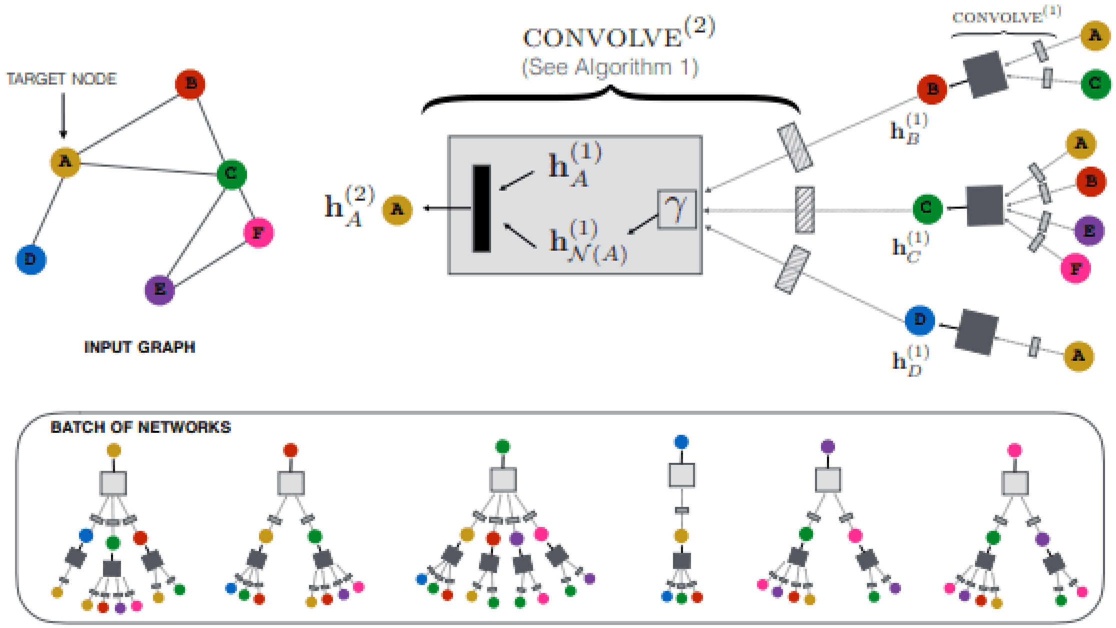
Excuses:
PyG & CUDA

CLASSIFIED

- PyTorch Geometric
  + Built-in GNN implementations, data-handling
  + Particular data-structure for storing graph data
  + Data-wrangling.is.not.fun.
- CUDA: CPU – GPU interface used in training
  + Debugging.is.not.fun.

GraphSAGE

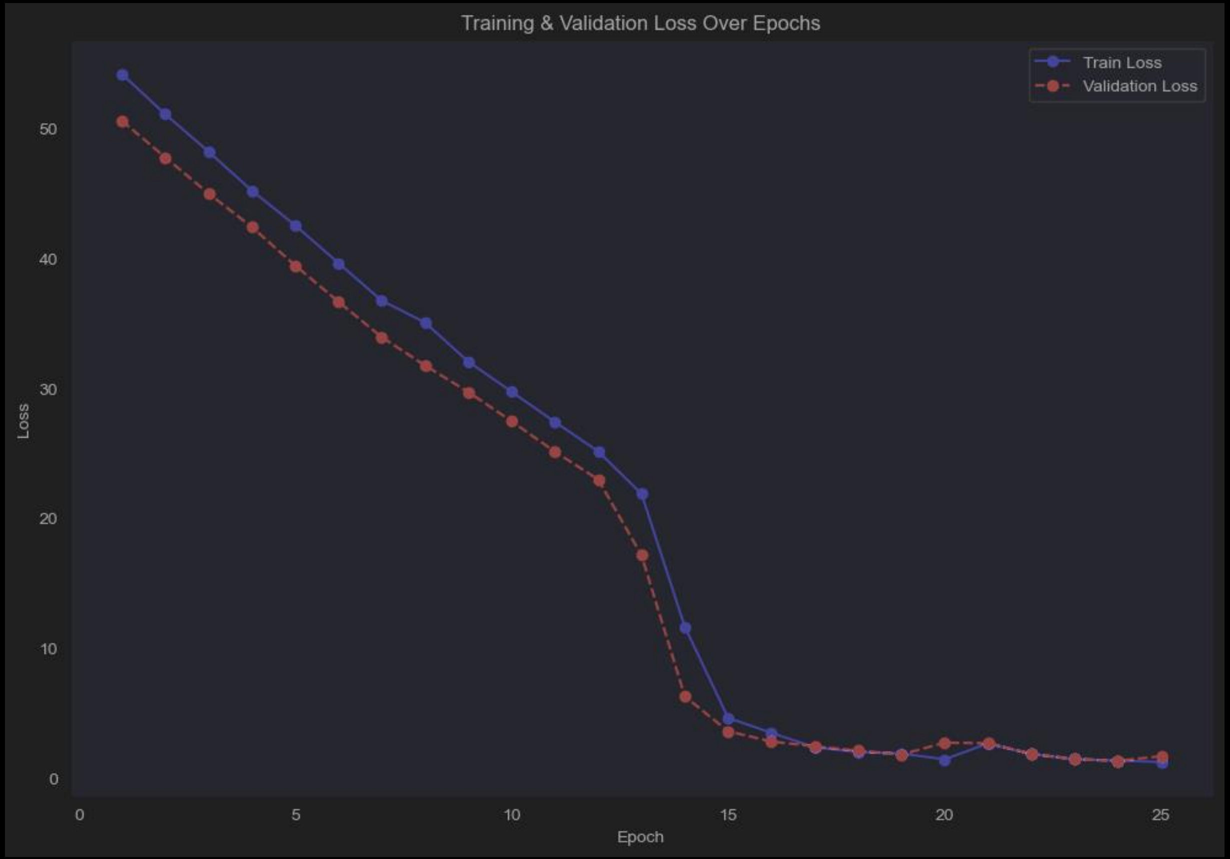




- Generates embeddings from graphs by sampling and aggregating information from a node's local neighborhood
- Fixed-size neighbourhood, batching mini-networks
- Inductive: can generalize to unseen nodes

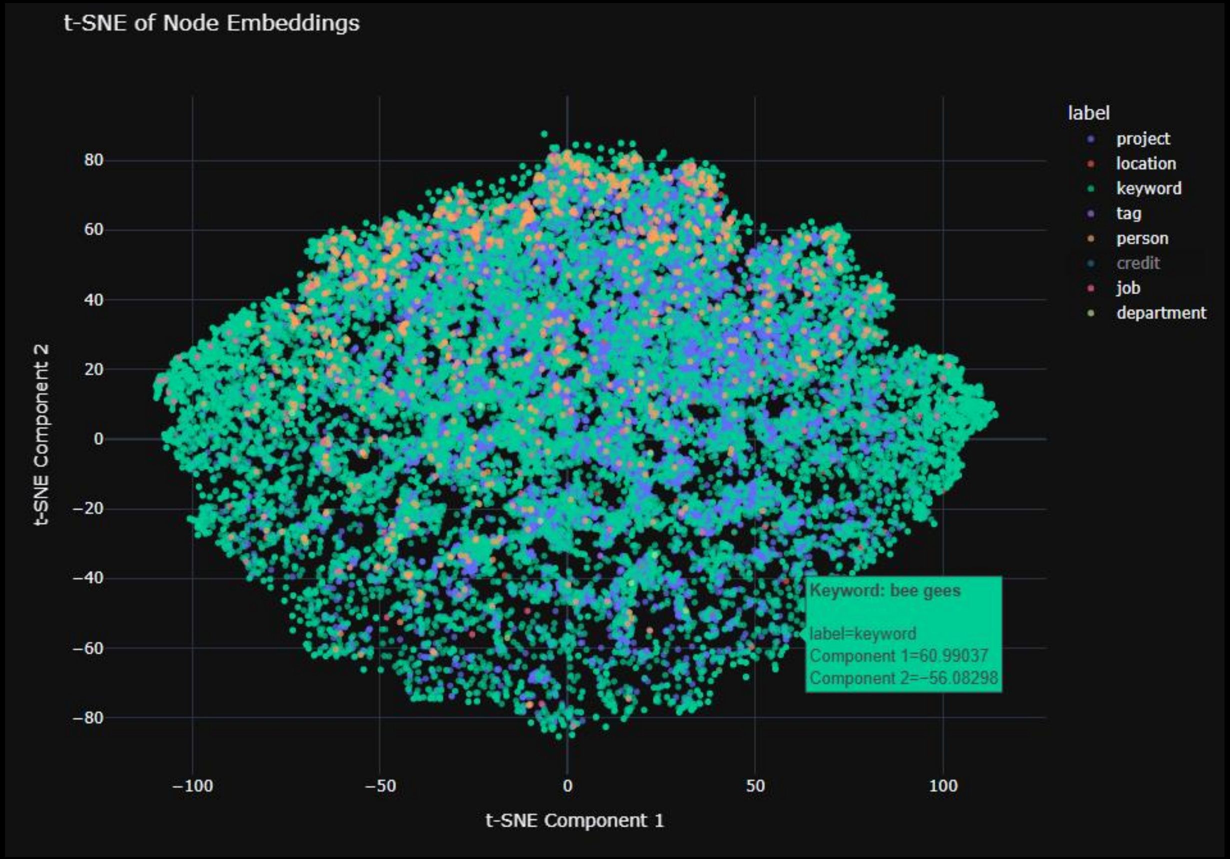
Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings



Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings

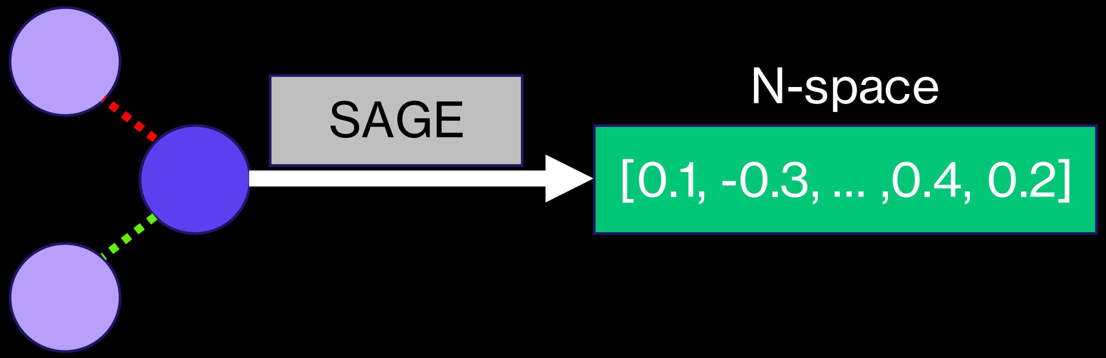


Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings
- Weirdly meh results…
- A serious conceptual problem

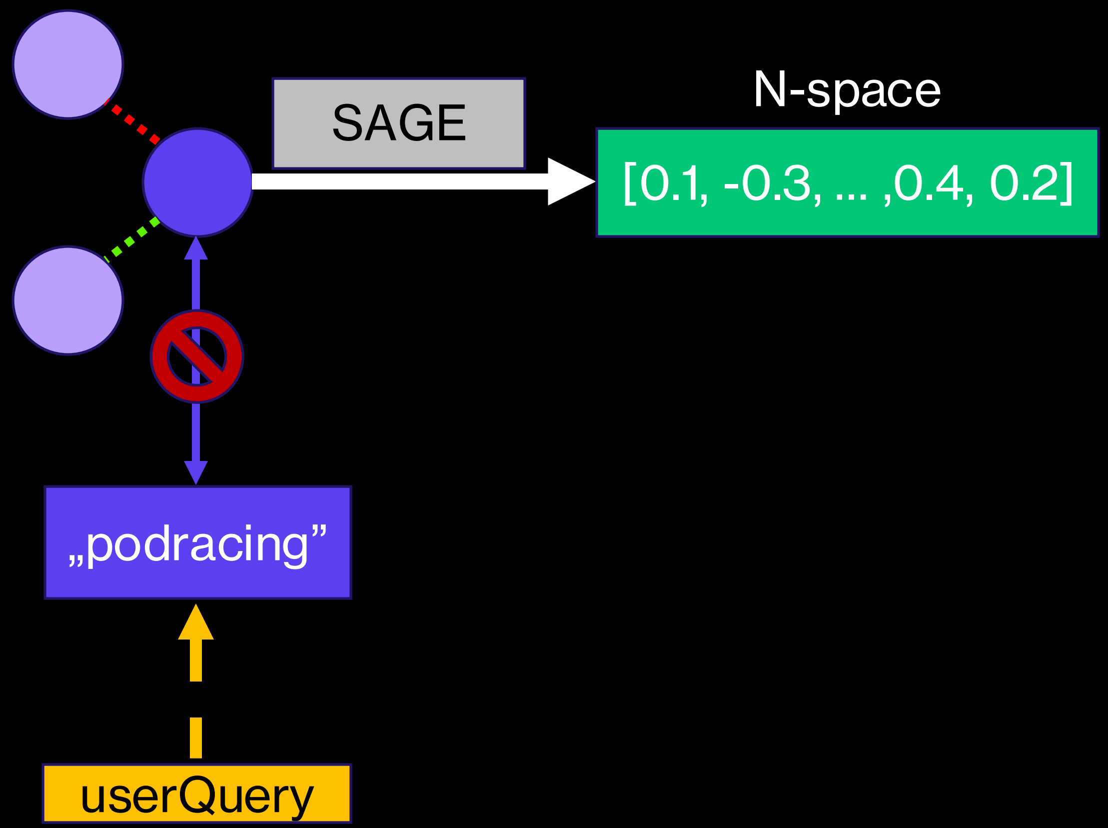
Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings
- Meh results…
- A serious conceptual problem



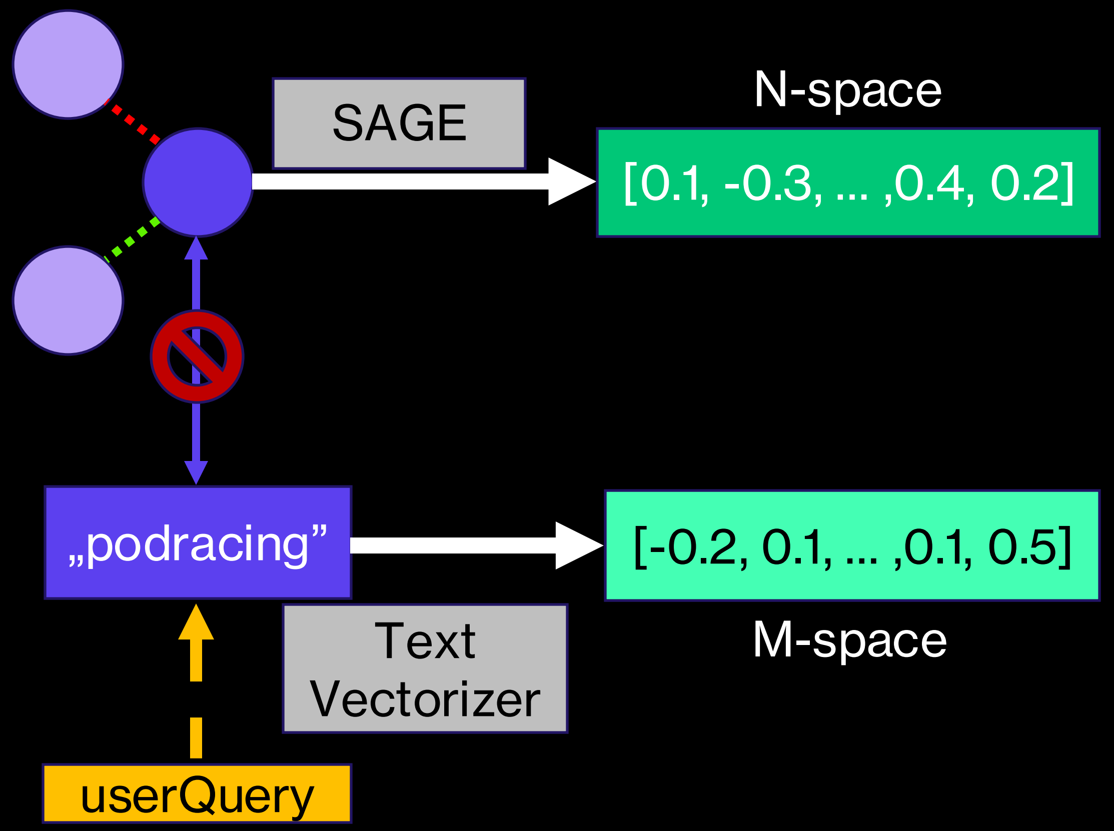
Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings
- Meh results…
- A serious conceptual problem



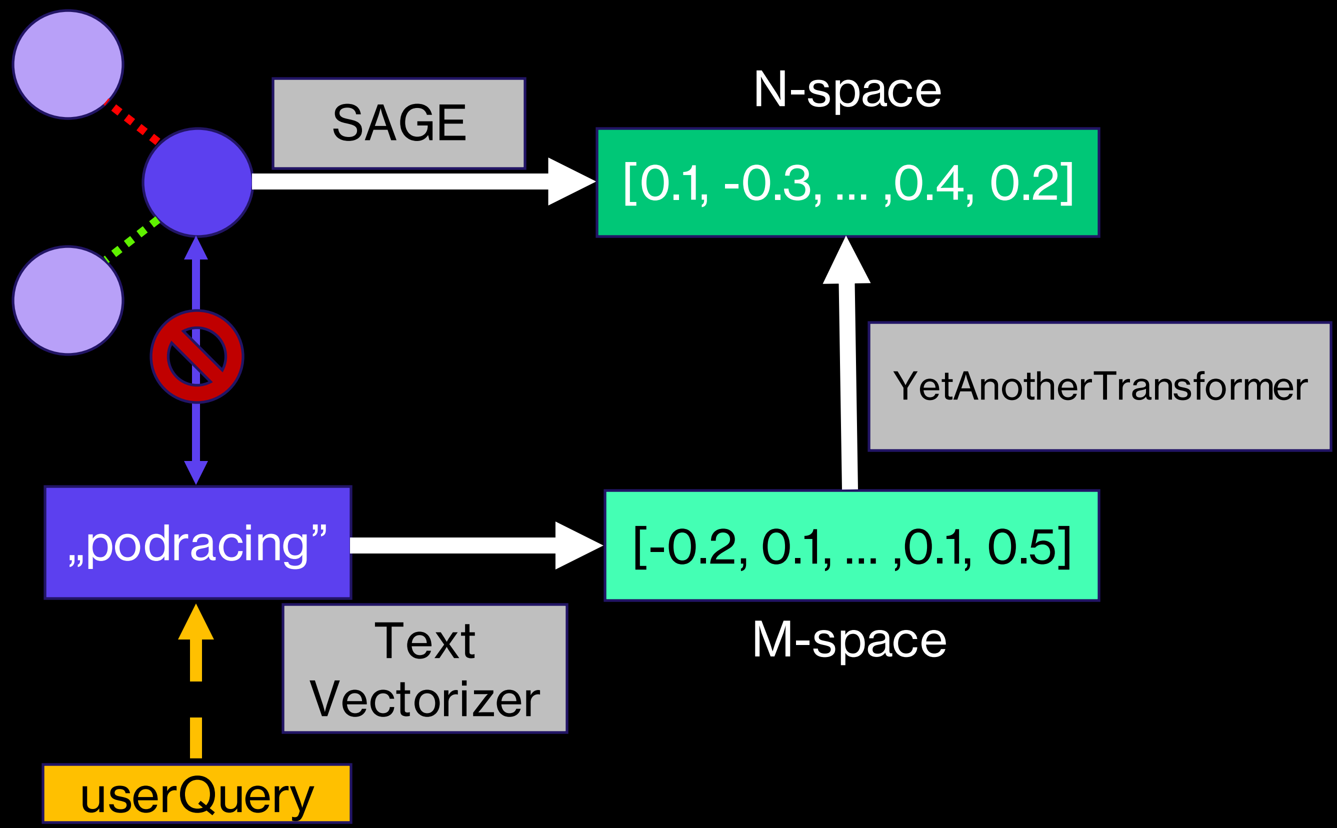
Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings
- Meh results…
- A serious conceptual problem



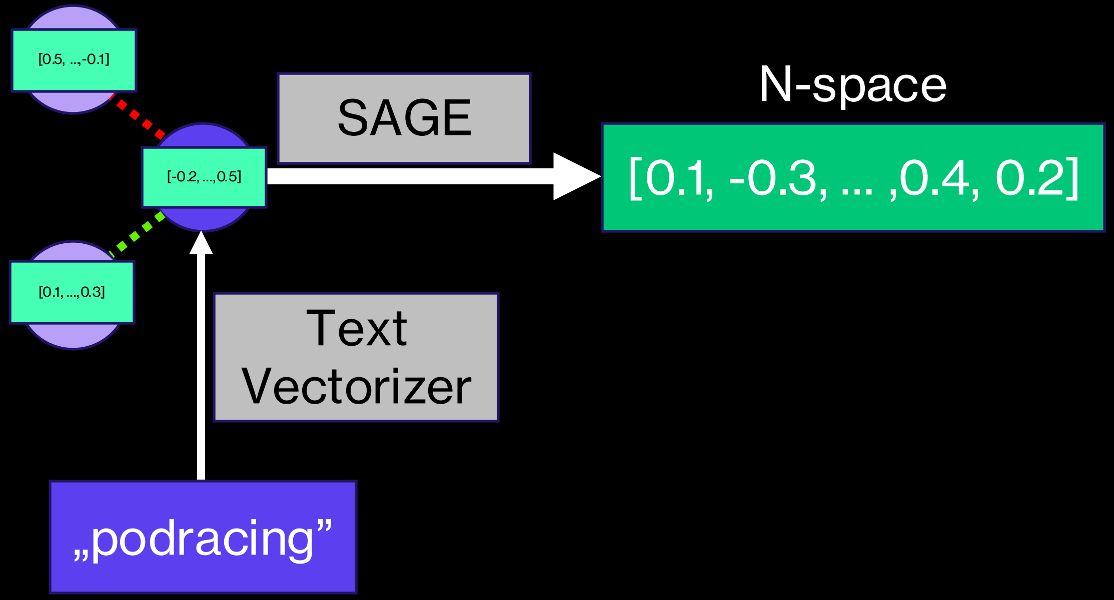
Experiment I.
GraphSage w/o text features

- Top 1.000 most prominent person
- More than 1.000.000 entity nodes (projects, keywords, etc.)
- Goal: learn embeddings by discriminating between true and false edges
- 128-dim embeddings
- Meh results…
- A serious conceptual problem



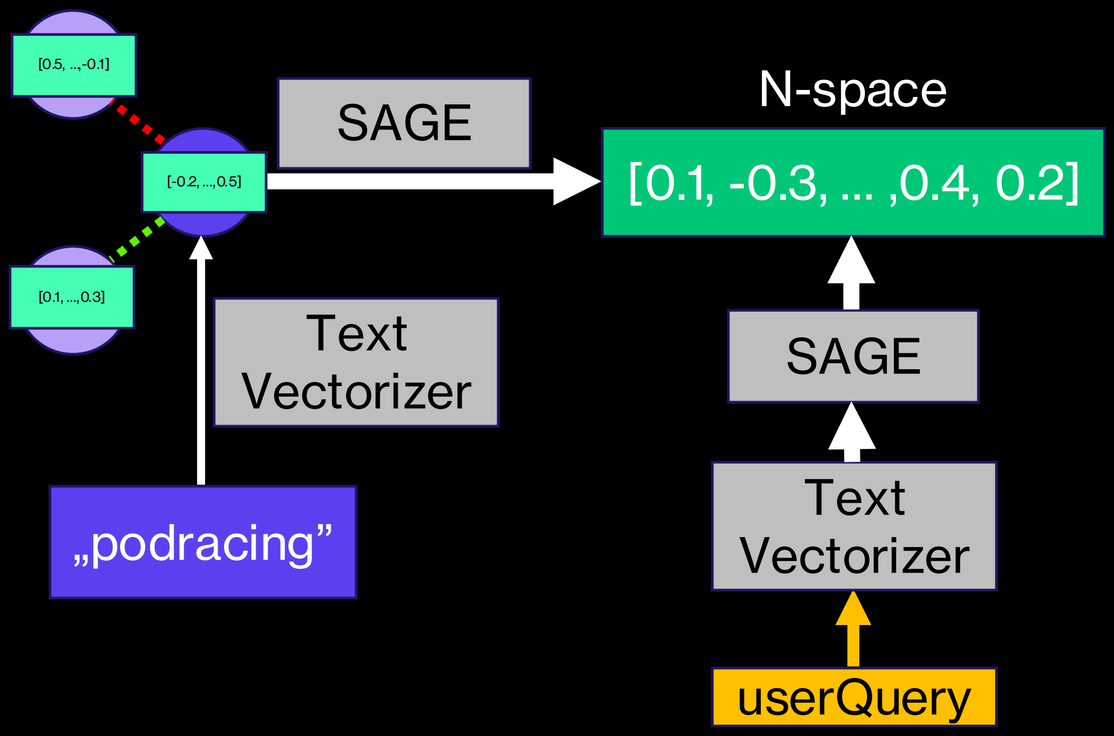
Experiment II.
GraphSage W/ text features

- Add original text/description as embeddings (using a pre-trained embeddings model)
- Goal: learn an embedding-transformation by discriminating between true and false edges



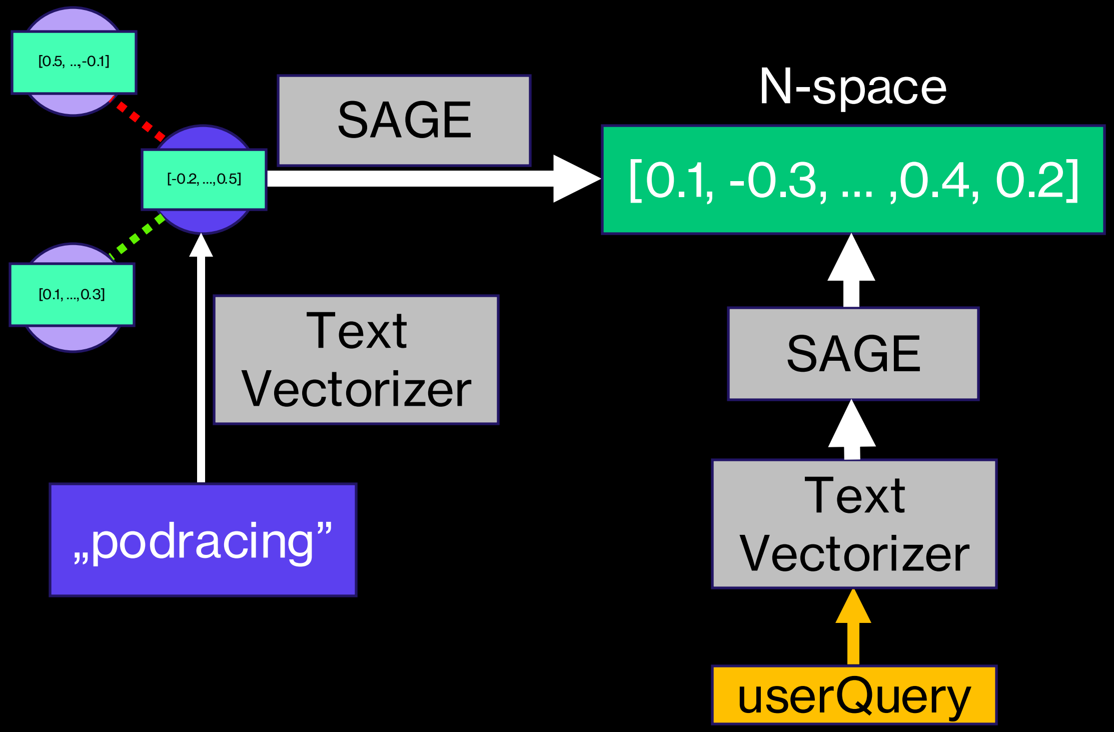
Experiment II.
GraphSage W/ text features

- Add original text/description as embeddings (using a pre-trained embeddings model)
- Goal: learn an embedding-transformation by discriminating between true and false edges



Experiment II.
GraphSage W/ text features

- Add original text/description as embeddings (using a pre-trained embeddings model)
- Goal: learn an embedding-transformation by discriminating between true and false edges
- But then... accidentally updated a package



Experiment II. GraphSage W/ text features

QUERY: SPACE BATTLE

NODE_TYPE: KEYWORD

„Still meh,

but not terribleh…”

/Friday afternoon/

node_id	short_info	score
33387	Keyword: marvel cinematic universe	0.778837
32001	Keyword: captain america	0.771900
29648	Keyword: space shuttle	0.759211
30668	Keyword: hulk	0.754716
22805	Keyword: jedi	0.754313
34130	Keyword: space exploration	0.747362
22812	Keyword: galactic war	0.744439
22801	Keyword: galaxy	0.743882
25573	Keyword: starfleet	0.740094
22388	Keyword: nasa	0.736259
32045	Keyword: lex luthor	0.733575
33748	Keyword: warp engine	0.732327
25604	Keyword: vulcan	0.732079
29799	Keyword: orcs	0.731463
35064	Keyword: captain atom	0.731034
25034	Keyword: space colony	0.729978
24889	Keyword: space marine	0.727638
32658	Keyword: iron man	0.726276
33386	Keyword: tony stark	0.726183
34803	Keyword: poseidon	0.724075
25252	Keyword: giant robot	0.723456
23489	Keyword: jurassic park	0.723297
22389	Keyword: spaceman	0.723165
31940	Keyword: mission to mars	0.722930
25601	Keyword: uss enterprise	0.719928
31325	Keyword: robot fighting	0.717569

Experiment II. GraphSage W/ text features

QUERY: STAR WARS EPISODE IV : <DESC.>

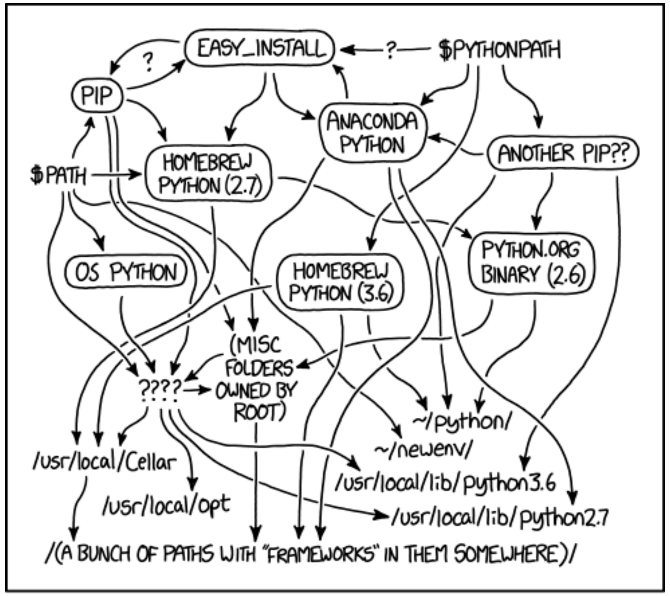
NODE_TYPE: PROJECT

„Still meh,

but not terribleh…”

/Friday afternoon/

node_id	short_info	score
11545	Project: Transformers: Dark of the Moon	0.743649
9793	Project: Home	0.701874
4955	Project: Hulk	0.699880
9690	Project: Star Trek	0.695946
1993	Project: Star Wars: Episode I - The Phantom Men…	0.695330
4116	Project: Star Wars: Episode II - Attack of the …	0.693225
909	Project: The Empire Strikes Back	0.688489
7474	Project: Star Wars: Episode III - Revenge of th…	0.682937
15726	Project: Captain America: Civil War	0.673755
7511	Project: Batman Begins	0.673558
210	Project: Star Wars	0.672327
18023	Project: Family Guy Presents: Something, Someth…	0.671329
21433	Project: Transformers: The Last Knight	0.668813
6070	Project: Spider-Man 2	0.667404
922	Project: Return of the Jedi	0.666997
1167	Project: The Lost World: Jurassic Park	0.664661
17339	Project: The Star Wars Holiday Special	0.664346
9418	Project: The Day the Earth Stood Still	0.663071
946	Project: The Terminator	0.662440
8579	Project: Spider-Man 3	0.659735
20750	Project: DC Showcase: Catwoman	0.659633
11778	Project: The Avengers	0.659184
15717	Project: Avengers: Age of Ultron	0.658772
9154	Project: The Incredible Hulk	0.657312
12245	Project: Mirror Mirror	0.656711
1627	Project: Blade	0.654354
9944	Project: 9	0.651828
3319	Project: Evolution	0.651016

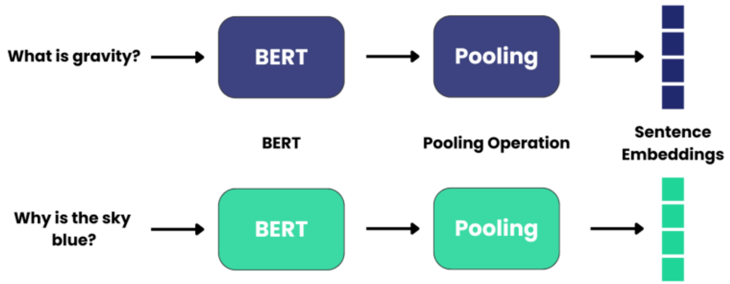


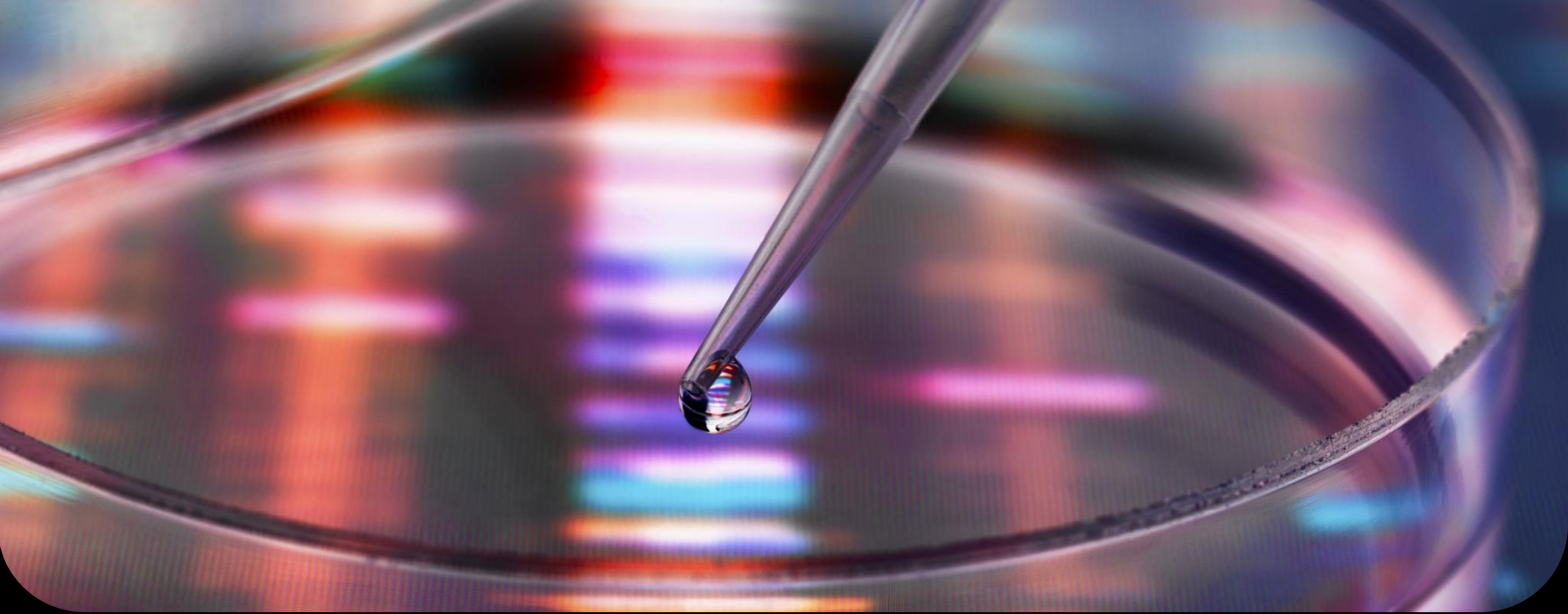
MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

I miss Docker…

SentenceTransformers (SBERT) & HuggingFace (HF)

- SBERT: Framework built on top of pre-trained transformer models and architectures for training & evaluation
- HF: A GitHub for AI models
- Goal: Fine-tuning models to generate embeddings tailored to specific datasets

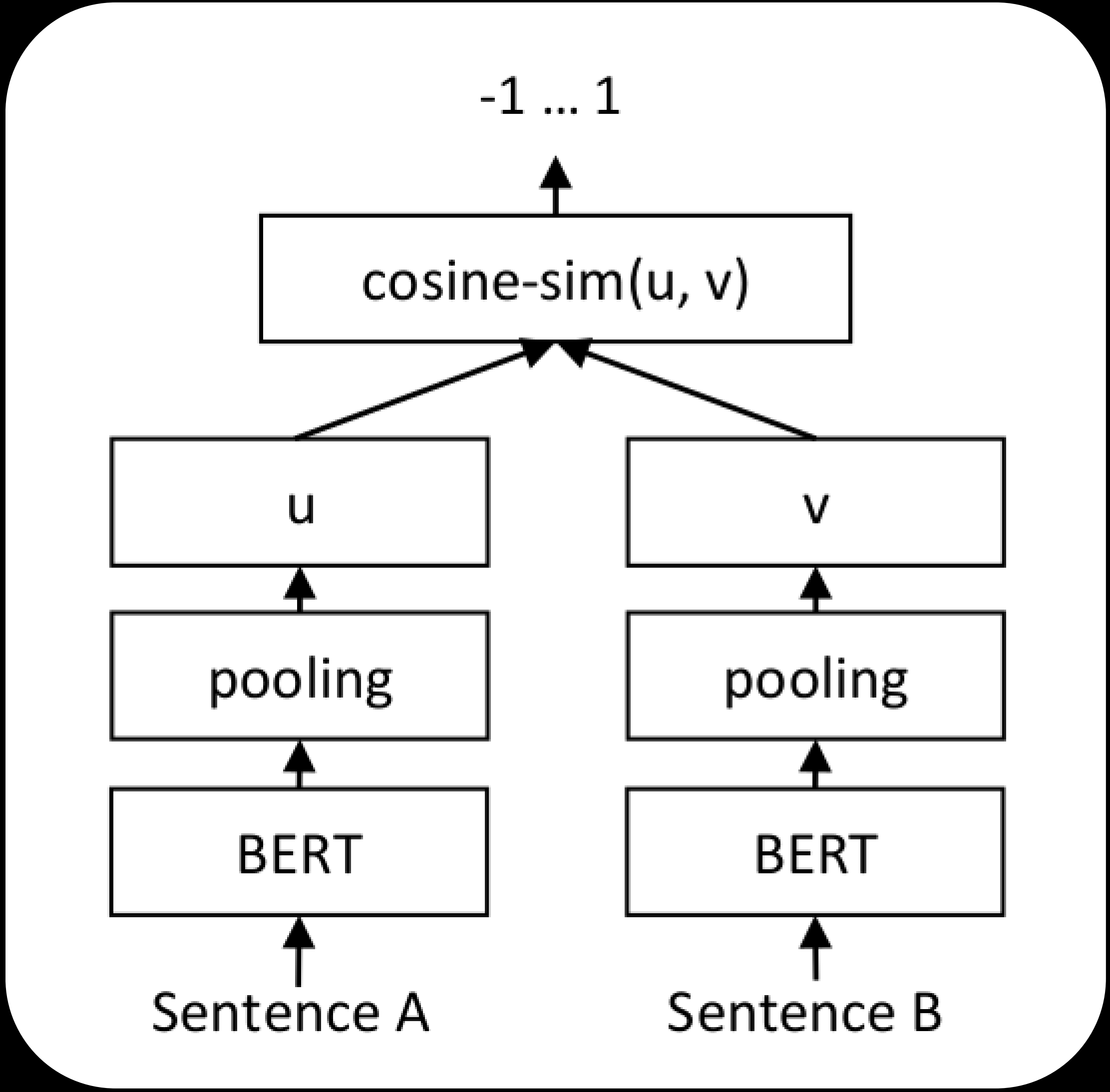




Fine-tuning methods

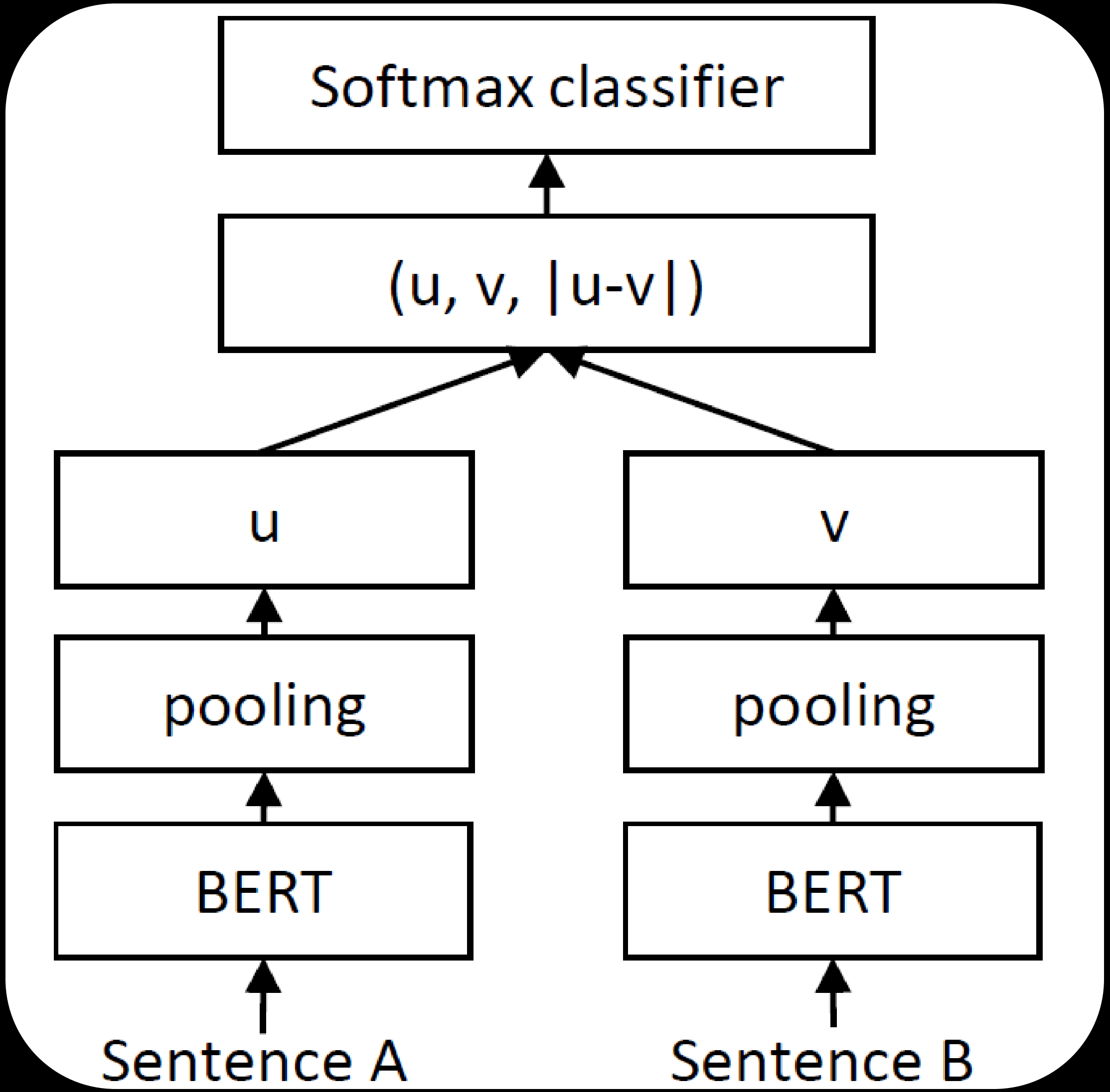
Semantic Textual Similarity (STS)

SentA	SentB	Label
My first sentence	My second sentence	**0.8**
Another pair	Unrelated sentence	**0.3**
Thesis	Antithesis	**-0.9**



Natural Language Inference (NLI)
- SoftMaxLoss

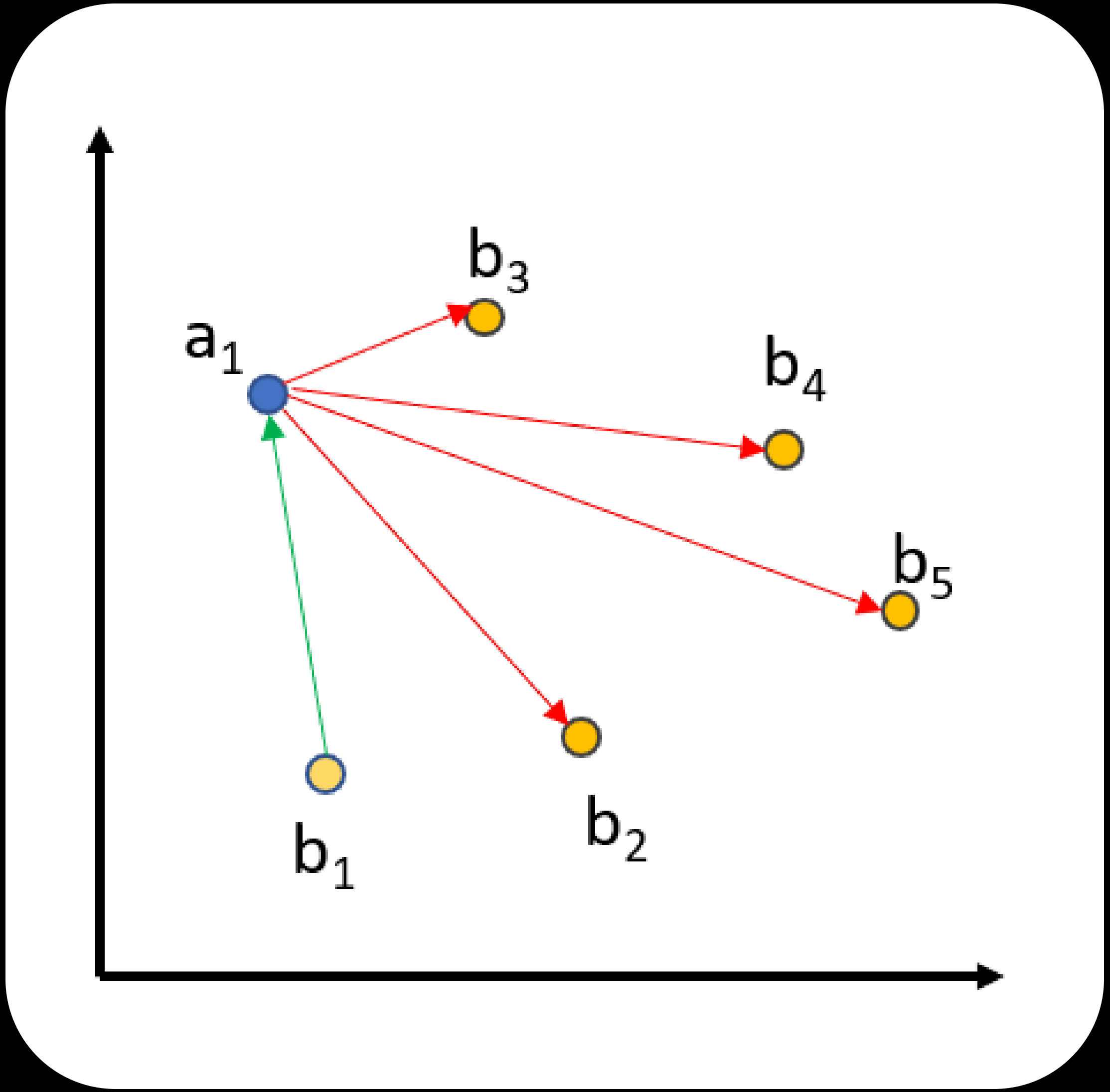
SentA	SentB	Label
A soccer game with multiple males playing.	Some men are playing a sport.	**entailment**
An older and younger man smiling.	Two men are smiling and laughing at the cats playing on the floor.	**neutral**
A man inspects the uniform of a figure in some East Asian country.	The man is sleeping.	**contradiction**



Natural Language Inference (NLI)

- MultipleNegativesRankingLoss

Anchor (a_i)	Positive (b_i)	Negative(b_j)
A person on a horse jumps over a broken down airplane.	A person is outdoors, on a horse.	A person is at a diner, ordering an omelette.
Children smiling and waving at camera	There are children present	The kids are frowning
A boy is jumping on skateboard in the middle of a red bridge.	The boy does a skateboarding trick.	The boy skates down the sidewalk.



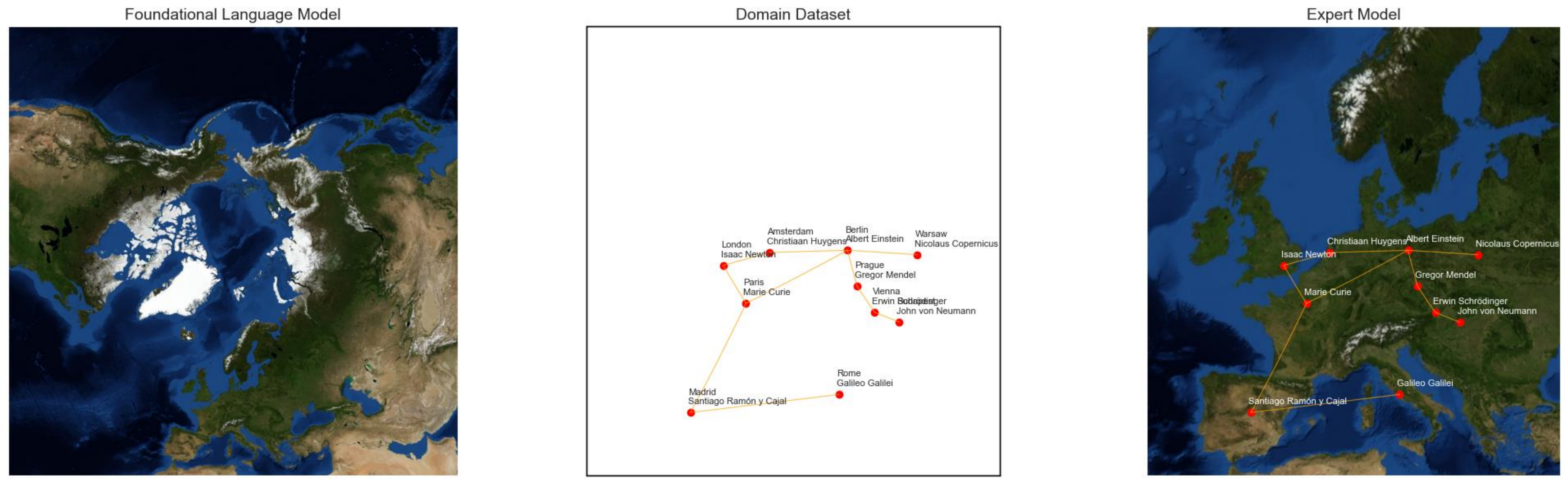
Natural Language Inference (NLI)
-TripletLoss
- MultipleNegativesRankingLoss

Anchor (a_i)	Positive (b_i)	Negative(b_j)
A person on a horse jumps over a broken down airplane.	A person is outdoors, on a horse.	A person is at a diner, ordering an omelette.
Children smiling and waving at camera	There are children present	The kids are frowning
A boy is jumping on skateboard in the middle of a red bridge.	The boy does a skateboarding trick.	The boy skates down the sidewalk.

Experiment III.
MPNET-IMDB

Rationale

foundational model + fine-tuning on domain data
→ „expert model”

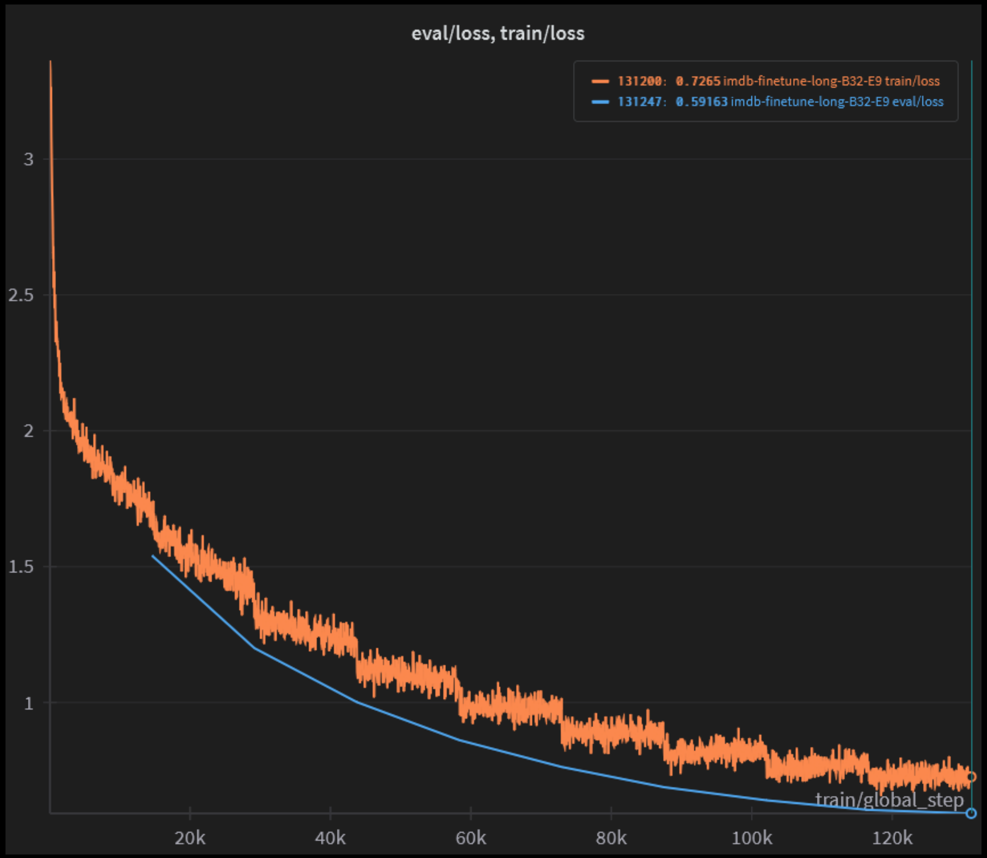


Experiment III.
MPNET-IMDB

- MultipleNegativesRankingLoss:
  + 32 triplets per batch (anchor-positive-negative)
  + negatives: random sampling*
  + 9 epochs (~20h)
  + monitor runs with wandb

anchor	positive
Dark Summer	Description: Follows the tale of a 17-year-old …
The Palm Beach Story	Movie; art directed by 'Hans Dreier'
Action	A genre of the movie 'Ready to Rumble'
The Exorcist	A 'Horror' movie
volcano	A keyword related to the movie 'Volcano'
Charlton Heston	Voice Actor in 'Alaska: Spirit of the Wild'
Killing Season	A movie related to 'blood'
Janet Hirshenson	Casting in 'Angels & Demons'
Even Cowgirls Get the Blues	A movie related to 'lbgt'
blood	A keyword related to the movie 'Innerspace'

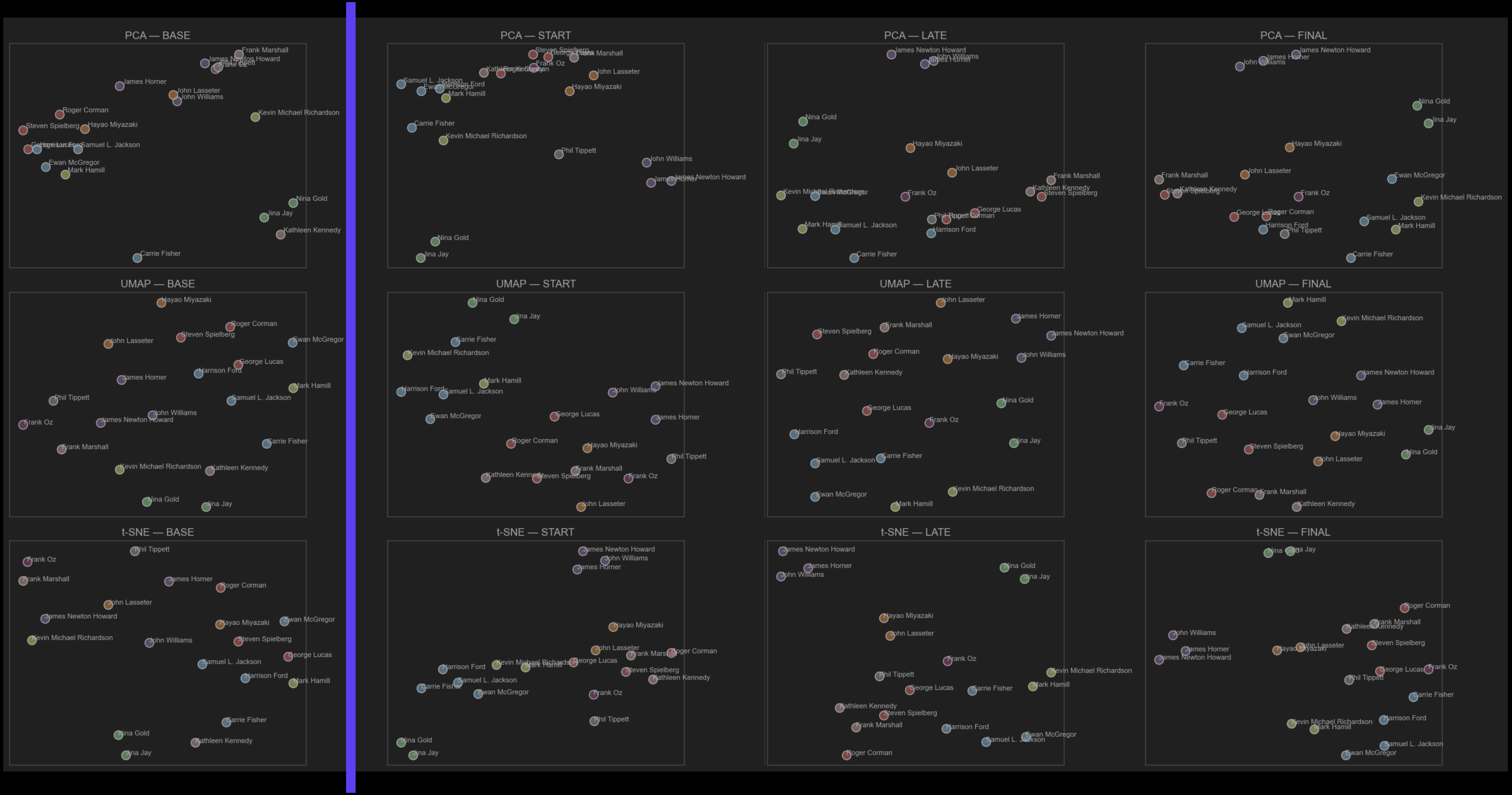
Weights & Biases



Meet the test subjects! 🤗

- Take names and roles
- Encode them using model-<CHECKPOINT>
- Let’s see how they compare!

Name	Role	Why They're Known For
George Lucas	Director	Creator of Star Wars and Indiana Jones franchises.
Steven Spielberg	Director	Known for Jurassic Park, E.T., Jaws, and Indiana Jones.
Kathleen Kennedy	Producer	President of Lucasfilm; Produced Star Wars sequels and Jurassic Park.
Frank Marshall	Producer	Produced Indiana Jones, Jurassic Park, Back to the Future.
Roger Corman	Director	Known for low-budget cult films; mentor to many famous directors.
Harrison Ford	Actor (Live Action)	Famous for roles as Han Solo (Star Wars) and Indiana Jones.
Carrie Fisher	Actor (Live Action)	Known for playing Princess Leia in Star Wars.
Samuel L. Jackson	Actor (Live Action)	Known for Pulp Fiction, Star Wars (Mace Windu), Marvel Cinematic Universe (Nick Fury).
Ewan McGregor	Actor (Live Action)	Known for Trainspotting, Obi-Wan Kenobi in Star Wars.
Mark Hamill	Voice Actor	Luke Skywalker in Star Wars; acclaimed Joker voice in Batman: The Animated Series.
John Williams	Film Composer	Legendary composer for Star Wars, Jaws, Jurassic Park, Harry Potter.
James Newton Howard	Film Composer	Known for The Dark Knight (with Hans Zimmer), The Hunger Games.
James Horner	Film Composer	Known for Titanic, Avatar, Braveheart.
Frank Oz	Puppeteer	Voice and puppeteer for Yoda (Star Wars); The Muppets.
Phil Tippett	Visual Effects Artist	Known for Star Wars stop-motion effects, Jurassic Park.
Kevin Michael Richardson	Voice Actor	Known for deep voice roles in The Simpsons, Family Guy, Lilo & Stitch.
Nina Gold	Casting Director	Known for casting Game of Thrones, Star Wars sequels.
Jina Jay	Casting Director	Known for casting Rogue One: A Star Wars Story, Tinker Tailor Soldier Spy.
John Lasseter	Animator / Animation Director	Co-founder of Pixar; directed Toy Story, Cars.
Hayao Miyazaki	Animator / Animation Director	Co-founder of Studio Ghibli; directed Spirited Away, My Neighbor Totoro.


PCA — BASE
PCA — START
PCA — LATE
PCA — FINAL
UMAP — BASE
UMAP — START
UMAP — LATE
UMAP — FINAL
t-SNE — BASE
t-SNE — START
t-SNE — LATE
t-SNE — FINAL


PCA — BASE
PCA — START
PCA — LATE
PCA — FINAL
UMAP — BASE
UMAP — START
UMAP — LATE
UMAP — FINAL
t-SNE — BASE
t-SNE — START
t-SNE — LATE
t-SNE — FINAL


PCA — BASE
Frank Marshall
James Newton Howard
James Horner
John Lasseter
John Williams
Roger Corman
Steven Spielberg
Hayao Miyazaki
Kevin Michael Richardson
Samuel L. Jackson
Ewan McGregor
Mark Hamill
Nina Gold
Jina Jay
Kathleen Kennedy
Carrie Fisher
PCA — START
Frank Marshall
John Lasseter
Hayao Miyazaki
Samuel L. Jackson
Mark Hamill
Carrie Fisher
Kevin Michael Richardson
Phil Tippett
John Williams
James Newton Howard
Nina Gold
Jina Jay
PCA — LATE
James Newton Howard
John Williams
Nina Gold
Jina Jay
Hayao Miyazaki
John Lasseter
Frank Marshall
Frank Oz
Kathleen Kennedy
Steven Spielberg
George Lucas
Samuel L. Jackson
Harrison Ford
Carrie Fisher
PCA — FINAL
James Newton Howard
Nina Gold
Jina Jay
Hayao Miyazaki
Frank Marshall
John Lasseter
Ewan McGregor
Frank Oz
Kevin Michael Richardson
Roger Corman
Samuel L. Jackson
Harrison Ford
Mark Hamill
Phil Tippett
Carrie Fisher
UMAP — BASE
Hayao Miyazaki
Roger Corman
John Lasseter
Steven Spielberg
Ewan McGregor
George Lucas
James Horner
Harrison Ford
Mark Hamill
Phil Tippett
Samuel L. Jackson
John Williams
Frank Oz
James Newton Howard
Carrie Fisher
Frank Marshall
Kevin Michael Richardson
Kathleen Kennedy
Nina Gold
Jina Jay
UMAP — START
Nina Gold
Jina Jay
Carrie Fisher
Kevin Michael Richardson
Mark Hamill
James Newton Howard
Harrison Ford
Samuel L. Jackson
John Williams
Ewan McGregor
George Lucas
James Horner
Roger Corman
Hayao Miyazaki
Phil Tippett
Frank Marshall
Steven Spielberg
Frank Oz
John Lasseter
UMAP — LATE
John Lasseter
James Horner
Frank Marshall
Steven Spielberg
James Newton Howard
Roger Corman
Hayao Miyazaki
John Williams
Phil Tippett
Kathleen Kennedy
George Lucas
Nina Gold
Frank Oz
Harrison Ford
Jina Jay
Carrie Fisher
Samuel L. Jackson
Kevin Michael Richardson
Ewan McGregor
Mark Hamill
UMAP — FINAL
Mark Hamill
Kevin Michael Richardson
Samuel L. Jackson
Ewan McGregor
Carrie Fisher
Harrison Ford
James Newton Howard
Frank Oz
John Williams
James Horner
George Lucas
Jina Jay
Hayao Miyazaki
Phil Tippett
Steven Spielberg
Nina Gold
John Lasseter
Roger Corman
Frank Marshall
Kathleen Kennedy
t-SNE — BASE
Phil Tippett
Frank Oz
Frank Marshall
James Horner
Roger Corman
John Lasseter
James Newton Howard
Hayao Miyazaki
Ewan McGregor
Kevin Michael Richardson
John Williams
Steven Spielberg
George Lucas
Samuel L. Jackson
Harrison Ford
Mark Hamill
Carrie Fisher
Nina Gold
Kathleen Kennedy
Jina Jay
t-SNE — START
James Newton Howard
John Williams
James Horner
Hayao Miyazaki
John Lasseter
Roger Corman
Harrison Ford
George Lucas
Steven Spielberg
Kathleen Kennedy
Samuel L. Jackson
Ewan McGregor
Frank Oz
Phil Tippett
Nina Gold
Jina Jay
t-SNE — LATE
James Newton Howard
James Horner
John Williams
Nina Gold
Jina Jay
Hayao Miyazaki
John Lasseter
Frank Oz
Kevin Michael Richardson
Phil Tippett
George Lucas
Carrie Fisher
Mark Hamill
Kathleen Kennedy
Steven Spielberg
Frank Marshall
Harrison Ford
Ewan McGregor
Roger Corman
t-SNE — FINAL
Roger Corman
Frank Marshall
John Williams
Steven Spielberg
James Horner
James Newton Howard
George Lucas
Frank Oz
Phil Tippett
Carrie Fisher
Kevin Michael Richardson
Harrison Ford
Mark Hamill
Samuel L. Jackson
Ewan McGregor

Retrieval test

```
query = "George Lucas"
node = "project"
```

Base

1. [project | base] The People vs. George Lucas (score: 0.7281)
2. [project | base] Star Wars (score: 0.6257)
3. [project | base] Star Wars: Episode I - The Phantom Menace (score: 0.5810)
4. [project | base] Star Wars: The Clone Wars (score: 0.5713)
5. [project | base] Rogue One: A Star Wars Story (score: 0.5528)
6. [project | base] Star Wars: Episode III - Revenge of the Sith (score: 0.5461)
7. [project | base] Star Wars: The Force Awakens (score: 0.5419)
8. [project | base] Empire of Dreams: The Story of the Star Wars Trilogy (score: 0.5309)
9. [project | base] The Star Wars Holiday Special (score: 0.5190)
10. [project | base] Robot Chicken: Star Wars (score: 0.5169)
11. [project | base] Star Wars: Episode II - Attack of the Clones (score: 0.4760)
12. **[project | base] The Good Dinosaur (score: 0.3482)**
13. **[project | base] Jurassic World (score: 0.3461)**
14. **[project | base] Godzilla, King of the Monsters! (score: 0.3458)**
15. **[project | base] Jurassic Park III (score: 0.3386)**

Fine-tuned (epoch1)

1. [project | finetuned] The People vs. George Lucas (score: 0.6409)
2. [project | finetuned] Empire of Dreams: The Story of the Star Wars Trilogy (score: 0.4891)
3. [project | finetuned] Star Wars (score: 0.4663)
4. [project | finetuned] Star Wars: Episode III - Revenge of the Sith (score: 0.3668)
5. [project | finetuned] Rogue One: A Star Wars Story (score: 0.3630)
6. [project | finetuned] Star Wars: Episode I - The Phantom Menace (score: 0.3559)
7. [project | finetuned] The Star Wars Holiday Special (score: 0.3415)
8. [project | finetuned] Star Wars: The Force Awakens (score: 0.3222)
9. [project | finetuned] Star Wars: Episode II - Attack of the Clones (score: 0.3046)
10. [project | finetuned] Star Wars: The Clone Wars (score: 0.2968)
11. **[project | finetuned] Fanboys (score: 0.2626)**
12. [project | finetuned] Robot Chicken: Star Wars (score: 0.2495)
13. **[project | finetuned] Carrie Fisher: Wishful Drinking (score: 0.1764)**
14. [project | finetuned] Before The Dinosaurs - Walking With Monsters (score: 0.1534)
15. **[project | finetuned] Family Guy Presents: Something, Something, Something, Dark Side (score: 0.1508)**

Retrieval test

```
query = „documentary movie about George Lucas"
node = "project"
```

Base

1. **[project | base] The People vs. George Lucas (score: 0.6060)**
2. **[project | base] The Star Wars Holiday Special (score: 0.5497)**
3. **[project | base] Empire of Dreams: The Story of the Star Wars Trilogy (score: 0.5318)**
4. [project | base] Rogue One: A Star Wars Story (score: 0.5276)
5. [project | base] Star Wars: The Clone Wars (score: 0.4885)
6. [project | base] Star Wars: Episode I - The Phantom Menace (score: 0.4796)
7. [project | base] Star Wars: Episode III - Revenge of the Sith (score: 0.4693)
8. [project | base] Star Wars (score: 0.4620)
9. [project | base] Star Wars: The Force Awakens (score: 0.4256)
10. [project | base] Robot Chicken: Star Wars (score: 0.4224)
11. [project | base] Star Wars: Episode II - Attack of the Clones (score: 0.4074)
12. [project | base] Jurassic World (score: 0.3545)
13. [project | base] Jurassic Park III (score: 0.3502)
14. [project | base] The Lost World: Jurassic Park (score: 0.3380)
15. [project | base] Jurassic Park (score: 0.3223)

Fine-tuned (epoch1)

1. **[project | finetuned] The People vs. George Lucas (score: 0.8249)**
2. **[project | finetuned] Empire of Dreams: The Story of the Star Wars Trilogy (score: 0.7673)**
3. **[project | finetuned] The Star Wars Holiday Special (score: 0.4955)**
4. **[project | finetuned] Carrie Fisher: Wishful Drinking (score: 0.4849)**
5. [project | finetuned] Before The Dinosaurs - Walking With Monsters (score: 0.4826)
6. [project | finetuned] We're Back! A Dinosaur's Story (score: 0.4577)
7. [project | finetuned] Star Wars (score: 0.4341)
8. [project | finetuned] Robot Chicken: Star Wars (score: 0.4274)
9. [project | finetuned] Rogue One: A Star Wars Story (score: 0.4244)
10. [project | finetuned] The Land Before Time VI: The Secret of Saurus Rock (score: 0.4151)
11. [project | finetuned] Star Wars: Episode III - Revenge of the Sith (score: 0.4105)
12. [project | finetuned] Star Wars: Episode I - The Phantom Menace (score: 0.4097)
13. [project | finetuned] Fanboys (score: 0.4040)
14. [project | finetuned] Star Wars: The Force Awakens (score: 0.3932)
15. [project | finetuned] Star Wars: The Clone Wars (score: 0.3720)

Retrieval test

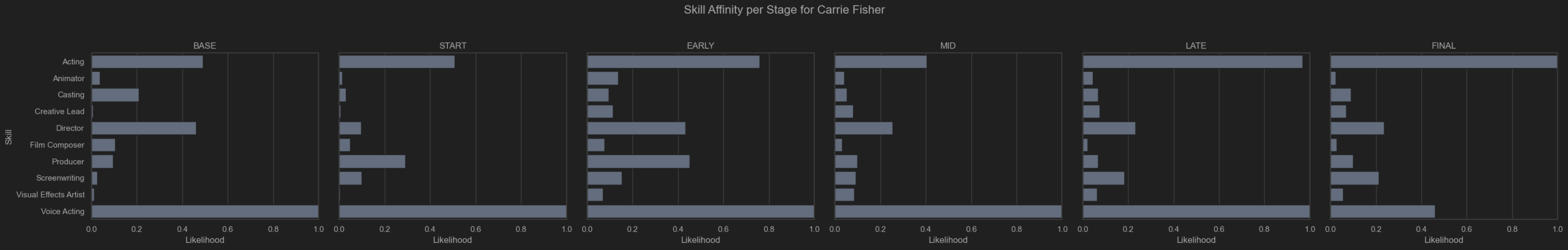
```
query = "a director of sci-fi movies"
node = "person"
```

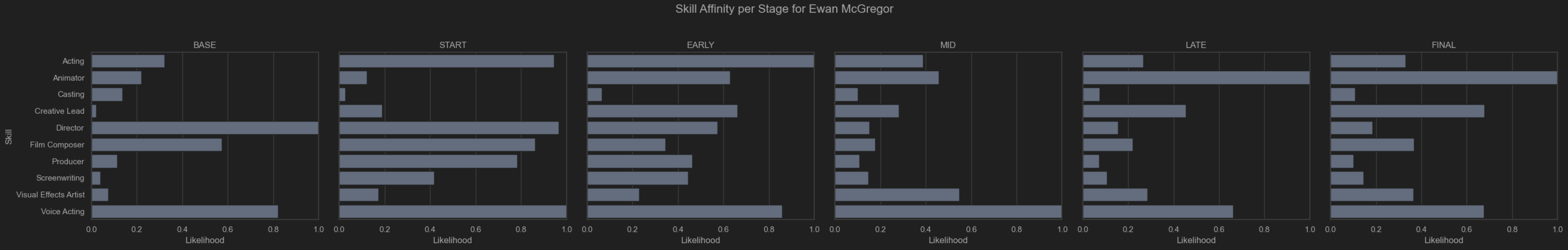
Base

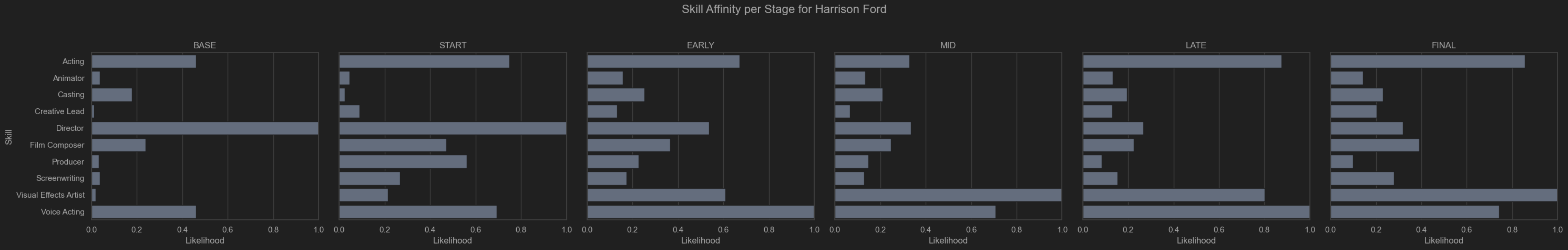
1. [person | base] Steven Spielberg (score: 0.5685)
2. [person | base] George Lucas (score: 0.4892)
3. [person | base] Max von Sydow (score: 0.4807)
4. [person | base] John Lasseter (score: 0.4721)
5. [person | base] Kevin Smith (score: 0.4530)
6. [person | base] Roger Corman (score: 0.4396)
7. [person | base] Hayao Miyazaki (score: 0.4256)
8. [person | base] Denis Leary (score: 0.4212)
9. [person | base] Harrison Ford (score: 0.4192)
10. [person | base] Forest Whitaker (score: 0.4080)
11. [person | base] Kenneth Branagh (score: 0.4023)
12. [person | base] Christopher Lee (score: 0.3943)
13. [person | base] John Williams (score: 0.3867)
14. [person | base] James Horner (score: 0.3807)
15. [person | base] Peter Cushing (score: 0.3800)

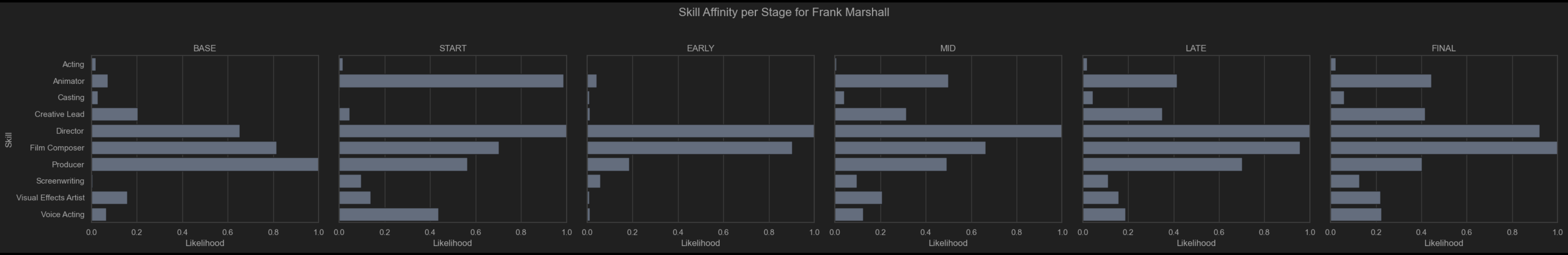
Fine-tuned (epoch1)

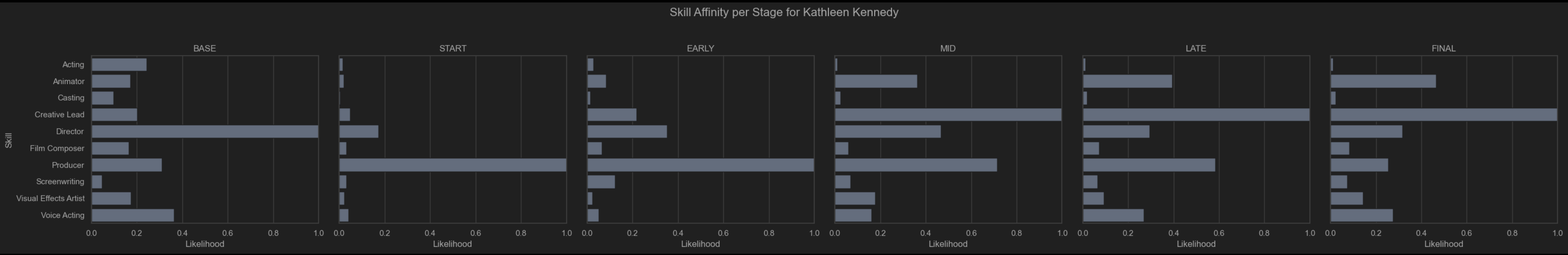
1. [person | finetuned] Bert I. Gordon (score: 0.6986)
2. [person | finetuned] Roger Corman (score: 0.6985)
3. [person | finetuned] Steven Spielberg (score: 0.6497)
4. [person | finetuned] Fred Olen Ray (score: 0.6420)
5. [person | finetuned] George Lucas (score: 0.6351)
6. [person | finetuned] Jim Wynorski (score: 0.5698)
7. [person | finetuned] Frank Marshall (score: 0.5560)
8. [person | finetuned] Phil Tippett (score: 0.5062)
9. [person | finetuned] Kevin Smith (score: 0.5042)
10. [person | finetuned] Malcolm McDowell (score: 0.5032)
11. [person | finetuned] Kevin Michael Richardson (score: 0.4977)
12. [person | finetuned] Denis Leary (score: 0.4970)
13. [person | finetuned] Joel Edgerton (score: 0.4876)
14. [person | finetuned] John Lasseter (score: 0.4870)
15. [person | finetuned] Hayao Miyazaki (score: 0.4869)


Skill Affinity per Stage for Carrie Fisher
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0 0.2 0.4 0.6 0.8 1.0
Likelihood

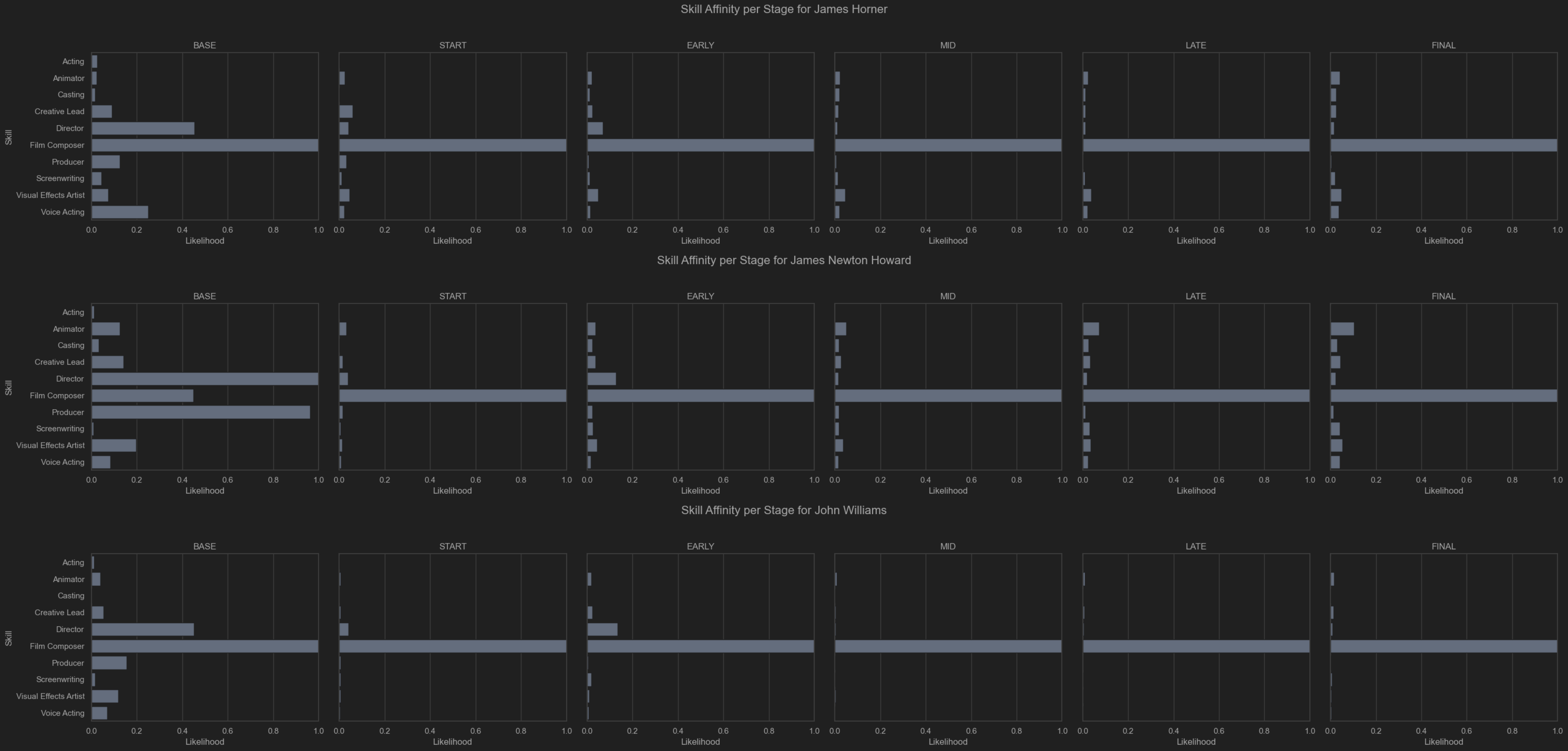

Skill Affinity per Stage for Ewan McGregor
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0 0.2 0.4 0.6 0.8 1.0
Likelihood

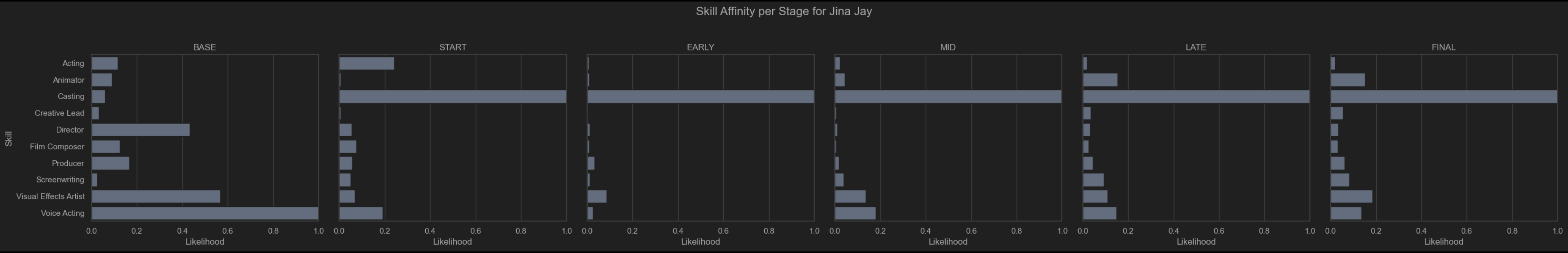

Skill Affinity per Stage for Harrison Ford
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0 0.2 0.4 0.6 0.8 1.0
Likelihood

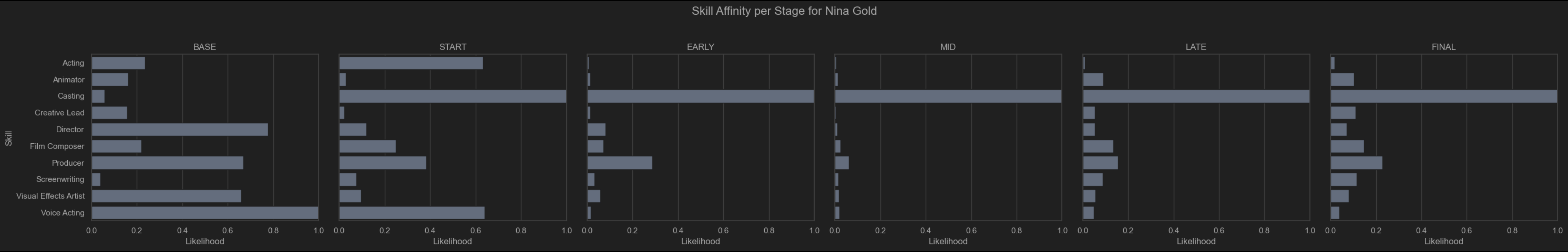

Skill Affinity per Stage for Frank Marshall
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood

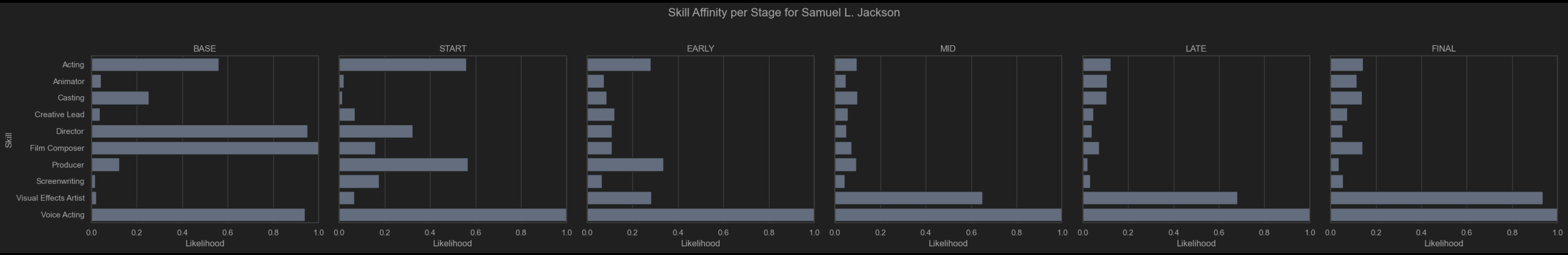

Skill Affinity per Stage for Kathleen Kennedy
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood

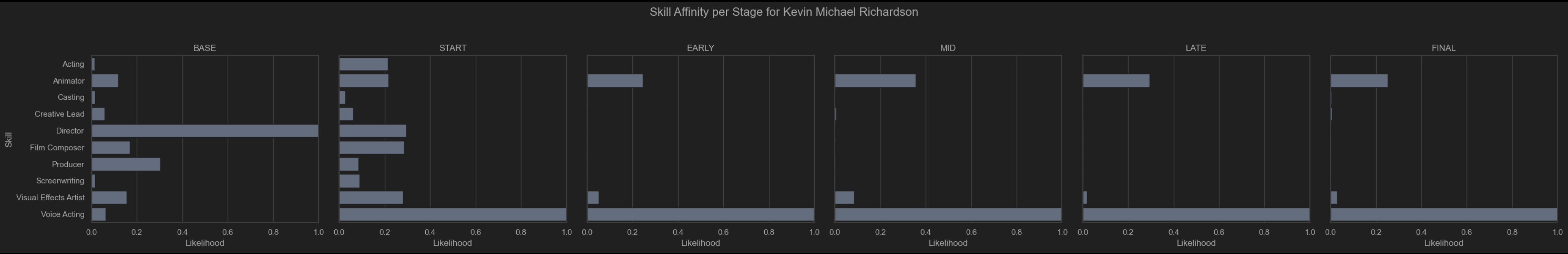

Skill Affinity per Stage for Steven Spielberg
Skill Affinity per Stage for Roger Corman
Skill Affinity per Stage for George Lucas
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0

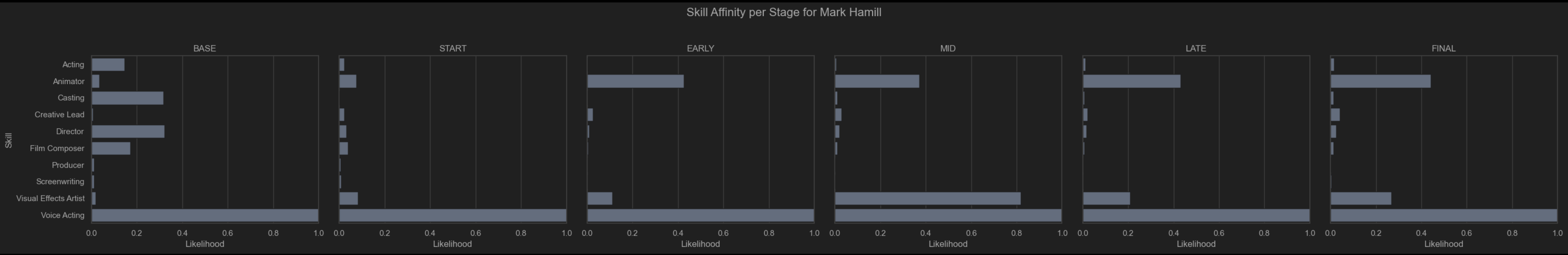

Skill Affinity per Stage for James Horner
Skill Affinity per Stage for James Newton Howard
Skill Affinity per Stage for John Williams
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0

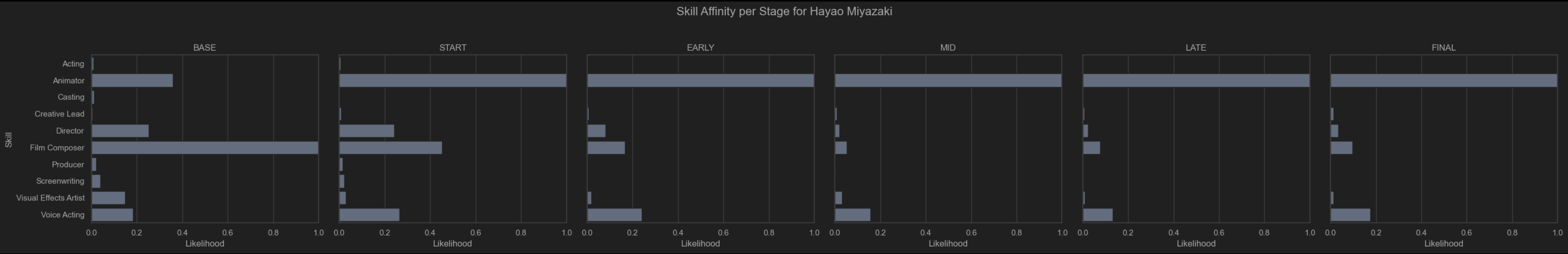

Skill Affinity per Stage for Jina Jay
BASE
START
EARLY
MID
LATE
FINAL
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Skill
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood


Skill Affinity per Stage for Nina Gold
BASE
START
EARLY
MID
LATE
FINAL
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Skill
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood

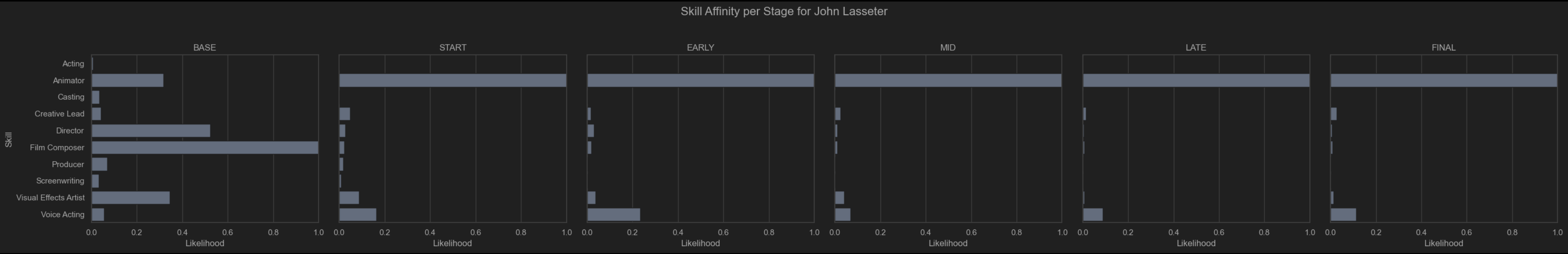

Skill Affinity per Stage for Samuel L. Jackson
BASE
START
EARLY
MID
LATE
FINAL
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Skill
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood


Skill Affinity per Stage for Kevin Michael Richardson
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood

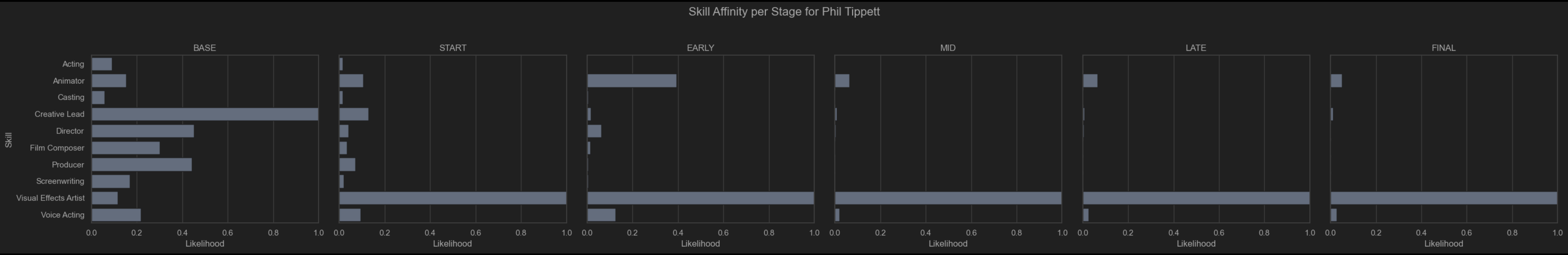

Skill Affinity per Stage for Mark Hamill
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0
0.2
0.4
0.6
0.8
1.0
Likelihood

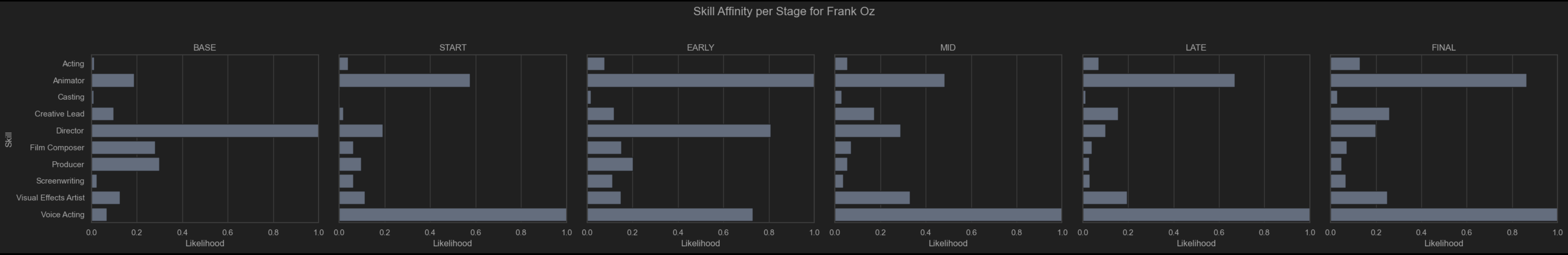


Skill Affinity per Stage for Hayao Miyazaki
BASE
START
EARLY
MID
LATE
FINAL
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Skill
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0



Skill Affinity per Stage for John Lasseter
BASE
START
EARLY
MID
LATE
FINAL
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
Skill
Likelihood
0.0
0.2
0.4
0.6
0.8
1.0


Skill Affinity per Stage for Phil Tippett
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0 0.2 0.4 0.6 0.8 1.0
Likelihood


Skill Affinity per Stage for Frank Oz
BASE
START
EARLY
MID
LATE
FINAL
Skill
Acting
Animator
Casting
Creative Lead
Director
Film Composer
Producer
Screenwriting
Visual Effects Artist
Voice Acting
0.0 0.2 0.4 0.6 0.8 1.0
Likelihood

Verdict

- Much more intuitive than GNNs
- Already has a comprehensive representation of the world
- Can separate entities and propagate information between ~~nodes~~ tokens
- Rooms for improvement
  + Increase batch size
  + Introduce multiple negatives
  + Introduce hard negatives
    - <person$_i$> | <person$_i$ description> VS <movie description>
    - <person$_i$> | <person$_i$ description> VS <person$_j$ description>