

Measuring Semantic Coherence of a Conversation

Svitlana Vakulenko, Maarten de Rijke, Michael Cochez,
Vadim Savenkov, Axel Polleres

WU

WIRTSCHAFTS
UNIVERSITÄT
WIEN VIENNA
UNIVERSITY OF
ECONOMICS
AND BUSINESS

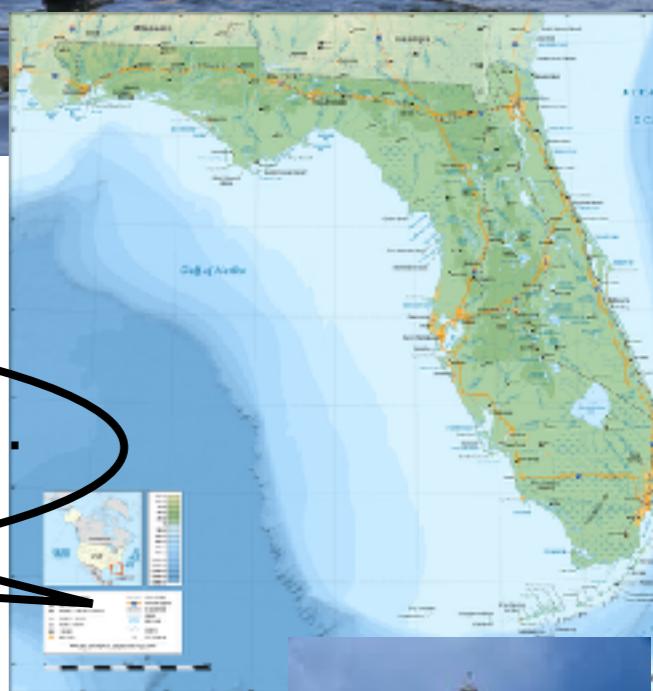


Semantic Coherence?

I think **Monterey** is a great conference location!



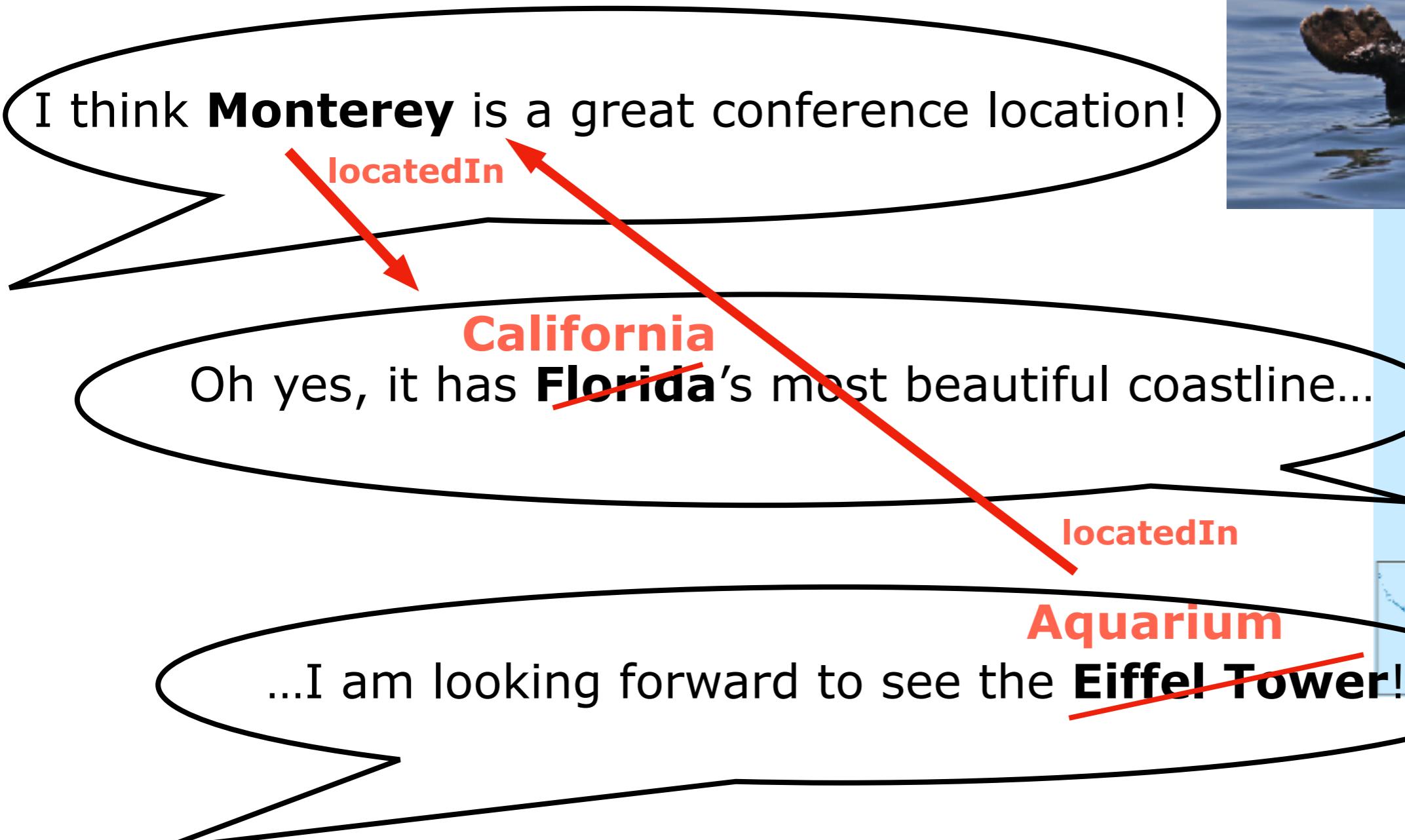
Oh yes, it has **Florida**'s most beautiful coastline...



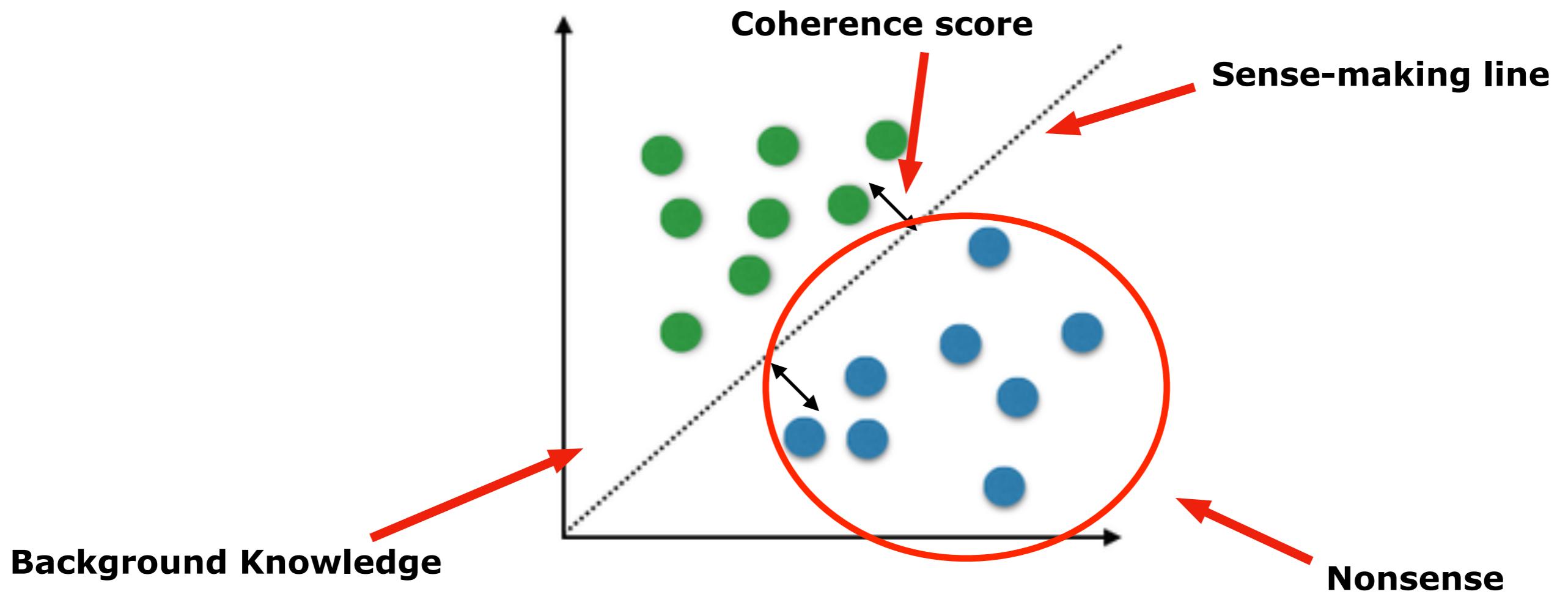
...I am looking forward to see the **Eiffel Tower**!



Semantic Coherence ~ Contextual Glue



Semantic Coherence IsA Classification Task



Applications

- **Conversational analysis**
 - reconstructing dialogs in a public chat
 - detecting topic shifts for segmentation



- **Conversational agents**
 - interpreting context
 - generating response



Contributions

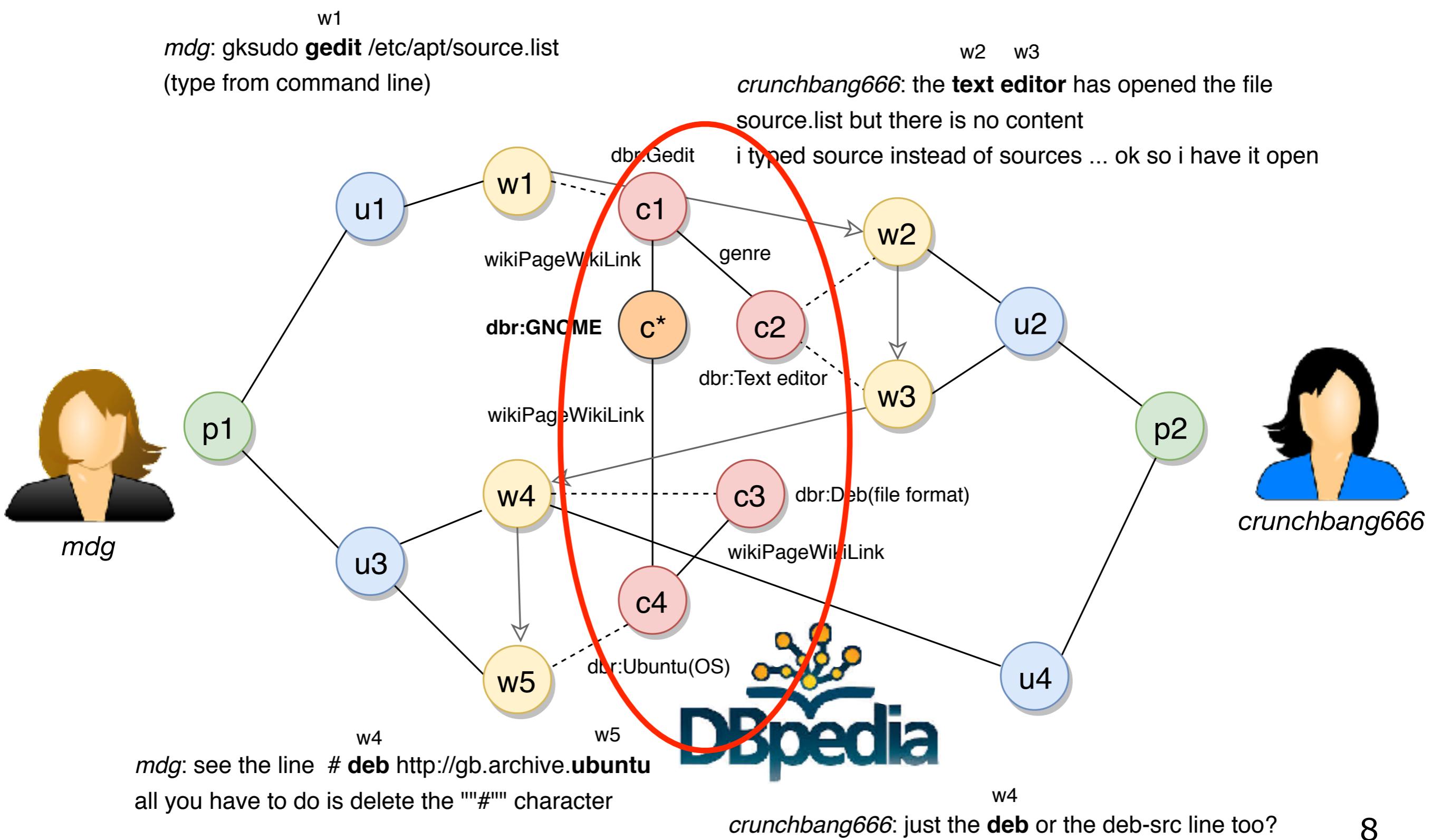
1. **Task** of measuring semantic coherence of a conversation
2. **Benchmark** for the semantic coherence task
3. **Approaches** and their evaluation:
 - 3.1. Subgraph induction approach
 - 3.2. Graph embeddings approach
 - 3.3. Word embeddings approach

Benchmark

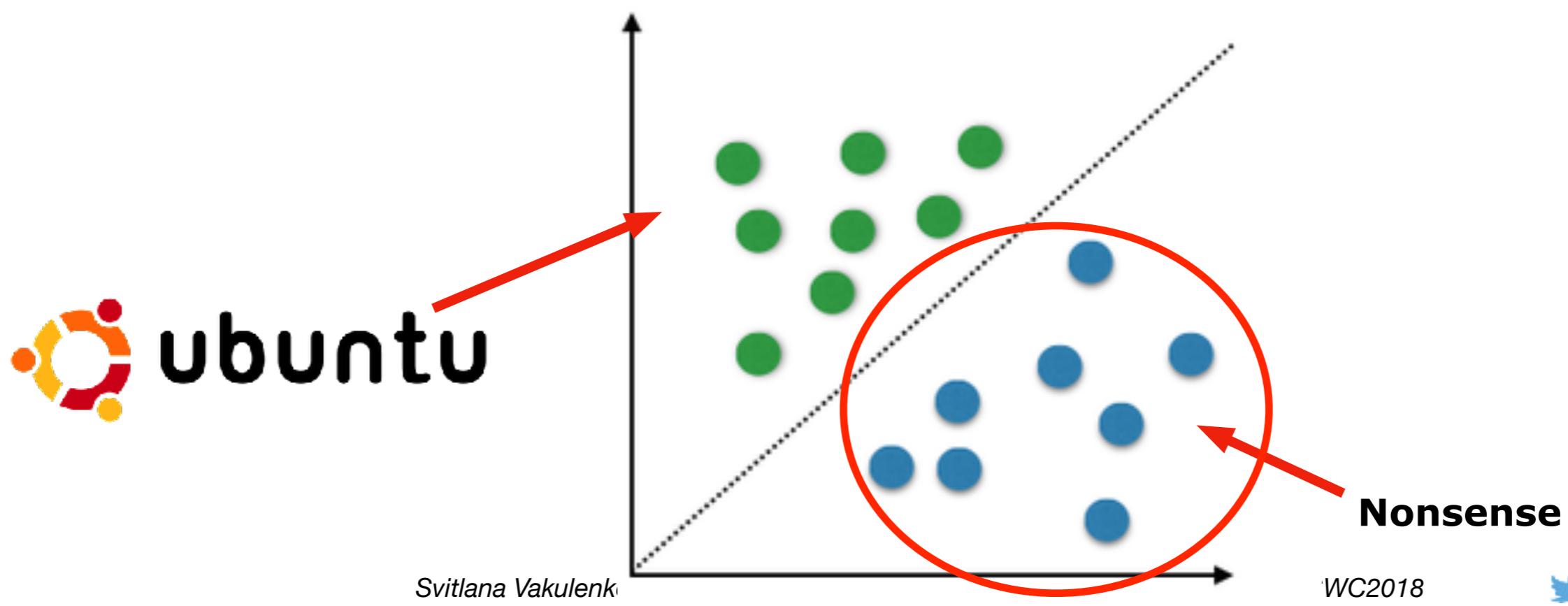
- **Conversational dataset**
 - Ubuntu Dialogue Corpus (IRC logs) ~2M
- **Knowledge representation** models
 - KGs: DBpedia+Wikidata HDT
 - KG embeddings: RDF2Vec, KGlove
 - Word embeddings: Word2Vec, Glove
- **Entity Linking:** DBpedia Spotlight API



Subgraph induction: Top-k Shortest Paths



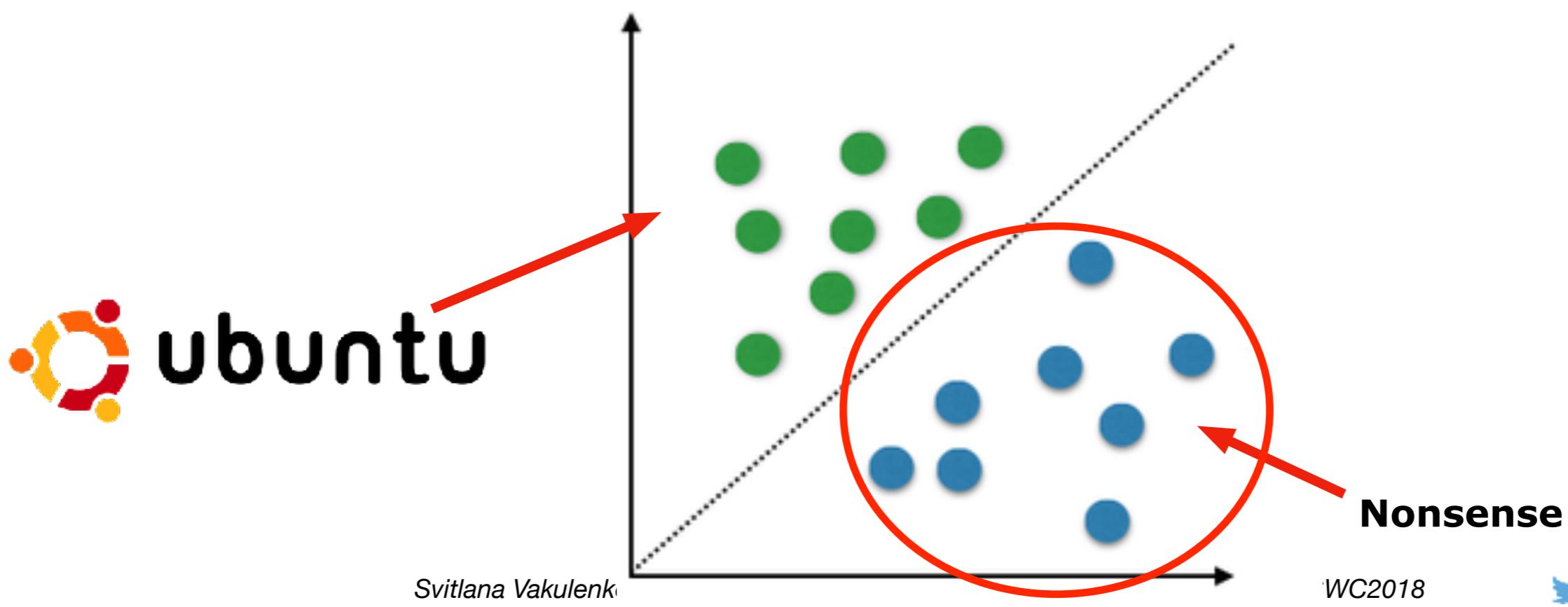
Benchmark: Incoherent Dialogues



Benchmark: Incoherent Dialogues

1. Vocabulary **sampling**

- 1.1. Random uniform
- 1.2. Vocabulary distribution



Benchmark: Incoherent Dialogues

1. Vocabulary **sampling**

■ 1.1.Random uniform

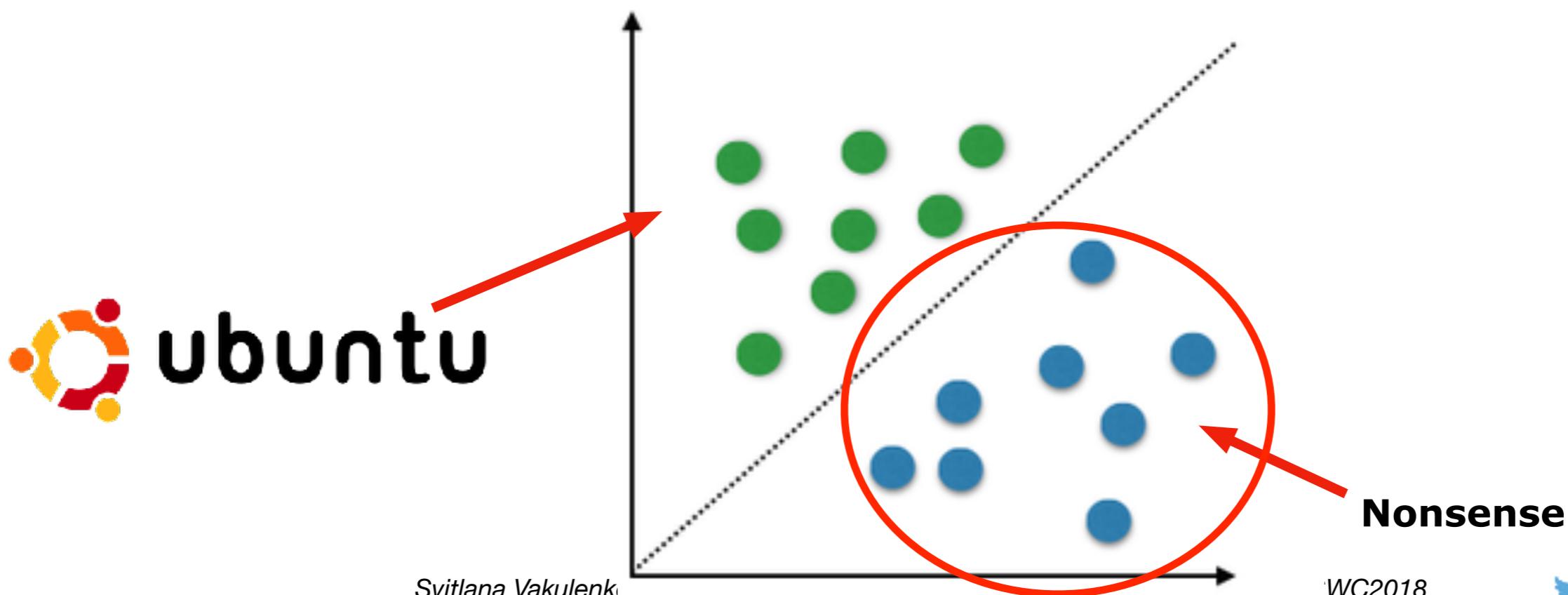
■ 1.2.Vocabulary distribution

2. Dialogue **permutations**

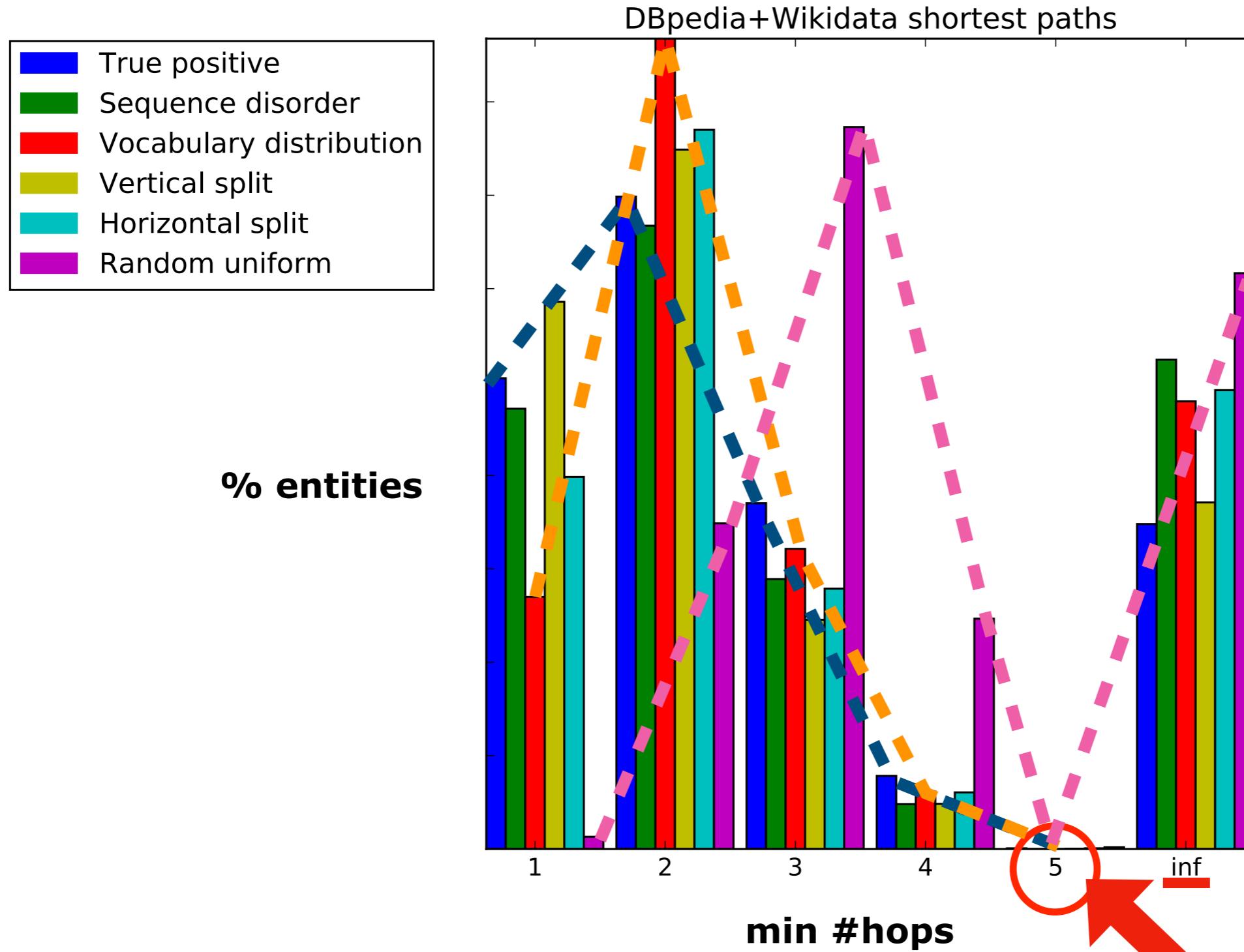
■ 2.1.Sequence disorder

■ 2.2.Horizontal split

■ 2.3.Vertical split

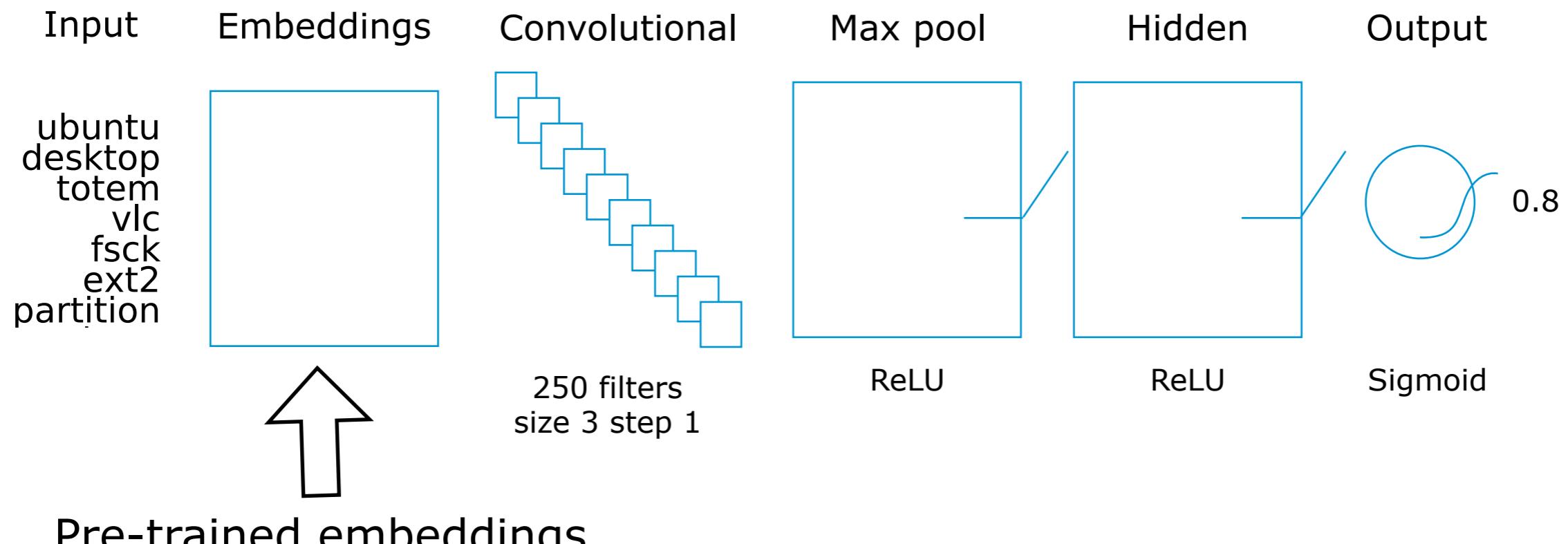


Subgraph Induction: Performance Bottleneck



Embeddings Classification Approach

- Convolutional Neural Network (CNN)
 - Input: sequence of words/entities
 - Output: **coherence score** [0;1]

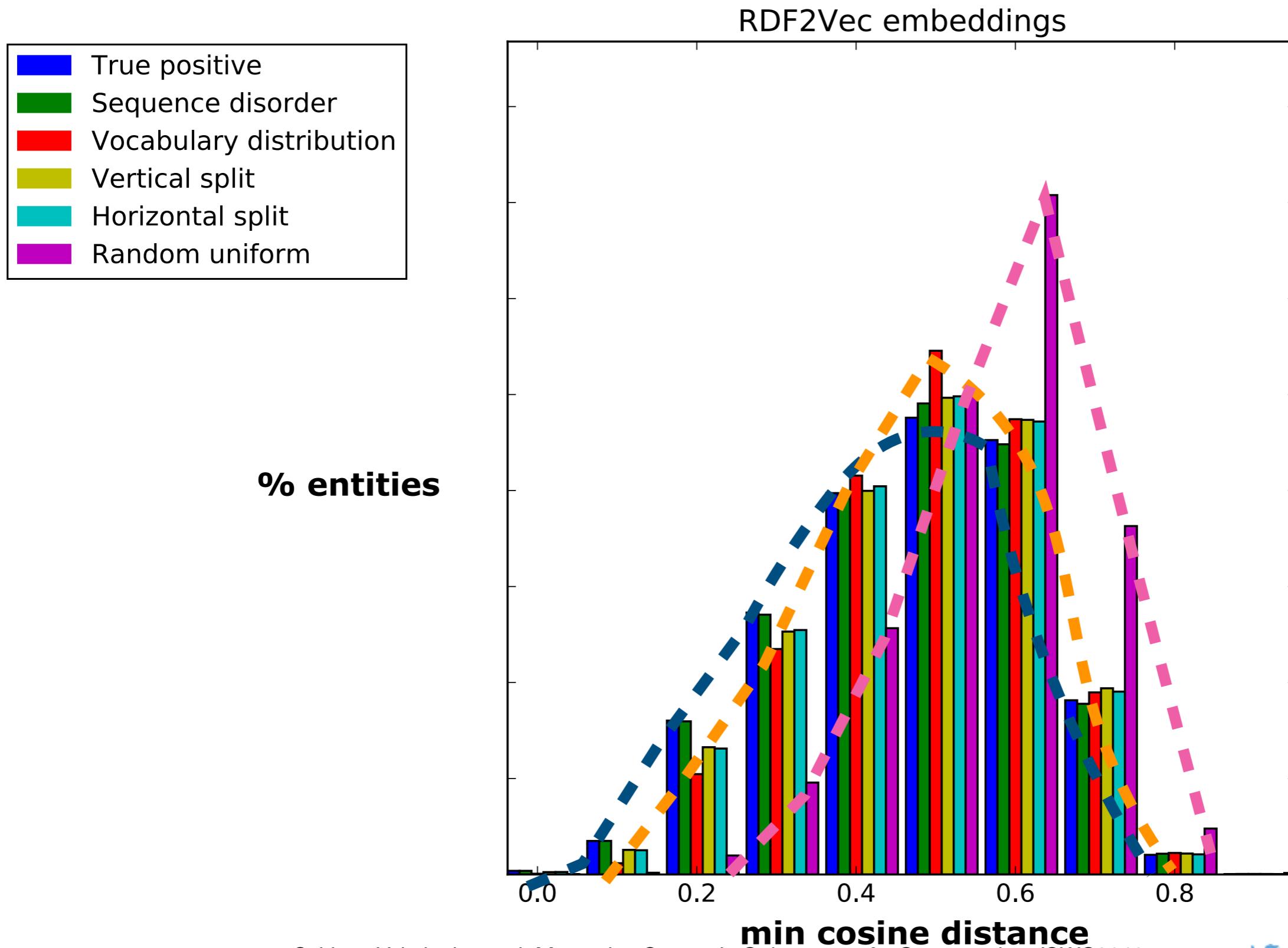


- Entities: RDF2Vec, KGlove
- Words: Word2Vec, Glove

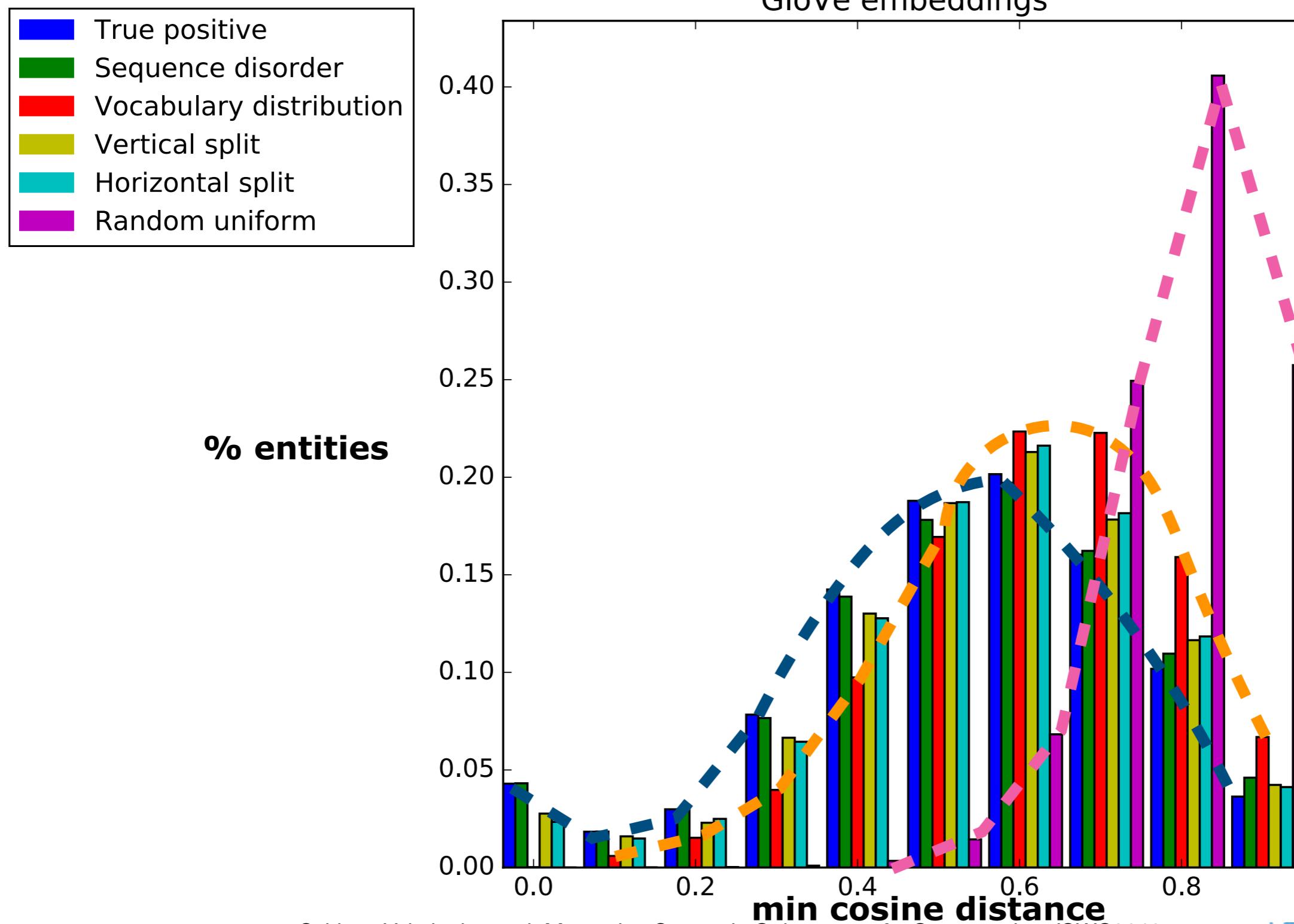
Results: Word Embeddings perform Best

| Embeddings | Data split | TPos | Accuracy | | | | | | | | | | | |
|-------------|------------|------|----------|-------------|-------------|-------------|------|-------------|------|-------------|------|-------------|-------------|-----|
| | | | RUF | Avg | VoD | Avg | SqD | Avg | VSp | Avg | HSp | Avg | Avg | Avg |
| Word2Vec | RUF | 0.99 | 0.99 | 0.99 | 0.02 | 0.50 | 0.02 | 0.50 | 0.01 | 0.50 | 0.01 | 0.50 | 0.60 | |
| | VoD | 0.89 | 0.62 | 0.75 | 0.90 | 0.89 | 0.53 | 0.71 | 0.18 | 0.54 | 0.20 | 0.54 | 0.69 | |
| | SqD | 0.75 | 0.65 | 0.70 | 0.88 | 0.81 | 0.81 | 0.78 | 0.27 | 0.51 | 0.29 | 0.52 | 0.66 | |
| | VSp | 0.59 | 0.50 | 0.55 | 0.82 | 0.71 | 0.41 | 0.50 | 0.59 | 0.59 | 0.61 | 0.60 | 0.59 | |
| | HSp | 0.62 | 0.39 | 0.50 | 0.71 | 0.66 | 0.38 | 0.50 | 0.55 | 0.58 | 0.63 | 0.63 | 0.58 | |
| GloVe | RUF | 0.99 | 0.99 | 0.99 | 0.00 | 0.50 | 0.01 | 0.50 | 0.00 | 0.50 | 0.00 | 0.50 | 0.60 | |
| | VoD | 0.93 | 0.38 | 0.66 | 0.93 | 0.93 | 0.39 | 0.66 | 0.19 | 0.56 | 0.08 | 0.51 | 0.66 | |
| | SqD | 0.76 | 0.71 | 0.73 | 0.91 | 0.84 | 0.82 | 0.79 | 0.16 | 0.46 | 0.15 | 0.45 | 0.66 | |
| | VSp | 0.60 | 0.25 | 0.42 | 0.92 | 0.76 | 0.43 | 0.51 | 0.65 | 0.62 | 0.66 | 0.63 | 0.59 | |
| | HSp | 0.71 | 0.34 | 0.52 | 0.81 | 0.76 | 0.30 | 0.50 | 0.55 | 0.63 | 0.66 | 0.68 | 0.62 | |
| rdf2vec PRS | RUF | 0.98 | 0.99 | 0.99 | 0.02 | 0.50 | 0.02 | 0.50 | 0.02 | 0.50 | 0.01 | 0.50 | 0.60 | |
| | VoD | 0.79 | 0.68 | 0.73 | 0.83 | 0.81 | 0.34 | 0.57 | 0.36 | 0.57 | 0.35 | 0.57 | 0.65 | |
| | SqD | 0.59 | 0.48 | 0.54 | 0.72 | 0.66 | 0.67 | 0.63 | 0.43 | 0.51 | 0.40 | 0.50 | 0.56 | |
| rdf2vec PR | HSp | 0.57 | 0.59 | 0.58 | 0.72 | 0.64 | 0.43 | 0.50 | 0.59 | 0.58 | 0.67 | 0.62 | 0.58 | |
| KGloVe Uni | RUF | 0.92 | 0.97 | 0.94 | 0.11 | 0.51 | 0.09 | 0.50 | 0.08 | 0.50 | 0.07 | 0.50 | 0.59 | |
| | VoD | 0.54 | 0.88 | 0.71 | 0.73 | 0.64 | 0.61 | 0.58 | 0.51 | 0.52 | 0.52 | 0.53 | 0.60 | |
| | SqD | 0.55 | 0.62 | 0.58 | 0.64 | 0.59 | 0.63 | 0.59 | 0.47 | 0.51 | 0.45 | 0.50 | 0.56 | |
| KGloVe PrO | HSp | 0.31 | 0.81 | 0.56 | 0.75 | 0.53 | 0.69 | 0.50 | 0.77 | 0.54 | 0.70 | 0.51 | 0.53 | |
| KGloVe PR | HSp | 0.47 | 0.69 | 0.58 | 0.61 | 0.54 | 0.54 | 0.50 | 0.57 | 0.52 | 0.65 | 0.56 | 0.54 | |

Results: KG Embeddings Classification



Results: Word Embeddings Classification

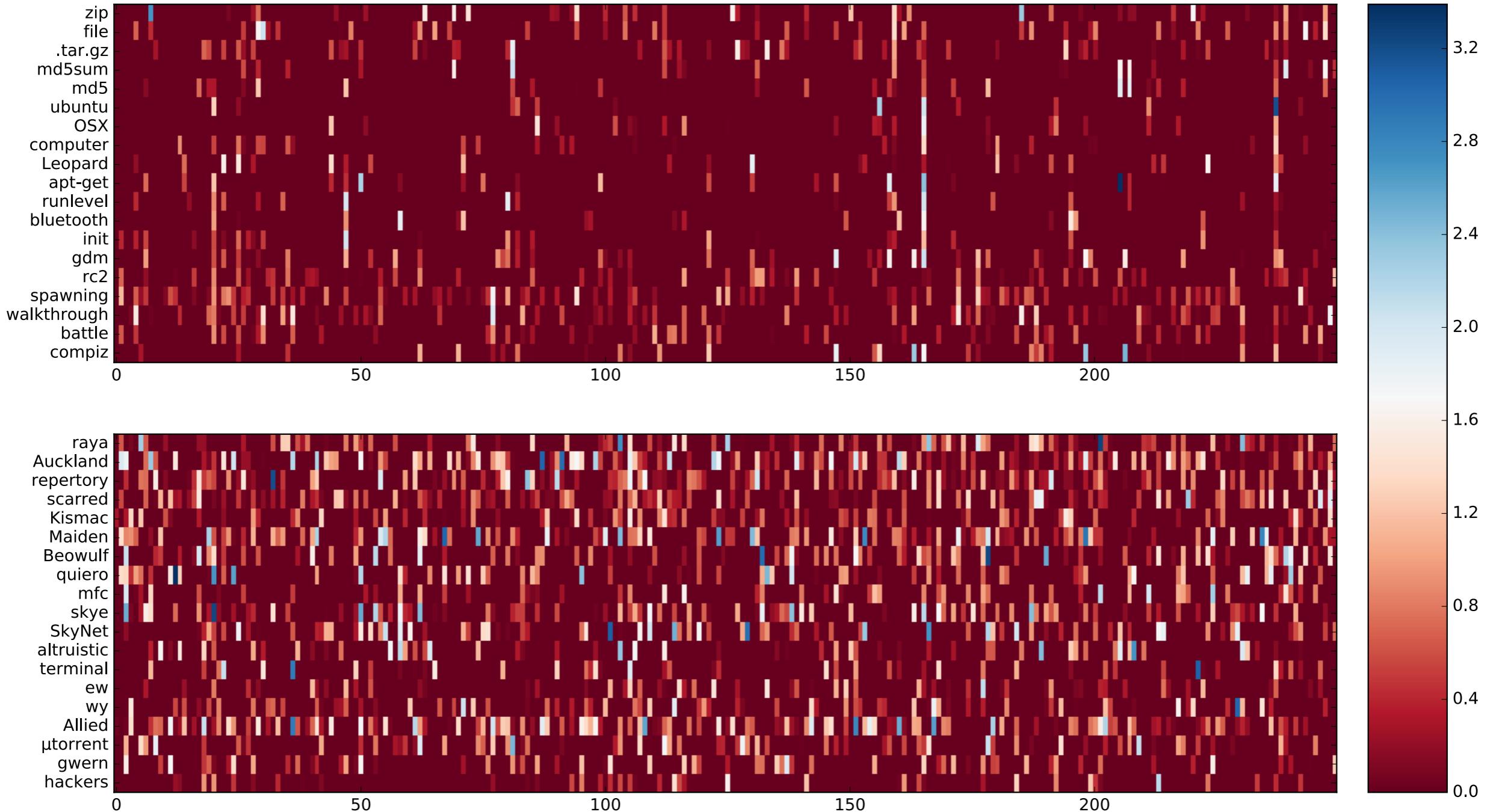


Results: Permutation IsA Difficult Task

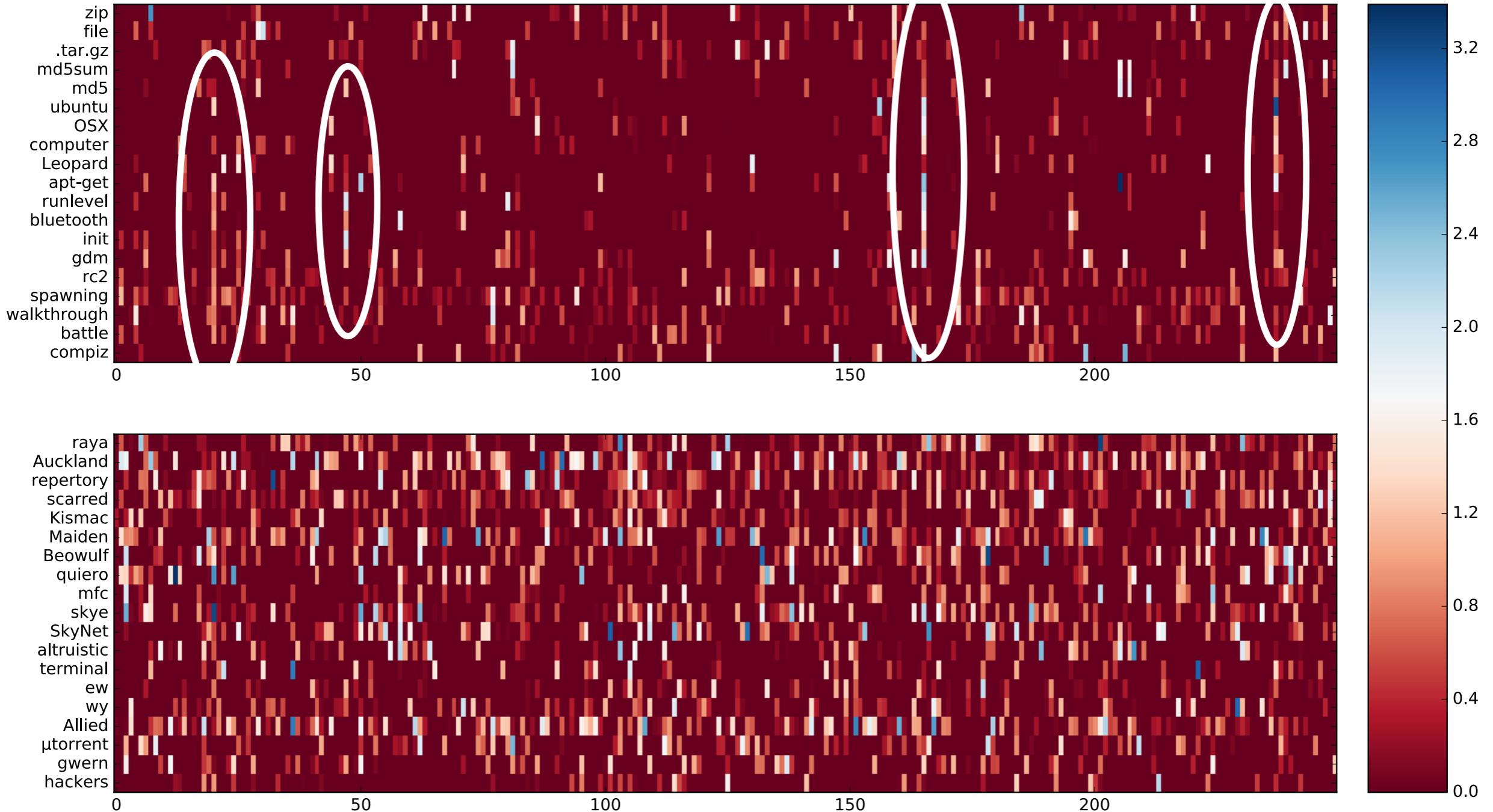


| Word | Embeddings | Data split | 1. sampling | | | | | | 2. permutations | | | | | | |
|----------|-------------|------------|-------------|----------|-------------|-------------|-------------|-------------|-----------------|----------|-------------|------|-------------|-------------|-----|
| | | | TPos | Accuracy | | | | | | Accuracy | | | | | |
| | | | | RUF | Avg | VoD | Avg | SqD | Avg | VSp | Avg | HSp | Avg | | Avg |
| Word2Vec | Word2Vec | RUF | 0.99 | 0.99 | 0.99 | 0.02 | 0.50 | 0.02 | 0.50 | 0.01 | 0.50 | 0.01 | 0.50 | 0.60 | |
| | | VoD | 0.89 | 0.62 | 0.75 | 0.90 | 0.89 | 0.53 | 0.71 | 0.18 | 0.54 | 0.20 | 0.54 | 0.69 | |
| | | SqD | 0.75 | 0.65 | 0.70 | 0.88 | 0.81 | 0.81 | 0.78 | 0.27 | 0.51 | 0.29 | 0.52 | 0.66 | |
| | | VSp | 0.59 | 0.50 | 0.55 | 0.82 | 0.71 | 0.41 | 0.50 | 0.59 | 0.59 | 0.61 | 0.60 | 0.59 | |
| | | HSp | 0.62 | 0.39 | 0.50 | 0.71 | 0.66 | 0.38 | 0.50 | 0.55 | 0.58 | 0.63 | 0.63 | 0.58 | |
| KG | GloVe | RUF | 0.99 | 0.99 | 0.99 | 0.00 | 0.50 | 0.01 | 0.50 | 0.00 | 0.50 | 0.00 | 0.50 | 0.60 | |
| | | VoD | 0.93 | 0.38 | 0.66 | 0.93 | 0.93 | 0.39 | 0.66 | 0.19 | 0.56 | 0.08 | 0.51 | 0.66 | |
| | | SqD | 0.76 | 0.71 | 0.73 | 0.91 | 0.84 | 0.82 | 0.79 | 0.16 | 0.46 | 0.15 | 0.45 | 0.66 | |
| | | VSp | 0.60 | 0.25 | 0.42 | <u>0.92</u> | 0.76 | 0.43 | 0.51 | 0.65 | 0.62 | 0.66 | 0.63 | 0.59 | |
| | | HSp | 0.71 | 0.34 | 0.52 | 0.81 | 0.76 | 0.30 | 0.50 | 0.55 | 0.63 | 0.66 | 0.68 | 0.62 | |
| KG | rdf2vec PRS | RUF | 0.98 | 0.99 | 0.99 | 0.02 | 0.50 | 0.02 | 0.50 | 0.02 | 0.50 | 0.01 | 0.50 | 0.60 | |
| | | VoD | 0.79 | 0.68 | 0.73 | 0.83 | 0.81 | 0.34 | 0.57 | 0.36 | 0.57 | 0.35 | 0.57 | 0.65 | |
| | | SqD | 0.59 | 0.48 | 0.54 | 0.72 | 0.66 | 0.67 | 0.63 | 0.43 | 0.51 | 0.40 | 0.50 | 0.56 | |
| | rdf2vec PR | HSp | 0.57 | 0.59 | 0.58 | 0.72 | 0.64 | 0.43 | 0.50 | 0.59 | 0.58 | 0.67 | 0.62 | 0.58 | |
| KG | KGloVe Uni | RUF | 0.92 | 0.97 | 0.94 | 0.11 | 0.51 | 0.09 | 0.50 | 0.08 | 0.50 | 0.07 | 0.50 | 0.59 | |
| | | VoD | 0.54 | 0.88 | 0.71 | 0.73 | 0.64 | 0.61 | 0.58 | 0.51 | 0.52 | 0.52 | 0.53 | 0.60 | |
| | | SqD | 0.55 | 0.62 | 0.58 | 0.64 | 0.59 | 0.63 | 0.59 | 0.47 | 0.51 | 0.45 | 0.50 | 0.56 | |
| | KGloVe PrO | HSp | 0.31 | 0.81 | 0.56 | 0.75 | 0.53 | 0.69 | 0.50 | 0.77 | 0.54 | 0.70 | 0.51 | 0.53 | |
| | | HSp | 0.47 | 0.69 | 0.58 | 0.61 | 0.54 | 0.54 | 0.50 | 0.57 | 0.52 | 0.65 | 0.56 | 0.54 | |

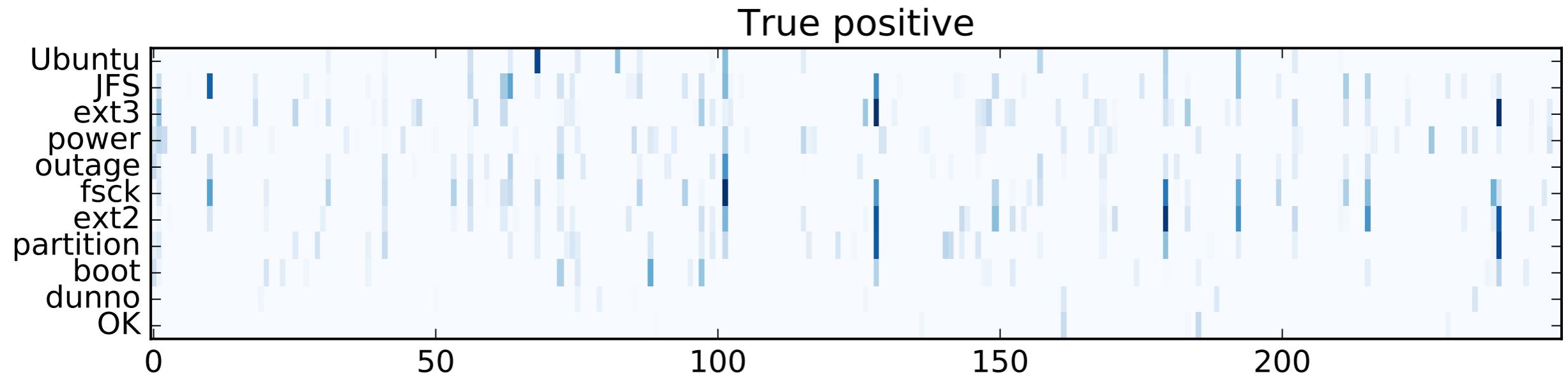
Coherence Patterns



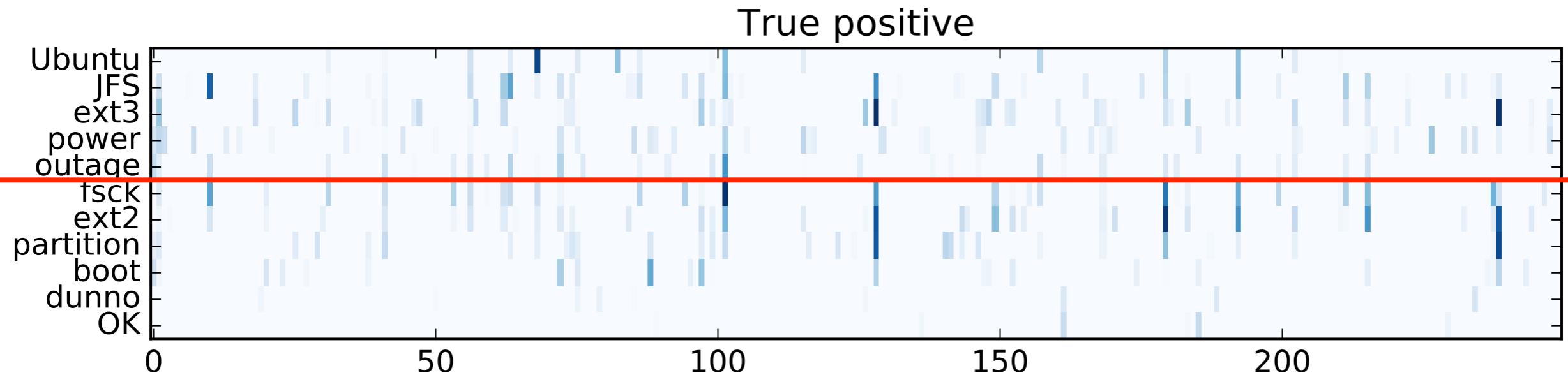
Coherence Patterns



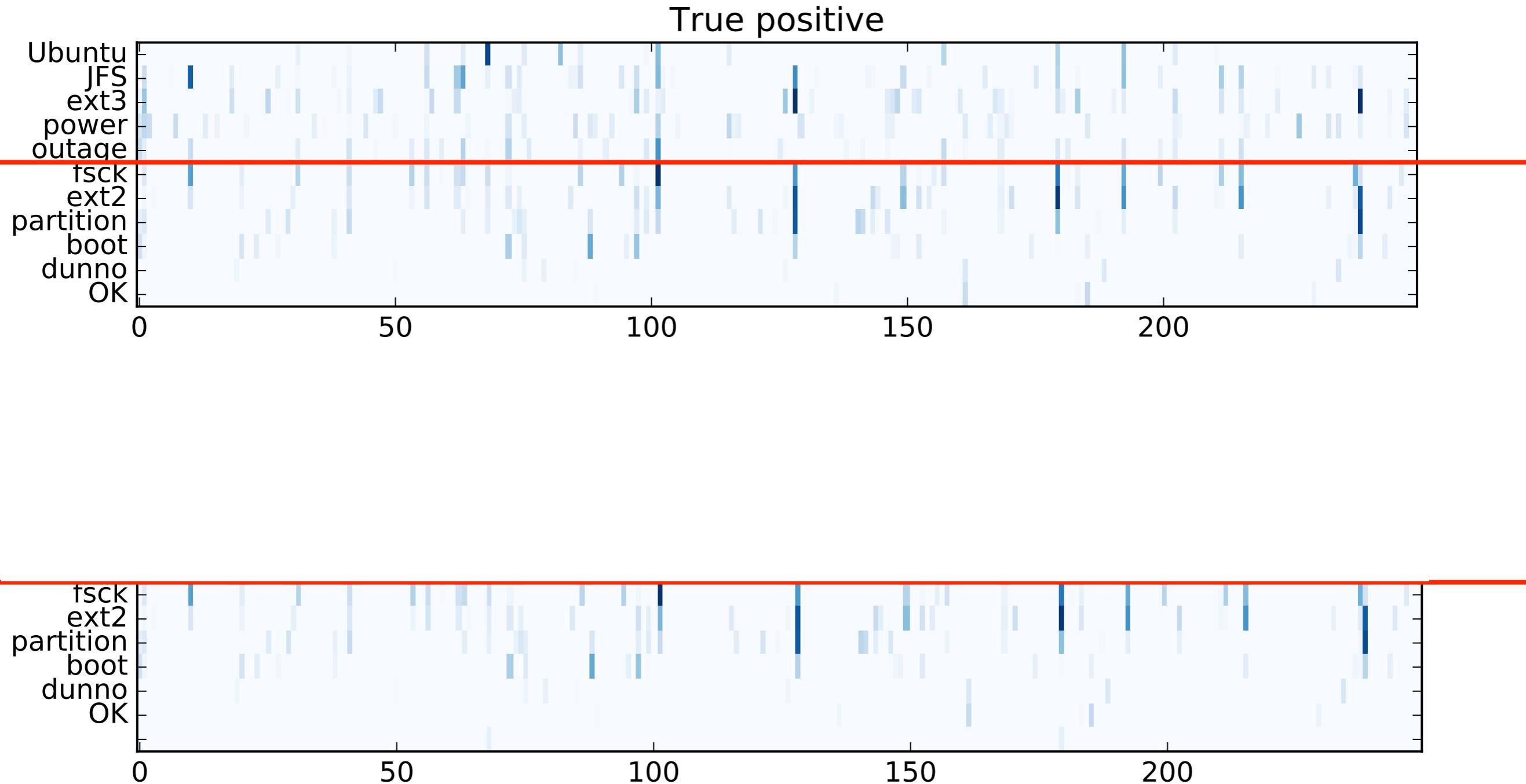
Horizontal Split



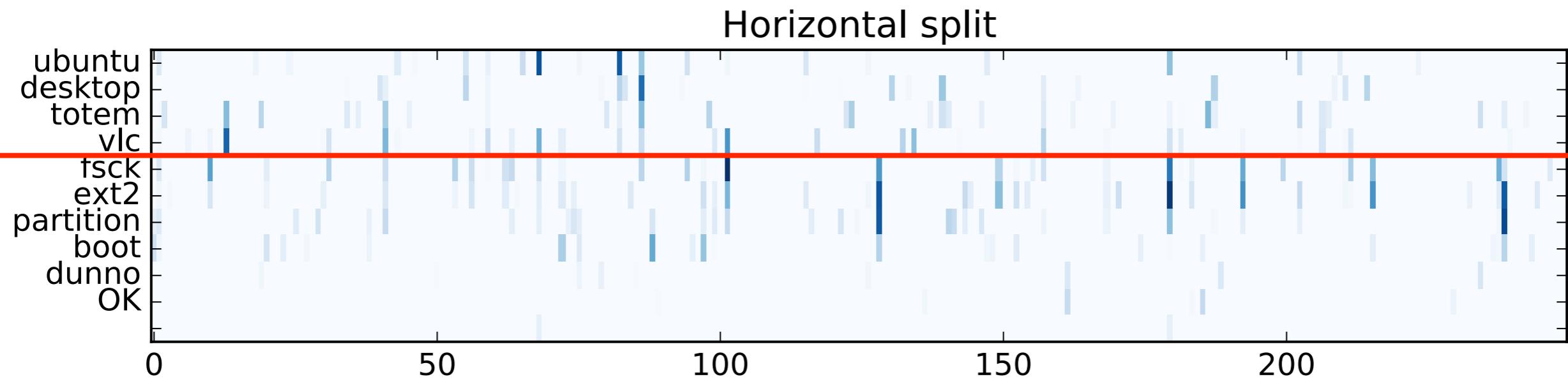
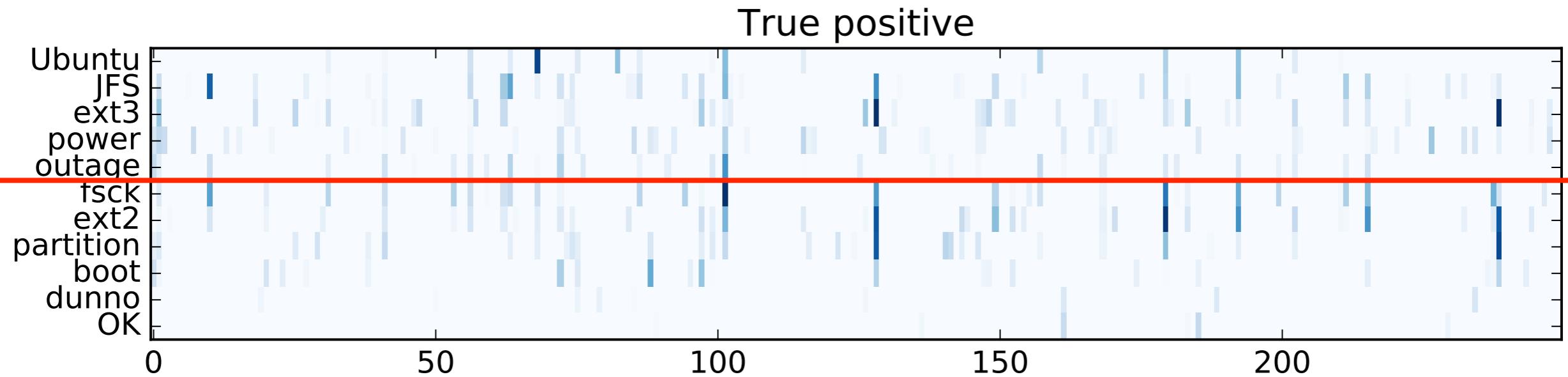
Horizontal Split



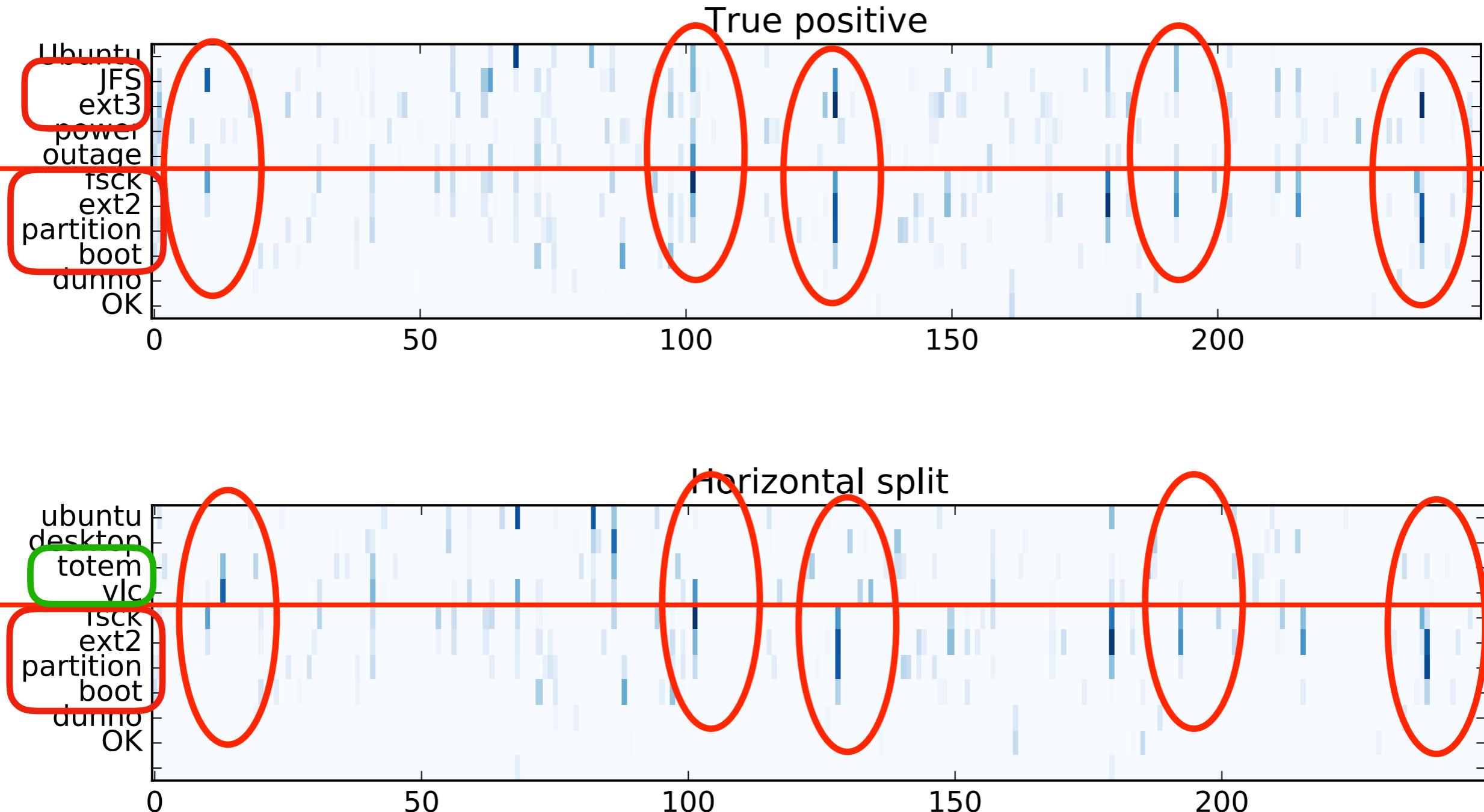
Horizontal Split



Horizontal Split



Coherence Patterns



Future Work

- Extend coherence measure beyond **KG entities**
- **Integration** of KG and word embeddings
- End-to-end training including **entity linking** layer

Open source: [https://github.com/svakulenkov/semantic coherence](https://github.com/svakulenkov/semantic_coherence)

Acknowledgements

THIS WORK WAS SUPPORTED BY THE FOLLOWING PROJECTS:

EU H2020 PROGRAMME UNDER THE MSCA-RISE AGREEMENT 645751 (RISE_BPM)

*PROJECT OPEN DATA FOR LOCAL COMMUNITIES FUNDED BY THE
AUSTRIAN FEDERAL MINISTRY OF TRANSPORT, INNOVATION AND
TECHNOLOGY (BMVIT) UNDER THE PROGRAM "ICT OF THE FUTURE",
BETWEEN NOVEMBER 2016 AND APRIL 2019.*

MORE INFORMATION [HTTPS://IKTDERZUKUNFT.AT/EN/](https://iktderzukunft.at/en/)



*Austrian Ministry
for Transport,
Innovation and Technology*