Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Alexander Panchenko

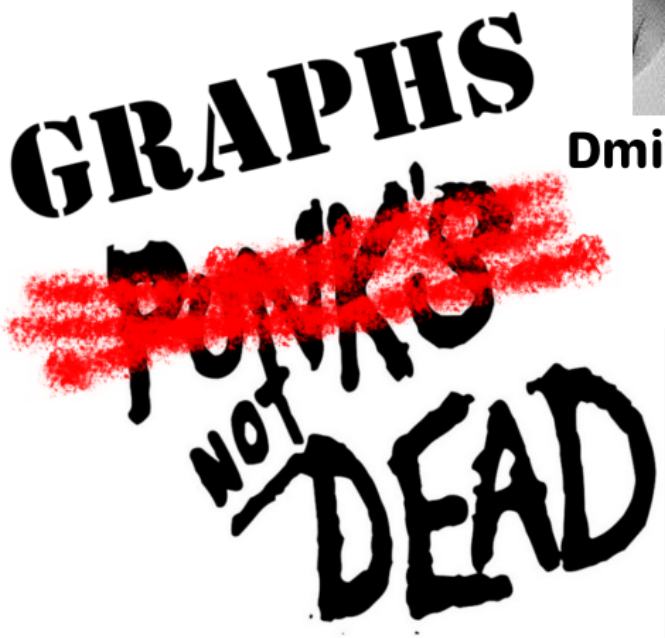
FROM UNSUPERVISED INDUCTION OF LINGUISTIC STRUCTURES FROM TEXT TOWARDS APPLICATIONS IN DEEP LEARNING

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

In close collaboration with ...

Chris Biemann

Dmitry 'Tsar' Ustalov



Stefano Faralli

Simone P. Ponzetto

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

In collaboration with ...

- **Andrei Kutuzov**
- **Eugen Ruppert**
- **Fide Marten**
- **Nikolay Arefyev**
- **Steffen Remus**
- **Martin Riedl**
- **Hubert Naets**
- **Maria Pelevina**
- **Anastasiya Lopukhina**
- **Konstantin Lopukhin**

Motivation

Deep Learning: everything is a vector

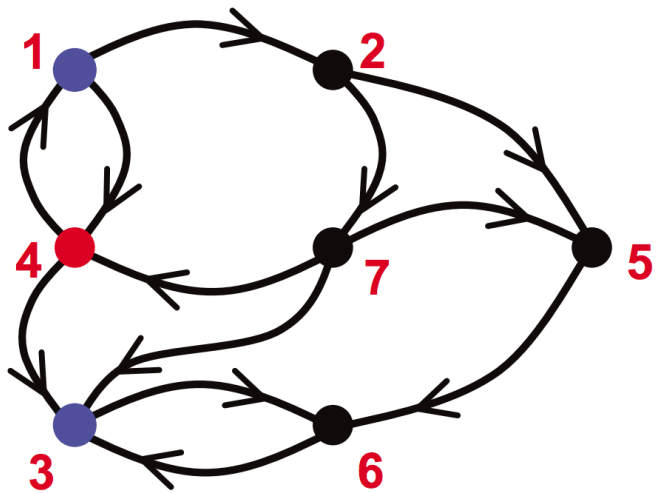
Levels of Linguistic Analysis

Image source: `https://commons.wikimedia.org/wiki/File:Major_levels_of_linguistic_structure.svg`

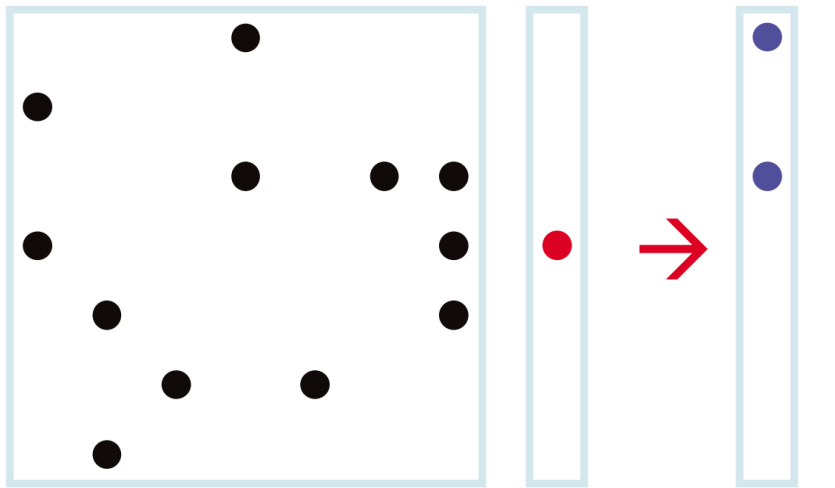
Linguistic Structures and Graphs

- (Written) language is a symbolic system

Graph Matrix Duality

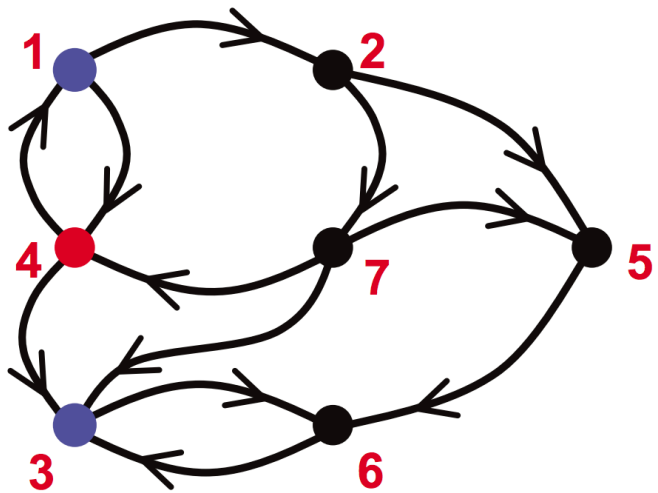


$$G = (V,E)$$

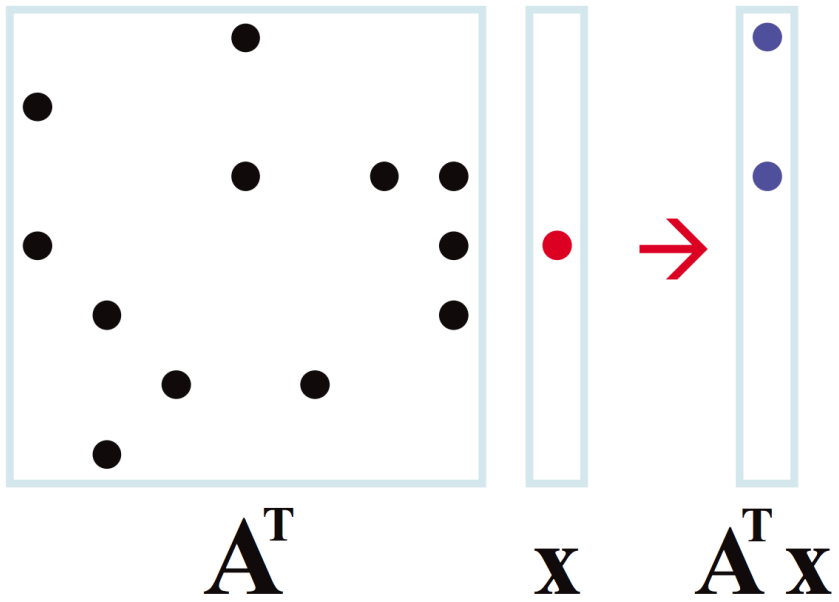


$$\mathbf{A}^T \quad \mathbf{x} \quad \mathbf{A}^T\mathbf{x}$$

Graph Matrix Duality



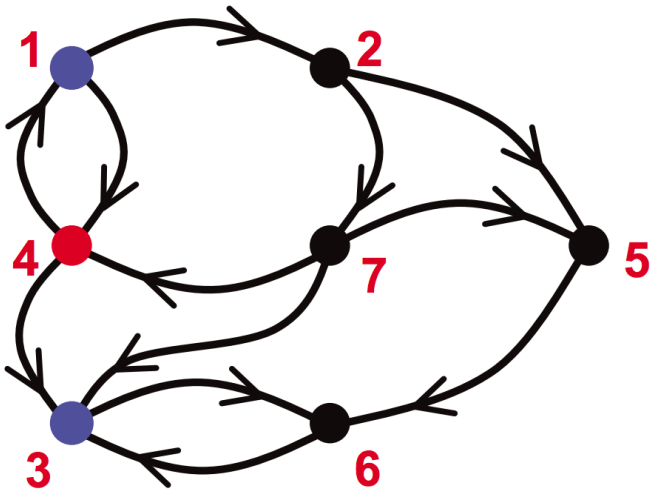
$$G = (V,E)$$



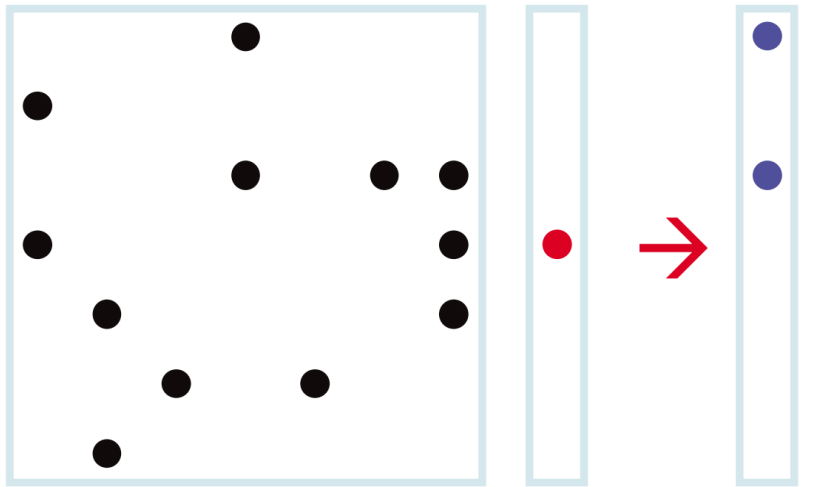
$$\mathbf{A}^T \quad \mathbf{x} \quad \mathbf{A}^T\mathbf{x}$$

- Adjacency matrix $\mathbf{A}$ is dual with the corresponding graph G.

Graph Matrix Duality



$$G = (V,E)$$



$$\mathbf{A}^{\mathbf{T}} \quad \mathbf{x} \quad \mathbf{A}^{\mathbf{T}}\mathbf{x}$$

- Adjacency matrix $\mathbf{A}$ is dual with the corresponding graph G.
- Vector matrix multiply $\mathbf{A}^T\mathbf{x}$ is dual with breadth-first search.

Goal

1. Learn the interpretable symbolic structures from text in an unsupervised way, which are more complex than tokens and lemmas.

Goal

1. Learn the interpretable symbolic structures from text in an unsupervised way, which are more complex than tokens and lemmas.
2. Represent the learned structures in the vector form.

Goal

1. Learn the interpretable symbolic structures from text in an unsupervised way, which are more complex than tokens and lemmas.
2. Represent the learned structures in the vector form.
3. Use the vector representations instead/in addition to word embedding the deep learning applications.

Goal

1. Learn the interpretable symbolic structures from text in an unsupervised way, which are more complex than tokens and lemmas.
2. Represent the learned structures in the vector form.
3. Use the vector representations instead/in addition to word embedding the deep learning applications.
4. More complex structures could improve performance, but also provide better interpretability of the deep learning models.