

# VA3: Síťová analýza jako nástroj studia racionality textu / vědeckého poznání

Petra Vidnerová

**TRUST Kick-off Meeting, 27. února, 2025**



Spolufinancováno  
Evropskou unií

**MSMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

Tato akce je organizována v rámci projektu „Vědění ve věku nedůvěry“, reg. č.: CZ.02.01.01/00/23\_025/0008711, financovaného z Evropského fondu pro regionální rozvoj.



Spolufinancováno  
Evropskou unií



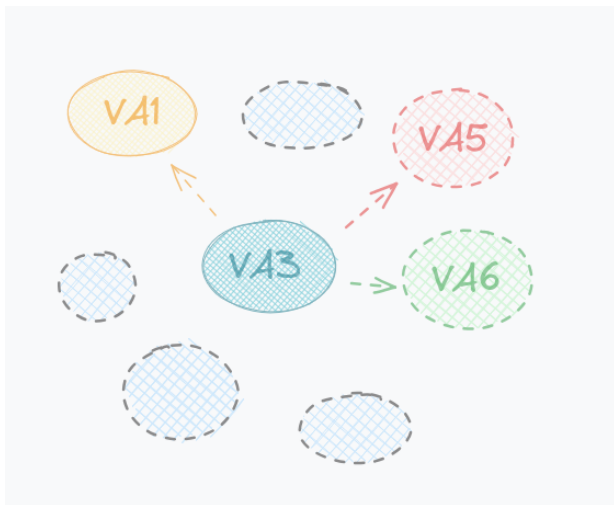
- ▶ Petra Vidnerová
- ▶ Roman Neruda
- ▶ Jaroslav Hlinka
- ▶ Pavel Šanda
- ▶ Zdeněk Kratochvíl
- ▶ Radim Kočandrlle
- ▶ Lenka Ambrožová
- ▶ Vojtěch Kaše



Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Jsme otevřeni spolupráci



Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Textová analýza

## Úlohy

- ▶ Klasifikace textů
- ▶ Analýza sentimentu
- ▶ Shluková analýza
- ▶ Identifikace klíčových slov

## Metody

- ▶ Klasické metody strojového učení
- ▶ Statistické metody
- ▶ Metody na interpretaci ML
- ▶ Nyní hlavně velké jazykové modely



Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Využití velkých jazykových modelů

## LLMs

- ▶ ChatGPT definice: Umělá inteligence pro zpracování přirozeného jazyka
- ▶ Používají neuronové sítě (transformery)
- ▶ Trénované na rozsáhlých textech
- ▶ Schopnost generovat a analyzovat text na základě instrukcí od uživatele

## Aplikace

- ▶ Klasifikace textu (např. sentiment, témata)
- ▶ Extrahování informací (pojmy, jména, vztahy)
- ▶ Shrnutí textu (automatická abstrakce obsahu)
- ▶ Generování textu (automatické odpovědi, doplňování textu)



Spolufinancováno  
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Využití velkých jazykových modelů

Here are the top ten keywords extracted from the document in JSON format:

json

Copy Edit

```
{
  "keywords": [
    "Network Text Analysis",
    "text network",
    "nodes",
    "edges",
    "semantic networks",
    "word co-occurrence",
    "centrality measures",
    "topic modeling",
    "graph algorithms",
    "knowledge graphs"
  ]
}
```

Let me know if you need any adjustments!



Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Využití velkých jazykových modelů

## Nevýhody LLMs

- ▶ Velká výpočetní náročnost
- ▶ Omezení na počet dotazů nebo placená verze

## Malé velké modely

- ▶ Redukované verze, možnost provozovat na běžném počítači nebo výpočetním clusteru na ústavu
- ▶ Mixtral-8x22B-Instruct-v0.1 (141B params)
- ▶ tiger-gemma-9b-v3 (9.24B params)

```
```json
{"keywords": ["network text analysis", "theoretical framework", "advantages", "disadvantages", "open-source tools", "text network", "information professionals", "research questions", "visual representation", "libraries"]}
```
```



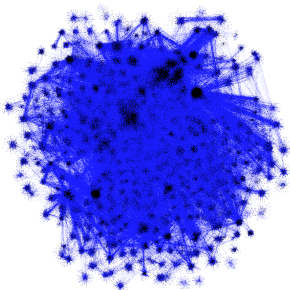
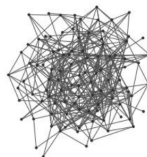
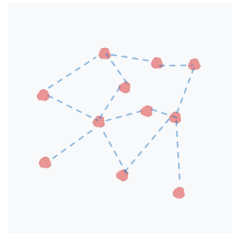
Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY



# Síťová analýza

- ▶ Technika modelování vztahů mezi entitami
- ▶ Síť (graf) - uzly, hrany
- ▶ Vztahy mezi dokumenty
- ▶ Vztahy mezi pojmy



Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Multi-agentní síťové modely

## Agent

- ▶ Entita, která vykonává akce nebo přijímá rozhodnutí na základě interakcí v síti (uzel v grafu)
- ▶ Interaguje s ostatními agenty
- ▶ Má vnitřní stav (paměť)

## Aplikace

- ▶ Modelování dynamiky v síti
- ▶ Šíření infekce, šíření informací, šíření klíčových slov



Spolufinancováno  
Evropskou unií

**MŠMT**  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Aplikace: síťová reprezentace textových pramenů

- ▶ V rámci VA3 primárně presokratických zlomků
- ▶ V rámci VA6 středověkých kvestií (spolu s O. Pavlíčkem)
- ▶ Novodobých akademických textů
  
- ▶ Použití klustrovacích metod na určení sémantických podčástí

## Výzkumné otázky

- ▶ Umožňuje tato reprezentace rozlišit různé modely myšlení?
- ▶ Může pomoci jako interpretační vodítko, jak postupovat při čtení celého textu (zejména u presokratiků)

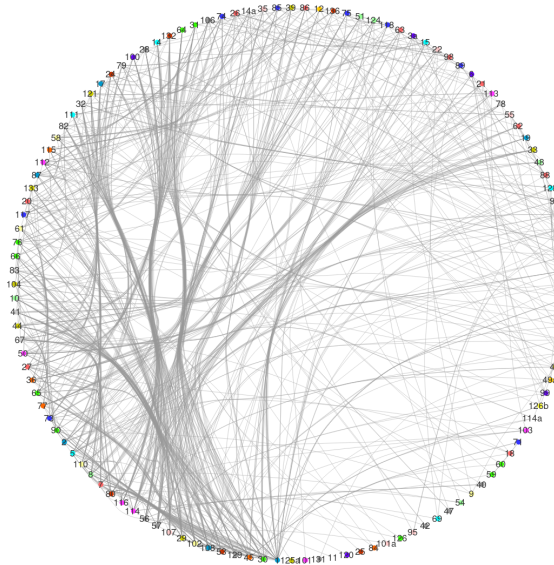


Spolufinancováno  
Evropskou unií



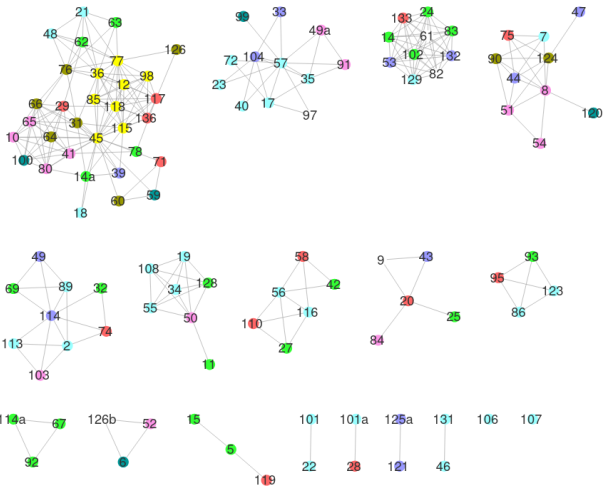
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Ukázka: síť propojení Herakleitových zlomků (uzel zlomek, hrana společné klíčové slovo)



Spolufinancováno  
Evropskou unií

# Ukázka: Shlukování do různých sémantických ostrovů



Spolufinancováno  
Evropskou unií

MT  
MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Ukázka: Shlukování klíčových slov



Spolufinancováno  
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

# Plány na nejbližší období

- ▶ Pečlivé otagování presokratických textů
- ▶ Studium výrazových prostředků ve filosofických počátcích formálně zdůvodněné argumentace (Parmenides/Xenofanes)
- ▶ (V rámci VA6) algoritmy pro segmentaci středověkých kvestií za pomoci neuronových sítí, pokus o OCR
  
- ▶ Průzkum možností volně dostupných LLM pro textovou analýzu



Spolufinancováno  
Evropskou unií