

Citační data a citační síť

Jakub Harašta

Vytěžení citací

- Dataset soudních rozhodnutí
 - Nejsou všechna; 1.1.1993 až 30.9.2018
- Identifikace segmentu ‚argumentace soudu‘
 - Není perfektní
- Identifikace referencí v segmentech
 - Není perfektní
- Evaluace, limity
 - IAA 70,62% (*strict*) / 81,08% (*lenient/overlap*) / 80,36% (*gold set*)
 - Predikce vs. IAA 65,22% (*strict*) / 70,86% (*lenient/overlap*)
 - Pipeline složená ze segmentace a pak rozpoznání reference vykazuje lepší výsledky

Čištění dat

- Identifikace typických chyb
- Manuální odstraňování typických chyb
- Násobná kontrola kvůli možným překlepům
- Párování referencí s rozhodnutími v datasetu

Dataset

ÚS cituje ÚS	137 308
ÚS cituje NS	8 445
ÚS cituje NSS	2 648
ÚS cituje ostatní	45 690

NS cituje NS	153 242
NS cituje ÚS	80 658
NS cituje NSS	804
NS cituje ostatní	112 564

NSS cituje NSS	91 223
NSS cituje ÚS	14 576
NSS cituje NS	1 340
NSS cituje ostatní	20 710

Vizualizace a analýza dat

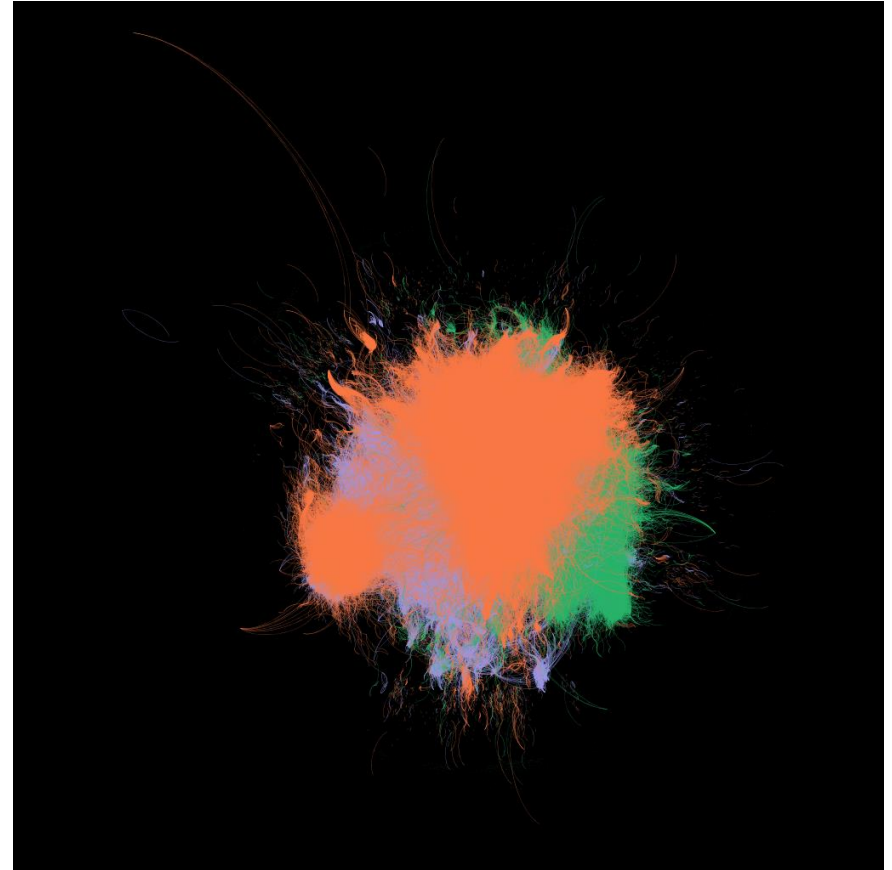
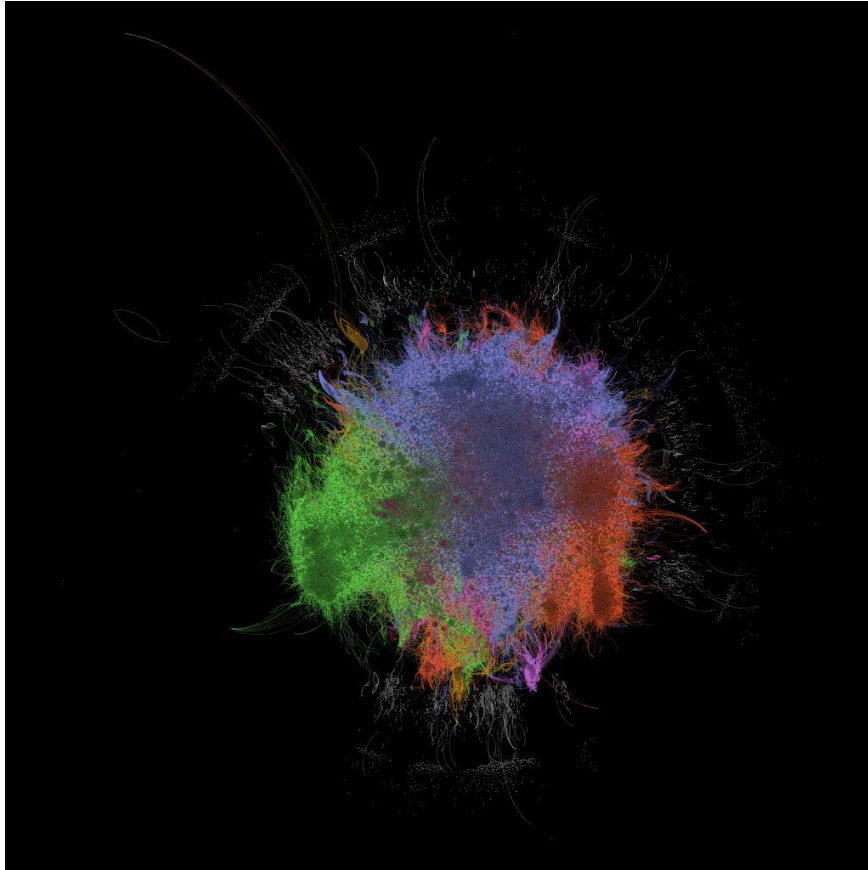
– Vizualizace

- To „wow“
- Nástroj: Gephi

– Analýza dat

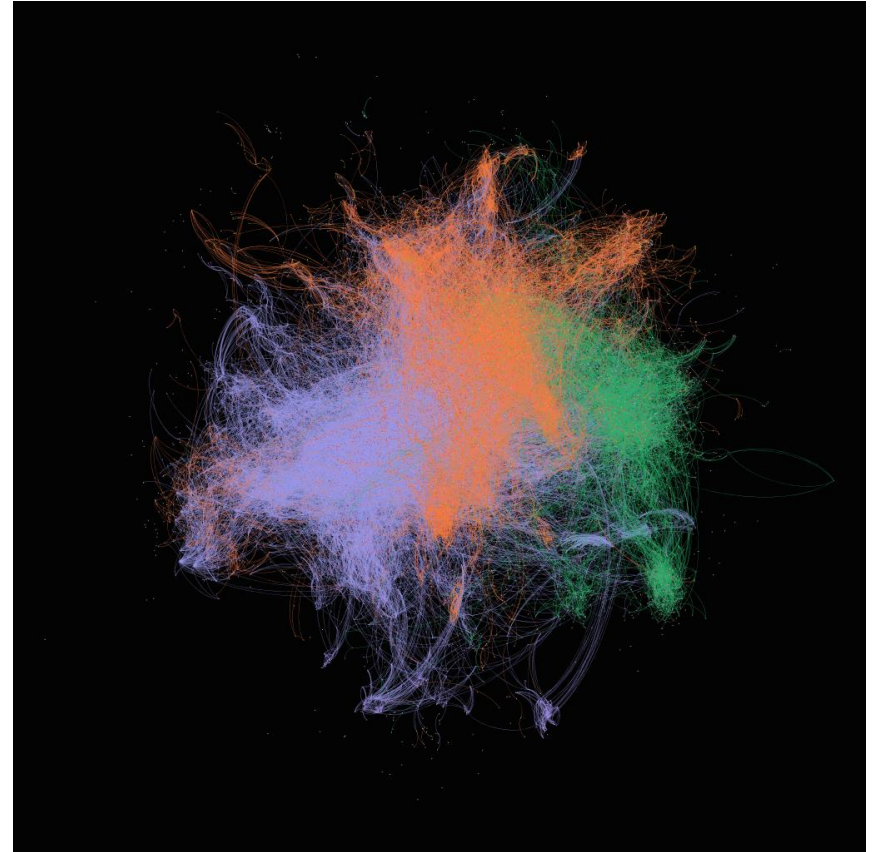
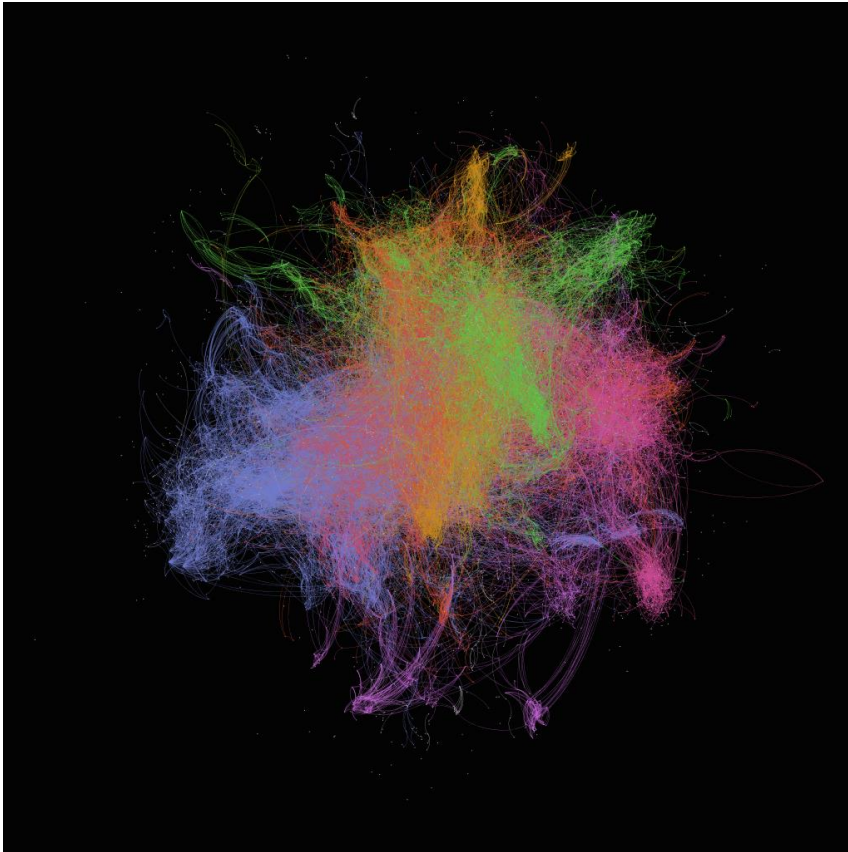
- To „zajímavé“
- Teorie: Smejkalová

Citační síť ($\ln > 1$)



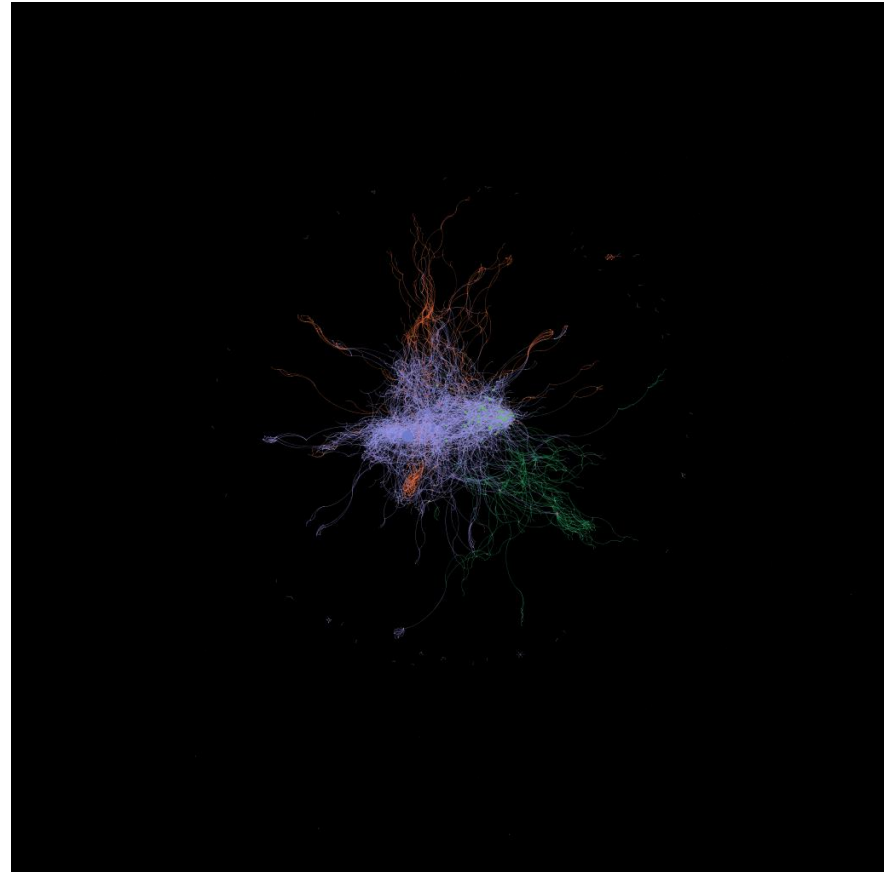
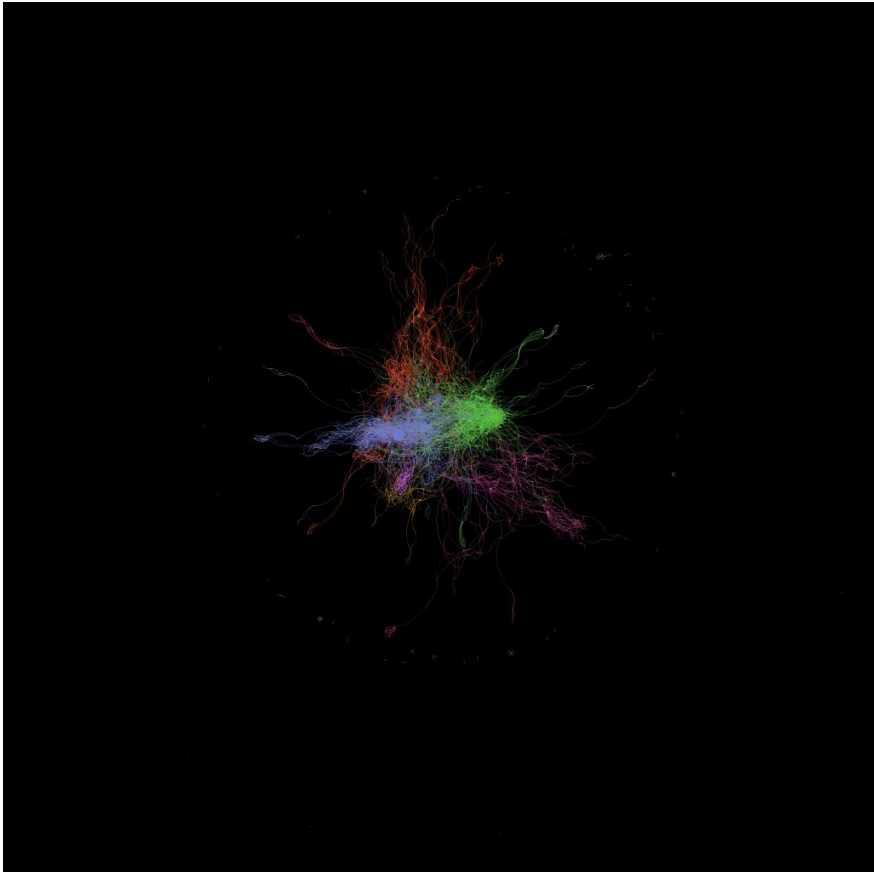
L: Modularita; P: Distribuce po soudech

Citační síť (In>6)



L: Modularita; P: Distribuce po soudech

Citační síť (In>50)



L: Modularita; P: Distribuce po soudech

Zpřístupnění datasetu

- Dataset bude podrobně popsán v monografii a před jejím vydáním zpřístupněn
- Využitelnost pro kvalitativní/kvantitativní studie

Tento projekt je podpořen Grantovou agenturou ČR v rámci projektu GA17-20645S.

jakub.harasta@law.muni.cz