

Discovering Senses of Social-Media Posts

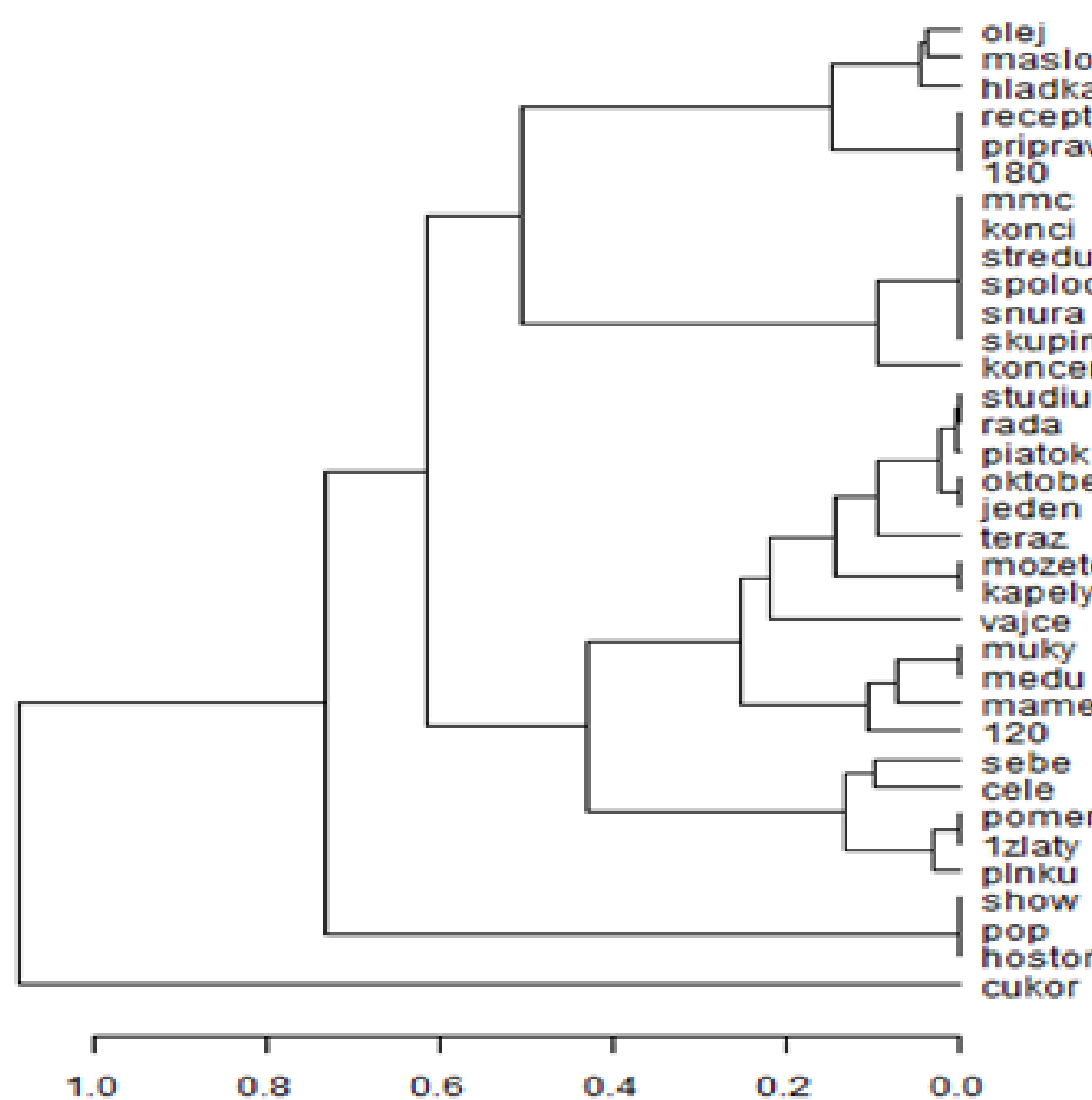
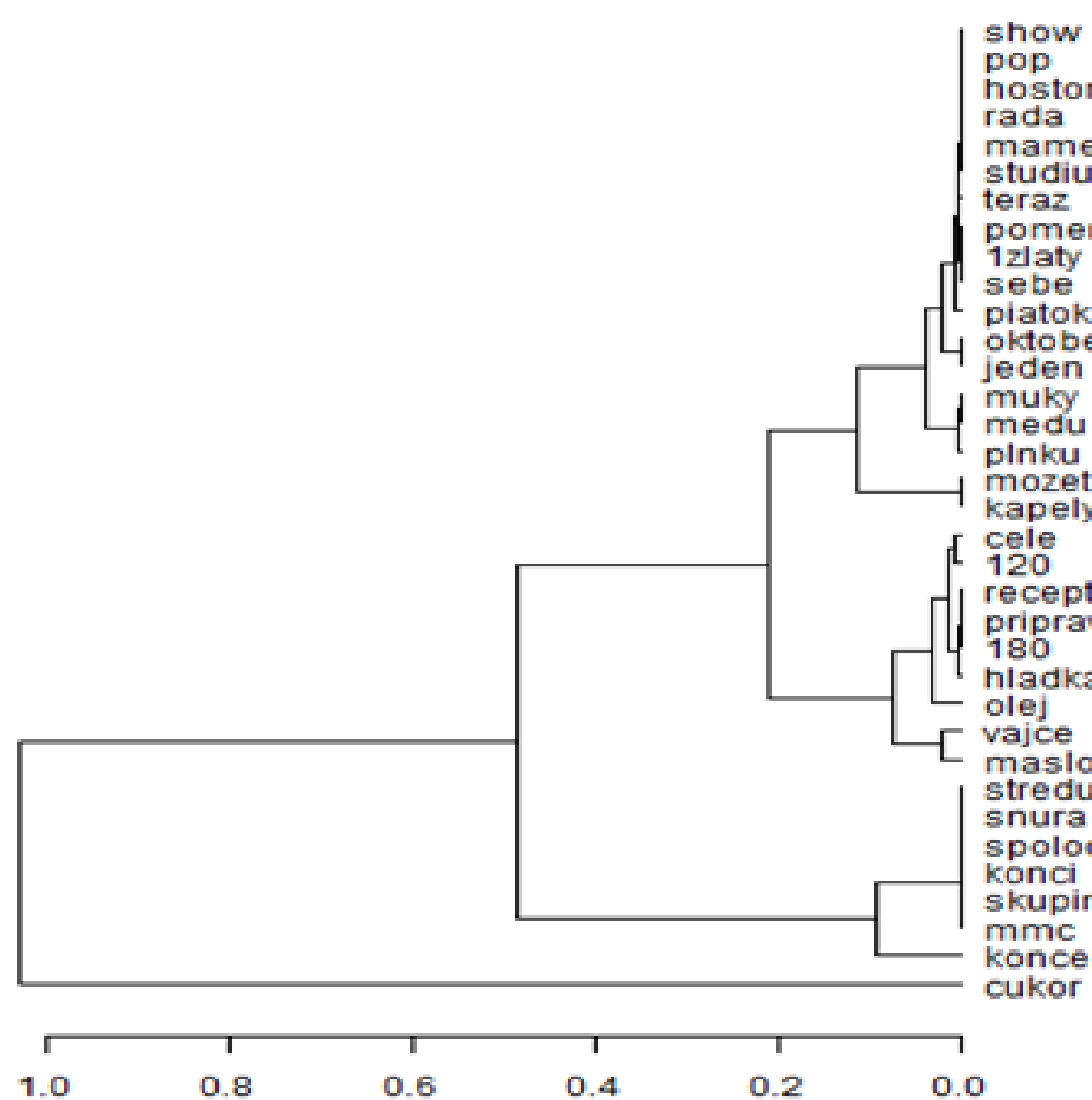
PETER BELAI, SUPERVISOR ING. MICHAL BARLA PHD.

Goal

Discover senses of the ambiguous word

puding.csv puding STEM - FALSE SVD - 2 TOPN - 35 LR -

puding.csv puding STEM - FALSE SVD - 4 TOPN - 35 LR -



Pre-processing

1. Choose messages containing chosen ambiguous word
2. Remove duplicate messages
3. Remove stop-words
4. Remove hyperlinks and special characters
5. (Optional) Lemmatization / stemming of the words

Algorithm

1. Get N words from the context window
2. Get top M words
3. Dimensionality reduction
4. Clustering

Top M words

1. Log-likelihood
2. Multiply by triangle function

	SVD Reduction					
	2	10	Number of themes – 2	Number of themes	Number of themes + 2	40
Precision	0.3894	0.5330	0.6491	0.6710	0.7723	0.6799
Recall	0.3557	0.5302	0.7427	0.78854	0.7896	0.6849
Branches	1.500	1.7813	1.9375	1.8125	1.5313	2.3438
Entire Branches	0.0313	0.2813	0.7188	0.8125	0.75	0.875

