

Polarized Topic Distributions: Comparing Reddit and Irish Times Headlines

Daria Protsenko, Nora Svensson Hahr, Alice Hamberger &
Fiammetta Rosenblatt

Table of Contents

1. Introduction and Background
2. Methodology
 - a. HDBSCAN Clustering
 - b. LDA
3. Results and Limitations
4. Conclusion



Introduction and Background

How does social & political polarization emerge from different news channels?

- Tackling polarization
 - Academia: media studies & sociology
 - Practice: reducing polarization
- increases mental & physical health
- reduces violence & discrimination

(McCoy and Murat, 2019)



Main Project Aim

Research Question:

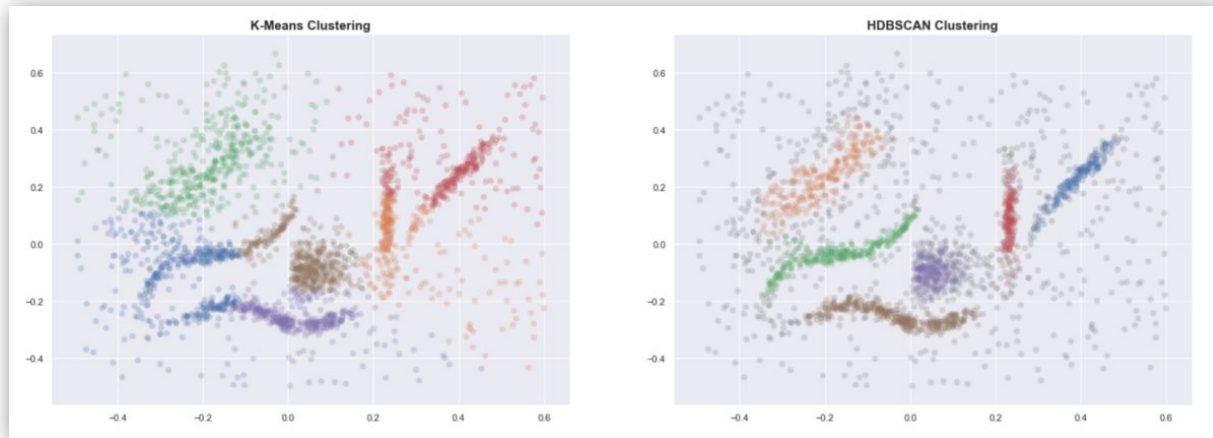
- Comparing the distribution of topics:
 - **informal** context (Reddit News Headlines) **vs.** **formal** context (Irish Times Headlines)
- Is there a signif. stat. diff. in proportion of polarizing topics across news sources headlines?

Hypothesis:

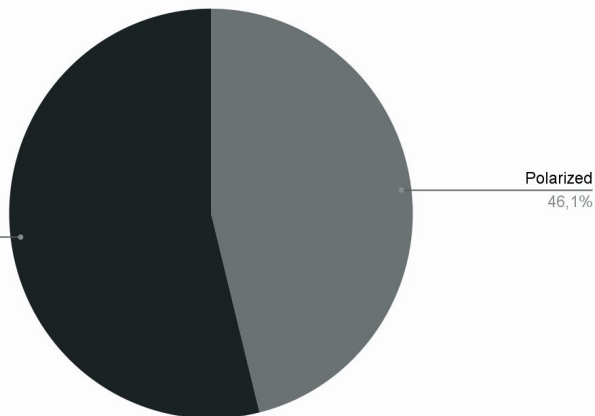
- Overrepresentation of polarized topics: **informal** context
- Null hypothesis: no significant difference in topic proportion of both headline sources

Methods I - HDBSCAN Clustering

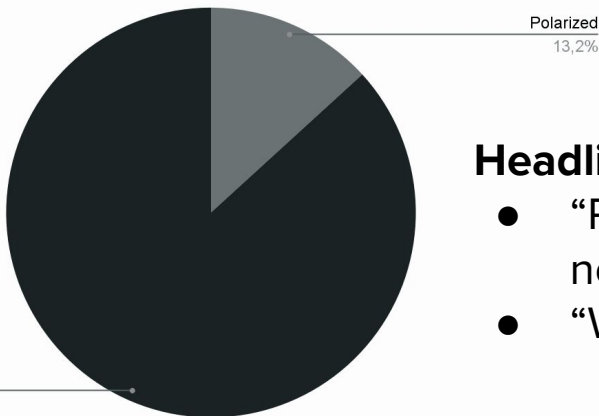
- Document embeddings
- **UMAP** Dimensionality Reduction
- Density Based Clustering (**HDBSCAN**)
- Optimizing clusters
- Labeling clusters after topic
- Deciding whether topic is polarizing



Reddit Worldnews Headlines



Irish Times Headlines



(Preliminary) Results HDBSCAN-Clustering

- Significantly **more polarized topics** on Reddit
- Some cluster **inconsistency**
- More **topic diversity** in Irish Times

Headlines out-of-cluster:

- “Political agenda excluded northern problem”
- “Warmest year record”

Olympics
Boycott
Beijing
Ceremony

Wall
Street
Record
High

Method II - Latent Dirichlet Allocation (LDA)

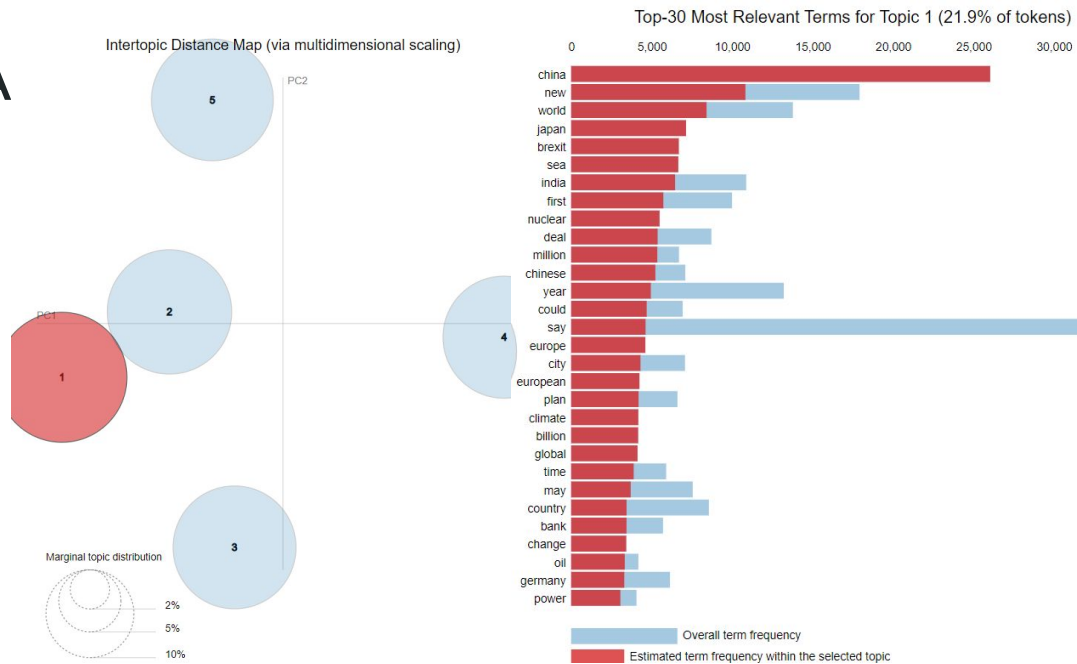
- Common tool in topic modelling (Manikonda et al., 2018).

Comparing Methods (why we did this one additionally):

- Empirical question to ask: which works better?
- Differ in purpose: **local vs global**
- Word2Vec considers **context**
- **Interpretability** of results

(Preliminary) Results - LDA

- PyLDAvis library
- Next steps: Filtering?



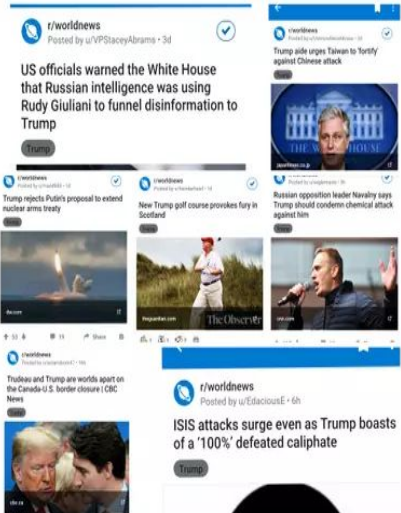
- Word Count



Conclusion

Yes,

There exists a significant difference in polarized topic coverage in the informal Reddit news source as opposed to the Irish Times (demonstrated by both methods).



Limitations and further reflections

Only data subset

“Polarized” labelling subjective

Future research sentiment analysis of text

Platform purpose & Technological affordances differ



References

- Angelov, D. (2020). *Top2Vec: Distributed Representations of Topics*. <https://arxiv.org/pdf/2008.09470.pdf>
- Arbaoui, B., De Swert, K., & van der Brug, W. (2016). Sensationalism in News Coverage: A Comparative Study in 14 Television Systems. *Communication Research*, 47(2), 299–320. <https://doi.org/10.1177/0093650216663364>
- Azeem, A. A., Hunter, J. A., & Ruffman, T. (2019). News headlines or ideological beliefs: What affects readers' interpretations of news stories about immigration, killing in the name of religion and other topical issues? A cross-cultural analysis. *New Zealand Journal of Psychology*, 48(1), 56–62. <https://link.gale.com/apps/doc/A630170770/AONE?u=amst&sid=bookmark-AONE&xid=9c6e20d7>
- Blom, J. N., & Hansen, K. R. (2015). Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76, 87–100. <https://doi.org/10.1016/j.pragma.2014.11.010>
- Chipidza, W., Akbaripourdibazar, E., Gwanzura, T., & Gatto, N. M. (2021). A topic analysis of traditional and social media news coverage of the early COVID-19 pandemic and implications for public health communication. *Disaster Medicine and Public Health Preparedness*, 1–23. <https://doi.org/10.1017/dmp.2021.65>
- Kroon, A. C., Trilling, D., & Raats, T. (2020). Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage. *Journalism & Mass Communication Quarterly*, 107769902093230. <https://doi.org/10.1177/1077699020932304>
- Manikonda, L., Beigi, G., Liu, H., & Kambhampati, S. (2018). *Twitter for Sparking a Movement, Reddit for Sharing the Moment: #metoo through the Lens of Social Media*. <https://arxiv.org/pdf/1803.08022.pdf>
- McCoy, J., & Somer, M. (2018). Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies. *The ANNALS of the American Academy of Political and Social Science*, 681(1), 234–271. <https://doi.org/10.1177/0002716218818782>
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861. <https://doi.org/10.21105/joss.00861>
- Wintrode, J. (2011). Using latent topic features to improve binary classification of spoken documents. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. <https://doi.org/10.1109/icassp.2011.5947615>
- Yan, R., & Gao, G. (2020). Topic Analysis by Exploring Headline Information. *Web Information Systems Engineering – WISE 2020*, 129–142. https://doi.org/10.1007/978-3-030-62008-0_9

Thank you for your attention!

Any questions ?