Mapping Shifts and Continuities in Media Discourse: A Proposal of a Pipeline

Anna Shirokanova (presenter), HSE University, and

Olga Silyutina, HSE University

IPSA AISP Virtual
26th World Congress of Political Science
26º Congresso Mundial de Ciência Política
10-15 | 07 | 2021

Outline

1. Motivation for the semi-automated pipeline
2. Our idea
3. Comparison with existing instruments
    a. Application 1: Internet Regulation
    b. Application 2: Labor Immigration
4. Conclusions: when and how to use, limitations, further steps

Motivation for the Semi-Automated Pipeline

- When there is a corpus that merits description
- the focus is on change over time
- How to trace changes in meanings and key terms?
- How to merge terms into topics?
- Thematic analysis.
- Now, what if the corpus is hundreds of texts? Use our pipeline.

Our idea

- a working instrument
- can describe media discourse across time
- keywords within topics can change over time
- trace shifts and continuities in discourse so that
- there is no need to match the topics manually by years

Key concepts

Assumption: ‘fluidity of discursive categories’ (e.g. cultural rights)

Topics are defined as *clusters of local semantic networks* that are meaningful from a particular historic standpoint.

Discourse streams represent the, they are ‘the same thing from period to period, although it need not remain one thing’

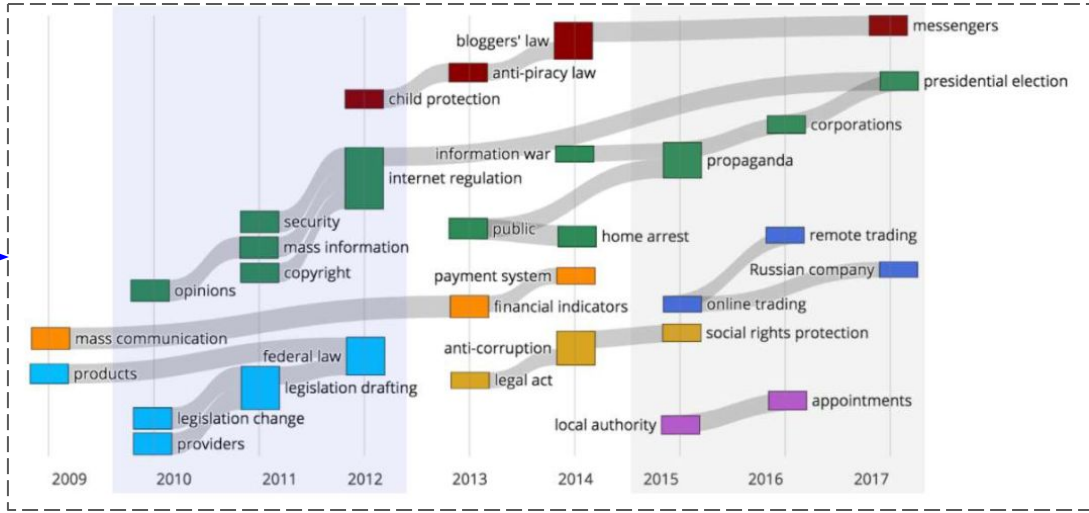
Our inspiration: **Rule, A., Cointet, J. P., & Bearman, P. S. (2015).** Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014. *Proceedings of the National Academy of Sciences*, *112*(35), 10837-10844. https://www.pnas.org/content/112/35/10837.short

In a nutshell

10k texts

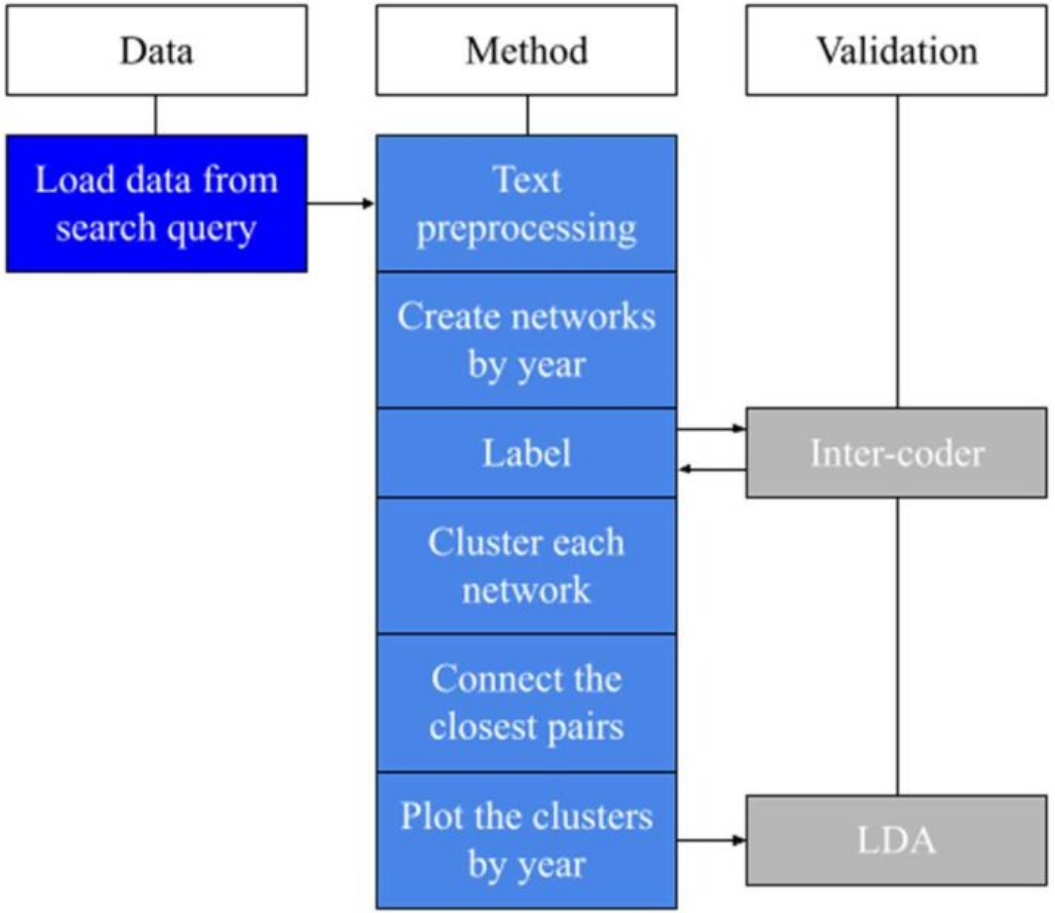
Selected by keywords

From media bank



River network of topics over years

The Pipeline



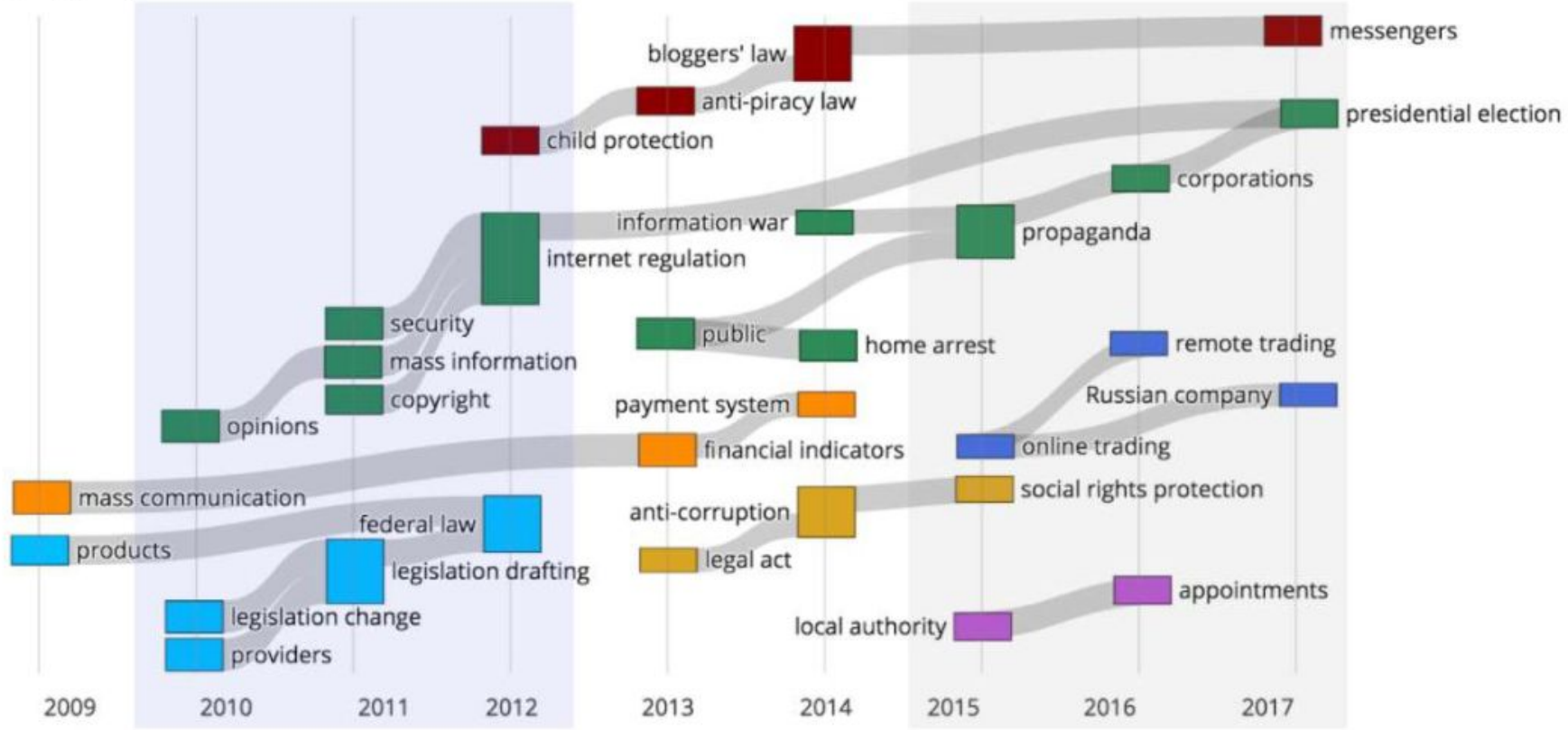
Comparison with existing instruments

- LDA (arbitrary # of topics)
- STM (time covariate is possible, but # of topics = const)
- Dictionary-based methods (words change meanings, while topics grow and fade)
- Rule et al. applied a more complex procedure over texts of comparable length over 200 years: word co-occurences and their interconnections (directly interpretable, non-exclusiveness of terms)

Our Contribution

- The pipeline can work with short texts (100 words)
- Can work with texts of different lengths
- Mutation of key terms is allowed
- Automatic connection of closest topic across time
- Few, interpretable steps (R + Python, no black-box models)
- Graphic output, a ‘river network’

Application 1: Internet Regulation

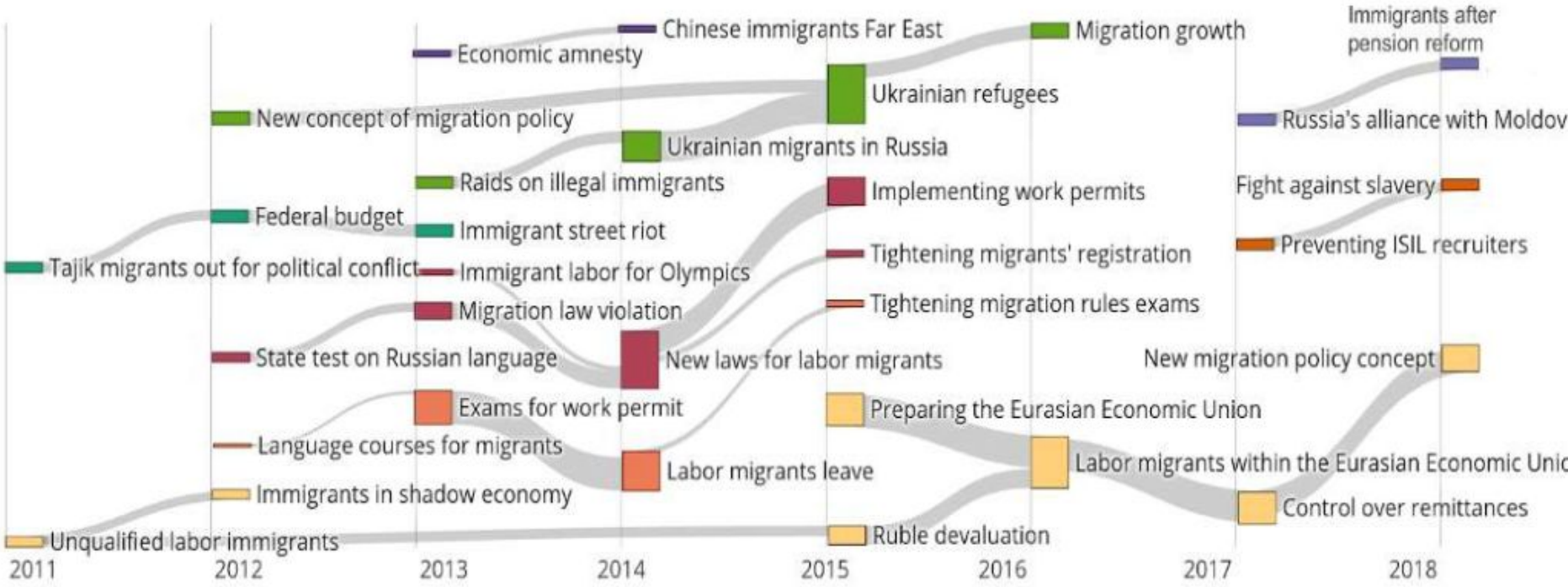


Case features

- Integrum media bank
- 7,000+ texts
- Russian (Cyrillic alphabet)
- 2011-2017
- Variety of topics grew twice over five years
- Result:

  We found 7 discursive streams: evolution of streams over time + 3 shorter periods

Application 2: Labor Immigration



Case features

- Public.ru media bank
- 2011-2018
- 5,000+ texts
- in Russian
- Countries sending migrants changed
- Migration rules changes twice
- Result:
  We discovered 8 topics with one major shift

Discussion

- This pipeline can deal with politically charged, evolving, and eventful topics
- Can work well in non-Latin scripts
- Fast processing of thousands of texts with trends over time is helpful at exploratory stages of media discourse analysis
- a varying number of topics and phrases rather than words
- Fewer steps than in the original method by Rule et al.

Discussion

Limitations:

- Depends on sufficient amount of coverage
- Human coders are to manually label the clusters: there are similarly-looking topics; some expertise is necessary for interpretation
- Reliability of clustering can be problematic (no benchmark for quality)

Next steps:

- Compare performance on a manually labelled corpus of 1k+
- Add covariates (e.g. partisanship) into pipeline
- Create a one-button app for non-coder users

Share Your Questions and Suggestions with us:

Contact:

ashirokanova@hse.ru

Twitter: @shirokaner

oyasilyutina@gmail.com

Telegram: @silyutinaolga

Github: github.com/olgasilyutina/ic2s2_internet_regulation/

Assistant on immigration project: Irina Busurkina, ipbusurkina@gmail.com