

CIVITAS

Councils In View: Italian Transcripts and Analysis System

Kick-off Meeting - November 2023

Spoke 2 - *Public Sector, Policy Design and Performance*

GRINS Foundation

Introduction

- Local policy making directly impacts the quality of public goods provided.
- **Municipal council meetings** represent a key component of local policy-making, as they provide forums for local politicians and officials to discuss and make decisions affecting their communities.

Project goal:

Create novel data on **transcripts** of **local council meetings** (LCMs) in Italy, to enable researchers and citizens to gain a comprehensive understanding of local decision-making processes.

- Project Members:
 - Giulia Romani (DSE - UNIBO)
 - Alessandro Saia (DSE - UNIBO)
 - Antonio Schiavone (DSE - UNIBO)

Novel Data on LCMs - Project Outline

- ▶ **[1] Mapping the availability of recordings on the web:** streaming video platforms, which provide to municipalities services of audio/video recordings of local council meetings:

- ▶ Youtube Lecce (LE)

- ▶ Civicam Imola (BO)

- ▶ Magnetofono Barrafranca (EN)

- ▶ VideoAssemblea Acerra (NA)

- ▶ Synedrio Fano (PU)

- ▶ ConsigliCloud Fiumicino (RO)

- ▶ **[2] Methods for data collection and elaboration:**

- ▶ **[2A]** Web scraping to map the availability of video streaming on the different platforms and to collect video **transcripts**.

- ▶ **[2B]** Text analysis tools to analyze **content** and **tone** of meetings.

Municipalities with Streaming Services in Italy

Around 1,250 municipalities in Italy broadcast their council meetings across various platforms, with the majority using **YouTube** for LCM streams.

Figure: Municipalities with streaming services

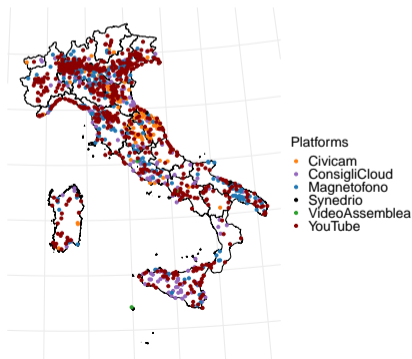
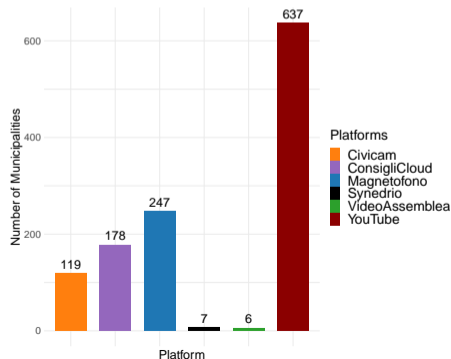


Figure: Numerosity by platform



Example from Youtube - LCM in Adria (Rovigo) - November 16, 2023



Diretta Consiglio Comunale



comune adria
991 subscribers

Subscribe



3



Share

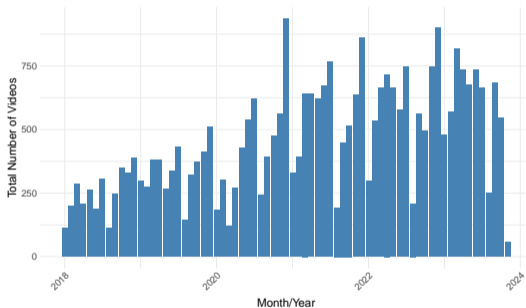
Save



Retrieval of Video (and Transcripts) from Youtube

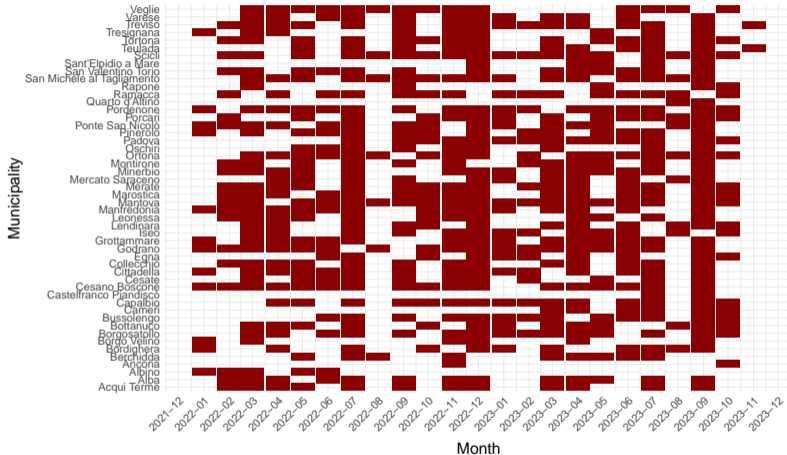
- *Developed a script for automated data retrieval from YouTube.*
 - Obtained access to the **YouTube Research Program**, which grants increased *API Quotas* for advanced data scraping.
 - Our script is both **flexible** and **scalable**, designed to accommodate the inclusion of new municipalities/LMCs efficiently.

Figure: Subtitles retrieved from YouTube



To date, we've collected **30,407 transcripts of LCMs from Youtube**, with an average length of 78,625 characters—equivalent to approximately 39 pages when written out.

Example - Monthly Publications for a Subset of Municipalities



Note: Each cell represents a municipality-month observation. A red cell indicates a month with an LMC uploaded to YouTube; an empty cell means no upload.

Next steps ...

[1] - Expanding coverage

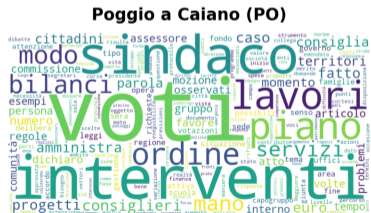
- Other platforms have LCMs recording available
 - Scraping these private platforms is significantly more complex than scraping YouTube.
 - Subtitles **not** available for all their videos.
 - Demands significant computational resources for caption generation in reasonable time and substantial disk space for storing the audio files for processing.

[2] - (Text and Audio) Analysis of LCMs

- Data cleaning of raw transcripts
- Data processing through automated and manual methods using **Natural Language Processing** techniques. For example:
 - Topic-analysis through Bag of Words or Latent Dirichlet Allocation (LDA) models
 - Investigate *emotions, language style* and *tone* in LCMs

Examples - What do they talk about?

Examples of **word cloud** visualizations of the most frequently discussed topics:

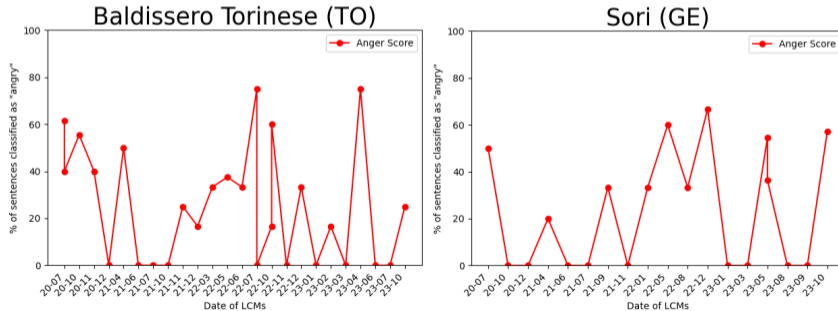


... schools, services, housing, but also votes, euros and balance sheets.

Examples - How do they talk?

Emotion analysis uses a transformer-based algorithm for emotion detection trained on FEEL-IT, a corpus of Italian Twitter posts annotated with four basic emotions: anger, fear, joy, and sadness (Bianchi et al., 2021).

Examples of **Anger Score** in LCMs: Percentage of sentences where *anger* is the prevalent emotion.



Thank you for your attention!

Appendix - Sources & Methods

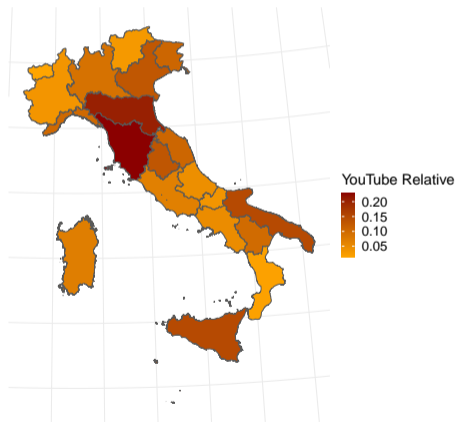
- ▶ **Data sources:** streaming video platforms, which provide to municipalities services of audio/video recordings of local council meetings:
 - ▶ Youtube Lecce (LE)
 - ▶ Civicam Imola (BO)
 - ▶ Magnetofono Barrafranca (EN)
 - ▶ VideoAssemblea Acerra (NA)
 - ▶ Synedrio Fano (PU)
 - ▶ ConsigliCloud Fiumicino (RO)
- ▶ **Methods for data collection and elaboration:**
 - ▶ Web scraping to map the availability of video streaming on the different platforms and to collect video subtitles from YouTube (→ **YouTube Researcher Program**).
 - ▶ Text analysis tools to provide some descriptive statistics and visualisation of the topics discussed in local council debates.

Appendix - Municipalities with Youtube streaming services

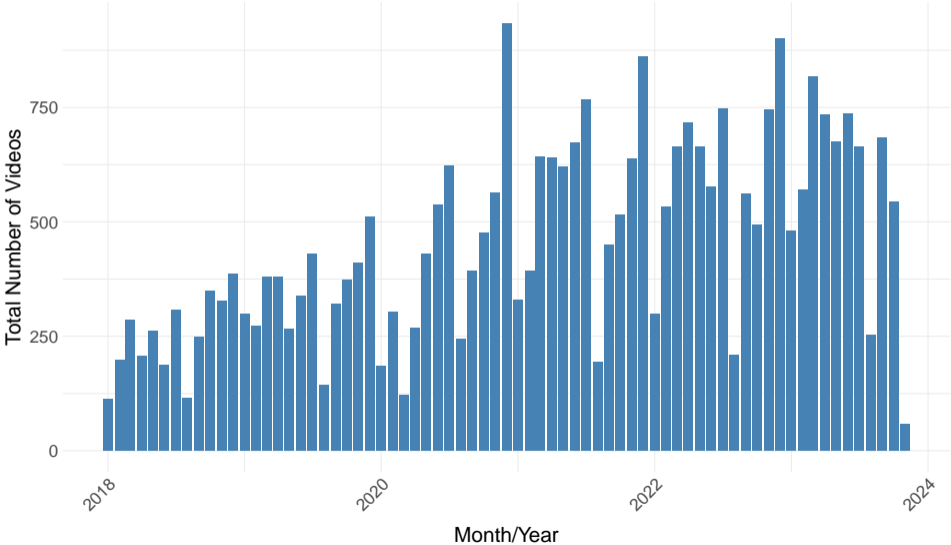
Figure: Distribution of YouTube streaming



Figure: Relative YouTube Coverage by Region



Appendix - All Subtitles retrieved in the 2022-2023



Appendix - Example - Word clouds by macro-region

Most frequently discussed topics in each macro-region:

