University of Konstanz
Department of Politics and Public Administration
Summer/Winter Semester 2024
Social Media Data Analysis
Max Pellert

Conspiracy narratives and public perception of 15-minute cities on YouTube

Project Report

Deadline: 21.08.2024

Julia King
01/911139
M.Sc. Social and Economic Data Science
julia.king@uni-konstanz.de

1 Motivation

The concept of a 15-minute-city is, in and of itself, nothing more than an architectural proposal for efficient and sustainable urban design. However, it rose to prominence during the past five years due to a staggering amount of online conspiracy theories and misinformation. While conspiracy theories often revolve around large events or hard-to-explain phenomena, like the moon landing or many celebrity or politician assassinations, 15-minute-cities are neither. 15-minute-cities, in contrast, are a suggestion in the research field of architecture and city planning, proposing to build new urban developments in a way that ensures residents can reach basic amenities within a 15-minute walking distance. The fact that a simple, academic proposal for city design was (and is) the center of large swathes of online conspiracy theories, is puzzling, and warrants investigation.

Understanding the puzzling emergence of 15-minute-city conspiracies requires investigation of the ways in which people discussed them publicly. This project aims, therefore, to track the development of theories revolving around this concept by analyzing the online discourse around it. This requires an in-depth understanding of the process in which conspiracy theories took over the discourse on 15-minute-cities. Consequently, this project collects metrics on the viewership of online videos revolving around 15-minute-cities, and to classify the discursive tone of both the videos themselves and the comments posted on them. Using this data, the following hypotheses are tested:

1. Videos produced before the first spike in public interest are more likely to be non-conspirative compared to those uploaded during and after the spike.
2. Comments under conspirative videos express higher levels of negative sentiment compared to comments under non-conspirative videos.
3. The engagement metrics (e.g., likes, comments, shares per view) of conspirative videos differ significantly from those of non-conspirative videos, with conspirative videos having higher engagement rates per view.

In order to test these hypotheses, this report will first explain the interfaces and resources used to collect the necessary data, before highlighting the data processing steps taken. Afterwards, it will outline and visualize the results of the statistical analyses performed in the project and discuss their validity and reliability. Finally, the report will identify key limitations and alternative explanations for the results.

The code and data used in the project as well as additional visualizations can be found in the github repository.

2 Data Retrieval

Explain the interfaces or resources you used to collect all data necessary for the project.

The following section will outline the various apis and data sources utilized in the project. Specifically, this project gathers data obtained from google trends (2024), which is then used to parametrize queries for the Youtube Data api (2024). Finally, the huggingface api (2024) is utilized to allow sentiment classification via LEIA.

In order to track the development of the conspiracy theory and test hypotheses 1 and 2, an investigation of the progression of public attention was necessary. In line with previous research (including Jun et al., 2018; Mavragani et al., 2018; Yeo and Knox, 2019), Google Trends data was chosen to model public interest and identify relevant phases of the conspiracy. Obtained via the pytrends package (dreyco676 & emlazzarin, 2023) in python (Van Rossum & Drake,

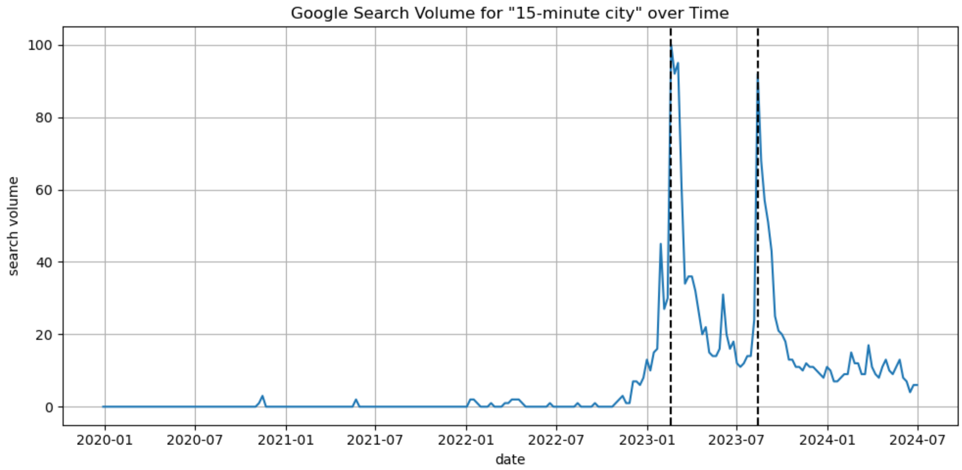


Figure 1: *Public interest in 15-minute cities measured by search volume*

2009), the data was limited to the United States of America, as the conspiracy itself was highly focussed on the country. Additionally, Google Trends distinguishes between search terms and topics, with topics also including variants of the given term. Thus, the query was specified for the 15-minute city topic.

The results of the Google Trends data collection, visualized in Figure 1, were crucial to the parametrization of the youtube api requests. Specifically, distinct phases centered around the presence or absence of a peak in public interest were algorithmically identified to ensure all relevant data was obtained while preventing excessive use of the api. For now, it is important to note that this procedure, the details of which will be discussed in Data Processing, also yielded a cutoff-point before the first relevant phase.

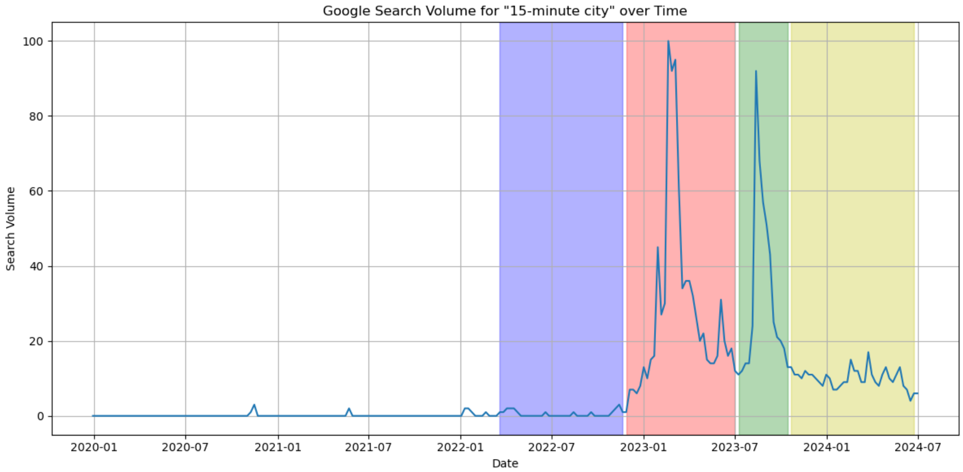
Following the google trends data collection, multiple endpoints of the YouTube Data Api were utilized to obtain relevant data on the most-viewed videos regarding 15-minute cities. First, the *search* endpoint was employed to identify the 50 most-viewed videos per week. This weekly structure, while quota-intensive, enabled the identification of temporal trends in the analysis stage and was thus crucial for this project. Afterwards, metadata on all of the identified videos was obtained using the *video* endpoint. Finally, multiple queries to the *commentThread* endpoint per video allowed for the collection of all top-level comments under all videos. In total, 119 weeks of videos were processed, resulting in a dataset containing information on 2943 videos and 204093 comments, a process that required multiple days due to the quota limits imposed by the YouTube Data API.

Finally, the huggingface api was utilized to obtain the LEIA-base transformer model (Aroyehun et al., 2023), which was later employed to classify the sentiments of the comments obtained via the YouTube Api. In addition to its performance in the assignments completed for this course, the LEIA-base model was chosen due to its ability to classify natural language sentiment in a fine-grained manner, distinguishing between Affection, Happiness, Anger, Sadness and Fear instead of a simple polarity score. This distinction of LEIA when compared to other sentiment analysis techniques such as VADER (cjhutto, 2020) is especially relevant in the context of conspiracy theories, as a comment expressing sadness implies a much lower threat level than a comment expressing anger. Thus, utilizing LEIA-base enables the generation of additional inferences into the conspiracy by providing a more fine-grained perspective.

3 Data Processing

Explain how you filtered data, normalized values, computed additional variables, etc.

Figure 2



Visualization of the generated phases

The Google Trends data was utilized to identify distinct phases in public interest, with the goal of identifying connections to the development of the conspiracy theory.

4 Analysis

Perform statistical analyses and visualizations that assess the question(s).

5 Conclusion

Evaluate answers to the question and their reliability.

6 Critique

Identify limitations and alternative explanations for your results.

Further research points: -what was the "spark" leading conspiracy-prone accounts to this topic? (e.g., inclusion of measure in 2020 campaign by Paris mayor Anne Hidalgo – but conspiracies mainly emerging 3 years later?) -Hypothesis: theories were driven by accounts already engaged in the "conspiracy bubble"

References

Aroyehun, S. T., Malik, L., Metzler, H., Haimerl, N., Di Natale, A., & Garcia, D. (2023). LEIA: Linguistic Embeddings for the Identification of Affect. *EPJ Data Science*, *12*(1), 1–21. https://doi.org/10.1140/epjds/s13688-023-00427-0

cjhutto. 2020. *vaderSentiment*. https://pypi.org/project/vaderSentiment/

dreyco676 & emlazzarin. 2023. *Pytrends*. https://pypi.org/project/pytrends/

Google. 2024. *15-Minute City*. https://trends.google.com/trends/explore?date=today%205-y&geo=US&q=%2Fg%2F11hk3_sx0c&hl=en-GB

Google for Developers. 2024. *Youtube Data Api V3*. https://developers.google.com/youtube/v3

Huggingface. 2024. *Transformers*. https://huggingface.co/docs

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, *130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Mavragani, A., Ochoa, G., & Tsagarakis, K. P. (2018). Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *Journal of Medical Internet Research*, *20*(11), e9366. https://doi.org/10.2196/jmir.9366

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Yeo, J., & Knox, C. C. (2019). Public Attention to a Local Disaster Versus Competing Focusing Events: Google Trends Analysis Following the 2016 Louisiana Flood. *Social Science Quarterly*, *100*(7), 2542–2554. https://doi.org/10.1111/ssqu.12666