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

Project Report

Link to GitHub repository

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 classifies the discursive tone of both the videos themselves and the comments posted on them. Using the data collected in this project, 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

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

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, 2009), the data was limited to the United States of America, as the conspiracy itself was highly focussed on the country. Additionally,

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

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 metrics 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. The data collection and analysis was limited to top-level comments, as those should provide the most accurate reflection of the audience reactions to the given video, instead of reactions to other comments.

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. Note that some weeks, especially those before the first spike in public attention, did not have 50 new videos uploaded, leading to a smaller number of videos collected for those weeks. Despite this, the dataset still provides more than adequate data for analyzing the progression of the conspiracy theory surrounding 15-minute cities.

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 insight into the conspiracy by providing a more fine-grained perspective.
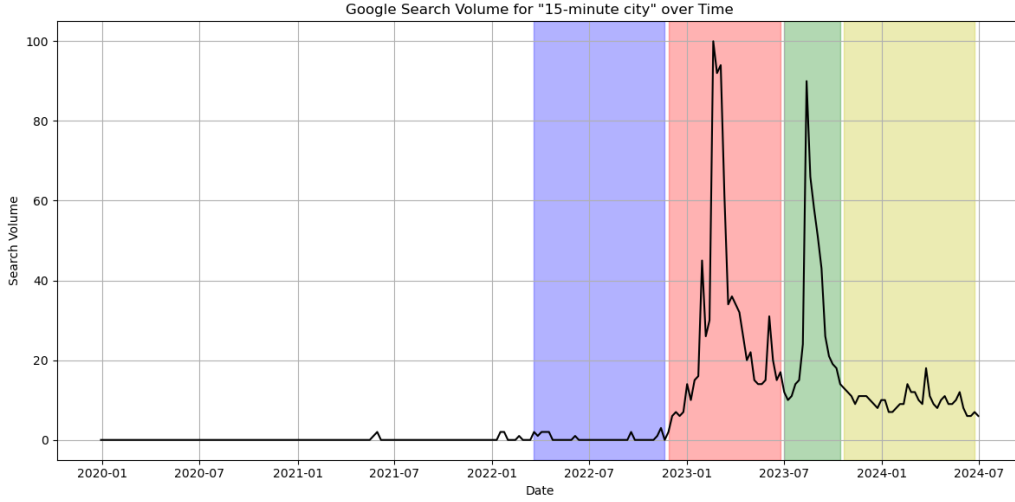
Figure 2: *Visualization of the generated phases*

3 Data Processing

The Google Trends data was utilized to identify distinct phases in public interest, enabling the progression of key variables during different stages of development of the conspiracy theory. Differences in rolling averages were used to identify the beginning and end of peaks, with steady increases indicating the beginning of a spike and the end of a steady decrease defining the end of a spike. As seen in Figure 2, four distinct phases were identified:

1. The phase before the first spike, which is hypothesized to contain mostly educational videos on the concept of 15-minute cities.

2. The first spike, hypothesized to be caused by the rise of conspirative content.

3. The second spike, hypothesized to be caused by mainstream media coverage.

4. The phase after both spikes, where search volume has returned to a steady level.

Following the identification of phases, a random sample from the results of the *videos* endpoint was drawn and designated for manual classification into conspirative and non-conspirative content. This classification was conducted using a tkinter app, which presented the rater with the title and description of a given video. A link to the full video was also provided for cases where unambiguous classification based on the previous attributed was not possible. In total, 400 videos were classified, allowing for a train-test split of 300/100 videos.

In order to facilitate automatic classification via supervised learning techniques, the video's titles and descriptions were preprocessed. This included the removal of links, newlines, non-alphanumeric characters and stopwords, as well as tokenizing and lemmatizing. These cleaned texts were included in a dataframe containing video-level information. This dataframe also included the phase, week of upload, video id, raw versions of the video's title and description as well as the number of views, likes and comments. Sadly, the number of times a video has been shared is no longer reported in the video metadata. Thus, hypothesis 3 can only be tested with regards to likes and comments per view.

In order to ensure the best possible performance, multiple supervised learning techniques were applied and evaluated. Specifically, the approaches employed by this project include the (Gaussian) Naive Bayes, Support Vector Machine and Random Forest algorithms from the sklearn python package (Pedregosa et al., 2011). All three models were fitted using the 300 tweets in the training data and evaluated based on their performance on the test set. The evaluations informed the final model choice and are not part of the substantial analysis and are thus presented in this section.

Based on the performance metrics visualized in Figure 3, the Naive Bayes classifier was selected, as it provided the highest accuracy as well as the highest F1-scores in the sample. However, based on the uncertainty estimates plotted in the figure, it is important to highlight that this performance was not indicative of a statistically significant advantage of Naive Bayes when compared to other algorithms, with especially the random forest algorithm often performing similarly. Thus, other training and testing samples may have resulted in the selection of a different model. The model was then used to automatically classify all 2943 videos, with the classifications being added to the video dataframe.

Finally, the emotion classification model LEIA was applied to obtain the sentiments expressed in each of the 204093 top-level comments. This also involved the creation of a new dataframe saving comment-level information including the phase, week and video id of the comment, as well as the comment id and text.

Figure 3: *Performance metrics of supervised learning models*

Notably, the comment texts were not cleaned before classification as to preserve important contextual information, which can be utilized by LEIA (in contrast to the supervised learning models above). The classification was done on a local machine over several hours.
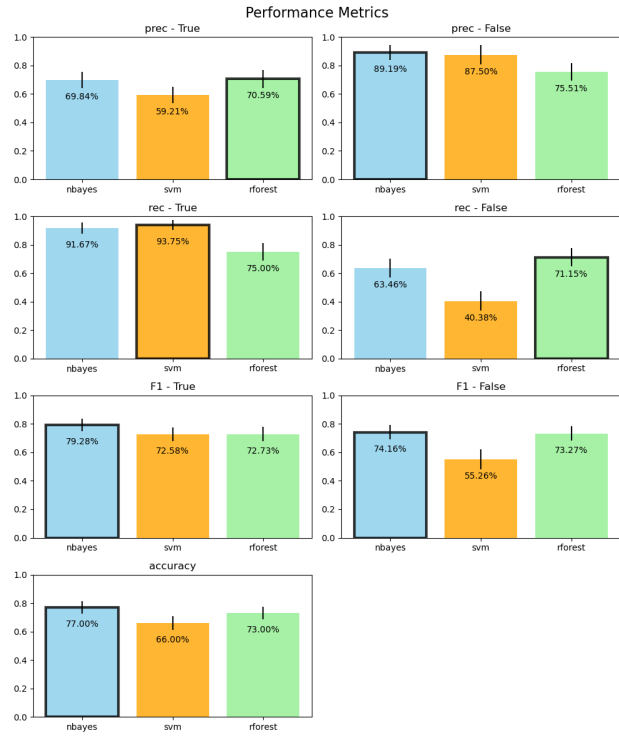
4 Analysis

The following analysis section will group the findings of the project by hypothesis.

4.H1 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.

To test this hypothesis, a contingency table of the variables *video phase* and *conspirative* was constructed. A normalized version of this table is visualized in Figure 4. In line with the hypothesis, the phase before the first spike in public interest featured the lowest share of conspirative content. Notably, the share of conspirative content remained above 50% for all following phases. This runs counter to the theory outlined in the Data Processing section, as an influx in mass media attention should've also increased the amount of informative

Figure 4: *Share of conspiracy narratives per phase*

content debunking the conspiracy theory. Nevertheless, the tendencies predicted by the hypothesis are found in the data.
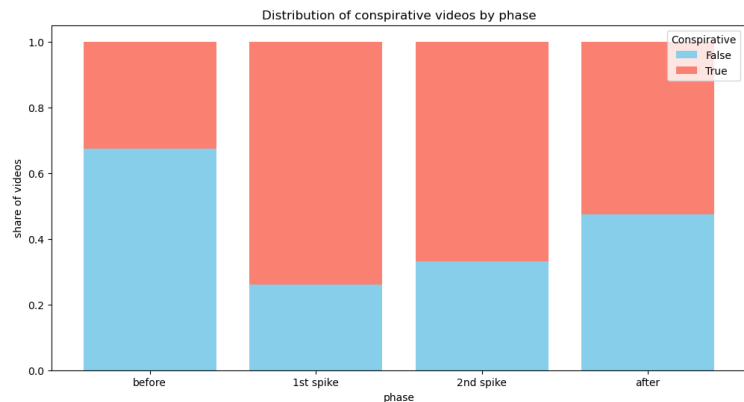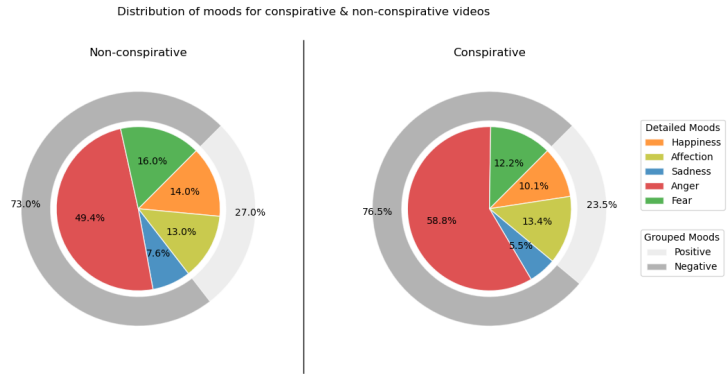
After this exploratory analysis, a test for statistical independence of two categorical variables was conducted. Due to the large sample size and differing group sizes, Pearson's Chi-squared test was chosen. Two variants of the test were conducted: Test 1a was based on a dataset distinguishing between all 4 phases, while test 1b was conducted by first grouping the data of phases 2, 3 and 4 together to more closely align with the hypothesis above. Both tests were highly significant, with Chi-square statistics of 210.93 ($p < 0.001$) and 108.13 ($p < 0.001$) respectively. Thus, this project was able to confidently reject the null hypothesis, showcasing a significant increase in conspirative content.

In addition to the per-phase analysis conducted in the above test, a continuous visualization of the progression of the rate of conspirative content was generated. This additional figure named *conspirative_time.png* is provided in the GitHub repository and similarly highlights the sharp increase in conspirative content around the first spike.

4.H2 Comments under conspirative videos express higher levels of negative sentiment compared to comments under non-conspirative videos.

To explore this hypothesized connection, the distribution of sentiments in the comments of conspirative and non-conspirative videos was visualized in figure 5. Significant differences in the distribution of detailed moods are observable. This is especially pronounced with regards to the share of comments expressing anger, which increases from 49% for non-conspirative videos to 59% for conspirative videos. However, the dif-

Figure 5: *Distribution of comment moods by conspirative classification*

ferences in grouped moods are much less pronounced, only amounting to about 3%. While the direction of this change is consistent with the hypothesis outlined above, the small magnitude of this difference is noteworthy.

Nevertheless, two Chi-squared test, one for the detailed sentiments and one with a polarity grouping matching the hypothesis above, were conducted. Once again, both tests revealed highly significant correlations, with Chi-square statistics of 1761.02 ($p < 0.001$) and 242.75 ($p < 0.001$) respectively. These results can be explained by the large sample size (204093 comments in total), as large samples can enable statistical significance even if the differences in group means are minimal.

4.H3 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 this hypothesis, per-view metrics were created and tested for normality. Notably, this required the exclusion of 122 videos, which featured obviously erroneous query results, including more likes than views. Visual inspection of the figures plotting the distributions of likes and comments per view (which can be found in the github repository in the image *engagement-metrics_view.png*) revealed strong non-normal distributions, even

		conspirative	
		True	False
metric	likes	0.052	0.026
per view	comments	0.004	0

Table 1: *Median engagement metrics*

when logged. Thus, table 1 shows the median engagement metrics per view. In line with our hypothesis, they are higher for conspirative videos.

Due to the aforementioned non-normality of the distributions, a nonparametric test was chosen instead of a traditional t-test. Specifically, the Mann-Whitney U rank test was employed, as it compares rank sums instead of means, allowing for non-normal distributions to be compared. Both tests yielded highly significant results, with U-statistics of 1169852.5 ($p < 0.001$) for likes and 1064695.5 ($p < 0.001$) for comments. Therefore, we find significant evidence of increased engagement with conspirative videos.

5 Conclusion

This project found statistically significant evidence for all three main hypotheses outlined in the research proposal. First, it established that conspirative content rapidly increased during the first spike in public attention. This sudden influx may indicate an audience overlap with preceeding conspiracy theories, as the 15-minute city conspiracy theory may have been shared in those channels. Further evidence for this can be found in the fact that conspirative videos in the dataset often refer to other conspiracy theories, including the New World Order conspiracy theory (see Spark, 2000).

Second, the project finds evidence of increased negative sentiment in the comments of conspirative videos. However, as outlined in the previous section, the substantive difference between the two groups was minimal when only observing sentiment valence, and significantly increased when introducing additional emotion dimensions. Thus, the reliability and generalizability of hypothesis 2 is severely diminished by the oversimplification of the emotion space when observing reactions to conspiracy theories. Future projects should utilize a more detailed operationalization.

Third, conspirative videos in the sample experienced significantly higher levels of engagement per view. This discovery is in line with previous research, which has shown other kinds of engagement inspired by conspiracy theories, including political engagement (Kim, 2022). However, the findings here only pertain to the most-viewed videos per week and can thus not be generalized to all conspirative videos.

6 Critique

As mentioned in the Analysis section, the calculation of per-view metrics revealed undeniably erroneous data provided by the YouTube Data api, including instances where likes exceed view counts by factors of up to 5. Notably, these errors are also present on the YouTube website (see this video for an example) and are thus unavoidable with regards to this project. Nevertheless, these discrepancies might impact the results, especially if such errors also occur in the non-detectable range, for instance with inflated like counts that do not exceed the view count.

Additionally, while the increase in google searches coincides with an increased share of conspirative YouTube videos, the causality of these events is unclear. Thus, it is possible that the sharp increase in conspirative content was not the product of organically occuring public attention but instead of thought leaders engaged in the wider "conspiracy bubble" promoting the 15-minute city conspiracy theory. Consequently, a qualitative analysis of early conspirative videos may reveal further insights into the emergence of conspiracy theories not pertaining to large events or hard-to-explain phenomena.

Finally, this project conducted extensive testing of the applied supervised learning techniques to ensure the usage of a well-performing model. However, the performance metrics (see Figure 3) reveal imbalances between the labels for all three algorithms, with significantly higher recall and lower precision for conspirative videos, which may have lead to an overestimation of conspirative content and could skew the results. Future projects should thus expand on model selection to aim for a closer match to the data.

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

Kim, Y. (2022). How conspiracy theories can stimulate political engagement. *Journal of Elections, Public Opinion and Parties*, *32*(1), 1–21. https://doi.org/10.1080/17457289.2019.1651321

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Spark, A. (2000). Conjuring Order: The New World Order and Conspiracy Theories of Globalization. *The Sociological Review*, *48*(2_suppl), 46–62. https://doi.org/10.1111/j.1467-954X.2000.tb03520.x

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