

SIRITVIS: Social Interaction Research Insights Topic Visualisation

Sagar Narwade¹, Gillian Kant², Benjamin Säfken¹, and Benjamin Leiding¹

¹ Technische Universität Clausthal, Clausthal-Zellerfeld, Germany ² Georg-August-Universität Göttingen, Göttingen, Germany

DOI: [10.21105/joss.06243](https://doi.org/10.21105/joss.06243)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Olivia Guest](#)

Reviewers:

- [@n3mo](#)
- [@cjbarrie](#)

Submitted: 27 October 2023

Published: 08 August 2024

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

SIRITVIS is a powerful tool designed to analyze text data from Reddit, Instagram, and other sources. It uses advanced techniques to clean data and identify main topics efficiently. The tool also includes features for visualizing data, making it easier to explore and understand the text.

One unique feature of SIRITVIS is its ability to map social media posts worldwide. This means it can show where people are talking about certain topics globally, helping to understand trends and conversations happening around the world. It also analyzes the mood or sentiment of these posts using a tool called VADER ([Hutto & Gilbert, 2014](#)).

SIRITVIS is very useful for researchers who want to study public discussions on social media. It is especially helpful for analyzing current events and global issues. The tool is easy to install using pip, a package manager. You can find detailed installation instructions on its GitHub page¹.

Statement of Need

Social media platforms like Reddit and Instagram have revolutionized the way we communicate, share information, and express our opinions. They are key places for public discussions, and by analyzing text data from these platforms, we can gain insights into what people are talking about, their feelings, and current trends. This information is valuable for fields like marketing, politics, and disaster response.

However, dealing with the large and constantly changing amount of text data from social media is challenging. To help with this, we have developed SIRITVIS, an open-source tool for analyzing social media text. SIRITVIS uses advanced topic modeling techniques to automatically identify and extract key topics from large text datasets, making it easier to explore and find meaningful patterns.

SIRITVIS supports several advanced topic models such as AVITM ([Srivastava & Sutton, 2017](#)), Latent Dirichlet Allocation (LDA) ([Blei et al., 2003](#)), Neural Latent Dirichlet Allocation (NeuralLDA) ([Srivastava & Sutton, 2017](#)), Product Latent Dirichlet Allocation (ProdLDA) ([Srivastava & Sutton, 2017](#)), and Contextualized Topic Models (CTM) ([Bianchi et al., 2021](#)). These models help in discovering key topics without prior knowledge of the data, making it easier to analyze and understand large collections of text.

SIRITVIS is designed to simplify the process of analyzing social media text data, making it accessible for various applications.

¹<https://github.com/CodeEagle22/SIRITVIS/>

SIRITVIS can extract and analyze data from platforms like Reddit and Instagram using their APIs (tools for accessing data). The software simplifies data extraction and then processes the raw data using advanced natural language processing (NLP) techniques to clean and prepare it for analysis. Users can adjust settings to fit their specific needs and expertise.

The package includes tools for evaluating the performance of trained models using various customizable metrics, ensuring users can assess the quality of their analysis. Additionally, SIRITVIS helps users identify and use the best models for their data analysis tasks.

To make understanding the data easier, SIRITVIS includes two powerful visualization tools: PyLDAvis and Word Cloud. PyLDAvis (Sievert & Shirley, 2014) creates dynamic and interactive visualizations of topic models, helping users explore and understand the relationships between topics and their keywords (see Figure 1). The Word Cloud tool (Mueller, 2023) visually displays the most frequently used words in a dataset, making it simple to spot key terms and patterns at a glance (see Figure 2). These tools provide a user-friendly way to uncover and communicate the insights hidden in text data.

SIRITVIS is particularly known for its ability to map the locations of Instagram posts and Reddit comments globally, showing the most popular topics in each area and their frequencies (see Figure 3). It also uses color-coding to show the sentiments (positive, negative, or neutral) of posts at each location (see Figure 4). Additionally, SIRITVIS allows users to explore specific keywords and see where they appear on the world map. This spatial information is very useful for understanding public discussions and supporting data-driven decision-making in various fields.

Comparing and Contrasting Available Toolsets

In recent years, understanding what people say on social media has gotten better. One new way to do this is called SIRITVIS. It's different from other tools like TTLocVis (Kant et al., 2020), TweetViz (Stojanovski et al., 2014), and Twitmo (Buchmüller et al., 2022).

SIRITVIS helps people look at social media posts in an easy way. It can understand lots of different topics and even figure out where the posts are coming from on a map. Plus, it works well with another tool called pyLDAvis, which helps show the results in a nice way.

One cool thing about SIRITVIS is that it has a bunch of ways to check if the results are good. These include things like checking if the topics are different enough from each other and if they make sense. This helps make sure the results are trustworthy.

SIRITVIS also comes with a tool called octis (Terragni et al., 2021), which helps measure how good the results are. It looks at things like if the topics are diverse, accurate, and if they make sense. This extra tool adds to SIRITVIS's reliability, ensuring that the results it gives are meaningful and useful.

SIRITVIS is made to be easy to use. It has a simple design and instructions that anyone can follow, whether they're new or experienced. And it can work with different types of data, so people can use it however they need to. This makes it useful for lots of different people who want to understand social media better.

Figures

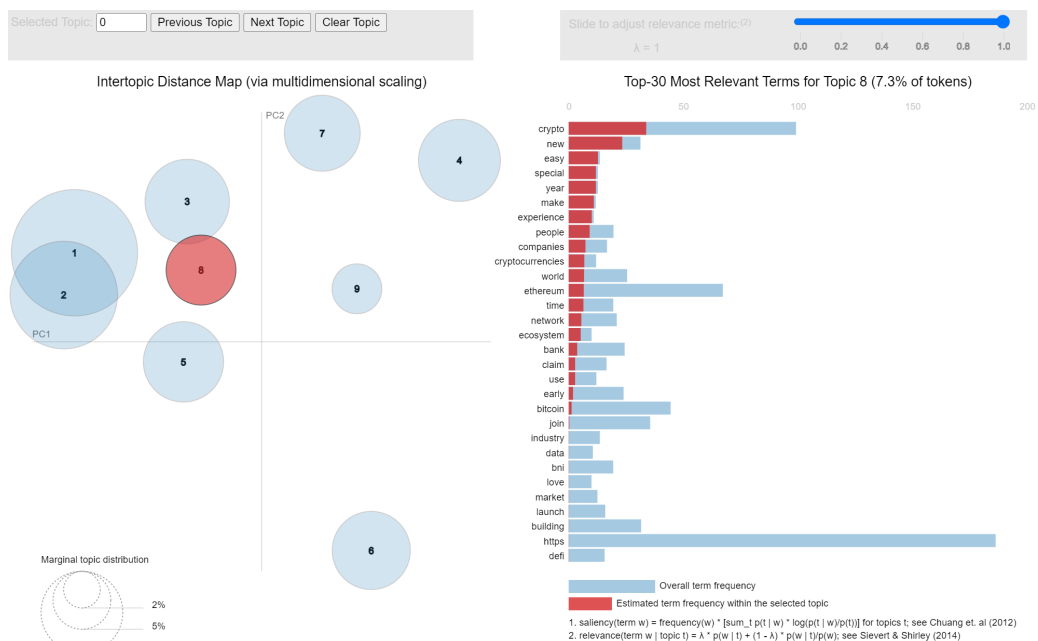


Figure 1: pyLDavis Visualisation

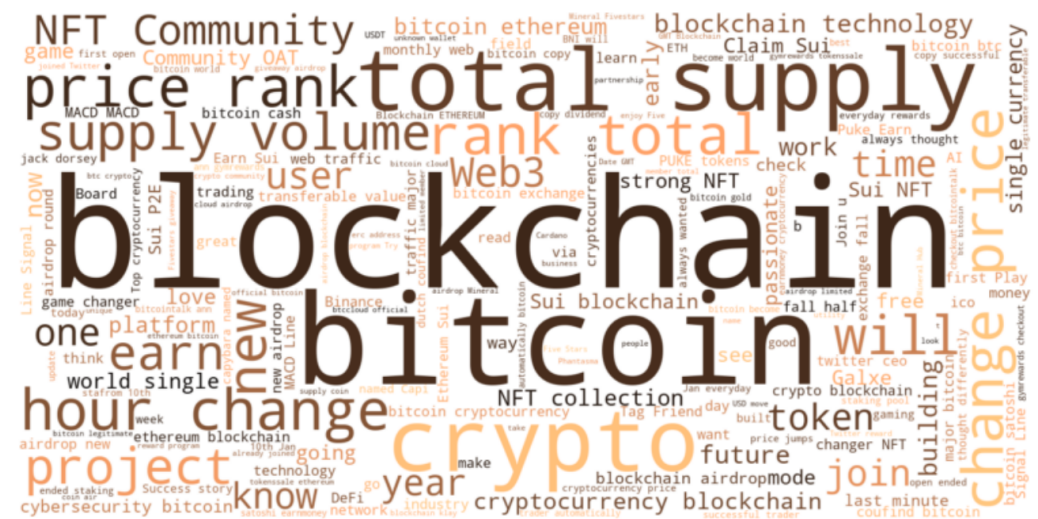


Figure 2: Word Cloud

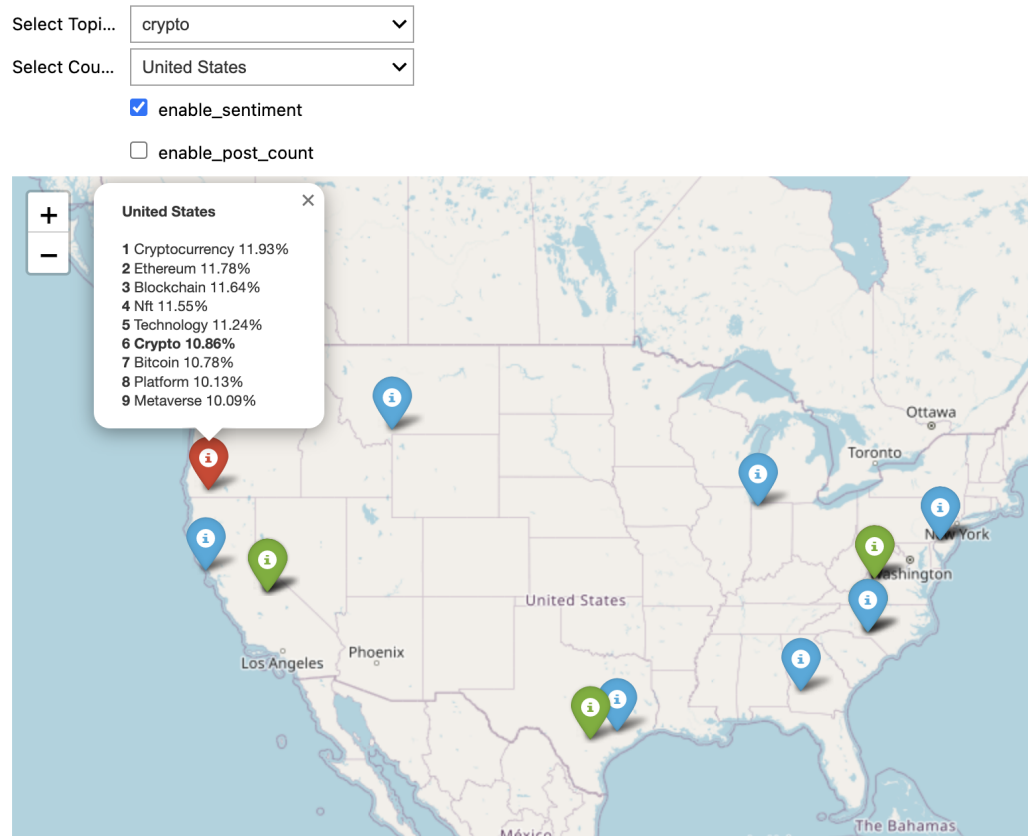


Figure 3: Topic Mapper Frequency Count

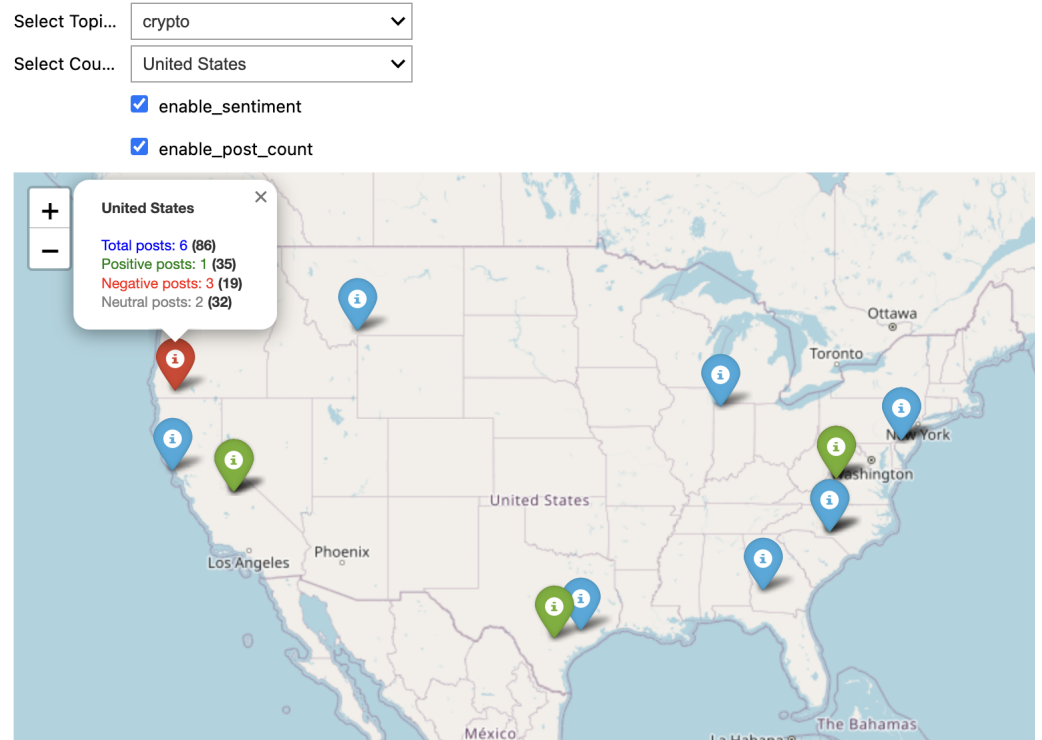


Figure 4: Topic Mapper Sentiment Count

Acknowledgements

We wish to formally acknowledge and express our gratitude for the substantial contributions of Christoph Weisser and Michael Schlee throughout the entire duration of this project.

References

- Bianchi, F., Terragni, S., Hovy, D., Nozza, D., & Fersini, E. (2021). Cross-lingual contextualized topic models with zero-shot learning. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1676–1683. <https://doi.org/10.18653/v1/2021.eacl-main.143>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Buchmüller, A., Kant, G., Weisser, C., Säfken, B., Kis-Katos, K., & Kneib, T. (2022). Twitmo: A twitter data topic modeling and visualization package for r. *arXiv Preprint arXiv:2207.11236*.
- Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Kant, G., Weisser, C., & Säfken, B. (2020). TTLocVis: A twitter topic location visualization package. *Journal of Open Source Software*, 5(54), 2507. <https://doi.org/10.21105/joss.02507>
- Mueller, A. C. (2023). *Wordcloud* (Version 1.9.1). <https://github.com/amueller/wordcloud>
- Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In J. Chuang, S. Green, M. Hearst, J. Heer, & P. Koehn (Eds.), *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (pp. 63–70). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-3110>
- Srivastava, A., & Sutton, C. (2017). *Autoencoding variational inference for topic models*. <https://arxiv.org/abs/1703.01488>
- Stojanovski, D., Dimitrovski, I., & Madjarov, G. (2014). Tweetviz: Twitter data visualization. *Proceedings of the Data Mining and Data Warehouses*, 1–4.
- Terragni, S., Fersini, E., Galuzzi, B. G., Tropeano, P., & Candelieri, A. (2021). OCTIS: Comparing and optimizing topic models is simple! In D. Gkatzia & D. Seddah (Eds.), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 263–270). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.eacl-demos.31>