2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

TrendTracker: Temporal, network-based exploration of long-term Twitter trends

1st John Ziegler, 2nd Johannes Sindlinger, 3rd Marina Walther, 4th Michael Gertz

Institute of Computer Science, Heidelberg University, Heidelberg, Germany
ziegler@informatik.uni-heidelberg.de, johannes.sindlinger@stud.uni-heidelberg.de,
m.walther@stud.uni-heidelberg.de, gertz@informatik.uni-heidelberg.de

***Abstract*—TrendTracker is a web application for the network-based and temporal exploration of long-term social media trends. Topical trends, represented as a series of hashtag co-occurrence networks, can interactively be explored while the user is provided with detailed trend analysis insights. This approach has several benefits compared to alternative trend visualization and exploration methods, such as ranked lists of trending keywords, as it provides the user with additional context-sensitive information. To showcase the TrendTracker application, we leverage a Twitter dataset of German political actors and demonstrate the system's capabilities in various ways. For example, the user is able to investigate a single trend from multiple perspectives, such as the trend's temporal development over time, including its topical shifts and changes in popularity. Also, given the network-based trend visualization, the user can intuitively understand the different facets of a trend and how these are interrelated. Thereby, individual hashtags and relationships can be tracked over time as well. Furthermore, the TrendTracker application allows the user to compare trends. This way, differences in the trends' temporal evolution or topical alignment can be uncovered. The demo is publicly available via the following URL: https://trend-tracker.ifi.uni-heidelberg.de.**

***Index Terms*—social media analytics, trend analysis, trend visualization, Twitter data**

I. Introduction

The study of trends in social media is highly relevant to many use cases. For example, from a business intelligence perspective, it is essential to track what the company's target group is interested in and how specific characteristics, e.g., environmental friendliness, of a product become more or less relevant over time. Also, in the field of political science, it is of great interest to study which political topics are discussed on social media, in *which context* and *how they evolve*. For these use cases, the temporal analysis of long-term trends extracted from social media is crucial. In our context, *long-term* trends must be understood in contrast to short-lived media content. We do not deal with breaking news but focus on topics prevalent in social media over a long period. Given such long-term trends, from an end-user perspective, a suitable visual representation of trending topics is crucial. Simple lists of trending topics are insufficient for several reasons, as described by Bhulai et al. [1]. They do not clearly indicate the relative importance of topics, and the topics' temporal evolution is not represented at all. Finally, multiple topics in these lists might be related and should therefore be clustered for a more descriptive representation as well [1]. The TrendTracker system presented in this demonstration overcomes these shortcomings by providing the user with trend scores over time and visualizing topics in the form of intuitively understandable temporal networks, giving the user additional context-sensitive information. Further advanced trend analysis features are provided. In summary, our contributions to the field of Twitter trend visualization are manifold:

1) The studied trends are visualized as temporal hashtag co-occurrence networks. This visualization method allows users to easily grasp the context of the respective trend and how different aspects of the topic are related.
2) A series of temporal network snapshots represents the long-term temporal development of a trend. Thereby, users can study how a trend, i.e., its focus or relevant aspects, changes over time.
3) Additional trend analysis insights are provided to the user, such as the trend's popularity over time or its relative trend relevance score.
4) Individual hashtags and their relationships can be highlighted in the trend networks and can be tracked over time, allowing to investigate their dynamics regarding the explored trend.
5) Multiple concurrent trends can be explored simultaneously in a comparative manner. Thereby, temporal and topical differences between trends can be uncovered.

Based on our TrendTracker system, we demonstrate how long-term Twitter trends can be explored in a temporal-sensitive context. For this, we leverage networks as the central visualization approach and give users capabilities to explore topically different trends and their development over time.

We thank the Klaus Tschira Foundation for funding this research in the framework of the EPINetz project: https://epinetz.de/.

ASONAM '23, November 6–9, 2023, Kusadasi, Turkey

ACM ISBN 979-8-4007-0409-3/23/11…$15.00
http://dx.doi.org/10.1145/3625007.3627335

Fig. 1. Landing page of the TrendTracker web application. Descriptive explanations facilitate the onboarding of the user.

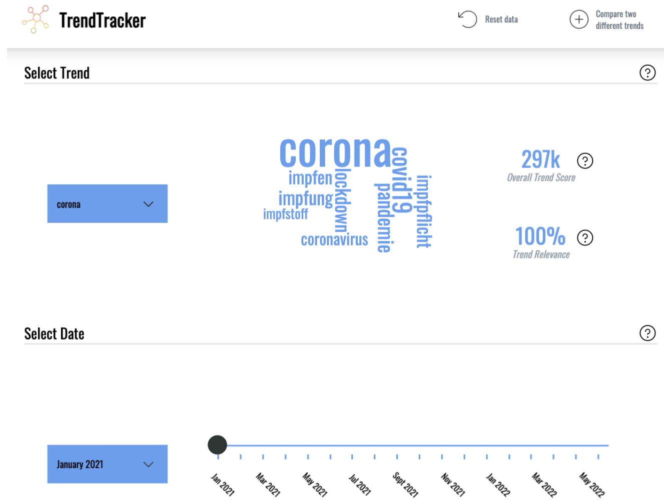

Fig. 2. Settings where users can configure the trend and time period they want to explore

II. Background

Most similar to our work are related approaches from the field of Twitter trend visualization. Several systems have been presented for that in the past, most notably the work by Doshi et al. [2]. They propose a system called *Tweetanalyzer* that allows users to explore real-time trends extracted from Twitter. As trends, they use the most frequently occurring hashtags and user names. Respective statistics are presented in the form of bar charts. Further, tweets related to named trends are displayed on a map to visualize their geographic location. In contrast to our TrendTracker system, they do not provide the user with network-based exploration capabilities, nor do they explore the long-term development of detected trends but are instead focused on their real-time occurrence. Further, one of the early works in the field of trend visualization based on Twitter data was published by Bhulai et al. [1]. To visualize trending topics in the most informative way, they use so-called ”dynamic squarified treemaps” that allow them to not only show the actual trends based on hashtags and terms but also incorporate information regarding the speed and acceleration of the trend development. Additionally, they cluster related topics for them to be more descriptive. Still, they do not incorporate information regarding the long-term development of trends and do not provide the user with network-based exploration capabilities. Further, Wanner et al. [3] propose another visualization technique to track Twitter topics. Using equal-sided triangles to visualize the occurrence of tweets within a timeline allows them to represent the unevenly distributed time-series data concisely. The colour of named shapes indicates the sentiment of respective tweets. By using multiple timelines at once, the user can also compare different topics or rather different time windows and is, therefore, able to investigate a topic’s temporal development. Even though the work of Wanner et al. [3] is dealing with the challenge of visualizing Twitter data, their focus is not on network-based methods and does not explicitly target the tracking of long-term trends as done by our TrendTracker system. More recently, Stojanovsky et al. [4] present a web application called *TweetViz* to explore Twitter data visually. They focus on user- and hashtag-oriented visualizations but do not deal with exploration capabilities for trends. Similarly, Kant et al. [5] publish a Python package called *TTLocVis* to generate, clean, analyze, and visualize Twitter data. Leveraging their package to visualize detected topics over time also allows to track their temporal prevalence. Still, they do not provide network-based exploration capabilities, nor do they deal with topical trends. The same is true for the *RIVA* social media analysis platform proposed by Wu et al. [6]. By counting the occurrence of hashtags, they are able to detect trends and visualize these in the form of pie charts. Nevertheless, further exploration capabilities are not provided.

III. Methodology

This section describes the dataset (cf. Section III-A) and analysis methodology (cf. Section III-B) of the German political Twitter data leveraged to showcase the TrendTracker application. For all the data analysis details, we refer the interested reader to the respective publication: ”No Mayfly: Detection and Analysis of Long-term Twitter Trends” [7]. Next to the data-related methodology, Section III-C gives a comprehensive summary of the system’s technical implementation.

A. Dataset

The used dataset is based on tweets from German political actors as provided by the EPINetz Twitter Politicians Dataset 2021 [8] and covers data from January 2021 until July 2022, in total approximately 1.8 million tweets. To retrieve the raw tweets, we rely on the Twitter search API v2[1] and extract timestamped information about the occurrence and co-occurrence of hashtags. In line with the works of Asur et al.

[1]Search Tweets introduction | Docs | Twitter Developer Platform: https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction; accessed 17/08/23

[9] and Budak et al. [10], we treat hashtags as representatives of topics and do not apply any additional topic extraction technique. In general, the methodology is not restricted to this dataset but can be applied to a broad set of use cases that deal with temporal occurrences of keywords.

B. Trend detection

Taking the timestamped information about the usage of hashtags as described in Section III-A, we construct hashtag co-occurrence networks, each covering the data of one month. To remove noise and focus on the most expressive hashtags, we remove all nodes with a degree lower than the network's median. Degrees follow a power law distribution. Therefore, we leverage the median as defined by Newman [11]. Additionally, edges are weighted by Pointwise Mutual Information to strengthen more semantically expressive relationships between hashtags [12]. In the next step, to cluster related topics as done by Bhulai et al. [1], we apply the Leiden community detection algorithm to each network [13]. The induced subgraphs of the ten most central hashtags of found communities make up the trend networks that the user can explore. To track topics over time, i.e., the communities of hashtags across the network snapshots, we apply the algorithm proposed by Lorenz et al. [14]. As trend scores, the community's cumulative sum of hashtag occurrences is taken. Finally, for the word clouds describing the complete trend, we leverage the joined networks of the temporal communities per trend and take the ten nodes with the highest PageRank [15] scores as representatives per trend. For example, Figure 2 shows the extracted word cloud related to the COVID-19 trend. For all networks, node sizes are adjusted according to their normalized PageRank centrality (cf. Figure 3). Computations are done using the `igraph` network analysis library [16]. On the TrendTracker website, the ten most prevalent trends over time are visualized.

C. Implementation

The TrendTracker web application is built with the SvelteKit framework[2], along with D3[3] and Chart.js[4], that are used for the interactive visualizations. Data about the long-term trends is retrieved from a Python REST API. The TrendTracker system is publicly available via the following URL: https://trend-tracker.ifi.uni-heidelberg.de.

IV. Analysis Workflows

By visiting the TrendTracker website, the user first reaches the landing page as shown in Figure 1. There, an introductory text explains the purpose of the application and facilitates the user's onboarding. Continuing the website visit, by scrolling down or clicking on the "Explore trends" link, the user reaches the actual data exploration part of the web app and is led to the analysis settings (cf. Figure 2). For explanatory purposes,

[2]SvelteKit • Web development, streamlined: https://kit.svelte.dev; accessed 17/08/23

[3]D3 by Observable | The JavaScript library for bespoke data visualization: https://d3js.org; accessed 17/08/23

[4]Chart.js | Open source HTML5 Charts for your website: https://www.chartjs.org; accessed 17/08/23

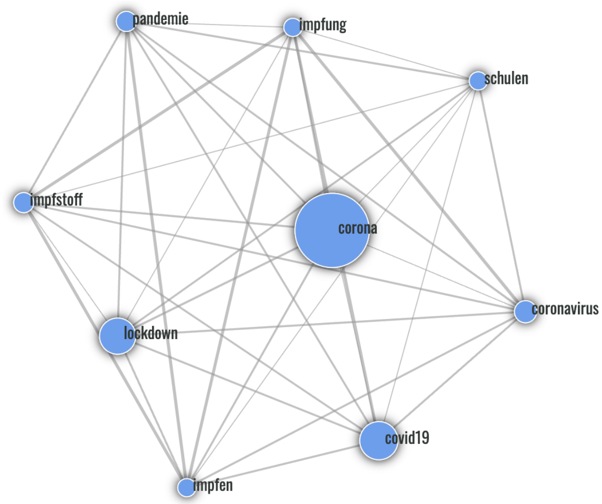

Fig. 3. Exemplary COVID-19 trend network

most application features come with a tooltip that is indicated by an encircled question mark.

A. Single trend exploration

Mentioned analysis settings (cf. Figure 2) allow users to select a given trend and time window. Thereby, additional information and statistics are provided for the trend selection, such as the word cloud visualizing the trend, the overall trend score and its trend relevance. This relevance score is derived by comparing the trend's popularity to the prevalence of the other trends. It, therefore, acts as a relative importance indicator which is a significant improvement compared to ranked lists, according to Bhulai et al. [1]. Furthermore, for the date selection, next to the dropdown list, the right-sided slider can be used to pick the analysis time window.

According to the selected trend and set time window, the trend network (cf. Figure 3) and temporal trend scores (cf. Figure 4) are presented to the user. These charts allow a network-based and temporal exploration of the trends. The built-in reactivity of the application leads to an immediate update of the statistics if the user adjusts the settings. Once a user scrolled past the configuration section, updating the settings is possible via the sticky header (cf. Figure 6). Further, the layout of the trend network can be adjusted to the user's preferences by dragging the respective nodes in the network.

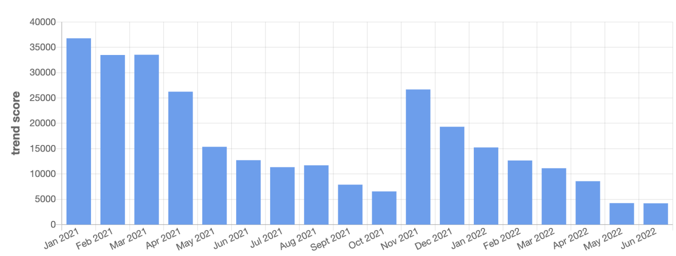

Fig. 4. Trend scores of the COVID-19 trend

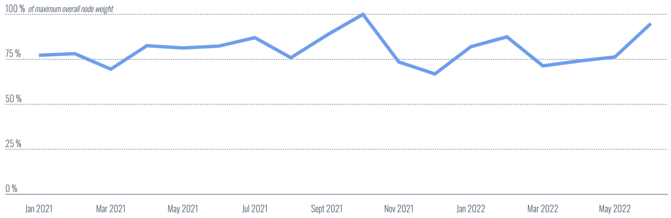

Fig. 5. Temporal node weights of the *corona* hashtag as part of the COVID-19 trend networks

B. Node and edge tracking

Not only can the users adjust the positioning of nodes in the trend networks, as mentioned above, but they can also track these across multiple network snapshots. By clicking on a node or a link, it is highlighted. If the user then changes the time window of the trend, the node/edge remains highlighted in the network in case it is still present. Thereby, the temporal tracking of individual aspects of a trend (i.e., represented by a hashtag) is facilitated. This way, the user can check whether some hashtags gain or lose importance for the trend over time. Besides the manual tracking of the highlighted entity, a line chart of temporal node/edge weights is also displayed. It shows the user the evolution of the importance of the highlighted entity. Figure 5 gives an example of such a line chart. It shows the node weights of the *corona* hashtag occurring in the COVID-19 trend networks.

C. Trend comparison

Some use cases also benefit from a direct comparison of two trends. Therefore, the TrendTrack application offers the capability to conduct two analyses in parallel. For both analyses, the trend and time window can be set individually. This way, the users are not limited to comparing the *same* trend at arbitrary points in time, but they can also contrast the state of *different* trends at the same or different times. Figure 6 shows such a comparison which contrasts the COVID-19 trend as of January 2021 with the state of the EU trend in March 2022. Generally, while the users conduct a trend comparison, the same features, such as node/edge highlighting and the temporal trend scores, are still available and are applied to both investigated trends.

V. Conclusion

This demonstration presents the TrendTracker web application to explore long-term social media trends. Its capabilities are demonstrated based on Twitter data of German political actors collected from January 2021 until July 2022. Users of the TrendTracker system can explore the temporal evolution of long-term prevalent topics in various ways. Especially the network-based visualizations give the user context-sensitive exploration capabilities, which is a significant benefit compared to existing Twitter trend visualization approaches. Additional features of the web application, such as the temporal trend scores, the trend comparison functionality or the entity highlighting, complement the application's capabilities.

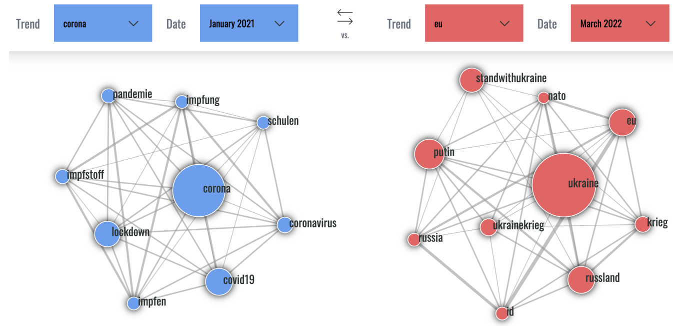

Fig. 6. Exemplary comparison of two trends. Both the trend and the time window can be adjusted individually.

References

[1] S. Bhulai, P. Kampstra, L. Kooiman, G. Koole, M. Deurloo, and B. Kok, "Trend visualization on Twitter: what's hot and what's not?" in *1st International Conference on Data Analytics*, 2012, pp. 43–48.

[2] Z. Doshi, S. Nadkarni, K. Ajmera, and N. Shah, "Tweeranalyzer: Twitter trend detection and visualization," in *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. IEEE, 2017, pp. 1–6.

[3] F. Wanner, A. Weiler, and T. Schreck, "Topic Tracker: Shape-based Visualization for Trend and Sentiment Tracking in Twitter," in *VisWeek*, 2012.

[4] D. Stojanovski, I. Dimitrovski, and G. Madjarov, "Tweetviz: Twitter data visualization," *Proceedings of the data mining and data warehouses*, vol. 1, no. 2, 2014.

[5] G. Kant, C. Weisser, and B. Säfken, "TTLocVis: A Twitter Topic Location Visualization Package," *Journal of Open Source Software*, vol. 5, no. 54, p. 2507, 2020.

[6] Y.-T. Wu, H.-Y. Hsieh, X. K. Sigalingging, K.-W. Su, and J.-S. Leu, "RIVA: A Real-Time Information Visualization and analysis platform for social media sentiment trend," in *2017 9th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2017, pp. 256–260.

[7] J. Ziegler and M. Gertz, "No Mayfly: Detection and Analysis of Long-term Twitter Trends," *BTW 2023*, 2023.

[8] T. König, W. J. Schünemann, A. Brand, J. Freyberg, and M. Gertz, "The EPINetz Twitter Politicians Dataset 2021. A New Resource for the Study of the German Twittersphere and Its Application for the 2021 Federal Elections," *Politische Vierteljahresschrift*, pp. 1–19, 2022.

[9] S. Asur, B. A. Huberman, G. Szabo, and C. Wang, "Trends in social media: Persistence and decay," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, 2011, pp. 434–437.

[10] C. Budak, D. Agrawal, and A. El Abbadi, "Structural trend analysis for online social networks," *Proceedings of the VLDB Endowment*, vol. 4, no. 10, pp. 646–656, 2011.

[11] M. E. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary physics*, vol. 46, no. 5, pp. 323–351, 2005.

[12] F. Role and M. Nadif, "Handling the impact of low frequency events on co-occurrence based measures of word similarity," in *Proceedings of the international conference on Knowledge Discovery and Information Retrieval (KDIR-2011). Scitepress*, 2011, pp. 218–223.

[13] V. A. Traag, L. Waltman, and N. J. Van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[14] P. Lorenz, F. Wolf, J. Braun, N. Djurdjevac Conrad, and P. Hövel, "Capturing the dynamics of hashtag-communities," in *International Conference on Complex Networks and their Applications*. Springer, 2017, pp. 401–413.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[16] G. Csardi, T. Nepusz *et al.*, "The igraph software package for complex network research," *InterJournal, complex systems*, vol. 1695, no. 5, pp. 1–9, 2006.