# Time-aware public opinion analytics of Hashtag movements

**Time-aware public opinion analytics of Hashtag movements**

In the past two decades, the measurement and analysis of social media data have grown in importance. Citizens are increasingly likely to use social media platforms to access political information and participate in civic and political protests or movements and express their political opinions. (Santos et al., 2021; Stieglitz & Dang-Xuan, 2013) These proclivities have corresponded with surging interest amongst political analysts and academics in forecasting political events based on how citizens behave and what they discuss on social media. Models for analysis include sophisticated inputs about the semantic and structural characteristics of social media trends. However, as social media platforms evolve and profiles, networks, and public feeds change, many classic studies of public opinion are increasingly hard to replicate. Public opinion calculated through social media admittedly has limited generalizability to understand the offline demographic; nevertheless, scholars have supported its use in comparative analyses of public opinion (Jaidka et al., 2019; Kou et al., 2017). A feature that is increasingly important to the role of social media platforms in the formation and maintenance of networked publics is the 'stream of consciousness' public feed as the live, ever-scrolling narrative of public opinion, delivered as posts, comments, trending topics, and events. Of course, each user is privy to only a sliver of the social media pulse which social media platforms have algorithmically curated based on the profiles they follow, the topics they consume, and the links they click (Gadde & Beykpour, 2018). Nevertheless, the sense of belonging to a collective is reinforced through their witnessing of the single, public feed common to everyone on the social media platform, and the common experience and understanding of the present time, emphasized by how most platforms timestamp content relative to the current moment ("2 minutes ago"), instead of when the content was posted (Leaver et al., 2020). The dynamics of public feeds are helpful to understand how information (and consequently, public opinion) spreads on social media (Vosoughi et al., 2018; Zhang et al., 2017), which can be modeling similar to viral diffusion, as person-to-person rather than a one-to-many contagion (Liang, 2018). They describe shifts in opinion related to developing events (Khosla et al., 2019) as well as the divergent discourse of counternarratives (Gallagher et al., 2018). They are increasingly valuable to model how exogenous events can affect online discourse, such as the emotions of political discourse (Ahmed et al., 2017). However, a pressing challenge remains that findings are often contingent on key researcher decisions (Bermingham & Smeaton, 2011; O'Connor et al., 2010), such as the operationalization of time in social media discourse: the unit of time to measure changing public attitudes. Even with temporal weighting (Skoric & Jaidka, 2023), electoral contexts are increasingly hard to obtain valid measurements of (Bossetta, 2018), and these approaches have not been validated on other general or trending topics of discourse.

The conversations on social media platforms are by no means static. The dynamic nature of user-to-user tie formation and delinkage alongside the shifting focus of public attention results in a continuous change of public opinion. These shifts in public opinion can typically be inferred from topical bursts, represented by hashtags in the social media realm. We call these forms 'hashtag archetypes,' and they represent the core focus of this paper. A hashtag is a phrase preceded by the # symbol and is a linguistic marker that expresses an ambient affiliation with a discussion topic (Geboers & Van De Wiele, 2020). Hashtags are used to discover emerging and breaking events through the concentration of public opinion (Kong et al., 2014). It is then crucial to differentiate types of hashtag archetypes to construct a better picture of the topic life cycle within the social communities (Dey et al., 2018). In this paper, we focus on three archetypes. The first is a topic constantly discussed in the background of exciting and emerging conversations. These types of topics are typically long-lasting events, such as the basis of the long-lasting coronavirus pandemic or the long-drawn 2022 Russia-Ukraine conflict. Hashtags such as "#coronavirus" and "#russiaukrainewar" hold the public's interest for a sustained period of time. The second hashtag archetype depicts a fade in the interest of the public towards a topic, typically characterized by a surge of conversations followed by a dearth of discourse. The last is the topics that first have an attention swell, have a fleeting moment where interest had been lost before the topic returns to public scrutiny.

In this study, we address these research gaps in the context of measuring topical social media discussions of different temporal forms. We pose the following research questions:

- How can we profile hashtag archetypes from social media discussions?
- How can we make accurate, dynamic, public opinion measurements from different hashtag archetypes?
- Do the public opinion measurements from hashtag archetypes correspond with offline measures of public opinion?

In the following section, we will discuss our data and analytical approach.

## Methods

### Primary data

We collected data from Twitter using the Twitter V1 Streaming API on three events: (1) **Pelosi Dataset** collected on Oct 30 - Dec 20 2022 with search terms (“pelosi attack" OR “depape" ); (2) **Vaccine dataset** collected on 1 Jan - 31 Oct 2022 with search terms ((“covid" OR “coronavirus") AND (“vax" OR “vaccine")) ; (3) **Russian-Ukraine dataset** collected on 24 Feb - 6 Nov 2022 with search terms (“russian invasion" OR “russian military" OR “invasion of ukraine"). These three events were selected to cover a wide range of online discourse, from politics to pandemics to a crisis. In total, we analyze over 9,880,057 tweets and 2,897,282 users.

### Survey data

For triangulating our opinion measures, we focused on the following sources:

- Presidential Polls dataset: Presidential Polls provided us with dynamic opinion measures for Nancy Pelosi[1], available at the weekly level, which offered a way to validate our measures from the Pelosi dataset.
- COVID19 beliefs dataset (Collis et al., 2022): The Global Survey on COVID19 Beliefs, Behaviors, and Norms, available at the country level, offered a way for us to validate our measures from the Vaccine dataset.

[1] https://projects.fivethirtyeight.com/polls/president-primary-r/2024/national/

- Gallup news dataset: We are in the process of acquiring access to responses on the Russia-Ukraine war in the Gallup News dataset[2], a licensed dataset which measures American and worldwide public opinion on the Russia-Ukraine war from a demographically representative sample of respondents.

## Approach

### *Profiling Hashtag Archetypes*

We profiled three key archetypes of hashtag longevity patterns: (1) **Constant:** the usage of the hashtag is generally constant throughout the discourse indicating a central topic or the evolution to a mainstream trend; (2) **Dies** where the amount of discourse fades away quickly over time; and (3) **Revives** where the topic oscillates between trending and fading away. A hashtag is profiled into one of the three archetypes of hashtags based on their frequency profile across the days, shown in Figure 1.

To differentiate between the three hashtag archetypes, we use the following heuristic. For each hashtag, we first plot a frequency-time graph. Next we use the SciPy library to find peaks of the frequency-time signal[3]. Then, we measure the distance between peaks and the amplitudes of each peaks to determine the hashtag type. (1) A hashtag is **constant** if there are no peaks or there are multiple peaks but the difference in amplitudes are very small. (2) A hashtag has **died** if there are multiple peaks but the amplitude of peaks have a decreasing gradient till at least more than half the original frequency. (3) A hashtag has **revived** if the subsequent peaks are at least more than half of the original frequency.

With the hashtags sorted into the three archetypes for each dataset, we identify the total number of days the hashtags are present in the dataset. This time window indicates the length this trend remained sticky and fresh in the mind of social media users. We then perform an ANOVA analysis across each of the datasets to compare the total number of days across the three types of hashtags. We look at the total number of days rather than the half life of a hashtag as per previous

[2] https://news.gallup.com/poll/513680/american-views-ukraine-war-charts.aspx

[3] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html

works (Glasgow & Fink, 2013) because a constant and a revived hashtag do not necessarily have appropriate half lives. Table 1 shows that the total number of tweets and total number of days observed of the three types of hashtags are significantly different ($p<0.05$ level). Therefore, our formulation of hashtag archetypes do split the hashtag types into distinct groups. Thus, we can use this formulation to examine the trends that stick around (constant hashtag), fizzle away (hashtag dies) or have seasonal trending patterns (hashtag revives).

In our paper, we will compare different operationalizations of opinion measurement and apply Granger causality estimation to validate our measures of social media-based public opinion against survey estimates.

## References


Ahmed, S., Jaidka, K., & Cho, J. (2017). Tweeting india's nirbhaya protest: A study of emotional dynamics in an online social movement. *Social Movement Studies*, *16*(4), 447–465.

Bermingham, A., & Smeaton, A. (2011). On using twitter to monitor political sentiment and predict election results. *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, 2–10.

Bossetta, M. (2018). The digital architectures of social media: Comparing political campaigning on facebook, twitter, instagram, and snapchat in the 2016 us election. *Journalism & mass communication quarterly*, *95*(2), 471–496.

Collis, A., Garimella, K., Moehring, A., Rahimian, M. A., Babalola, S., Gobat, N. H., Shattuck, D., Stolow, J., Aral, S., & Eckles, D. (2022). Global survey on covid-19 beliefs, behaviours and norms. *Nature Human Behaviour*, *6*(9), 1310–1317.

Dey, K., Kaushik, S., Garg, K., & Shrivastava, R. (2018). Topic lifecycle on social networks: Analyzing the effects of semantic continuity and social communities. *European Conference on Information Retrieval*, 29–42.

Gadde, V., & Beykpour, K. (2018). Setting the record straight on shadow banning. *Twitter Blog*. https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html

Gallagher, R. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2018). Divergent discourse between protests and counter-protests:# blacklivesmatter and# alllivesmatter. *PloS one*, *13*(4), e0195644.

Geboers, M. A., & Van De Wiele, C. T. (2020). Machine vision and social media images: Why hashtags matter. *Social Media+ Society*, *6*(2), 2056305120928485.

Glasgow, K., & Fink, C. (2013). Hashtag lifespan and social networks during the london riots. *Social Computing, Behavioral-Cultural Modeling and Prediction: 6th International Conference, SBP 2013, Washington, DC, USA, April 2-5, 2013. Proceedings 6*, 311–320.

Jaidka, K., Ahmed, S., Skoric, M., & Hilbert, M. (2019). Predicting elections from social media: A three-country, three-method comparative study. *Asian Journal of Communication*, *29*(3), 252–273.

Khosla, S., Chhaya, N., Jindal, S., Saha, O., & Srivastava, M. (2019). Do events change opinions on social media? studying the 2016 us presidential debates. *International Conference on Social Informatics*, 287–297.

Kong, S., Mei, Q., Feng, L., Ye, F., & Zhao, Z. (2014). Predicting bursts and popularity of hashtags in real-time. *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 927–930.

Kou, Y., Kow, Y. M., Gui, X., & Cheng, W. (2017). One social movement, two social media sites: A comparative study of public discourses. *Computer Supported Cooperative Work (CSCW)*, *26*, 807–836.

Leaver, T., Highfield, T., & Abidin, C. (2020). *Instagram: Visual social media cultures*. John Wiley & Sons.

Liang, H. (2018). Broadcast versus viral spreading: The structure of diffusion cascades and selective sharing on social media. *Journal of Communication*, *68*(3), 525–546.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Fourth international AAAI conference on weblogs and social media*.

Santos, J. S., Bernardini, F., & Paes, A. (2021). A survey on the use of data and opinion mining in social media to political electoral outcomes prediction. *Social Network Analysis and Mining*, *11*(1), 103.

Skoric, M., & Jaidka, K. (2023). Electoral predictions from social media data. *Skoric, Marko and Jaidka, Kokil, Electoral Predictions from Social Media Data (February 1, 2023). In A. Ceron (Ed.), Elgar Encyclopedia of Technology and Politics*.

Stieglitz, S., & Dang-Xuan, L. (2013). Social media and political communication: A social media analytics framework. *Social network analysis and mining*, *3*, 1277–1291.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151.

Zhang, Y., Moe, W. W., & Schweidel, D. A. (2017). Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, *34*(1), 100–119.

| Dataset | Total Tweets | | | Total Days | | |
|---|---|---|---|---|---|---|
| | **Degrees of Freedom** | **F value** | **Pr(>F)** | **Degrees of Freedom** | **F value** | **Pr(>F)** |
| Pelosi Dataset | 2 | 912.73 | 0.0 | 2 | 11279.10 | 0.0 |
| Vaccine Dataset | 2 | 2368.91 | 0.0 | 2 | 6509.82 | 0.0 |
| Russia-Ukraine Dataset | 2 | 453.50 | 9.4E-7 | 2 | 24484.70 | 0.0 |

**Table 1**

*ANOVA results comparing across the three hashtags archetypes*

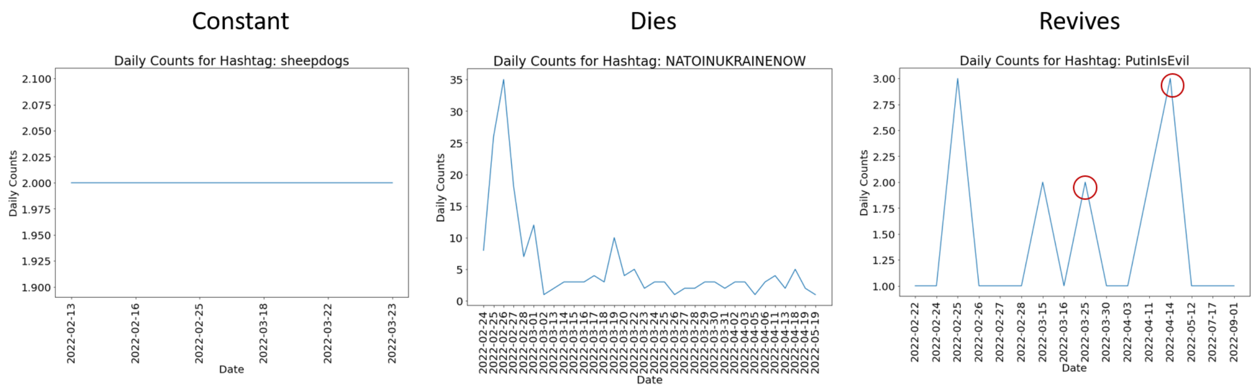


**Figure 1**

*Frequency profiles of the three archetypes of hashtags. “cahispanic" is a constant hashtag; “pelosiattack" is a hashtag that dies, and “NYTimes" is a hashtag that revives. increase font size*