Dynamic Social Media Monitoring for Fast-Evolving Online Discussions

Maya Srikanth[1], Anqi Liu[1], Nicholas Adams-Cohen[2], Jian Cao[1], R. Michael Alvarez[1] and Animashree Anandkumar[1]

[1]California Institute of Technology, [2]Stanford University

Motivation

- Social media discussions are dynamic
- Evolving keywords and hashtags allow hate speech to propagate without detection[1]
- Data streaming methods rely on static and outdated keywords
- **Proposal: Build dynamic monitor to track fast-changing discussions and detect online abuse**

Dynamic Method

Algorithm

- Begin with initial keyword set s_0 at $t = 0$.
- Repeat until $t = T$:
  - Use keyword set s_t to stream dataset K_t
  - $G_t \leftarrow$ obtain 50-dimension GloVe[2] embeddings trained on K_t. $P_t \leftarrow$ update time series models with latest frequency data from corpus.
  - For each word $s \in s_t$: **
    - Find n closest neighbors via cosine distance
    - $C_s \leftarrow$ choose relevant neighbor keywords
    - $C_{s'} \leftarrow$ discard hashtags with declining trend or low corpus counts
    - $s_t \leftarrow C_{s'}$

** use interface UI (Figure 3)

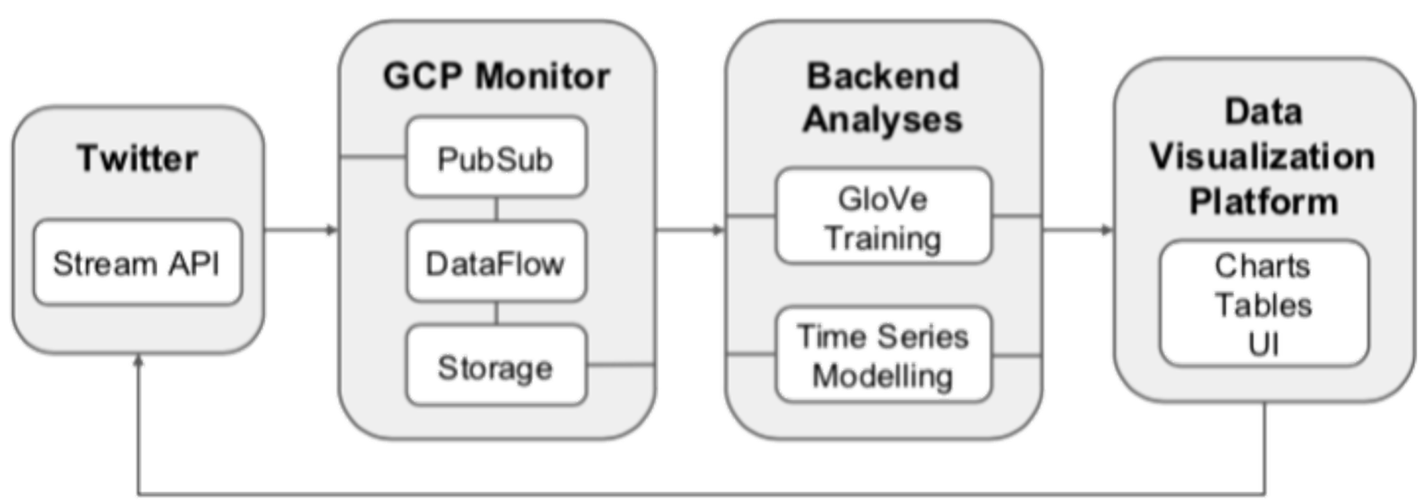


Figure 1. Data Visualization Platform Workflow.

Data Vis Platform Application

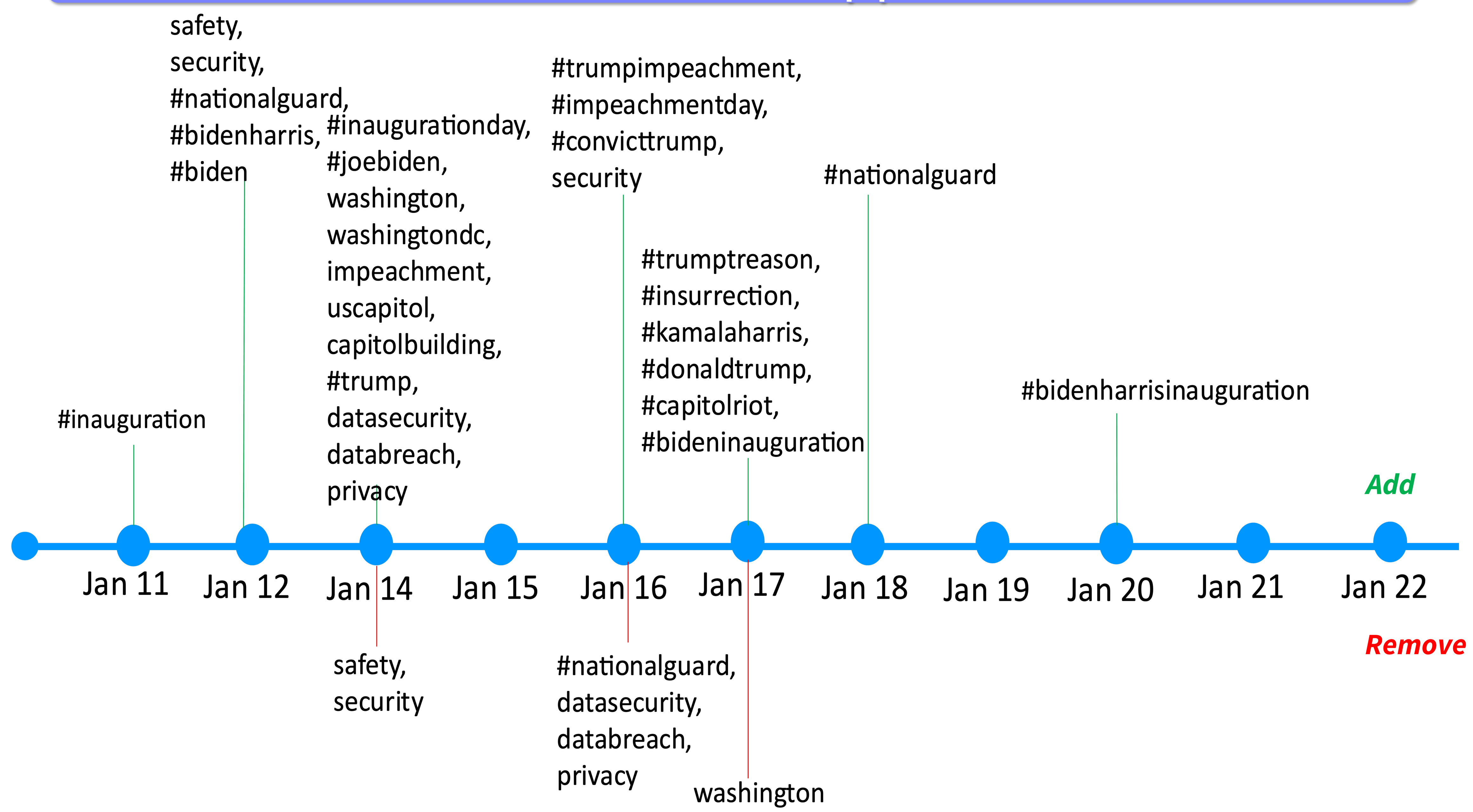
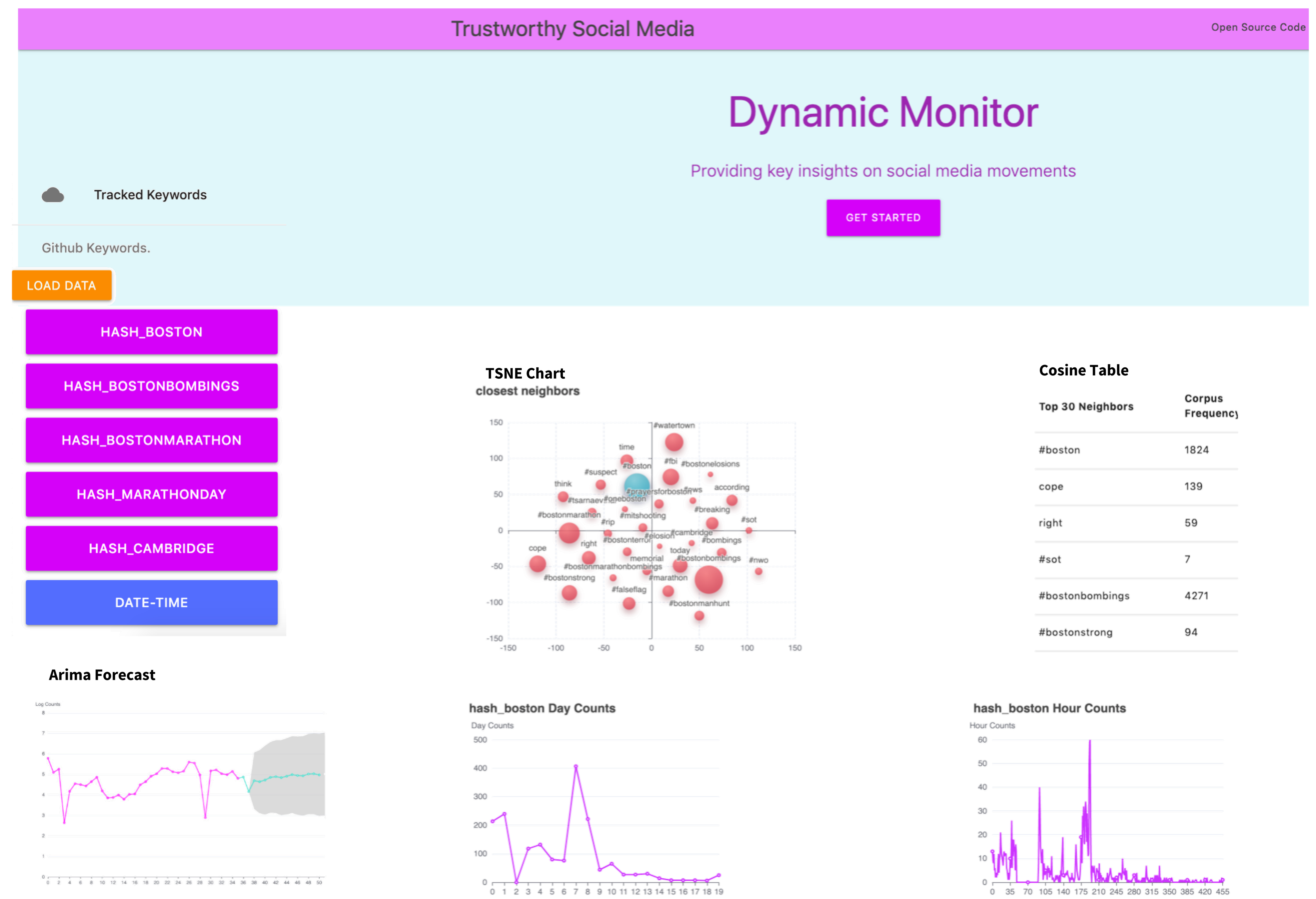


Figure 2. Dynamic keyword updates (add & remove) in real-time case study of 2021 #inauguration discussions on Twitter.

Figure 3. Frontend of Data Vis Platform used in dynamic method. **Features include**:

- **Tracked keywords**
- **Cosine table** – 30 nearest neighbors, sorted by cosine distance & frequency
- **TSNE chart** – 30 nearest neighbors
- **Arima forecast** – forecast of raw keyword counts with optimal Arima model



Case Studies Overview

- **2021 Presidential Inauguration on Twitter (algorithm + interface, real-time)**
  - Used dynamic method to study Twitter real-time discussions concerning the presidential inauguration.
- Figure 2 shows evolution of keyword set used for data collection
  - Dynamic Method based on embeddings & frequencies captures more discussion than static set of keywords

- **2017 #Metoo (algorithm, historical simulation)**
  - Simulated dynamic method on 12 months of historical #MeToo data. Results summarized in Table 1.
  - **Dynamic** ($n = 15$ keywords): uses embeddings and frequency data from previous month to pull data
  - **Last-top**: uses top 15 most frequent hashtags in previous month to pull data
  - **Static**: uses top 15 hashtags in January to pull data throughout all months

	Jaccard Similarity	Avg. F1 Weighted	Avg. F1 Unweighted
Dynamic	.5406	.6976	.7083
Last-Top	.508	.6665	.6041
Static	.4594	.6199	.5166

Table 1. Quantitative results from simulating dynamic method and 2 baseline methods on millions of historical #MeToo data. Dynamic method earns higher avg. F1 score than frequency-based monitors. F1 score and Jaccard similarity are calculated with respect to a ground truth set of 20 most popular hashtags in the entire universe of monthly #MeToo tweets. Weighting accounts for size of monthly data.

[1] Liu, A., Srikanth, M., et. all. 2019. Finding social media trolls: Dynamic keyword selection methods for rapidly-evolving online debates. In *AI For Social Good Workshop, NeurIPS*.
[2] Pennington, Jeffrey, Richard Socher and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. pp. 1532–1543.