Bursty Event Detection Throughout Histories

Debjyoti Paul, Yanqing Peng, Feifei Li

deb, ypeng, lifeifei @cs.utah.edu

InitialDLab, School of Computing
University of Utah

What is a *bursty event*?

The burstiness of an event at any time t is defined by the *acceleration* of event mentionings for a defined time span of observation. *Burstiness* measures the increase in incoming rate of event mentionings, rather than the incoming rate itself. This measure quickly identifies important events that are emerging.

What is bursty event detection *throughout histories*?

Existing works focus on the detection of current trending events, but it is important to be able to go back in time and explore bursty events throughout the history. Our system can answer the following type of questions in realtime with space-efficient probabilistic data structure and algorithms.

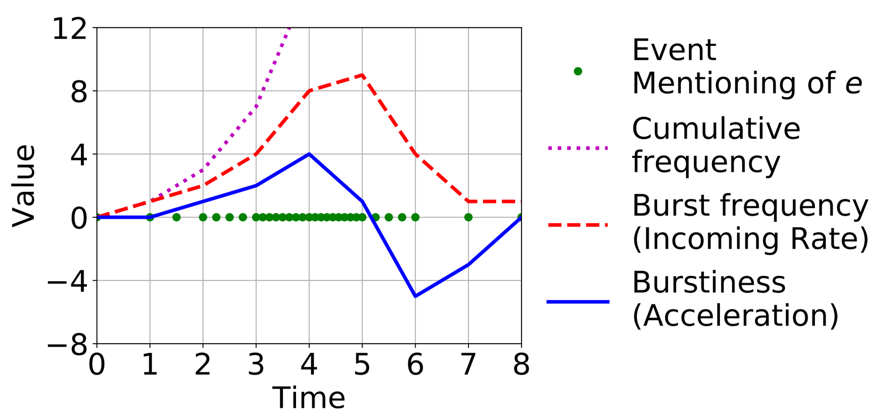
(i) What are the bursty events in the first week of October in 2016?
(ii) Is "Anthem Protest" a bursty event in second week of September in 2017?

Key Contributions

1. The paper first presents a formal definition of burstiness $b(t)$ for any historical time t.
2. PBE: The probabilistic data structure estimates burstiness $b(t)$ in a single event stream.
3. Two PBE's, PBE-1: Approximation with buffering, PBE-2: Approximation without buffering.
4. CM-PBE: Estimate burstiness $b_e(t)$ of any event e at any time t over a multi-event stream.
5. Bursty Event Detection Query: Find all the bursty events for a given a time range $[start, end]$.
6. Extensive empirical investigation using Olympic Rio 2016 and US Election 2016 tweets.
7. Space efficient and query efficient probabilistic data structure with theoretical guarantees.

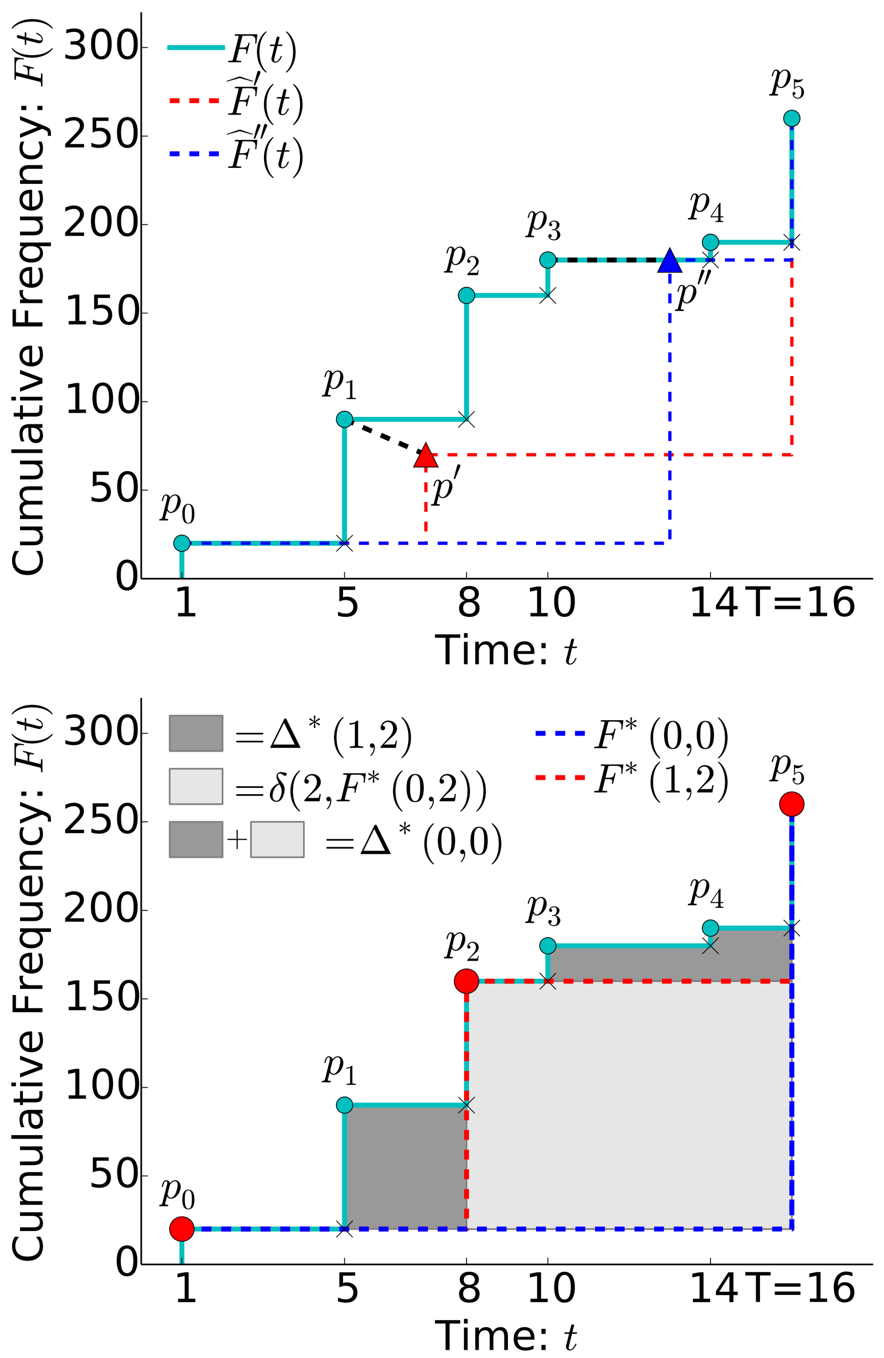
Preliminaries

Notation	Meaning
e	an event or an event id
$M = \{(m_1, t_1), (m_2, t_2), \cdots\}$	stream of timestamped messages
$h : m_i \rightarrow [1, K]$	map a message to an even id
$S = \{(a_1, t_1), (a_2, t_2), \cdots\}$	event stream with (id, timestamp)
S_e	event stream of e (timestamp only)
$S[t_1, t_2]$	substream of S in time range $[t_1, t_2]$
τ	burst span
$F_e(t)$	cumulative frequency of e in $S[0, t]$
$f_e(t_1, t_2)$	frequency of e in $S[t_1, t_2]$
$bf_e(t) = f_e(t - \tau, t)$	burst frequency (incoming rate) of e at t
$b_e(t) = bf_e(t) - bf_e(t - \tau)$	burstiness (acceleration) of e at t



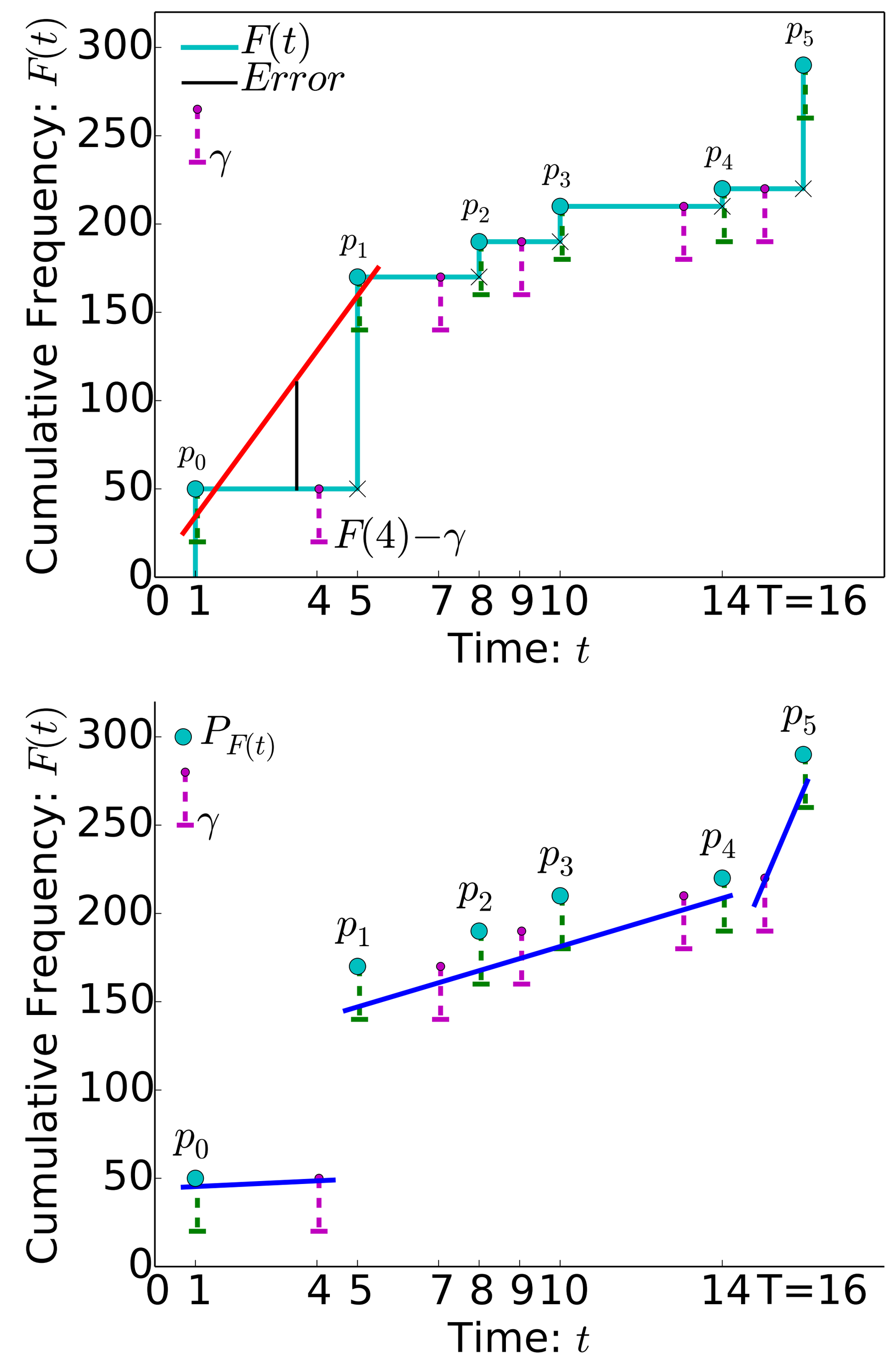
An example of burst where $\tau = 1$.

PBE-1 Buffered



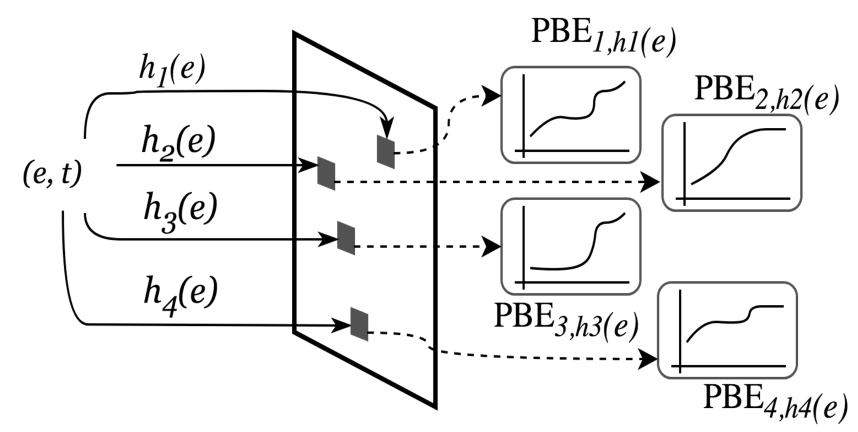
Principle for PBE-1 approximation.

PBE-2 Non-Buffered

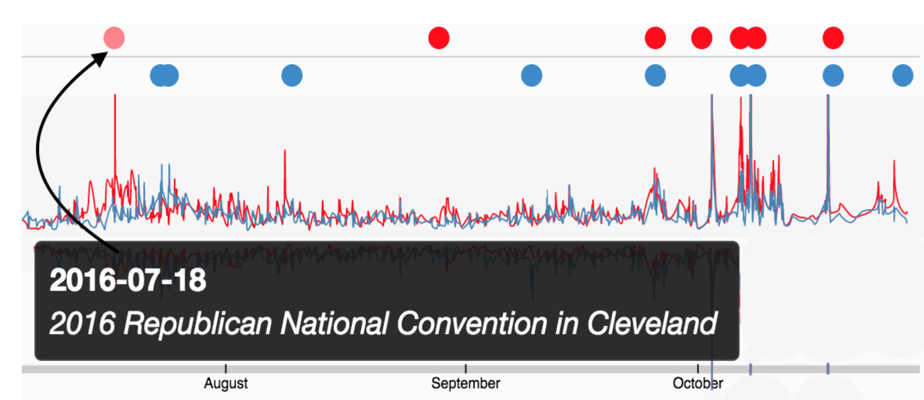


Principle for PBE-2 approximation.

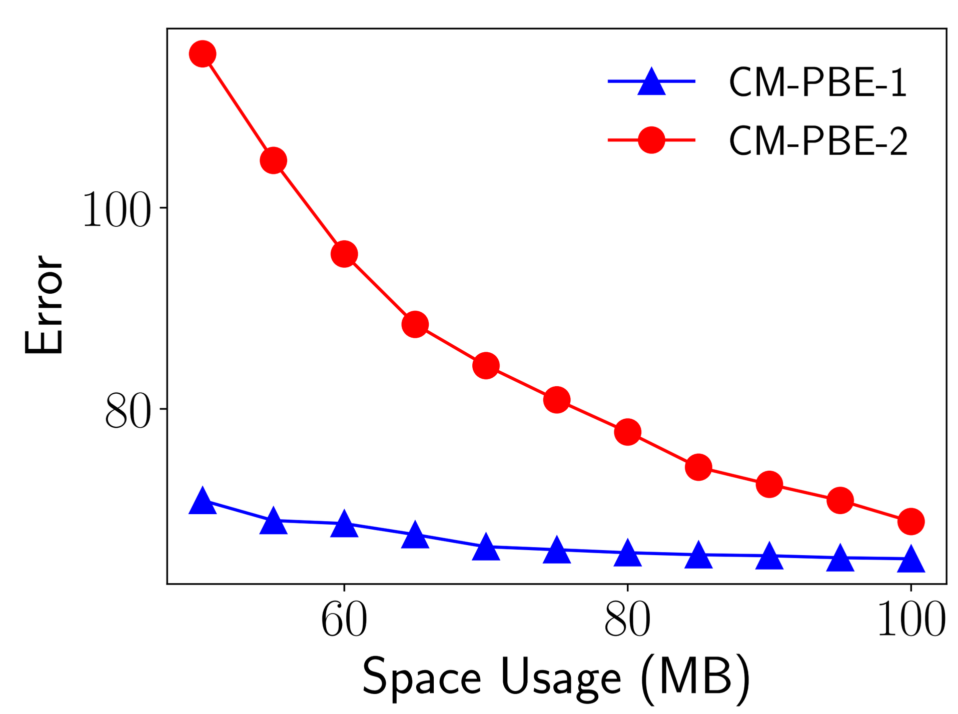
CM-PBE & Results



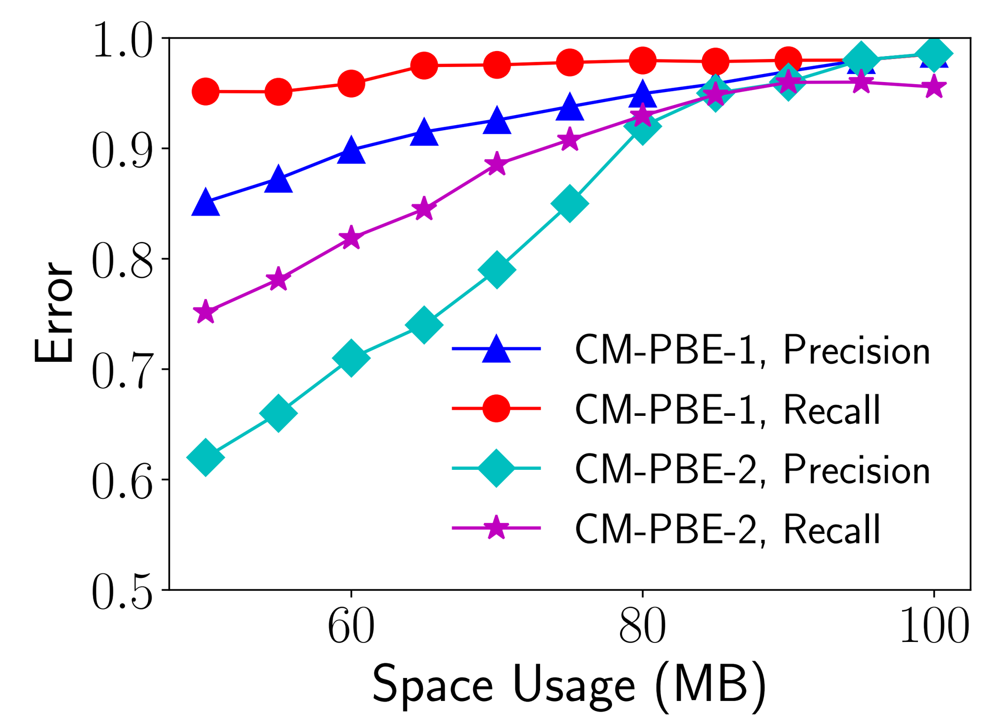
Pictorial representation CM-PBE



Bursty Events from USElection 2016



Absolute Error vs Space OlympicRio



Precision-Recall vs Space OlympicRio