Crisis Sub-Events on Social Media: A Case Study of Wildfires

Shan Jiang*†, William Groves‡, Sam Anzaroot‡, Alejandro Jaimes‡

*Work done during an internship at Dataminr

†Northeastern University, ‡Dataminr

sjiang@ccs.neu.edu, {wgroves, sanzaroot, ajaimes}@dataminr.com

Motivation

- Social media becoming a powerful tool during crisis events;
- A series of *sub-events* occurring as a major crisis unfolds;
- Understanding sub-events is crucial for crisis management.

Contribution

- A framework to identify sub-events on social media;
- A case study of sub-events after California wildfires.

Analytical Framework

- Manually curating a small set of keywords to query tweets;
- Messages reporting sub-events are (n, v) pairs, e.g., fire reported;
- Dependency parsing and traversal to extract (n, v) pairs;
- Word2vec representation of (n, v) pairs;
- Clustering similar (n, v) pairs as a sub-event;
- Labeling tweets with sub-events and running application tasks.

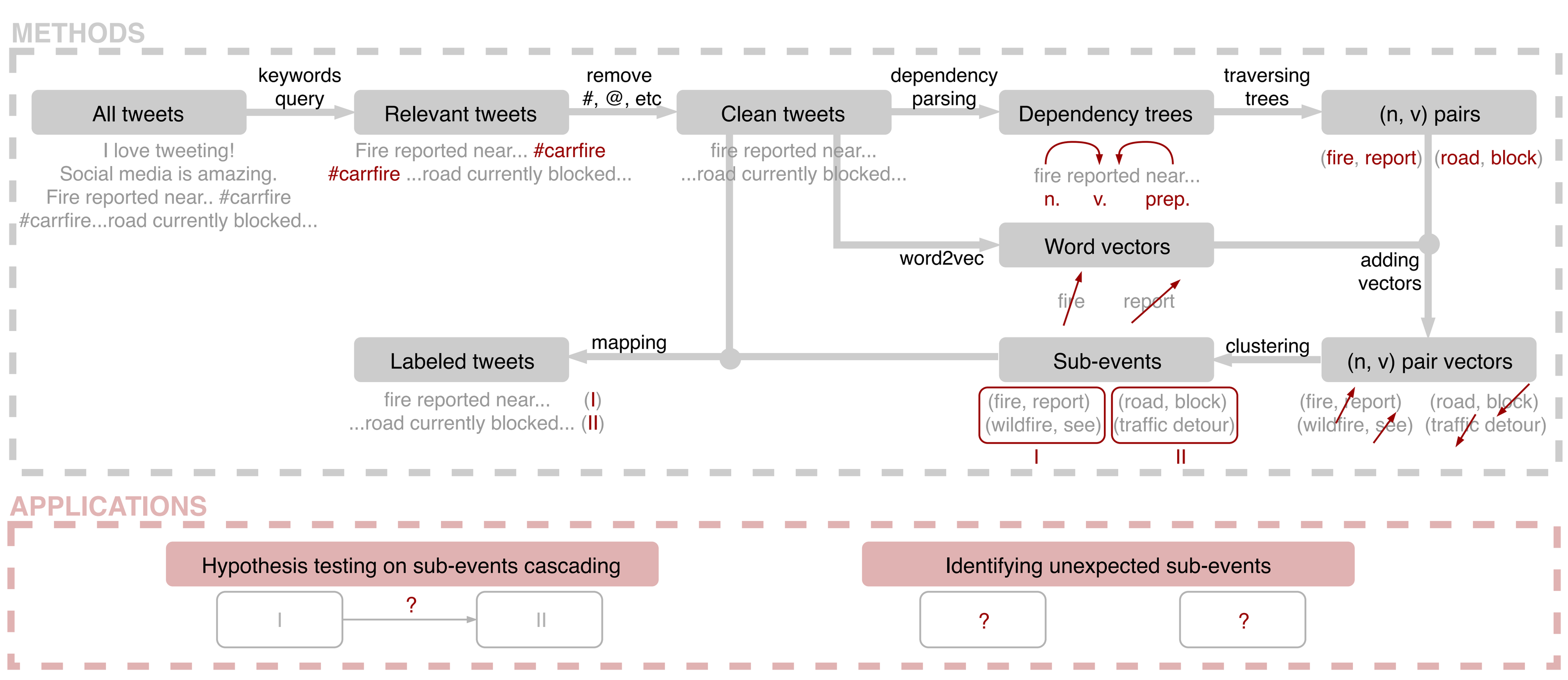


Figure 1: Methods and applications.

Wildfire Data

- 4 largest wildfires in California during 2018 to 2019;
- Wildfires names, curated queries, tweet numbers shown in Table 1;
- Distribution figures fitted using kernel density estimates;
- Wildfire starting dates marked by red lines;
- Selected time periods marked by shaded areas.

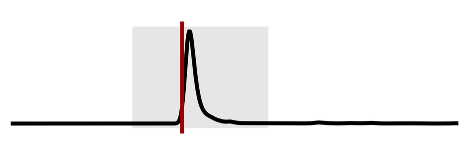
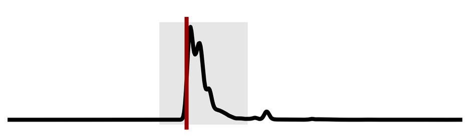
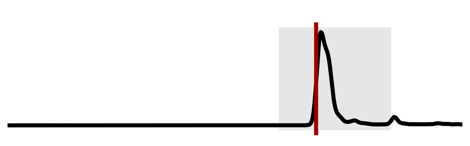
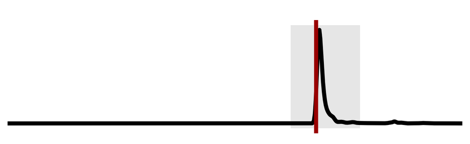
Name	Query	Tweet	Distribution
Carr Fire	#carrfire OR ((#carr OR carr) AND (#fire OR fire OR #wildfire OR wildfire))	321K	
Mendocino Fire	#ranchfire OR #riverfire OR #mendocinocomplexfire OR ((#mendocinolakecomplex OR #mendocinocomplex) AND (#fire OR fire))	47K	
Camp Fire	#cafire OR #calfire OR ((#campfire OR #campfires OR #fire OR fire OR #wildfire OR wildfire) AND california)	1,014K	
Woolsey Fire	#woolseyfire OR #woolseyfires OR ((#woolsey OR woolsey) AND (#fire OR fire OR #wildfire OR wildfire))	580K	

Table 1: Wildfire tweets and data statistics.

Hypothesis Testing

- Cascading network of sub-events hypothesized in previous work;
- Reconstructing the network for our analytical purposes;
- Manually mapping node to a seed (n, v) pair;
- Finding similar (n, v) pairs as sub-events;
- Visualizing sub-events as wordclouds in Figure 2-3;
- Cascading, i.e., edges, are tested using time lags.

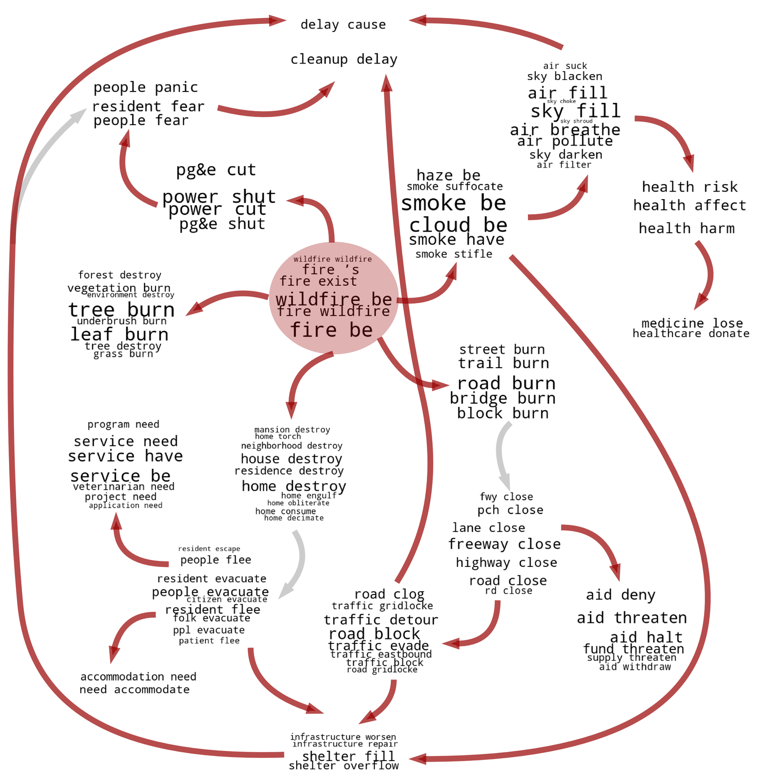


Figure 2: Evidence from Camp Fire. All 18 (100%) sub-events are identified and 20 of 23 (87%) cascades are supported, including complete cascading chains, e.g., fire induces smoke, which causes air pollution, which later harms health and eventually affects the healthcare system. There are 3 cascades which are not supported, e.g., evacuation is not after home destruction, no significant lag between road burn and close.

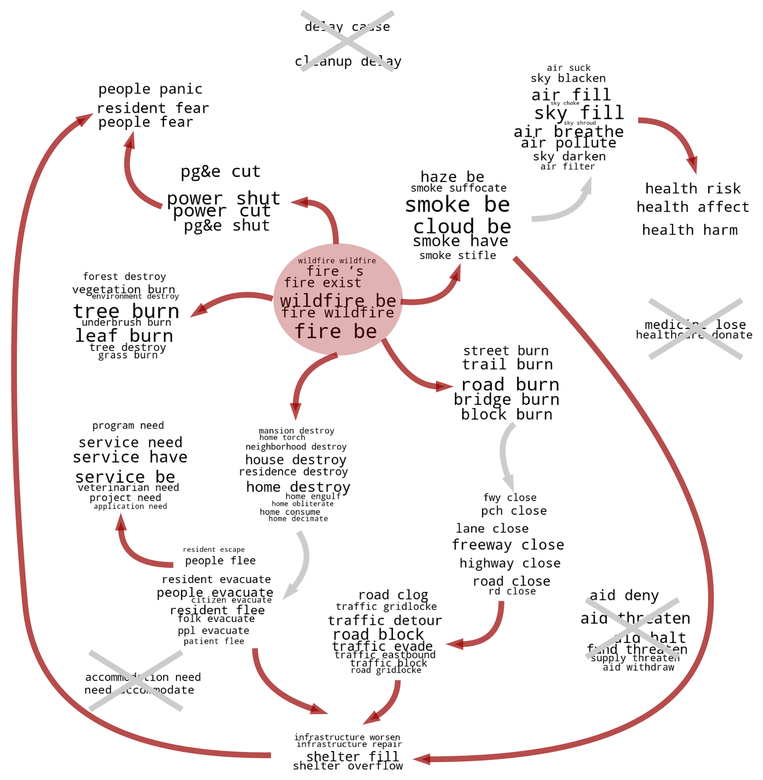


Figure 3: Evidence from Carr Fire. 14 of 18 (78%) sub-events are identified and 13 of 16 (81%) cascades, minus the ones from or to unidentified sub-events, are supported. We observe a high degree of alignment between evidences from Carr Fire and Camp Fire for both supported and unsupported cascades, e.g., fire induces power issue and then panic, and evacuation does not happen after home destruction.

Unexpected Sub-Events

- Defining "unexpectedness" by cosine similarities of (n, v) pairs;
- Filtering out sub-events that are related to the known;
- Clustering remaining (n, v) pairs;
- Examples of "unexpected" sub-events shown in Figure 4-11.

Figure 4: Contain.

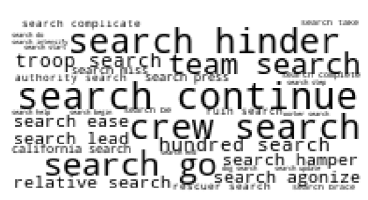


Figure 5: Search.

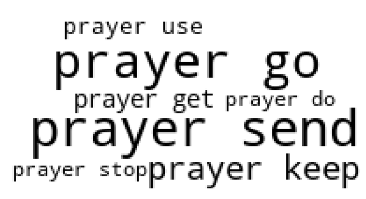


Figure 6: Prayer.

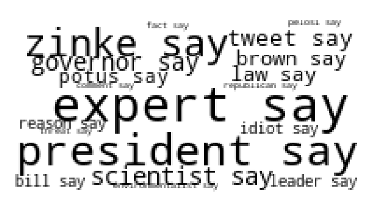


Figure 7: Official.

Figure 8: Law.

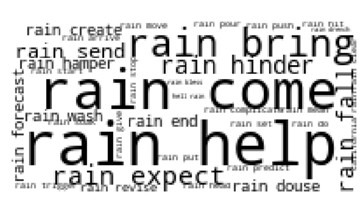


Figure 9: Rain.

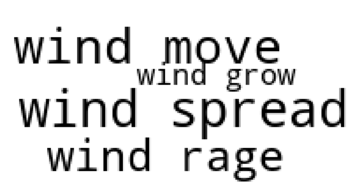


Figure 10: Wind.

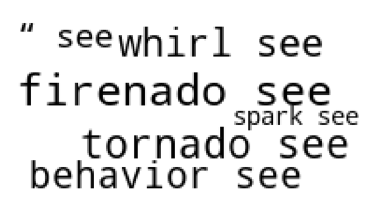


Figure 11: Whirl.

Ongoing Work

- Qualitative and quantitive evaluation using domain experts;
- Extending methods to understand sub-events of other crises;
- Minimizing human input;
- Incorporating systematic parameter optimization;
- Building an end-to-end model to replace pipeline methods.