



**UNIVERSITÉ
DE GENÈVE**



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



POLITECNICO
MILANO 1863



**Université
de Paris**



Crowd4SDG

Study on timeliness of disaster-related data

Barbara Pernici

Politecnico di Milano

April 6, 2022



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944



Crowd4SDG

Motivation

- Driver: SDG 13, Climate action
- Towards improving indicator 13.1.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population
- Support emergency response with evidence-based content from social media
- Exploration focuses on:
 - the **onset** of flood events
 - timeliness** of the delivery
 - spatial** representation of the contents



Crowd4SDG

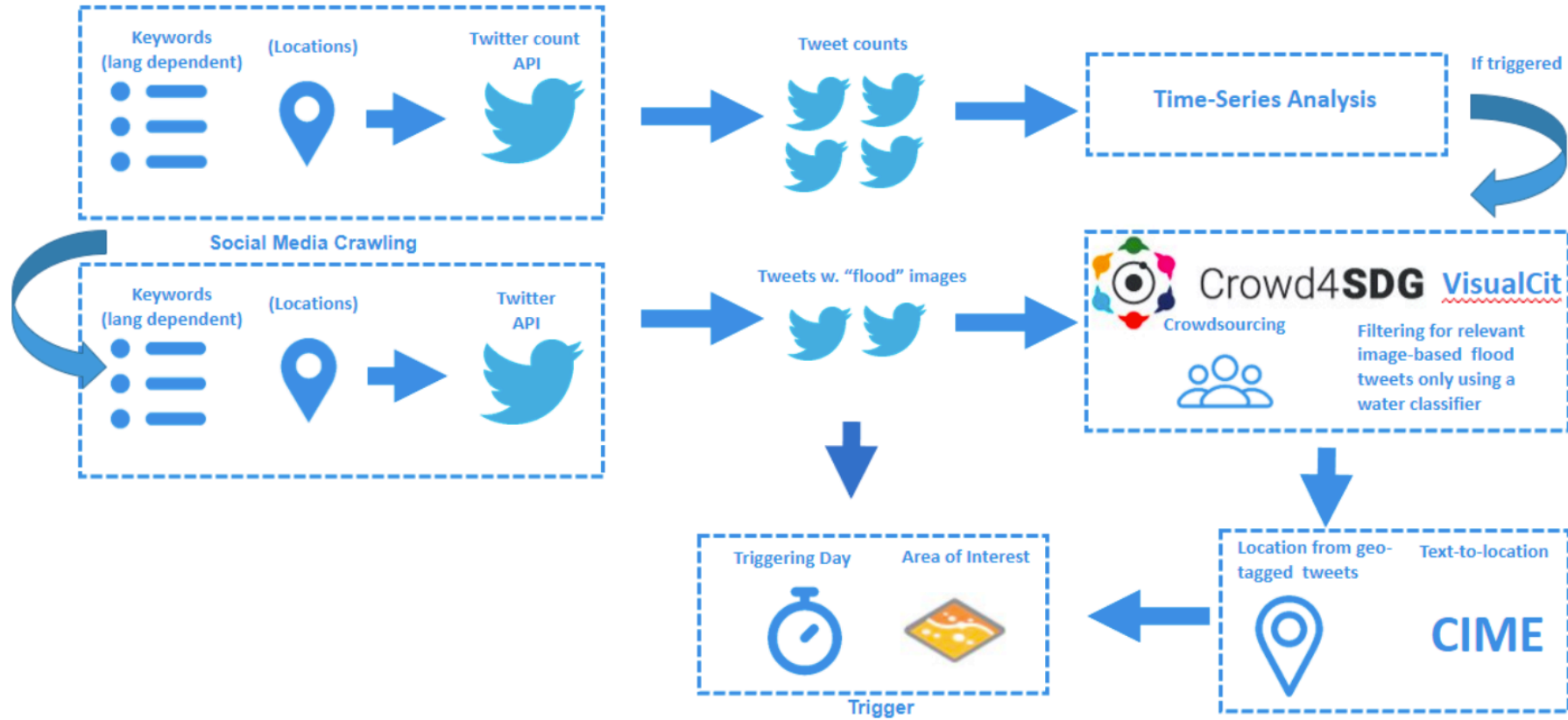
Data collection process on Twitter

1. Twitter sensing with **language-specific** small word dictionaries
2. Selection of **potentially relevant** image contents with VisualCit
3. **Text analysis** of outputs for (a) tackling text-only tweets and (b) dictionary refinement
4. **Geolocation** of posts with CIME
5. Draw **intensity maps** aggregated by administrative region



Crowd4SDG

Main phases of TriggerCit

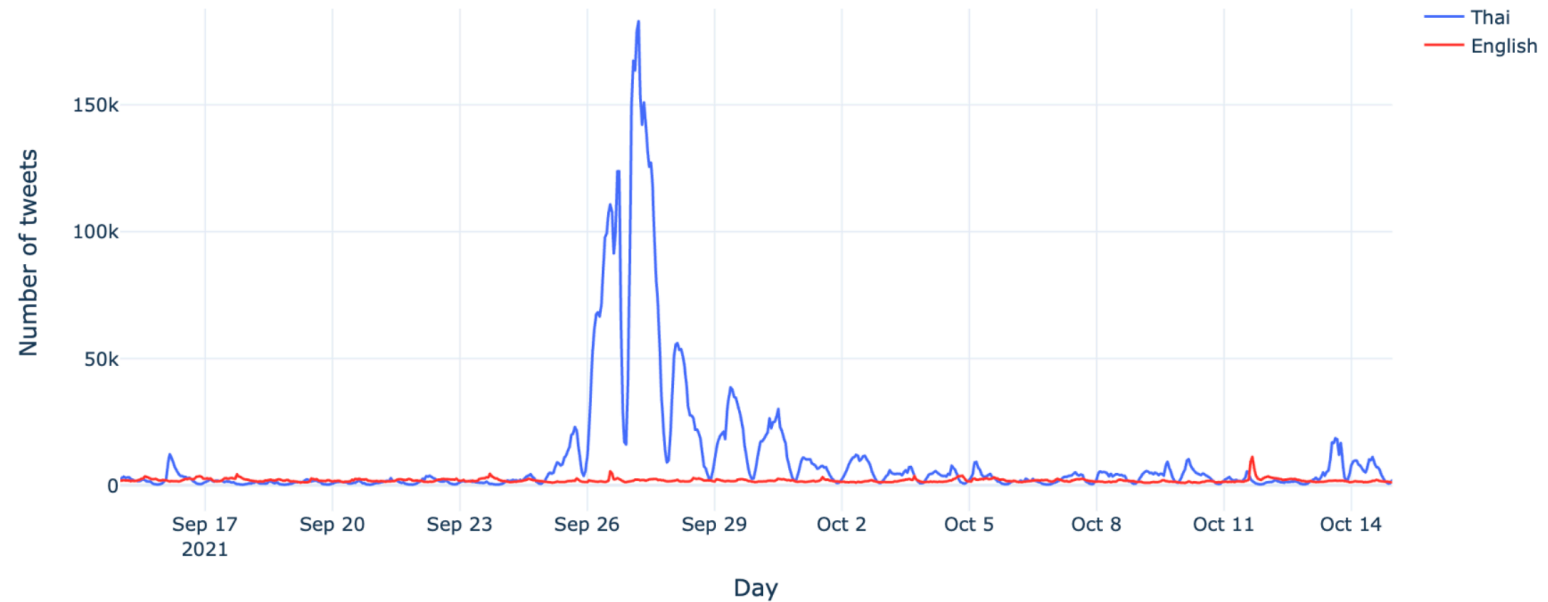




Crowd4SDG

Multilingual context and event detection

We focused on very small, no-cost dictionaries.
Language-based effect of events on social media volumes





Crowd4SDG

Example of the filtering effect

The combined effect of pipeline components removes noise from the undifferentiated input stream, letting through social media posts containing useful and descriptive information



(a) Removed items

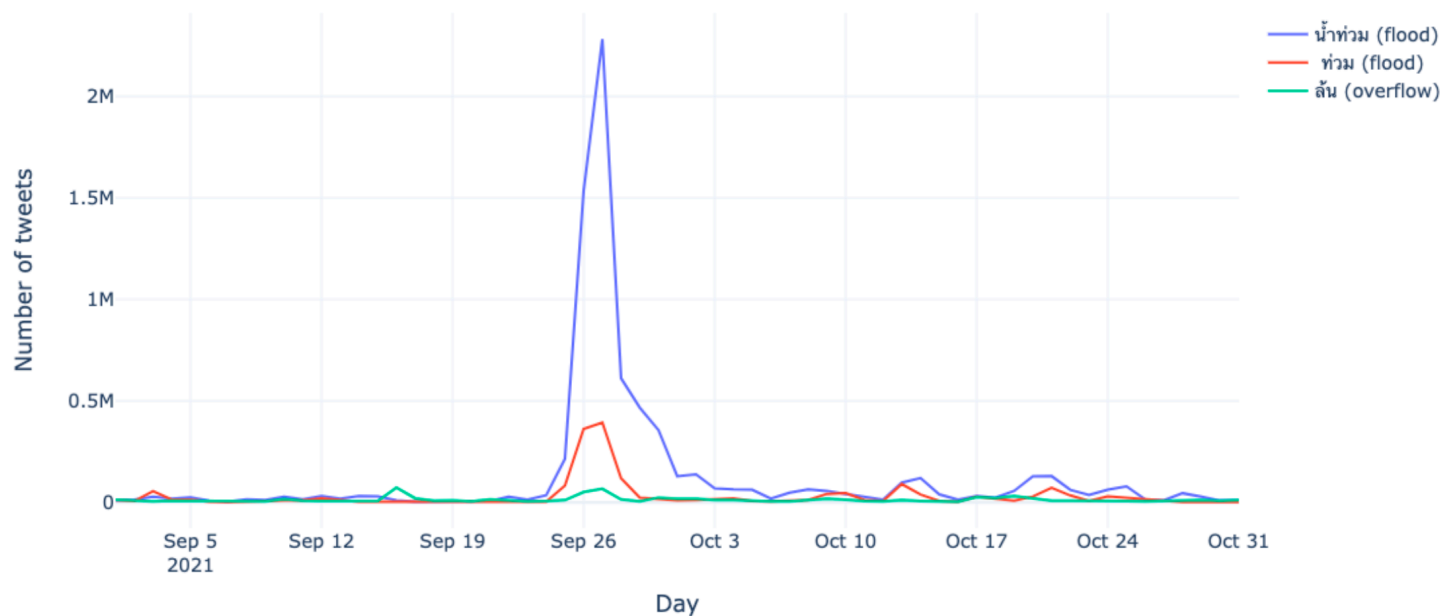


(b) Kept items



Experimental results – Thailand case study

The event onset is clearly identified



Tweet counts for seed dictionary entries

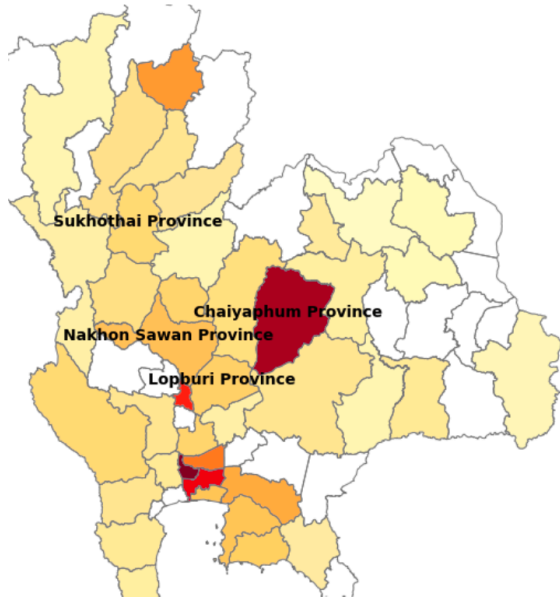
Data	Count
All tweets, September 26-27	4'145'447
No retweets	66'868
Containing images	6'292
Native Twitter locations	227
Overall images	8'774
Passed VisualCit filters	3'056
Places geolocated by CIME	1'671

Data volumes through processing steps

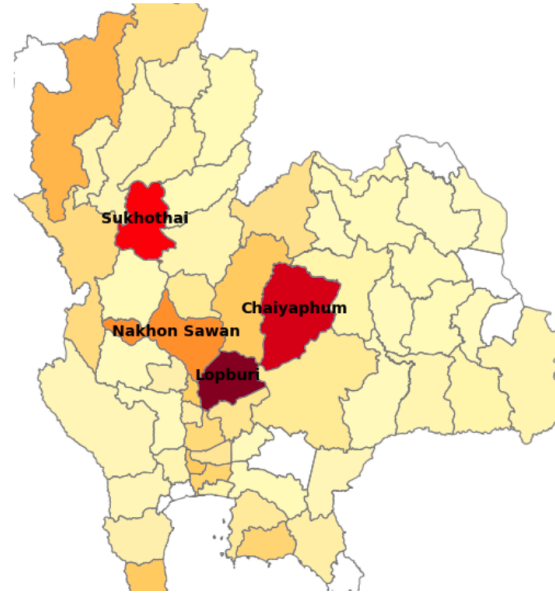


Crowd4SDG

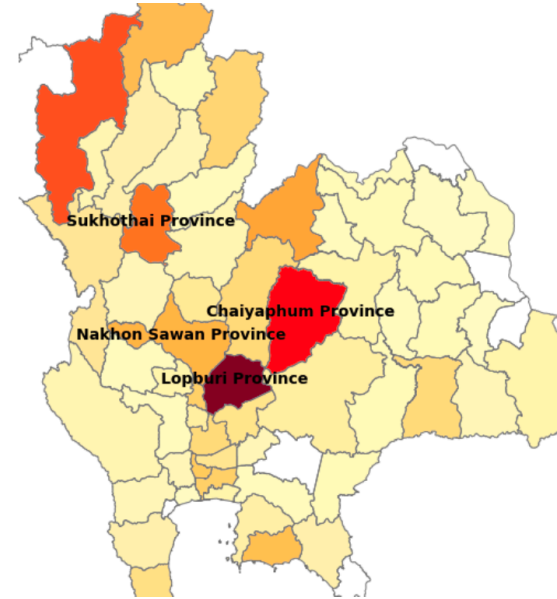
Experimental results – Thailand case study



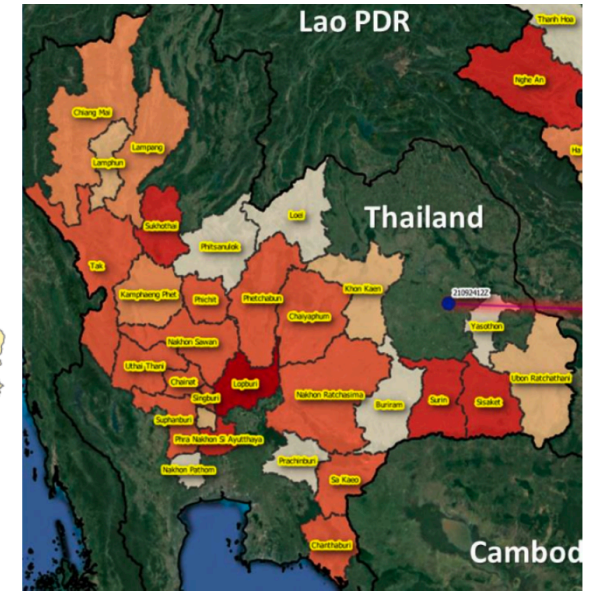
(a) Native Twitter geolocation



(b) VisualCit geolocation



(c) VisualCit geolocation with extended tweet set



(d) Affected people

Geolocations / inhabitants ratio by region (a) Twitter native geolocations, (b) Twitter native + CIME geolocated, (c) Twitter native + CIME geolocated from extended dataset (with images + promising text-only tweets), and (d) Number of affected persons by region at September, 28th (source: ReliefWeb)



Crowd4SDG

Experimental results – Timeliness

Time of the events as reported from different sources

Dates / Countries		Thailand	Nepal
Activation-based	UNOSAT activation	28/09/2021	30/06/2021
	Copernicus EMS activation	None	None
Manual	GDACS Disaster Alerts	27/09/2021 (Green Alert)	28/06/2021 (Green Alert)
Forecast	GloFAS	24/09/2021	28/06/2021
Manual	FloodList reported news	27/09/2021	04/07/2021
	TriggerCit	26/09/2021	02/07/2021



Crowd4SDG

Performance

Average execution time of the pipeline on a single server machine:

1. **171ms** for processing a tweet
2. **224ms** for geolocating a tweet

Level	Admin level	Count
Province	4	7
District	6	8
Municipality or subdistrict	8	176
Village or community	10	9
Other points	15	1'265

Administrative level of the geolocations, Thailand case study



Crowd4SDG

What we are currently working on

- Use **validated** event data as a ground truth, to benchmark and enhance the architecture components, evaluating uncertainty in space and time
- Explore the robustness of a supervised approach for automated event **triggering**
- Additional for data layers (additional systems, sensors, models, media) for an integrated **multi-modal** approach



References

- Carlo Bono, Barbara Pernici, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, Mehmet Oğuz Mülâyim, Edoardo Nemni, TriggerCit: Early Flood Alerting using Twitter and Geolocation - a comparison with alternative sources, 2022, accepted for presentation at ISCRAM 2022, preprint: <https://doi.org/10.48550/arXiv.2202.12014>
- V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J. L. Fernandez-Marquez, A. Ravi Shankar, M. Carman and B. Pernici, Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021
- G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts, Geoinformatica, Issue 1/2022



Thank you!

www.crowd4sdg.eu

