Better Situational Information on Haze Crises in Southeast Asia and their Impacts on Human Mobility in Singapore

Imaduddin Amin
Pulse Lab Jakarta
imaduddin.amin@un.or.id

Muhammad Rheza
Pulse Lab Jakarta
muhammad.rheza@un.or.id

Jonggun Lee
Pulse Lab Jakarta
jonggun.lee@un.or.id

ABSTRACT

Peatland fires and haze events in Southeast Asia are disasters with national and international implications, including Indonesia, Malaysia, and Singapore. The phenomenon directly affects to economic, social, and environmental losses, which indirectly affect other quality of life aspects. Despite the extensive efforts of many organizations, the situation persists. Currently, Indonesian disaster management authorities manage forest and peatland fire and haze events based on satellite data as well as static data on population density and distribution. In order to better protect vulnerable populations and the environment, the authorities are looking for more timely data and more information on the dynamics of the disaster, especially the situation on the ground. United Nations Global Pulse's Haze Gazer[1] enhances disaster risk management efforts by providing real-time situational information from diverse data sources, including peatland hotspot data, meteorological data, official statistics, social media signals, and mobile data. We show the potential of new type of data sources to complement the information from existing data sets. We demonstrate how text analysis from social media allow us to capture citizen' feedback and report on the latest situation on the ground. Furthermore, the location information from social media can be used to understand citizen mobility. Specifically for this challenges, we would like to extend the finding by including mobile data that relatively have better spatial and temporal resolution compared to others.

ACM Reference format:
Imaduddin Amin, Muhammad Rheza, and Jonggun Lee. 2017. Better Situational Information on Haze Crises in Southeast Asia and their Impacts on Human Mobility in Singapore. In *Proceedings of CIKM, Singapore, Nov 2017 (CIKM'17),* 4 pages.
DOI: 10.1145

1 PROBLEM STATEMENT

Haze crises in Southeast Asia, which include the almost annual wildfires and haze events in Indonesia, have an international dimension. These crises have clear and immediately quantifiable short-term effects, such as the closure of markets and cancellation of transportation, as well as hidden but significant effects on mid- and long-term scales, such as the deterioration of the quality of life, negative impacts on education outcomes and cardiovascular health issues.

Current practices concerning the development of situational information on haze crises mostly involve the collection of data on hotspot locations (from satellites and ground-level observation), air quality information (for example periodically updated air quality data sets on different sizes of particles) and meteorological information (such as wind direction). Considering the complexity of the dynamics of haze situations, the resolution and timeliness of these data sets, however, are neither sufficient for effectively addressing the problem nor helpful to citizens in minimizing disruption to their daily lives.

The extent of the spread of haze across a region over time is affected by many factors, including the size of the wildfire, the characteristics of wind and the flows of air, among others. When such factors work together in an unexpected manner, it is not feasible to spatio-temporally simulate the event computationally within an acceptable time scale. Moreover, human behavior, especially changes to human behavior in response to a situation, cannot be fully understood without rich contextual information.

The work outlined in this paper aims to improve situational information on haze crises and share it directly with affected citizens from across the region, especially Singapore. Specifically, we propose a data analytics and visualization system which provides unique insights on the impact of haze on the ground in real-time by combining (i) a set of information on the origins of the haze events, namely peatland fire hotspot data in Indonesia, (ii) social signals from Singapore, including, text-, image- and video-based social media posts, and (iii) the mobility information including the dynamics of the behavioral changes of human mobility in Singapore from mobile data.

2 DATA SOURCES

Haze Gazer utilizes multiple dataset such as official data, satellite data, social media and mobility information to provide the latest condition of haze situations. The following shows a list of data sources we use in building a data analytics and visualization system.

- **Peatland hotspot information**: In this prototype, we use hotspots information available on the Global Forest Watch platform as a main source. In addition, Haze Gazer allows a user to identify their preferable data sources. For example, the Haze Gazer that is installed in President Staff office of Indonesia uses data from National Institute of Aeronautics and Space (LAPAN) [2] , the Indonesian government space agency, as a primary sources for peatland hotspot information.
- **Information of meteorological updates** : We included two most important meteorological information related to haze situation. First, the air quality information that taken from haze.gov.sg for Singapore region and The World

[1] http://unglobalpulse.org/blog/haze-gazer-crisis-analysis-tool
http://www.youtube.com/watch?v=tcuH6Mu_CmM

[2] https://www.lapan.go.id/

CIKM'17, Singapore
2017. 978-x-xxxx-xxxx-x/YY/MM.
DOI: 10.1145

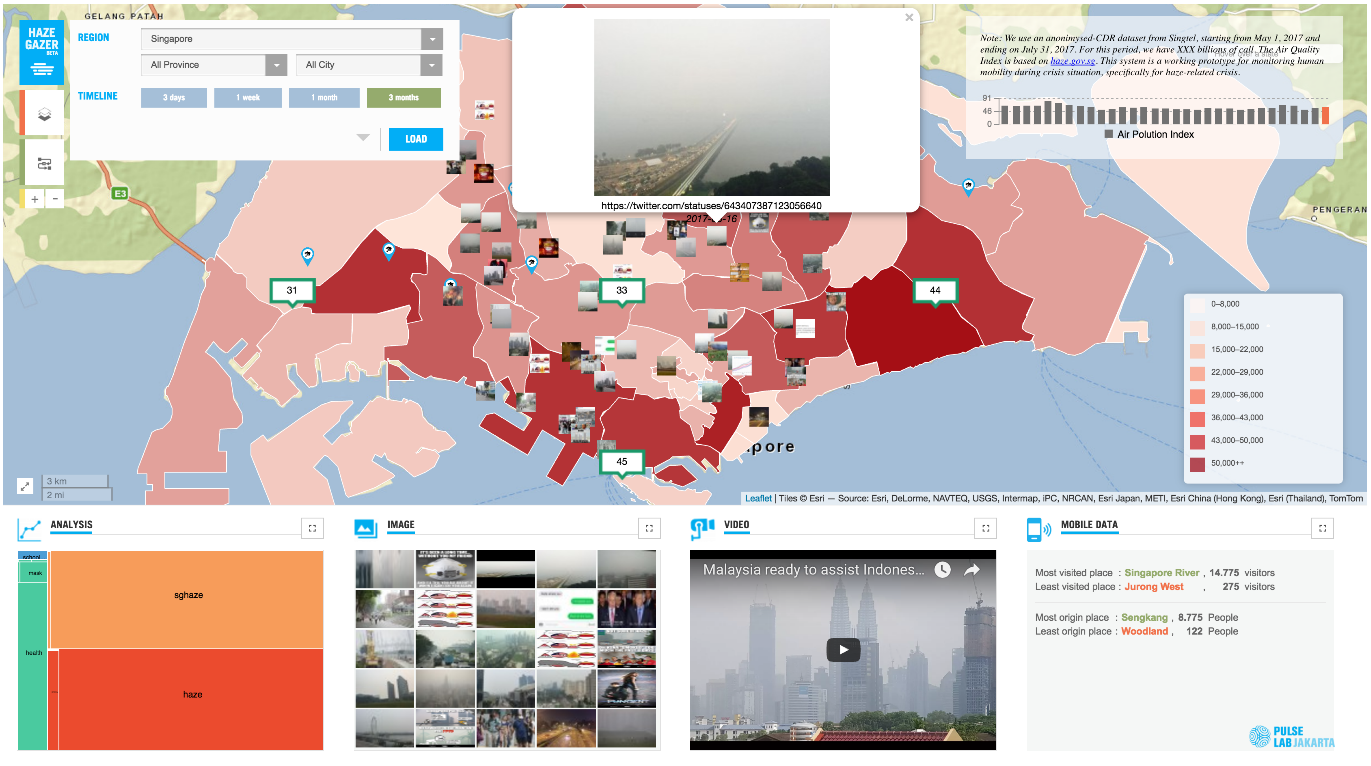

Figure 1: Haze Gazer Home Interface

Air Quality Index Projects [3] for other surrounding countries including Indonesia and Malaysia. Second, a wind direction data that are available from National Centers for Environmental Prediction (NCEP) central operations [4].

- **Public information** : We retrieve official statistics data from data.gov.sg to identify the most vulnerable cohorts of the population, up to planning area level. We also include public spaces location such as school and hospital locations with POI's are taken from OpenStreetMap footnotehttps://www.openstreetmap.org.
- **Social signals** : The sources for understanding real-time social signals is included from different sources and types to provide different kinds of insights on haze situation. Text information collected from Twitter is analysed to get insight about the latest update of haze condition, opinions and citizen feedback to government. Multimedia data from Instagram, Twitpict, and Youtube shows the latest condition of haze situation that later can be used by the authority to infer the severity of haze conditions.
- **Mobility information** : We include two types of of mobility information provided by DataSpark. First, a visit information that available from discrete API and second, a user mobility that taken from Origin-Destination API.

[3]https://www.aqicn.org/
[4]http://www.nco.ncep.noaa.gov

3 METHODOLOGY

We use the new data sources, including social signals and mobility information, as a supplementary information for the ground truth information from satellite and meteorological data. Firstly, we collect all social media messages related to haze situation using a set of keywords or taxonomy. We identify the haze-related keywords in Bahasa, Malay and English to explore the cause, event, impact and emergency response effort related to a specific fire or haze event in Singapore. We then inferred social signal from messages that captured using those keywords and found strong correlation between social signals with peatland hotspot information. The social signals getting stronger when the number of hotspot is increasing [1].

In addition, we use location information from social media to understand the changes of citizen mobility behaviour. For this purpose, we defined mobility characteristics for a citizen such as centroid, distance and spread. A centroid is defined as a average latitude and longitude of user activity, distance is measured as a euclidian distance between two centroids, and spread is calculated by comparing each signal with the centroids.

Using that mobility characteristics, we show that a severe haze condition leads a change of citizen' regular behavior [2]. For example: a set of citizen move far away during the severe haze period, while another group will stay at home but reducing their short-distance mobility activities. By comparing to air quality data, it is shown that citizen will avoid affected area with bad air quality and move to relatively safer area that has better air quality index.

For this challenge, we aims to extend our result by using the mobile data as it has better resolution compare to social media data.

To this end, we then measure the characteristics of citizen mobility behaviour changes.

3.1 Typical Citizen Behaviour

In order to understand about the data, first we plot the activity at difference level of spatial and temporal aggregation. For this study; region, planning area and sub-zones are three levels of spatial aggregation while daily, hourly and 15 minutes interval are three levels of temporal aggregation that available from data set.

From the data, some of sub-zones and 15 minutes interval aggregation level return a small number of data that were not passed the minimum requirements of data adequacy and data minimization for privacy concern. Thus, for this study, two spatial aggregation (region and planning area) and two temporal aggregation (hourly and daily) were selected.

We found out that generally the citizen behavior pattern is repeated in weekly cycle. Thus, part of our analysis and visualization is comparing the data in weekly interval. For instance, next week Monday data will be compared with current Monday week data and last week data and so on.

3.2 Citizen Mobility Behaviour Changes

We look at citizen mobility behaviour changes by using two types of information from the dataset namely footfall and OD-pair. Footfall includes the information of number of visitor in particular time and place while OD-pair includes the information of movement of people from a point (origin) to another (destination).

Our previous study shows that the citizen' behaviour changes related to air-quality condition. People tend to move to cleaner area, or they will stay at home and prefer to reduce their mobility. Hence, hypothetically we expect that some places will have more visitor while others have less, influenced by the air quality dynamics. In addition, we also expect that the OD-pair will also have some changes compare its typical behaviour.

Given the data period on May 2017 until July 2017, all of air-quality measurement shows the moderate to good condition so we could not find significant changes in citizen mobility. This is understandable as according to authority, people can do their normal activity when air quality is moderate or good condition.

4 APPLICATION DESIGN

We designed our dashboard as a multi-tiered architecture consisting of automated data acquisition, processing and delivery from multiple data sources. In addition, the system is designed to allow easy incorporation of additional data themes and application to other regions.

At the presentation level, the dashboard enable a user to interactively manage an active layers to visualize information of her interest. For instance, a user can activate a hotspot location layer with population layer to have sense of affected population or activate hotspot location and wind direction to understand where affected area of the haze.

Figure 1 shows the main page of dashboard where layers information are available on the main area and the additional information are available on four boxes at the bottom of the page. Currently, four additional information highlight the analysis derived from text, images, video and mobile data.

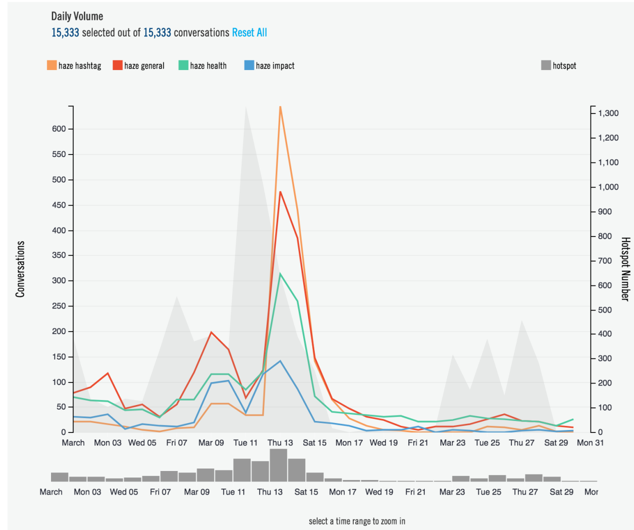

Figure 2: Additional Information of Social Media conversation compared to Haze condition

A user can explore more detail insights on additional information by open of each box. For example, a figure 2 shows the dynamic of conversation in social media and its comparison with peatland hotspot information. In this report, we would like to highlight more a visualization and analysis that using mobile data as describe in following sub section.

4.1 How Mobile Data is used at this Platform

We display two main information from mobile data such as (a) visiting places, and (b) origin destination or citizen mobility. For this purpose, we provide four main insights as following.

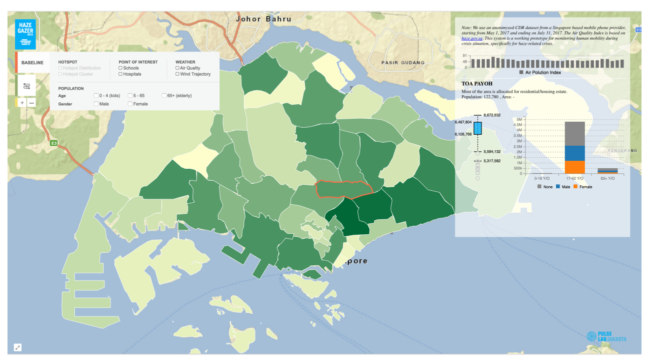

Figure 3: Maps presenting number of visitor for each planning area

4.1.1 Visiting a Place. This information is available under the main menu of hazegazer, presenting the distribution of visitor in each planning area. Detail information is available for each planning area, covering the information presented in Figure 3 such as (a)

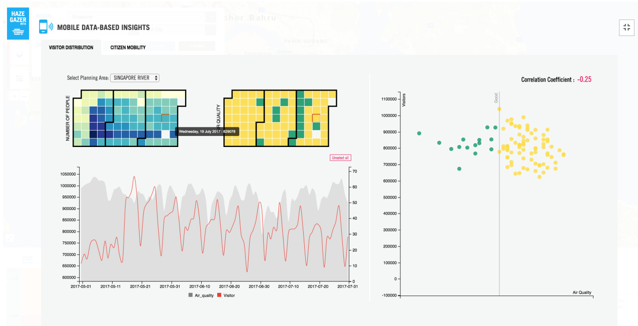

Figure 5: Calendar view to present typical day of visiting activity

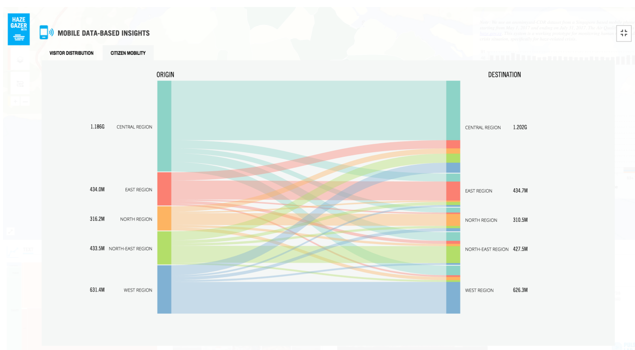

Figure 6: Origin Destination Pair

air quality historical data, (b) a short description of planning area, (c) number of total visits compare to selected period, presented in a box plot, and (d) number of visits segregated by their gender and age range. For example in Figure 3, it shows the data on May 4, 2017 at Toa Pavoh area. The air pollution index on the day is shown as orange bar in the air quality chart. The number of visitors at Toa Pavoh on the day is below the 25% of percentile during selection period. On the other words, it shows that on the day, less people visits Toa Pavoh area compared to the number of visits during the selected period.

4.1.2 Moving from a place. This information is presented from origin destination data where color on the map show the number of citizen who left from that area. Similar to visiting a place, a user can easily access historical data of air quality and select specific date to see the data of user's interest as shown in Figure 4. The detail information is showing origin destination pair, from the chosen area. For example, in the figure 4, it shows that people who left Bedok area are mostly going to Tampines area.

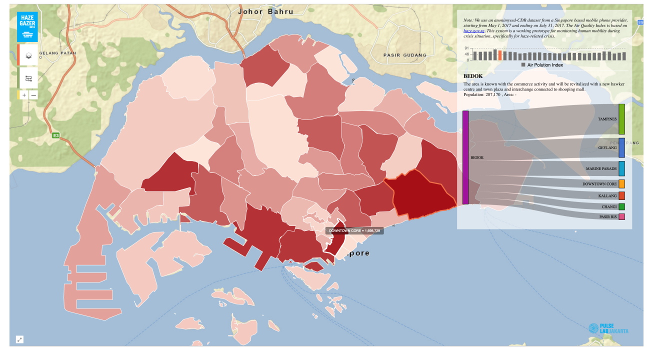

Figure 4: Maps presenting number of citizen move for each planning area

4.1.3 Typical day of visiting activity. This information shows the comparison of visiting activity of five regions in Singapore in timeline manners. As discussed previously, a pattern can be seen in weekly cycle, thus we try to compare the same day for each week. This calendar view allows a user to see which date is non-typical that different from another week as shown in Figure 5. The darker the color, the more people are visiting the area. We also see the distribution of visitor relatively to air-quality measurement. In this case, air quality is always under moderate to good condition which means there will be no significant changes in user regularity. A correlation is calculated and presented to give some sense of the area as we expect two types of correlation. A negative correlation means that less people will visit on the bad air quality and vice verse.

4.1.4 Origin Destination activity. This information shows the origin-destination of visitors per region or planning area level as shown in Figure 6. For instance, it shows the proportion of visitor's flow going from and to particular region.

5 IMPACT

In summary, Haze Gazer enhances disaster management efforts by providing real-time insights on the:

- Locations of fire and haze hotspots;
- Strength of haze in population centres;
- Locations of the most vulnerable cohorts of the population; and most importantly,
- Response strategies of affected populations, including movement patterns and in-situ behavioural changes.

This project clearly aims to provide better insights to affected citizens against haze crises in Southeast Asia in a short-term but it hopes in a mid-term to raise the awareness of the importance of this issue.

In addition, this project also obviously provides public authorities with enhanced data visualization in which they could acquire a more integrated information regarding the haze situation and the citizen' behavior on the ground. At the end, it hopes to empower the citizens and their communities, to build their resilience by having (near) real-time information.

REFERENCES

[1] Mark Kibanov, Gerd Stumme, Imaduddin Amin, and Jong Gun Lee. 2015. Supporting Peat Fire Management using Social Media. *2nd GESIS Computational Social Science Winter Symposium 2015 (CSS WS 2015)* (2015).

[2] Mark Kibanov, Gerd Stumme, Imaduddin Amin, and Jong Gun Lee. 2017. Mining Social Media to Inform Peatland Fire and Haze Disaster Management. *Social Network Analysis and Mining* Issue 1 (2017).