

Twitter and perceptions of crisis related stress

Methodological white paper¹

December 8, 2011

1 Abstract

The purpose of this research project is to determine which indicators might be present in social media data that could shed light on how populations cope with global crises, such as commodity price volatility or the continuing global economic crisis. As an initial investigation, this project was limited to the analysis of publicly available data from Twitter for July 2010 through October 2011. The work was further limited to tweets in Javanese/Bahasa Indonesia and English. The topics of focus included the affordability/availability of food, fuel, housing and loans. By classifying a populations' tweets into several categories associated with relevant topics, it was possible to perform quantitative analysis to better understand populations' vulnerabilities: detecting anomalies such as spikes or drops in the number of tweets about particular topics (e.g. comments about power outages in Indonesia or student loans in U.S.), observing weekly and monthly trends in Twitter conversations (e.g. discussions around debt in U.S.), finding patterns in the volume of particular topics over time (e.g. discussions around housing in U.S.), comparing the proportions of different sub-topics to understand shifts in trends over time (e.g. the ratio of tweets about formal loans vs. informal loans in Indonesia) or relating trends in Twitter conversations with external indicators (e.g. conversations around the price of rice in Indonesia mimicking the official inflation statistics). This research has pointed to the strong potential use of Twitter data for understanding the immediate worries, fears and concerns of populations, but at the same time, the research suggested that it is a poor source of data for gauging people's long term aspirations. There are several remaining challenges, in particular that Twitter has a specific culture and demographic which needs to be better understood to strengthen any analysis of this type. Overall, this exploratory research shows some of the potential of Twitter data for exploring people's perceptions of crisis-related stress and suggests research lines and methodologies for further investigations.

2 Context and Project Objectives

Introduction and Description of Project Objectives

¹ This methods white paper arose from an on-going series of collaborative research projects conducted by the United Nations Global Pulse in 2011. [Global Pulse](http://www.unglobalpulse.org) is an innovation initiative of the Executive Office of the UN Secretary-General, which seeks to harness the opportunities in digital data to strengthen evidence-based decision-making. This research was designed to better understand where digital data can add value to existing policy analysis, and to contribute to future applications of digital data to global development. This project was conducted in collaboration with [Crimson Hexagon](http://www.crimsonhexagon.com). For more information on this project or the other projects in this series, please visit: <http://www.unglobalpulse.org/research>.

United Nations Global Pulse (UNGP) is dedicated to better understanding how new types of data can strengthen available information on how people are impacted by global crises. This project, conducted in partnership with Crimson Hexagon, sought to lay a foundation to use what UNGP believes could represent a powerful source of new data: the global conversation taking place over social media. In particular, this research focuses on understanding what global crises “look like” on Twitter, one of the fastest growing social media platforms in the world.

Twitter is an online service dedicated to social networking and “microblogging,” allowing its users to send and read “tweets,” or posts containing up to 140 characters. It was created in March 2006 and today has more than 100 million users. From July 2010 to November 2011, the volume of tweets making up Twitter’s public firehose² has grown from about 60 million per day to nearly 200 million per day. Along with its growth, Twitter has also expanded its global reach, with countries other than the U.S. representing approximately 71% of all Twitter use.³ According to Crimson Hexagon data, Indonesia ranks fourth in countries worldwide in terms of overall volume, garnering more than 5.5 million location-tagged tweets per day.⁴ This volume of social media data represents an enormous opportunity for research. Over the past few years the U.S. and Indonesia have both faced significant challenges due to the global economic crisis, including in key segments of the economy, such as finance and housing in the U.S. or food in Indonesia.

Due to this combination—the enormous volume of user-generated content and the macroeconomic shocks of the past two years—the UNGP and Crimson Hexagon determined that the U.S. and Indonesia would be appropriate locations for the focus of this project.

The primary objectives for the project are:

- To learn which types of policy-relevant questions may be answered in part based on conversations that happen over Twitter;
- To strengthen common knowledge around the methodologies needed for policy-relevant and accurate insights from Twitter data;
- To gain meaningful insight into how populations in Indonesia and the United States are coping with key areas of volatility such as commodity prices, debt and housing;
- To lay the groundwork for further analysis of social media in areas of public concern.

In the course of achieving these objectives, Crimson Hexagon and UNGP also documented the lessons learned from the project, including:

- Which topics of public concern are more/less suitable for ongoing analysis using social media?
- How do local social norms affect what type of information is shared using social media? How does this affect the ability to perform this analysis across countries?

Technology Overview and Previous Use

Crimson Hexagon’s technology for analysis is based on an algorithm originally developed by Professor Gary King, director of Harvard University’s Institute for Quantitative Social Science. The algorithm

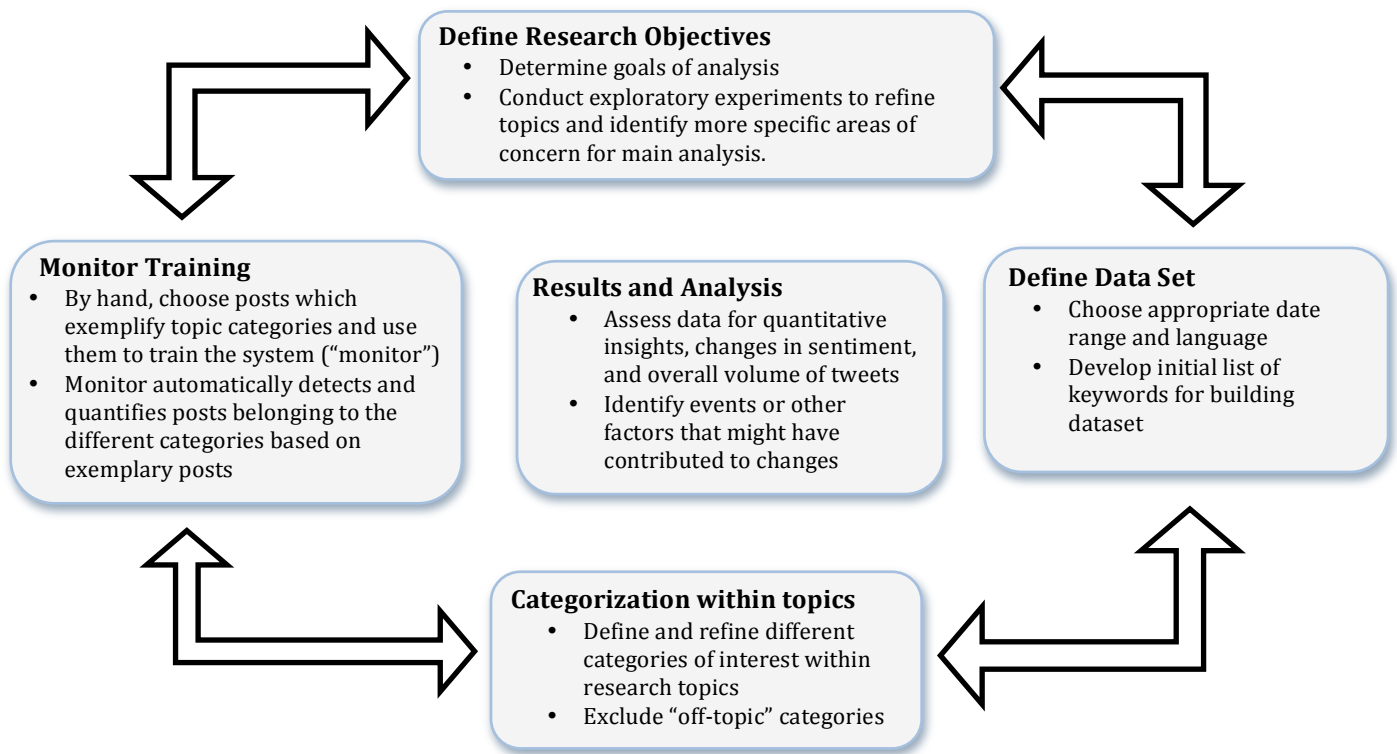
² The firehose represents all publically available tweets.

³ This figure is based on an approximation of Crimson Hexagon’s Twitter content with location tagged

⁴ The total number of tweets from Indonesia is certainly higher than this, as only a portion of tweets are tagged with location information

provides a means to measure the proportions of specific opinions or themes that are present in large, text-based data sets. In addition to Twitter data, potential sources for analysis in social media include blogs, on-line forums or chat rooms, publically available Facebook data, comments, and news. Additionally, the algorithm is capable of analyzing text-based data from non-social media sources such as open-ended survey responses, interview transcripts and other written records. Monitoring textual data is essentially quantifying massive amounts of qualitative information by identifying statistical patterns in language used to express opinions on various topics. Typically, this is used in both commercial and social science applications for numerous global brands, agencies, and media organizations. For example, commercial entities monitor consumer opinion to inform marketing and product strategies, and media outlets often seek to monitor responses to major news stories as well as longer-term issues, enabling them to better engage and connect with their audience.

Workflow Overview



Describing Monitors

The overall tools used for data collection, categorization and analysis are called “monitors.” The following steps outline the process of setting up, running, and using the monitors.

Step 1: Define Goals

First, researchers determine and document the goals and drivers of desired analysis. This step provides a framework for analyses and ensures that monitors are aligned with project objectives. Monitors are

configured to holistically assess conversations based on specific project objectives, and often initial inquiries point to ways of adapting the monitors to better suit the key project objectives. As such, the entire process is iterative and requires both flexibility and focus.

For example, in this project UNGP was initially interested in looking broadly at how people perceive the future—what they perceived their risks to be, and what they were worried about or excited about. However, as we moved through steps two and three, we discovered we could get a much more refined analysis from the available data if we based the study on particular sectors. Given that in the first step, the project team had clearly identified the importance of specific results over more general findings, the team refined the scope to examine how people discuss key issues of concern—food, fuel, housing, and financing/debt.

Monitors were built within each of these four topics to reflect themes of cost pressures expressed in online conversation. For each topic, one monitor was built in English and another in Bahasa Indonesia/Javanese.

Step 2: Defining the Data Set

Crimson Hexagon collects and stores publicly available social media data from a variety of sources. Documents (called “posts”) are collected daily and are maintained in a database, which allows users to investigate content retroactively. For this project, the source of data used was Twitter content only. As noted above, Crimson Hexagon has received and stored publicly available Twitter content from the full Twitter firehose since July 2010. The dataset in this study therefore includes all publicly available tweets from July 2010 until October 2011. The massive growth in Twitter over this period is reflected in an increasing volume of Tweets in each of the monitors.

This represents the entire universe of tweets. The following steps are designed to further refine the dataset to include a smaller subset of tweets in the actual monitor.

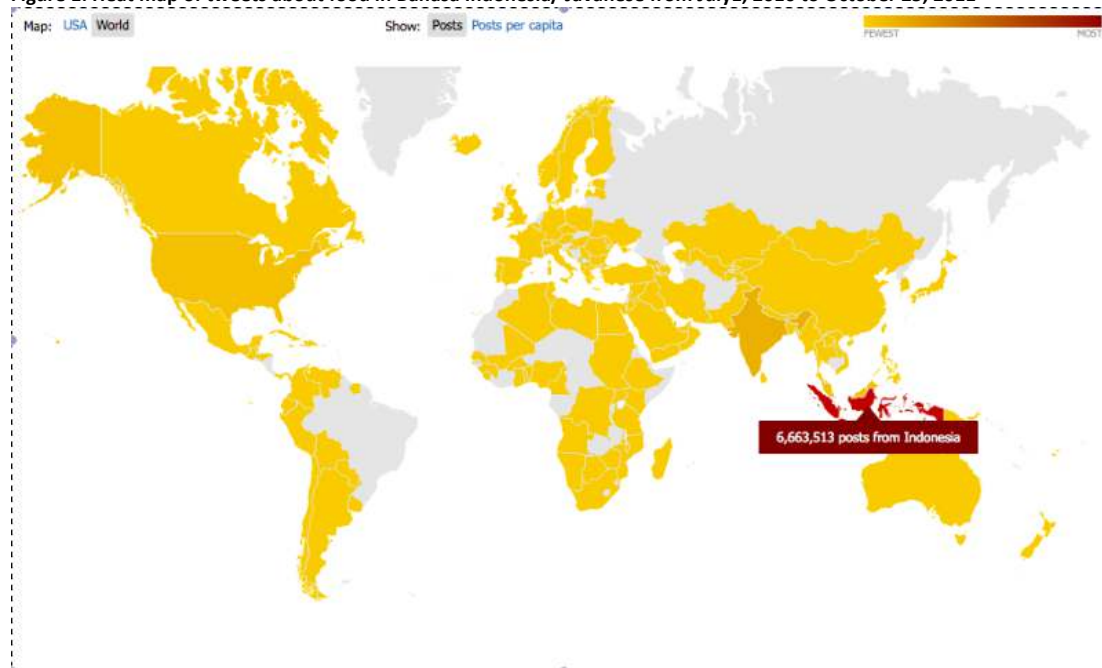
Filtering the Data Set

In order to focus the monitors on tweets from the desired locations and languages, specific filters were used to capture content from Indonesia (in Bahasa Indonesia and in Javanese) and from the U.S. (in English). While the information provided in the Twitter firehose enables language and location tagging, it is incomplete (see above). For the purposes of this analysis, we relied on keyword filtering using common terms to identify tweets in Bahasa Indonesia and Javanese, rather than simply relying on defined language and location tags from Twitter. Since Bahasa Indonesia and Javanese are languages essentially exclusive to Indonesia, we can assume that for the most part those tweeting in Bahasa Indonesia/Javanese are Indonesian. Thus, all tweets in these languages were included in the dataset. (Note: the heat map on the next page shows the global distribution of Indonesian Twitter activity. While conversation is not exclusively located in Indonesia, it is heavily concentrated there.)

For the English language portion of this analysis, we filtered the conversation to include only those tweets that were positively geo-located in the U.S. in order to ensure that our analysis focused on American opinions. While this greatly reduced the size of the data set, it offered certainty on the geographic source of the data.

Next, to analyze content that might be relevant to the subject of analysis, a broad keyword filter was used. The aim of this filter was simply to identify which tweets from the firehose might be on topic. For example, when analyzing conversations about fuel prices, the keywords “gas” or “gasoline” might be used so that only tweets containing one of those words are analyzed. Obviously, many mentions of the words gas or fuel might not be relevant to the subject of the monitor. In this case, the monitor is trained to recognize these as “off topic” (see below).

Figure 1: Heat map of tweets about food in Bahasa Indonesia/ Javanese from July1, 2010 to October 25, 2011



The following is an example of the keyword strings used in the Food monitors:

Table 1: Keyword strings in Food monitors

United States	Bahasa Indonesia/Javanese
("buy groceries" OR "buy food" OR "afford food" OR "afford groceries") AND location:USA	(kalau OR kl OR saya OR sy OR tetapi OR tapi OR tp OR kamu OR untuk OR utk OR adalah OR dalam OR dlm OR oleh OR banyak OR dengan OR dgn OR atau OR juga OR jg OR antara OR dapat OR dpt OR bagi OR hanya OR atas OR punya OR lain OR kowe OR kw OR wis OR wes OR arep OR sampun OR pun OR iki OR kuwi OR kene OR kono OR ora OR iso OR nek OR neng OR ku OR aku) AND (makan OR mkn OR maam OR maem OR ma'am OR mangan OR mknn OR makanan OR panganan OR ngemil OR cemilan OR nasi OR beras OR sega OR sego OR sekul OR sembako) AND - (.my OR .ph OR kad) AND Language:id

This process was repeated for each topic defined in the research objective. This project relied on eight distinct datasets, one for each location in each of our four topics.

Step 3: Categorization

Before analysis could begin, it was necessary to analyze the data to establish the category structure within each dataset. The Crimson Hexagon algorithm analyzes the proportions of data that fall into a series of user-defined categories. As such, it is imperative that the analyst “train the algorithm” (see below) and define a category structure that meets his or her objectives and fits the data available.

Table 2 shows an example of how the Gas/Fuel data set was broken into categories for both the U.S. and Indonesia. Note that categorization should also be done concurrently with the monitor training, described in Step 4, so that categories can be refined based on priorities as reflected in the actual Twitter conversations.

For example, in the debt/finance category for the United States, we originally sought to define different types of debt, including the difficulty in paying back house loans or the accessibility of new loans. What we discovered is that for many “types” of debt, people are not very likely to discuss their ability to pay in the public sphere. We thus had to refine our categorization appropriately. We found that people do discuss some types of debt very openly—in particular student loans. Therefore, we learned that this methodology was not necessarily going to shed light on impending foreclosure, but may be well-suited to monitoring trends in student loan debt.

Thus, the category structure for each monitor was refined several times to identify the most interesting and relevant themes of conversation. The example below shows that the exact categorization structure will vary based on the context, and the context will clearly be very different in different places. Thus, within the same broad topic of “Gas and Fuel” the categorization varied dramatically in the U.S. and Indonesia. The final category structure for Gas and Fuel is described in Table 1 (see Annex I for the full category breakdown of all topics).

Step 4: Monitor Training and Analysis

The process of “training” monitors involves both manual and automated processes. The first step is for a researcher to manually classify randomly selected posts according to the category structure defined in step three. The goal here is to select and code posts that are the best examples of each category, which are then used to train the monitor. Posts that are not clear or can fit into more than one category are skipped during training. When each category has enough training posts, (e.g. around 20), the monitor is ready to run, meaning the algorithm begins its analysis of the historical data. The algorithm analyzes all of the data and provides the proportion of tweets related to each theme using the information from the training set to determine the statistical pattern of conversation for each category. The following is an example of the process used to produce monitors in the topic of Gas and Fuel.

Gas and Fuel

The goal of looking at gas and fuel was to explore how rising prices in gas and fuel were affecting populations; so the initial categories focused specifically on questions around price. Through a preliminary analysis of gas and fuel discussion, researchers observed that public concerns in this area differ considerably between Indonesia and the U.S. In Indonesia, conversations included discussions of affordability and availability of various types of fuel, including gasoline, diesel oil, and kerosene. In the U.S., gasoline prices dominated conversations. For the purposes of this project, researchers determined

that it would be useful to differentiate the nature of the conversation in these two countries and dive into these two distinct subtopics of gas and fuel. In addition, many of the concerns related to gas and fuel in Indonesia were not related to price at all, but rather other issues such as safety, availability and outages.

The following guidelines were used for the two monitors under the topic “Gas and Fuel”:

Table 2: Setup process for Gas and Fuel monitors

Indonesia: Fuel – Bahasa + Java

1. Source: Twitter only
2. Date range: July 1, 2010 to present
3. Bahasa Indonesia/Javanese-language content only
4. Keywords: common words and phrases in language referring to the topic, news author exclusions

Categories:

- Afford
- Can’t Afford
- Expensive
- Substitution (fuel switching)
- Fuel Safety
- Power Outages
- Lines for Fuel
- Off-topic/Irrelevant

U.S.: Gas Prices – English

1. Source: Twitter only
2. Date range: July 1, 2010 to present
3. English-language content only
4. Keywords: (gas OR gasoline) AND location:USA

Categories:

- High-level Gas Price Discussion
 - News/media discussion
 - Consumer discussion about changing prices
- Affordability of Gas Price
 - I purchased gas
 - I can’t afford gas/too expensive
 - Someone gifted me gas
 - Consumer posts on gas prices in specific location
- Off-topic/Irrelevant

By building monitors to reflect distinct subtopics within the overarching topic of gas and fuel, researchers were able to more specifically address relevant and country-specific concerns.

3 Analysis

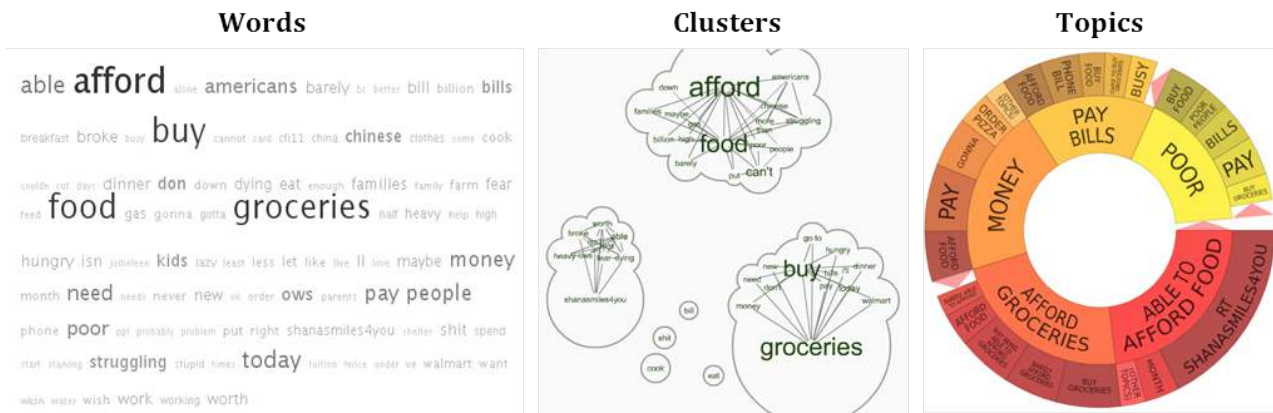
During data analysis within the monitors, the classification algorithm quantifies the breakdown of each category represented in the training set. As shown in Figure 2, the platform is able to provide a quantitative overview of the results of the analysis.

Over a given period of time, researchers can also investigate trending topics and explore stories behind the spikes by looking through specific tweets. Figures 3 demonstrates this type of work.

Figure 2: Summary and Opinion Analysis tabs of Crimson Hexagon ForSight platform



Figure 3: Exploring stories using Words, Clusters, and Topics features of ForSight



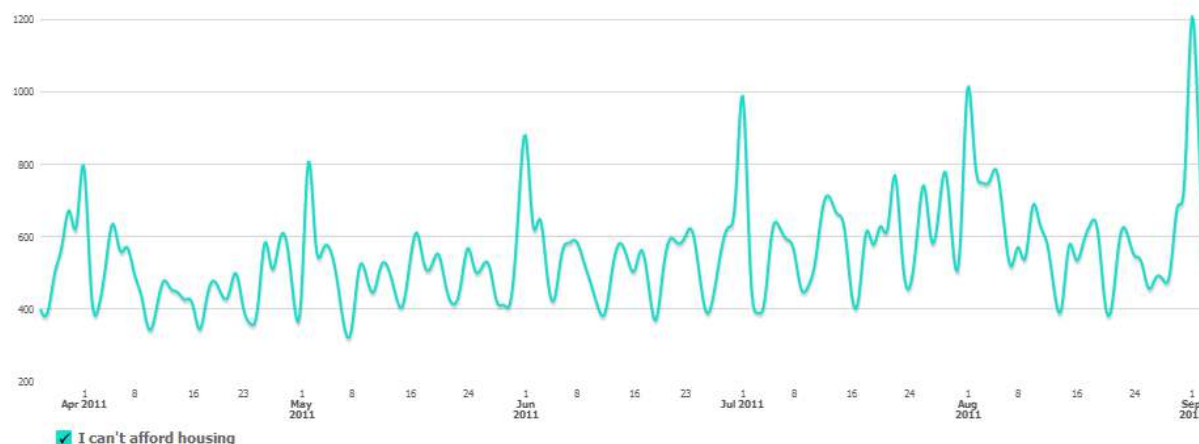
In this research, Global Pulse and Crimson Hexagon primarily explored the *types* of analytical methodologies that might be useful for understanding the data and transforming it to policy-actionable information. In order to set up the basis, a special effort was made to understand what people tweet about, and thus comprehend the universe of tweets related to food, fuel, finance and housing and generate the different categorizations.

Before we explore the methodologies and the data, a note on privacy is warranted. Privacy concerns are an important consideration in any social science research project. It is important to note that this form of social media analytics relies on *aggregate* analyses of large amounts of public data, as opposed to tracking of *individual* data or information. Sometimes, specific tweets are read to get some context or to make sense of an anomaly. Wherever individual tweets are analyzed, they represent data that is also available by accessing a publically available page within Twitter.com. Hereafter we explain some of the methodologies and results we found.

Baseline assessment

For all of the analysis, understanding the baseline is critical. First, the overall growth of Twitter—both the growth in Twitter users and the growth in tweets— over the relevant time period in each location needs to be understood. Second, identifying “normal” tweeting patterns around each of the issues is key for finding changes in trends. In some cases, the baseline trend of conversations is not flat but presents clear periodic patterns. Figure 4 below represents one intuitive example.

Figure 4: Volume of conversations about affording housing in the U.S. from April 29, 2011 to September 3, 2011.

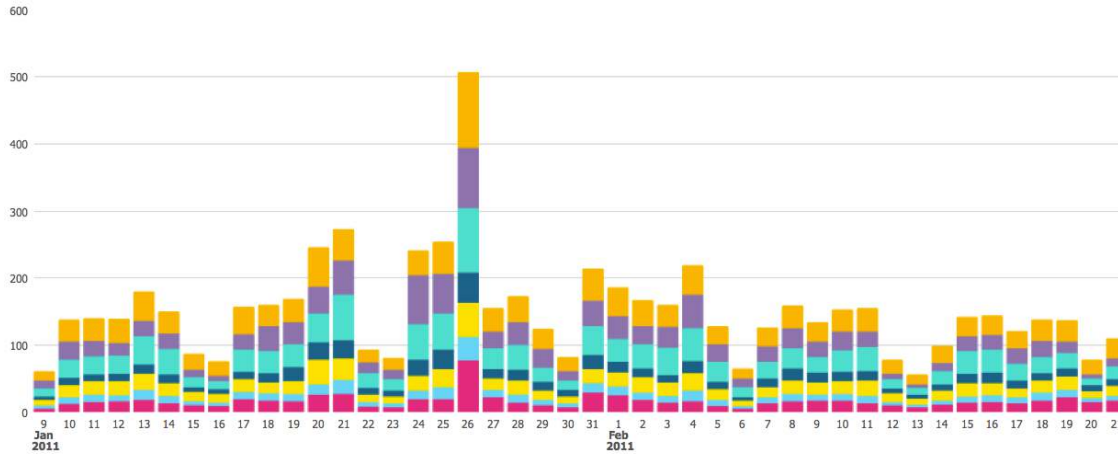


The volume of conversations about “affording housing” spikes consistently on the first day of every month. This result is intuitive since most people pay rent or other housing related bills around the first of the month. There is also another weekly pattern – with fewer conversations over the weekend - which is less pronounced but no less consistent. This simple example shows how the signal representing the baseline conversations around housing is modulated by a weekly and a monthly frequency.

Anomaly detection

Often we noticed spikes or drops in the volume of a particular topic. This usually happens on a daily timescale and is often driven by a particular event or news item. Figure 5 represents one such example.

Figure 5: Student loan conversations in U.S. from January 9, 2011 to February 21, 2011



This figure shows that the volume of tweets about student loans doubles in one day. The spike is on January 26, the day after President Obama’s State of the Union address (many of these tweets will have started after midnight on the same night of the speech). Examining the trending topics among the tweets, it is clear that many in the “Twitterverse” felt that President Obama failed to adequately address issues related to student loans.

One would expect in an emergency situation, detecting anomalies would be a key analytical tool to understand actionable impacts in real time. As an example, Figure 6 represents the number of tweets related to power outages and lack of fuel in Indonesia. Figure 7 shows tweets about being in a line to buy fuel. While the main reason driving this indicator is people going out at night at the peak hour, it shows the real-time potential of this kind of information for understanding immediate preoccupations.

Figure 6: Power outages in Indonesia

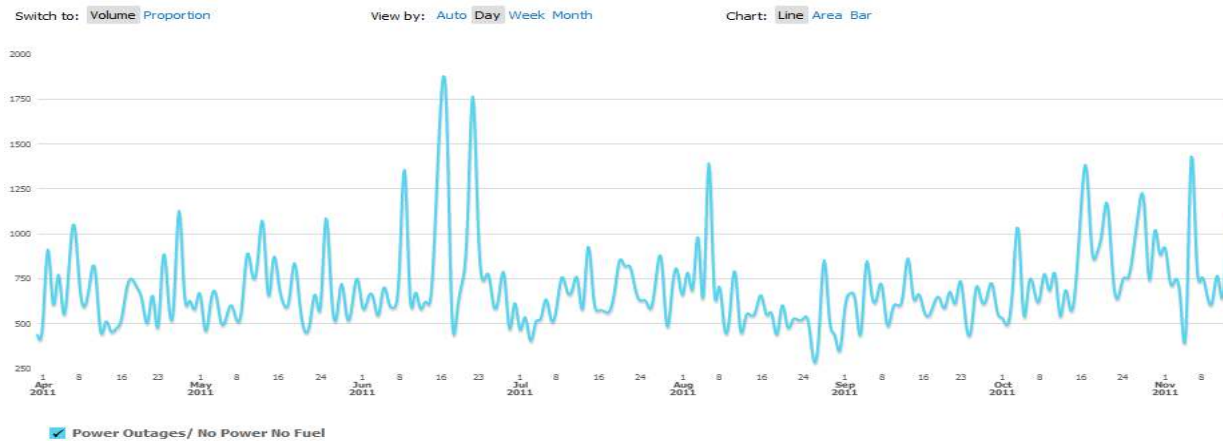
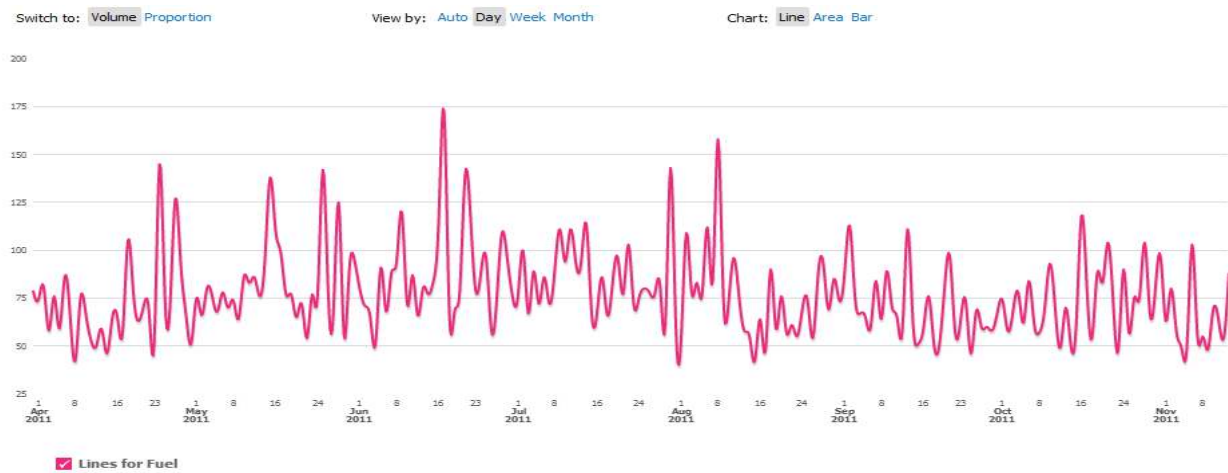
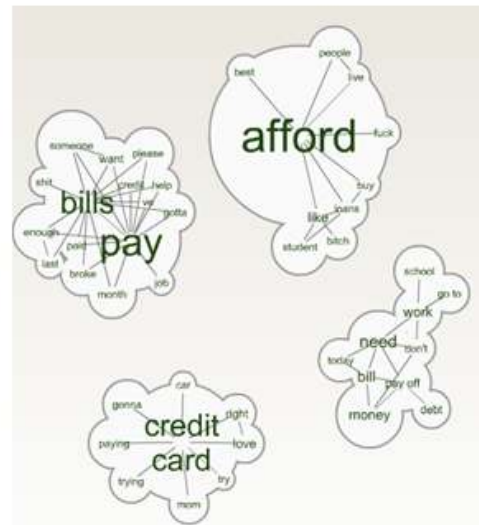
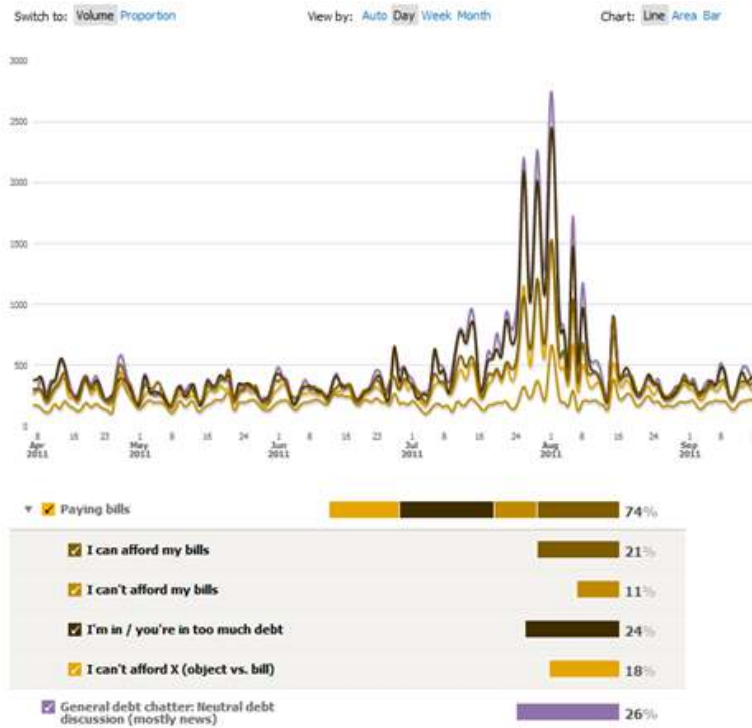


Figure 7: Lines for fuel in Indonesia



However, what we call anomalies might happen in a longer timescale – weeks or months. In Figure 8, we show that conversations around finance in the US, modulated by the baseline weekly pattern of fewer discussions on the weekends, show an increase of conversations from the 15th July to the 15th August motivated by the US debt ceiling debate.

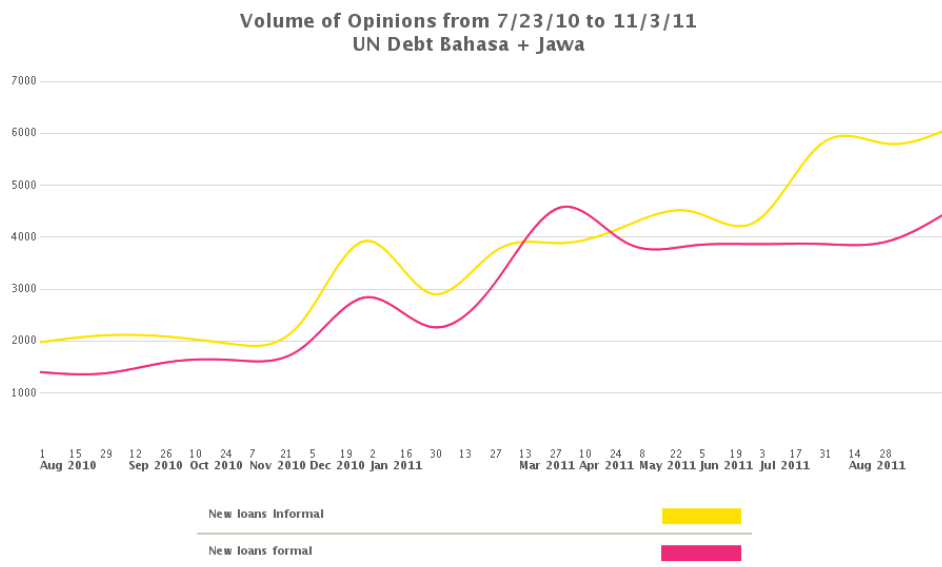
Figure 8: Increased conversations during US Debt Ceiling Debate



Proportions between categories within the same topic

We also looked at the proportions of tweets in different categories on the same topic. Changes in the relative shares of sentiment within one topic could point to changing circumstances in the population—both positive and negative. One interesting example is from discussions of debt in Indonesia. As shown in Figure 9, the proportion of tweets about informal debt is rising compared to the proportion of tweets about formal debt. Understanding the implications of this observation requires further investigation. It could mean that formal mechanisms for accessing loans are being eroded, indicating increased stress in the population. Conversely, it could also mean that people are relying less on formal mechanisms because their informal lending networks have strengthened.

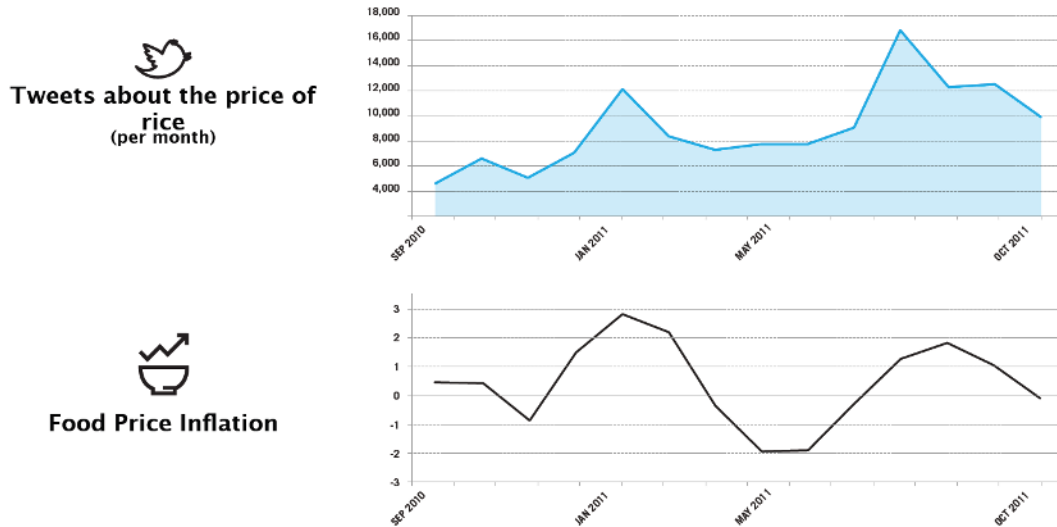
Figure 9: Proportional analysis of Debt Conversations in Indonesia



Cross validation with external data sources

The long term evolution of a particular category of tweets may also underline long-term trends on the importance of an issue that might be represented by another external indicator. In the following example we show the number of tweets per month commenting about the price of rice in Indonesia. While there has been a continuous growth in the quantity of tweets, we see two periods – around February 2011 and September 2011 – when more conversations took place. Interestingly these increases follow the official inflation for the food basket, indicating that when prices rise, people notice and express their concerns. Further investigation is required to gain a more qualitative insight into the messages underlying these increasing concerns: the drivers of conversation in these tweets include commentary on rice being expensive; people expressing relief when they are able to find cheap food; specific changes in the prices of rice per measuring unit (karung/liter/etc.), and what percent of one’s budget is spent purchasing rice.

Figure 10: Volume of tweets per month about the price of rice from October 2010 to October 2011 in Bahasa Indonesia/ Javanese and monthly inflation rate for the food basket in Indonesia from October 2010 to October 2011.



5 Challenges, Lessons Learned and Next Steps

Topics of Twitter posts vary widely, and may include news, life-casting⁵, consumer opinions, general conversation, advertisements, spam, and simple spreading of information and content. This range of online communication must be taken into account to determine the needs and the scope of any project. In our initial scoping, we found that the most straightforward analysis was based on daily anomaly detection. However, other medium-long term insights were gleaned from overall growth and changes in proportions when looking at complementary topics of discussion. It is important also to establish the baseline for detecting the anomalies, therefore taking into account the intrinsic increase of Twitter usage or periodical usage patterns for some topics of discussion. We also found that this particular form of social media may be more suited to some topics of study than others.

Hereafter we discuss some of the interesting lessons learned and challenges faced, which in some cases also led to what we believe are rich future lines of study. These challenges are relevant for any future work on Twitter data, both within UNGP objectives and more generally.

Demographics

In order to conduct better policy relevant research on how a population is coping with a crisis, better demographic information on who is tweeting is necessary. This might include some combination of

⁵ Life-casting refers to the practice of using social media to broadcast daily life events.

traditional statistics, such as expanding the statistics already collected by the International Telecommunications Union, and more sophisticated work mining Twitter itself⁶.

Geography

The geo-location of tweets may provide interesting information about global networks, migration, and the social graph. It may be useful to better target deeper investigations when Twitter data suggests that populations are experiencing crisis impacts.

Key Influencers

In this project, we are looking at overall numbers of tweets. In future work, we would like to better understand how to leverage some of the attributes that are particular to Twitter culture, such as number of Twitter followers and indeed retweets. In the first case, not all tweets are the same. Some tweets reach 20,000 other twitterers while others reach only 5. By better understanding the role and reach of highly networked individuals, there could be an opportunity to identify nodes on Twitter that impact conversation. In the second case, while retweets may have the potential to influence a monitor on any particular project, they also could reveal a great deal about how information spreads, and what type of information is spreading.

Time Horizon

In the course of this project, the “now-casting” nature of Twitter data was confirmed. While the data is largely perception data, it primarily sheds light on perceptions of the moment, and may be less suited to understanding how people perceive the future. However, now-casting can reveal a great deal, particularly in the case of emergencies; furthermore, aggregating these perceptions over time can create a baseline such that even slower-moving crises may be detected.

Twitter Culture: Conversations in the Public Sphere

We also learned more about the dynamics of what people are willing to share on Twitter and what they may be less willing to put in the public sphere. This was different between Indonesia and the US, but did not necessarily conform to expectation—the example cited earlier of debt discussions in the U.S. is a case in point. For this reason, one of the primary lessons learned was about the *process* involved in working with Twitter data in general, and probably with social media tools more broadly. Project objectives seek to capture human behaviors but also must adapt to how human interactions are already taking place. This is nothing different than standard social science research, which emphasizes an understanding of local culture. However, what is relevant here is that Twitter has its own dynamic culture, which may change over time and varies by topic, location, and other factors.

The Justin Bieber Effect: Signal to Noise Ratio

Some irrelevant trends in Twitter can drive the results. This effect is particularly pronounced when we searched for trends in broad conceptual categories. For example, when we initially started this project, we cast a wide net, looking at what causes might be connected to the use of words like “afford,” “worry”, and “excited.” What we found was that Justin Bieber was trending in every category. Focusing

⁶ For early experimental work on inferring Twitter demographics see “Understanding the demographics of Twitter users” by A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11), Barcelona, Spain (2011).

on the more specific topics was more fruitful here, such as “food” or “housing.” However, getting too specific – for example, looking for the discussions around “milk” – did not result in enough data. Finding the balance between specificity and broad trends was key. In addition, even when generally this balance was achieved, there is still potential for a lot of “noise.” For example, jokes that get retweeted or “go viral” can bias results for a particular category. In addition, events or big news items can drive the attention of all of Twitter—for example, in all of our monitors in the US, there was a spike in Twitter when Osama bin Laden’s death was announced.

The lessons-learned from this research also led to specific ideas on future lines of work.

First, one of the most interesting challenges, as previously discussed, is the balance between specificity and volume. More specific queries and monitors lead to more specific and typically more actionable results, as we saw with student loans. But, analyses that are too specific risk overlooking broader trends. In order to fully grasp the right balance for this project and future research, additional thought must be given to how the data will be used and what level of specificity is best suited to policy-actionable data in all the categories – food, fuel, housing and finance.

Second, when analyzing patterns and anomalies, looking for non-intuitive patterns could potentially point to previously unknown patterns of impact in a population’s behavior. Because of the time period of our data-set, we are confined to looking at patterns of one month or less. However, over time pattern detection could include seasonal, annual, or other trends. This research should be conducted in close collaboration with practitioners, to discover relevant non-intuitive trends.

Third, while focusing on populations’ tweets about certain topics is a powerful approach, we also see a great opportunity for tracking specific programs through Twitter data. In this project, particular events - like the State of the Union Address – were reflected as a spike or anomaly in our topic categories. We could also have built the monitors around particular events or United Nations’ programs themselves, similar to how a newspaper may track interest in a story. For example, it might be interesting for development agencies to monitor a programs based on how they appear in tweets, alongside their other monitoring activities.

6 Conclusion

Overall, this initial research has shown the potential of Twitter analysis to explore people’s perceptions of crisis-related stress. We focused on food, energy, finance and housing topics in the U.S. and Indonesia. While the most straight forward application of this data is to understand people’s reactions to specific events in a daily scale, there is also potential for trend detection on a weekly or monthly scale. The main challenges in these early stages are the creation of baseline data and the understanding of the demographics of people who tweet. What we can affirm is that people use Twitter to express their concerns about crisis-related events – though differently from country to country – and people speak about broad concerns such as their financial situation as well as about basic needs such as food, energy, and housing.

Annex I

The following shows an example of the category breakdown in Bahasa Indonesia/Javanese and English monitors across the four topics selected:

Topic	Monitor	Categories
Food	Bahasa + Jawa – Food	<ul style="list-style-type: none"> - Afford - Can't afford - Expensive <ul style="list-style-type: none"> - Unavailable - Neutral
	English – Food	<ul style="list-style-type: none"> - Afford - Can't afford <ul style="list-style-type: none"> - Neutral
Energy	Bahasa + Jawa – Fuel	<ul style="list-style-type: none"> - Afford - Can't afford - Expensive - Substitution (fuel switching) <ul style="list-style-type: none"> - Fuel safety - Power outages - No power/No fuel - Lines for fuel
	English – Gas Prices	<ul style="list-style-type: none"> - High-level gas price discussion (News/Media, Consumers discussing changing prices) <ul style="list-style-type: none"> - Affordability of gas price discussion (Purchased gas, Can't afford gas/ expensive, Someone gifted me gas, Consumer posts on location)
Finance	English – Consumer Debt	<ul style="list-style-type: none"> - Paying bills (Can afford/Can't afford bills, In Debt, Can't afford object) <ul style="list-style-type: none"> - Neutral/General debt discussion (mostly news)
	English – Student Loans	<ul style="list-style-type: none"> - Have student loan debt - Hate Sallie Mae - Student loan company is a nuisance - Will be paying student loans forever <ul style="list-style-type: none"> - Paying off/paid off loans - Tough to pay back loans - Facts/News about student loans
Housing	Bahasa + Jawa – Housing	<ul style="list-style-type: none"> - Afford to buy - Can't afford to buy - Afford rent <ul style="list-style-type: none"> - Wish I had a house - Selling my house
	English – Housing	<ul style="list-style-type: none"> - Want to buy/sell and buy new - Want to refinance mortgage - Can afford housing <ul style="list-style-type: none"> - Can't afford housing - Discussion of others' ability to afford - News/Media mentions