“Interactive Crisis Mapping” Proposal

Background and Motivation.

Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

I am generally interested in the content of messages and how meta-narratives are formed within culture. As it relates to crises, there are stakeholders – actors and victims – within the geo-political structure containing the unfolding events. The stakeholders generate primary messages about the crisis and its projected effects from their viewpoint and interests. Additionally, there are secondary messages which are perpetuated from outside the crisis, some of which are a near mirror of the primary and others are representative of external viewpoints and interests. "More" coherent and resonant messages will tend to have larger coverage footprints. "Less" coherent messages will tend to be lost among the noise of other media and will have a smaller footprint. One of the outcomes of this project will be to provide some fairly robust tools needed to amplify the signal-to-noise ratio of various messages being perpetuated so that relevant media can be mined for deeper insights. Further, this project will expose, at times indirectly, the extent of media coverage given to a given crisis over time and potentially even down to the attention given to individual messages. By correlating winning and losing stakeholders to the coverage given in attempts to bolster or weaken their respective messages, an interested entity can derive strategies to employ in a future crisis in hope of securing better outcomes in the humanitarian and political responses applied to the crisis.

Project Objectives.

Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.

Overview and Hypothesis

This project will explore the extent and effects of media coverage on a crisis, distinguishing between traditional and social media providers and further, to the extent data is available, between stakeholders and non-stakeholders. For anyone interested in the effectiveness of humanitarian or disaster engagement / response, our visual products will facilitate user understanding of which events garner media attention; in turn, a user can use our product to identify strategies how to achieve better results.

A crisis for our purposes is a destructive event in human history inspiring extraordinary social and/or media attention. We hypothesize that media coverage is correlated to successful crisis resolution such that intensive coverage results in and from more stakeholder energy being exerted to efficiently manage a crisis. Also, we hypothesize that there are detectable indicators throughout the lifespan of a crisis, from onset to resolution, and we further hypothesize that by systematically analyzing and integrating patterns found in multiple crises, future indicators projected from an ongoing crisis may be predicted at a precision correlated to the quality and quantity of comparable analysis. Predictions will be constrained to represent anticipated conformance to or deviation relative the lifespan of other compared crises.

Audience and Focus

Our product is able to assist users in better understanding a given crisis, the content of coverage, and the saturation of messages within analyzed coverage. High- and mid-level visuals from the project should be useful to all consumers of crisis media, to include those with even a passing interest in humanitarian and geo-political issues. Additionally, low level visuals should be useful to producers of crisis media, those directly connected to the crisis domain, and interested researchers.

Framework and Questions

Within the project, we intend to establish a framework able to provide scalable context for analyzing the events within a crisis in both absolute terms and relative to other crises. Our framework will result in a consumable visual product that should allow a user to explore the following questions upon engagement with the product, understanding that a user will only be able to explore certain questions through deep analysis -- we will balance the needs of more casual media consumers with those working with the crisis domain as we establish our project feature set and milestone goals:

- Which crises are available for comparison?
- Where and when did each crisis occur?
- What is the scale of the crises in both relative and absolute terms?
- What is the corpus of data backing this visualization?
- How can additional crisis data be added?
- How can crisis data be filtered for more precise comparisons and better context shaping?
- What were the key events / story points in each crisis?

- What is the coverage given to each detectable event / story point within a crisis, to include the length and intensity of the tail of coverage after an event? And, how does the intensity of coverage correlate to the intensity of the event?
- What is the sentiment analysis of a crisis, down to individual events / story points?
- What is the saturation of messages from within analyzed coverage?
- Are there any prominent stakeholders with a sizeable media footprint? If so, what message are they promoting?
- Are there any prominent external advocates with a sizeable media footprint? If so, what message are they promoting?
- Given a crisis x, how does its sequence frequency compare in relative terms to the sequence frequency of comparable crises of size n (i.e. is it moving through its lifespan at a slower or faster rate)?
- Given a crisis x that is earlier in its lifespan than comparable crises of size n, what predicted paths will the crisis take?
- Are there any macro patterns that emerge from the analysis of multiple crises -- e.g. given crisis x of characteristic set y and media coverage profile z, it is predicted to follow path p?

Crises of Interest

At a minimum, our project will document the time line of four crises, two historical and two ongoing and try to investigate how each crisis was covered and the general sentiment around the crisis. We intend to setup a framework to manage the inclusion of additional crisis data and may be able to assess more than our minimum in the final presentation. However, we want to ensure quality over quantity from the outset and thus will only focus on the agreed upon minimum until all other project goals are satisfied. Here is the preliminary list of crises for our presentation; they are subject to change pending outcomes of our research:

- Natural Disaster:
  - Baseline: Typhoon in Philippines (Nov '13).
  - Ongoing: Pakistan starvation from drought (starting ~Jan '14).
- Political Crises:
  - Baseline: Protests in Turkey (~May ‘13).
  - Ongoing: Ukrainian protests (starting ~Feb ‘14).

Data.

From where and how are you collecting your data? If appropriate, provide a link to your data sources.

Understanding the Crises

- Research using packaged crisis overview material from major news outlets such as BBC, CNN, and NYT.
- Key information will be manually supplied to the products as Story Points, geo-temporally bound.
- Consider images and video correlating to events.
- Possible data from Dhairya's work and contacts.

Traditional Media

- Faroo API Search (http://www.faroo.com/hp/api/api.html -- 1 million queries per month) or similar.
- Google News Search with identified terms and temporal binding for the crisis, possibly limiting domain.
- General News Aggregators: http://blog.programmableweb.com/2012/02/01/81-news-apis-digg-fanfeedr-and-clearforest/.
- AP: http://developer.ap.org/files/AP_Metadata_Services_Developer_Guide.pdf,http://developer.ap.org/home.
- TBD RSS feeds.
- Direct searches of identified news outlets.

Social Media

- Faroo API Search (http://www.faroo.com/hp/api/api.html -- 1 million queries per month) or similar.
- Reddit.
- TBD RSS feeds.
- Twitter (1 week window).
- Possible historic data for Twitter, such as GNIP (http://gnip.com/ -- will require educational account) or similar.

Data Processing.

Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

The pattern of engagement will generally be very similar for API accessed data as well as non-API – the steps will be as follows:

1. Form query based on understanding of params supported by source and submit.
2. Get response and parse result metadata.

3. Crawl URLs from results.
4. Add URL content to indexer.
5. Apply any manual annotations as needed.

The exception to this general pattern will be any semi-structured to structured data sources we are able to include. For these sources, we will be able to form more refined queries and be able to leverage any structured annotations more directly. For Twitter, if we are able to obtain historic data, we would be able to consider authors, followers, hashtags, retweets, and more granular sentiment / keyword markup. Where available data supports measurability, we will look to establish popularity of a given post / article and its length of importance.

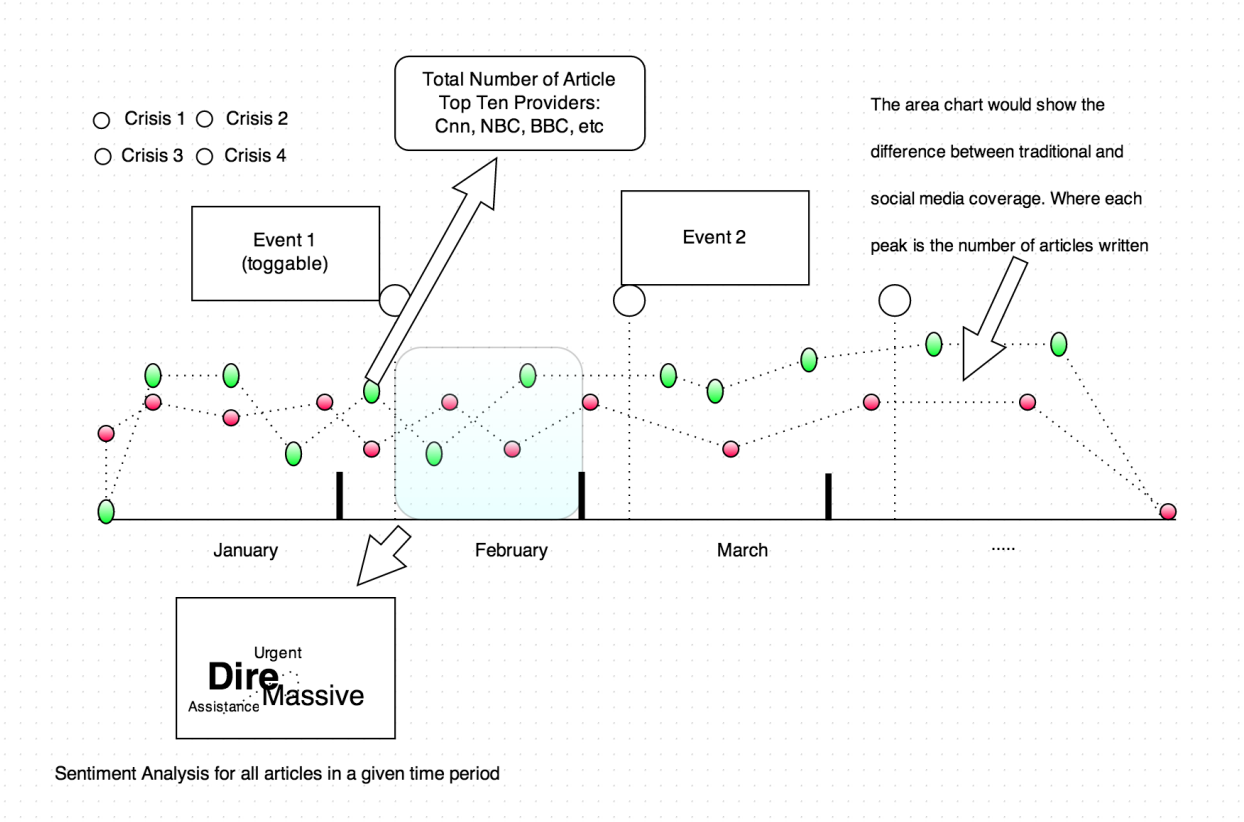
Visualization.

How will you display your data? Provide some general ideas that you have for the visualization design. Include sketches of your design.

- High-Level -- Geo-Bubble chart.
  - Shows magnitudes of crisis damage vs. media coverage, possible using multi-variable scaling. This will be geo-located on a map of the earth.

  [Note: Low-level sketch is due Week 8 according to internal schedule]
- Mid-Level -- Timeline able to show relative and absolute time scales.
  - Brush to show keywords / sentiment in a Word Cloud or similar.
  - Story points.
  - Compare crisis overlay (media coverage).

  Here is an initial sketch from our team discussion; this will be refined for Week 8 internal deadline.



- Low-Level: Deeper engagement with data.
  - Access to actual articles.
  - Statistics of article / corpus based on index metrics.
  - Moods of article / corpus based on index metrics.

  [Note: Low-level sketch is due Week 8 according to internal schedule]

Must-Have Features.

These are features without which you would consider your project to be a failure.

High-Level:

- Geo-Bubble high-level chart showing available crises, traditional media coverage, social media coverage. Encode by type of crisis and type of coverage.

Mid-Level:

- Story Points would need to be added for context.
- Multi-Line chart able to compare selected crises.
- Keyword / Sentiment Word Cloud able to express majority opinions.

Low-Level:

- Access to read individual posts / articles.

Optional Features.

Those features which you consider would be nice to have, but not critical.

High-Level:

- Possibly encode more interactivity with the bubble-chart such that mid-level chart responds to selections.
- Consider facet filtering (e.g. exclude / include select providers) to better shape context.
- Expose ability to add more crisis data via UI (e.g. provide Google News Search Terms and temporal information and have it auto-parsed).

Mid-Level:

- Normalization by crisis maturation patterns for comparison (i.e. compare a crisis to a provided standard model).
- Sentiment timeline: where did opinion shift? Any correlation to events?
- Rise and Fall of voices / messages: show when and to what extent messages and proponents received attention.

Low-Level:

- Statistics of article / corpus based on index metrics.
- Moods of article / corpus based on index metrics.

Project Schedule.

Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

Week 6 (March 2-8)

- Internal Deadlines (due Saturday, March 8): Refine project concept.

Week 7 (March 9-15)

- ***Milestone-1: Thursday, March 13: Project proposal due (part of Homework 3)***
- Internal Deadlines (due Saturday, March 11): Finalize group project proposal document portion.

Week 8 (March 16-22)

- ***Spring Recess***
- Internal Deadlines (due Saturday, March 22): Finalize datasets, validate data ingestion (ETL), identify story points, and initial UI mockups.

Week 9 (March 23-29)

- Internal Deadlines (due Saturday, March 29): Project in GitHub, project hosted with initial data, and harden Milestone-2 scope.

Week 10 (March 30 - April 5)

- Internal Deadlines (due Saturday, April 5): Checkpoint toward Milestone-2

Week 11 (April 6-12)

- ***Milestone-2: Thursday, April 10: Functional project prototype due***
- Internal Deadlines (due Saturday, April 12): Harden Milestone-4 scope.

Week 12 (April 13-19)

- ***Milestone-3: Week of April 14: Project review with the TFs***
- Internal Deadlines (due Saturday, April 19): Checkpoint toward Milestone-4.

Week 13 (April 20-26)

- Internal Deadlines (due Saturday, April 26): End of major development, focus on presentation, code QA, and documentation.

Week 14 (April 27 – May 3)

- ***Milestone-4: Thursday, May 1: Projects due (including screencast)***
- Internal Deadlines (due Tuesday, April 29): Screencast complete and only documentation change up through delivery.

Week 15 (May 4-10)

- ***Milestone-5: Thursday, May 8: Best project presentations and prizes***