TweetTracker Mini-Guide
Naval Postgraduate School
Military Scholars Socio-Technical Research
About this deck

This is a short introduction to Arizona State University’s TweetTracker
What is TweetTracker?

1. A sophisticated “crawling” system for accumulating tweets from Twitter
2. Has internal visualization built in:
   - Mapping for the first 7500 tweets
   - Rapid trending
   - Easy search and export
   - Exportable data in many formats: .tsv, .xml, .kmz, JSON

Crawling Explained

- A “job” is set up (aka “query” or “request”). It goes on and on. You don’t have to set it again.
- A “crawler” is created and launched.
- In about two hours, results trickle in to the server or other reservoir.
- Then users can grab, search, filter, export, process, fuse data—and set up and look at trends on any term they wish.
How much data will you get?

- Does it get ALL the data?
  - **No.** Twitter generates 20MB of data per SECOND! You couldn’t possibly handle all the data!
  - No one knows how much of the data you will get. You’re limited to 1% of ALL the data that Twitter produces for that day. TweetTracker usually produces 10-50K tweets per HOUR on a significant disasters, occasionally more. This is sufficient for most purposes.
  - During massacres in Syria in 2011, ASU noticed that they got FEWER tweets than expected. This was probably due to many people, all over the world, requesting the same data at the same time.

- **MOST people request and READ tweets.** SOME people retweet those tweets
  Few people (comparatively) WRITE them.
Logging into TweetTracker

http://blogtrackers.fulton.asu.edu/twttracker/#

The first time, you register. After registering, log in.

TweetTracker
A Twitter data aggregation, event monitoring, and event analysis platform to track emerging Humanitarian Assistance/Disaster Relief events and crises around the world in near real-time.

Click here to register the first time
Existing Jobs

These are existing jobs. You can choose one or several. Notice the slider bar at the bottom. There are MANY jobs.
The Afghan Media and Afghan Elections Jobs

On 26 March, 2014, two jobs were set up to monitor the Afghan Elections.

The objective: to show how to use TweetTracker to help identify rumor and reports of violence during the election period.

Two jobs were established:

**Afghan Elections:** primarily tracking hashtags associated with the election

**Afghan Media:** a special filter that contains a small number of hashtags and a large number of media sites in Afghanistan (like Tolonews, Pajhwok, ArianaTV) and reporters on the ground in Afghanistan. This is the “capable crowd” -- the people whose business it is to know everything that is going on in the region and reporting on it via news, blogs and Twitter.

The examples here come from Afghan Media because:

- It’s very small and fast.
- It’s less “junky” -- less filled with chit-chat, bots and other irrelevant information
- It has interesting results.
- It can be compared with the general filter for instructional purposes.
Tracking Tweets

TweetTracking has **three modes**

- Tweet Viewing (Tweetalyzer) including the Global Entities Tab
- Search/Export (for further analysis, reducing the dataset)
- Trending (visualization of topics over time)
• Tweetalyzer is the default mode

• It will show you ALL the tweets you’ve collected

• The more data, the slower and noisier the results.

• If you want the whole results of a busy day or a long period of time, then get patient (it could take awhile -- 10-15 minutes)
Choose an Event

- Check a box (Yes, you can check more than one. More on this later)
- Next, we’ll choose the time mode:
  - Streaming
  - Movie
  - Fixed Date

These are clickable “buttons”
Streaming Mode

Streaming Mode is useful if you are tracking an ACTIVE crisis.

* Tweets come in “while you wait”

* You can set the time back a few hours, which shows you data FROM a particular point.

Example: You set a filter up at midnight, when you heard about the earthquake. You go to bed, then get up to see what’s happening. You can set the TIME to when you last looked at it.
Using Streaming for a “Hot” Topic

Streaming Mode is most often used for a topic that is “breaking.”

Example: Afghan Media Job

The instructor set the filter for monitoring events on Twitter for the Afghan election. However much the action is going to take place from 7pm til 6 in the morning.

At 10 am, the instructor opens the streaming job and then SET the time on the job BACK to 0200 hours, (the time she went to bed).

Now this job would contain all the new and current tweets AND the tweets from 2am onward.
Note that the time of the tweets on top is 10 am. The earlier tweets are below.

Now you can look from the MOST recent, the the LEAST recent.

But you can also look at the AGGREGATE results in the “Global Entities Tab”
The Exercise

To recreate the Saturday morning after the election, we’re going to use “Fixed Date” mode.

Select the job “Afghan Media”

Now hit the FIXED DATE MODE button
Choose Time Window

- Use the calendar to choose your date start date (April 5, 2014)
- Put in start date, time using military time (leave hh and mm blank and you’ll get the whole day). (0200)
- Put in the END DATE and TIME (April 5, 2014 -- 1030)
- Hit LOAD TWEET button. Wait for your tweets to load.
More about the screen

You can HIDE THE MAP to see more tweets.

Number of tweets in the sample is here.

One page of 25K tweets at a time is provided. For more, use the next and previous page buttons.

You can click on these “buttons” to reorder the tweets-- By user, by date, by the tweet itself, by geo-tag, verification, number of followers AND retweets. They’re all clickable. And you can click on the tweet itself.

<table>
<thead>
<tr>
<th>User</th>
<th>Date</th>
<th>Tweet</th>
<th>GeoTagged</th>
<th>Verified</th>
<th>Followers</th>
<th>Retweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>stephensmys</td>
<td>Jan, 28, 2014</td>
<td>RT @AndreaWoo: This is the Mexican ski team's uniform. (&quot;Team&quot; being Hubertus von Hohenlohe - also a German prince) #Olympics <a href="http://t.co/">http://t.co/</a>...</td>
<td>false</td>
<td>false</td>
<td>208</td>
<td>0</td>
</tr>
<tr>
<td>jenn_wead</td>
<td>Jan, 28, 2014</td>
<td>RT @CrossFitLA: So happy for and proud of @SalMasekela on his way to cover #sochi #Olympics! <a href="http://t.co/uq822rXoup">http://t.co/uq822rXoup</a></td>
<td>false</td>
<td>false</td>
<td>252</td>
<td>0</td>
</tr>
<tr>
<td>berryblazebuso</td>
<td>Jan, 28, 2014</td>
<td>RT @LukasteersUnite: I totally respect all the athletes who are</td>
<td>false</td>
<td>false</td>
<td>243</td>
<td>0</td>
</tr>
</tbody>
</table>
OTAAT vs MTAAT

Most tweet tracking software platforms help people look at One Tweet At A Time (OTAAT). Twitter Timelines provide us with excellent methods of looking at each tweet - About 800 of the latest tweets are provided to the user, more than enough for individuals.

Organizations often use larger aggregators like Hootsuite or Tweetdeck to enable users to track multiple issues and discussions (one tweet at a time, many searches). Great for public affairs.

DTNA offers visualizations of individual tweets but aggregates them visually. A great advantage (to be covered later).

TweetTracker offers aggregations and summary reports over thousands and thousands of tweets, or “Many Tweets At A Time” -- MTAAT. This is what the Global Entities tab is about.
Global Entities

- Global Entities tracks viral information
- LINKS: all the links in the set, with their frequencies.
- USERS: most frequently mentioned twitter accounts (good for spotting civil authorities, media, important bloggers)
- The hashtags set in the job appear in GREEN -- OTHER important hashtags are in black—helps you find the hashtags you didn’t know about! Also, they’re the natural biases in the sample.

<table>
<thead>
<tr>
<th>Links</th>
<th>Frequency</th>
<th>Mentioned Users</th>
<th>Frequency</th>
<th>Hashtags</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://tolonews.com/elections2014/">http://tolonews.com/elections2014/</a></td>
<td>15</td>
<td>tolonews</td>
<td>2894</td>
<td>afghanelections</td>
<td>2063</td>
</tr>
<tr>
<td><a href="http://dl/y/LeqPfVU">http://dl/y/LeqPfVU</a></td>
<td>11</td>
<td>HnajaFzoda</td>
<td>270</td>
<td>afghanistan</td>
<td>1473</td>
</tr>
<tr>
<td><a href="http://www.youtube.com/watch?v=HI46OGH-CCA">http://www.youtube.com/watch?v=HI46OGH-CCA</a></td>
<td>5</td>
<td>bashirgawkin</td>
<td>139</td>
<td>kabul</td>
<td>327</td>
</tr>
<tr>
<td><a href="http://tolonews.com/en/tolonews-live-stream">http://tolonews.com/en/tolonews-live-stream</a></td>
<td>5</td>
<td>saimamohsin</td>
<td>123</td>
<td>skhns</td>
<td>130</td>
</tr>
<tr>
<td><a href="http://twitrax.com/carhe">http://twitrax.com/carhe</a></td>
<td>4</td>
<td>seadmohteni</td>
<td>100</td>
<td>afgh</td>
<td>119</td>
</tr>
<tr>
<td><a href="http://www.tolonews.com/tolonews-live-stream">http://www.tolonews.com/tolonews-live-stream</a></td>
<td>3</td>
<td>hashtelabdally</td>
<td>82</td>
<td>lec</td>
<td>92</td>
</tr>
<tr>
<td><a href="http://bastabaf.com/9262">http://bastabaf.com/9262</a></td>
<td>2</td>
<td>empcs</td>
<td>81</td>
<td>kandahar</td>
<td>88</td>
</tr>
<tr>
<td><a href="http://m.wsj.com/articles/BL-DISPATCH-3974">http://m.wsj.com/articles/BL-DISPATCH-3974</a></td>
<td>2</td>
<td>kahzni</td>
<td>67</td>
<td>afghan</td>
<td>96</td>
</tr>
</tbody>
</table>
**Links** takes the URLs out of all the tweets in the sample. It provides the FREQUENCY for each link.

This shows you VIRAL INFORMATION flowing through the site.

Be careful. In some data sets there are “bots” trying to get as many clicks on their website as possible. (TweetTracker can help you find some of those). 99 times out of 100, this is not a significant problem though. But it is there. It will be covered later.

**TT** counts the number of times people are mentioned (in slang, people speak of @mentions).

This will provide you with information on who are the most trusted sources, the people with the biggest networks in your sample.

This helps you find bots, popular sites you may want to follow or track, and the best sources for information.

### Hashtags

Hashtags are the topic counters for topics beginning with the # mark.

In many countries, no one uses hashtags at all! They were invented before Twitter’s keyword indexing was more primitive.

Hashtags CAN show you breaking news rapidly and new topics that are becoming significant that you don’t know about—often indicating a breaking crisis.
When experienced Trackers see “place names” (Ghazni, Kandahar), then that is one indicator that something may be going on in that location.

Violent acts often use place names as a hashtag.
So now we're going to go look for what's happening in Kandahar and Ghazni.
Click on the Search/Export Button.
Next, enter in the SEARCH TERMS and click SEARCH.
Go back to Tweetalyzer.

You should get 164 tweets, with a Word Cloud that looks like this:
Click on one of the words in the Word Cloud!
The Afghan Elections job has over 82K tweets for the time period from 2am to noon on 5 April 2014.

Call up that job. Let’s see what the general public discussed.

Next, we’ll use Trend line analysis to take a look at the traffic by topic during that time period.
Enter Keywords in the Box.

THEN GO ALL THE WAY DOWN HERE!
Hit the TRENDS DATE SELECTION!!
Move the Slider to HOUR

Then hit the PROCESS BUTTON
AND there you are, what topics were hot (from your keyword list) and WHEN they were hot.

You can go back and play with the data to find the information you want by considering the most productive time frames.

This can be used to figure out a breaking incidence, such as a case of breaking violence.
That’s it for the Mini-Guide

You can now play with the data and learn how to manipulate it for different views.

Future guides will show you more about the MTAAT method of Tweet Tracking.