

SOCIAL MEDIA SEMINAR

Ryan Wesslen / Project Mosaic

Part I: April 20, 2017

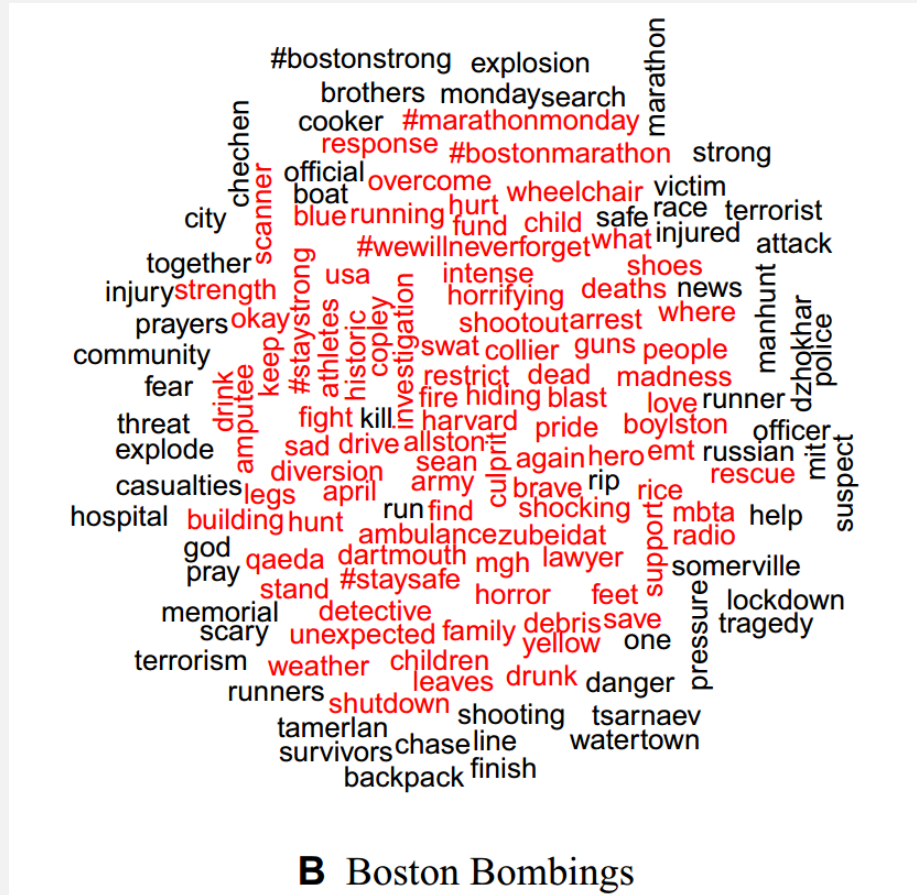
OVERVIEW

- Twitter Querying Crash Course: 10 minutes
- Filtering Rules (How to Query): 30 minutes
 - Example: 10 minutes
 - Try your own search: 20 minutes
- Review, Next Week Homework & Helpful Links: 10 minutes

TWITTER QUERYING CRASH COURSE

- Public API
 - RESTful API => Last 2-3 weeks; up to ~3,200 tweets
 - Streaming API => Future data; up to 1% (~5MM/day)
- Firehose (Gnip)
 - Historical PowerTrack => For **short window events**
 - University wide monthly limits: 10 million AND 200 days
 - Raw (JSON), noisy, large datasets (next week)

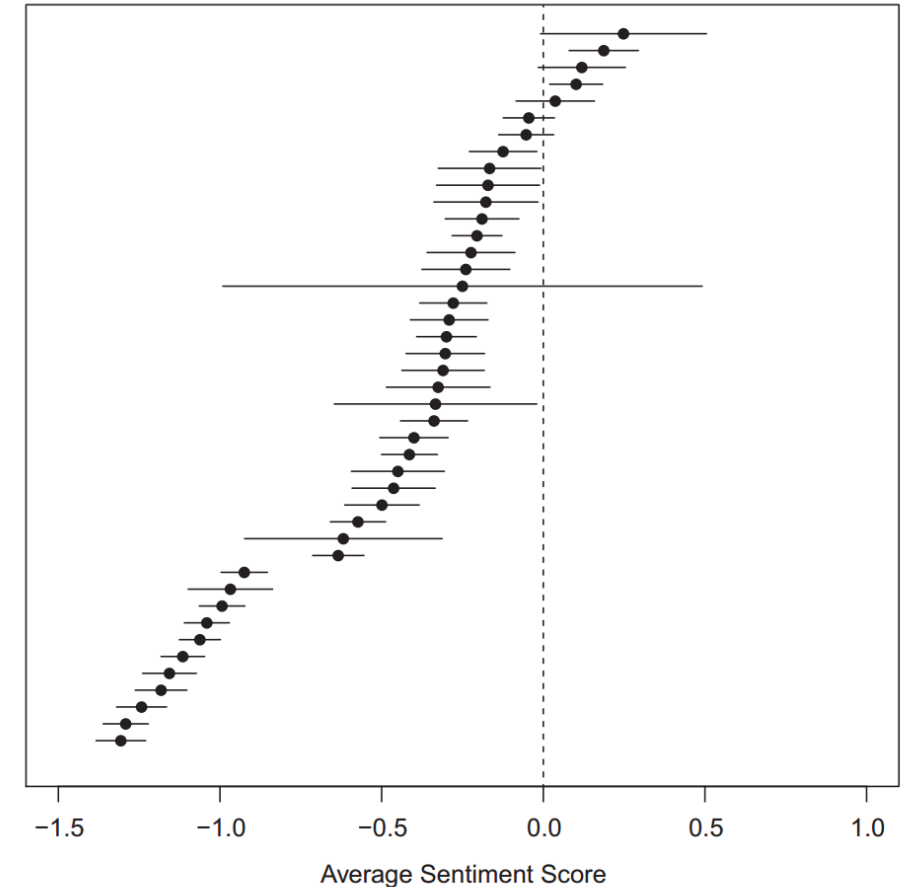
WHY ARE FILTERING RULES IMPORTANT?



Note: Word clouds of keywords were selected by human users; those selected by one and only one respondent are in red (or gray if printed in black and white). The position of each word within the cloud is arbitrary.

Source: King, Lam & Roberts (2017): Computer-Assisted Keyword and Document Set Discovery from Unstructured Text

FIGURE 2 Average Sentiment of 43 Document Sets



Note: Each document set was selected by a different keyword list, with point estimates (as dots) and 95% confidence intervals (horizontal lines) shown.

FILTERING RULES PROCESS

Step 1: Choose Time Horizon

Step 2: Iterative query process (“algorithm”)

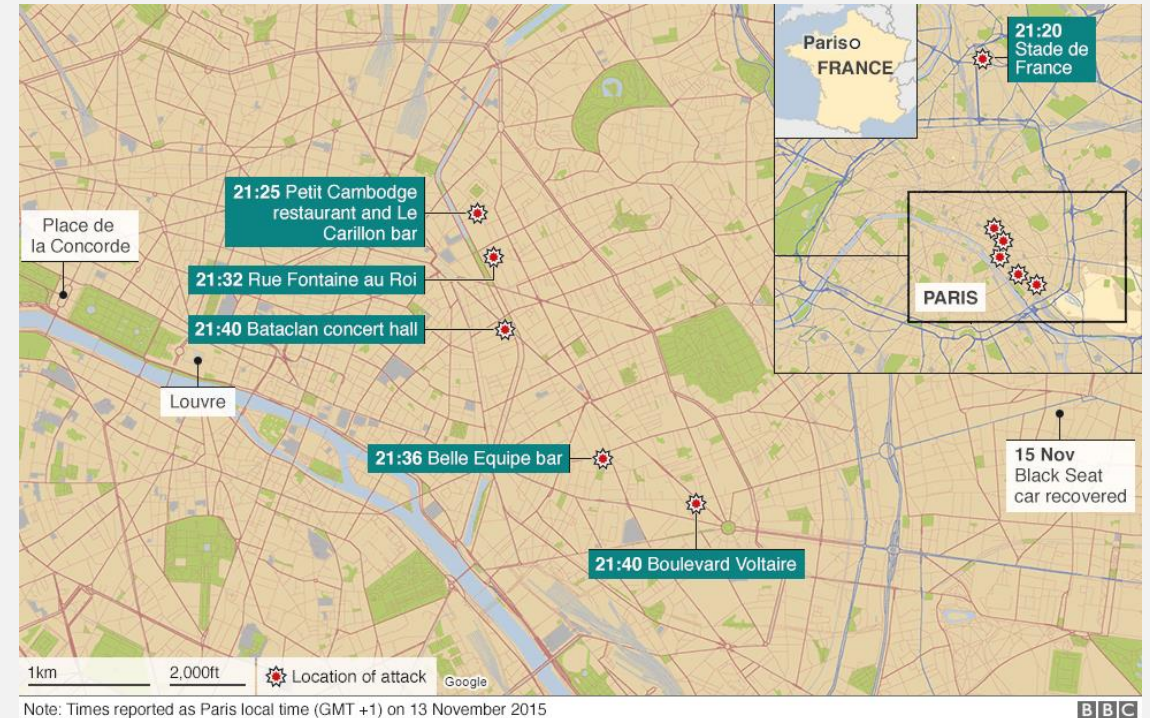
- Trial and error. Repeat until satisfied.
- Use tools (next page) to help.

Step 3: Translate query into JSON filtering rules.

FILTERING RULES & TOOLS TO HELP

- [Gnip Historical PowerTrack Filtering Rules](#)
 - What filter operators (e.g., keyword, user, geolocation) are available.
- [Google Trends](#)
 - Helpful for popular long term events.
- [Statweestics](#)
 - Can show trends for last six months; only works for top 500 terms.
- [Twitter Advanced Search](#)
 - Used to research examples. For recent/future, can use public API for data.

EXAMPLE: NOV 2015 PARIS ATTACKS

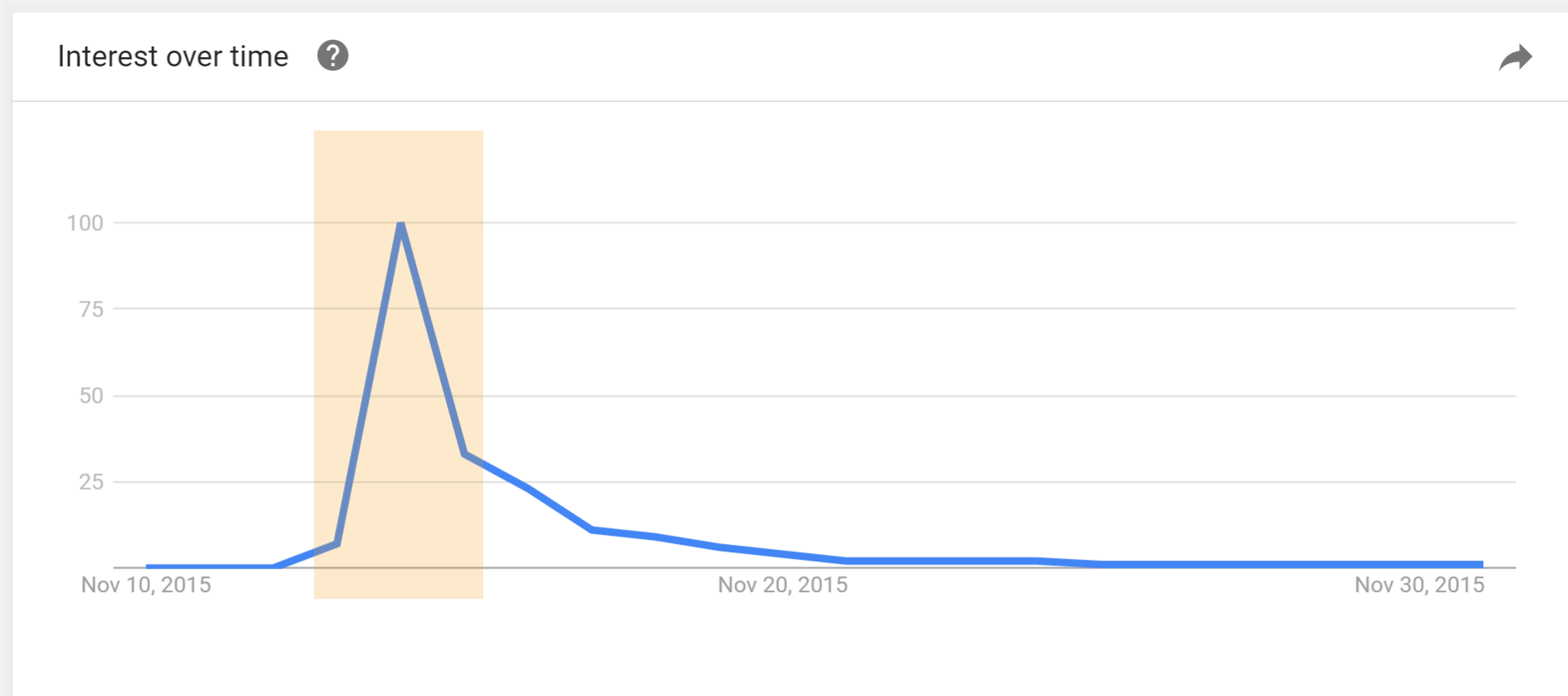


Photos: <http://www.bbc.com/news/world-europe-34818994>

Example from King, Lam & Roberts (2017): Computer-Assisted Keyword and Document Set Discovery from Unstructured Text

STEP 1: CHOOSE TIME HORIZON

Use [Google Trends](#): Nov 13 – Nov 15, 2015



STEP 2: ITERATIVE QUERY PROCESS

1. Search [#parisattacks](#) on Twitter Advanced Search
2. Examine results.
 - A. Find [#prayersforparis](#). Add as an OR operator.
 - B. Find paris – but need to filter.
 - I. [paris AND -\(#parisattacks OR #prayersforparis\)](#)
 - II. Add in additional terms as AND with an OR list
3. Combine for total query:
[#parisattacks OR #prayersforparis OR \(paris AND \(attack OR solidarity OR pray OR terror\)\)](#)

STEP 3: CREATE FILTERING RULES JSON

```
{
  "title": "ParisAttacks",
  "publisher": "twitter",
  "streamType": "track_v2",
  "dataFormat": "activity_streams",
  "fromDate": "201511130000",
  "toDate": "201511160000",
  "serviceUsername": "twitterUsername",
  "rules": [
    {
      "value": "(#parisattacks OR #prayersforparis OR (paris (attack OR solidarity OR pray OR terror)))",
      "tag": "parisattacks"
    }
  ]
}
```

Must be a unique name!

1. YYYYMMDDHHMM format
2. GMT/UTC time
3. fromDate inclusive; toDate exclusive ([link](#))

AND operator is implied ([link](#))

Can create multiple tags
to partition the dataset
(can help in preprocessing!)

To download this sample json, click [here](#).

Pro tip: Install [Notepad++](#) and/or use [Json Formatter](#)

INDIVIDUAL EXERCISE

Now you try to create your own query.

For additional filter rules (operators) like geolocation or by user, see the [Gnip Historical PowerTrack Filtering Rules](#)

HOMework FOR NEXT WEEK

Homework:

- Refine your Json searches.
- If you want to me to run estimates, email me your Json rules by **Tuesday April 24 11:59PM**.
- Can send up to three Json files.
 - Remember: Each must have a unique title (see slide 8).
- Important! Test your json file for formatting here before sending:
<https://jsonformatter.curiousconcept.com/>
- Files with invalid formatting can not be run!

HELPFUL LINKS ON GNIP DATA

- [Gnip Twitter Metadata Dictionary](#)
- [Identifying and understanding Retweets](#)
- [Filtering By Location](#)
 - [Geo Metadata](#)
- [Converting Json to CSV \(Ruby Code\)](#)
- [Consuming, Parsing and Processing Tweets with Python](#)

Strongly suggested reading material for next week

LINKS ON PUBLIC API & ANALYSIS

- [Social Media \(Twitter, Facebook\) Data Acquisition in R](#)
- [Twitter Text Analysis with R Workshop](#)
- [R Interactive Visualizations Tutorials \(Uses Gnip Data\)](#)
- [Beer in Charlotte on Twitter Tutorial](#)
- [GitHub Repository](#)