Broker Bots: Analyzing automated activity during High Impact Events on Twitter

Sudip Mittal

M.Tech Thesis Defense
6th June 2014
Committee

- Dr. Ponnurangam Kumaraguru (IIIT-Delhi)(Chair)
- Dr. Vinayak Naik (IIIT-Delhi)
- Dr. Sundeep Oberoi (TCS)
Introduction

• Twitter 101
• High impact events
• Automated activity/Twitter bot
• Research Motivation
Problem Statement

• Analyze and study automated activity (bots) during High Impact Events on Twitter.
Twitter 101 - Users
Twitter 101 - Tweet

PCWorld @pcworld - 39m
This chip makes @EpsonAmerica WorkForce printers faster than color lasers.
pcwrd.us/1p3d4R0 by @MRiofrioPCW pic.twitter.com/smuuz62qH7W
High Impact Events

• **Events (some example):**
  
  • Government policy changes,
  
  • Elections,
  
  • Natural calamities: earthquakes, tornados, etc.
  
  • Celebrity gossip.

• **Criteria for High Impact Events:**
  
  • Great political and economic impact;
  
  • May also have moderate/high damage to life and/or property.
Sample Posts - Boston Bombings (2013)

• I saw people’s legs blown off. Horrific. Two explosions. Runners were coming in and saw unspeakable horror. – Jackie Bruno (@JackieBrunoNECN) April 15, 2013

• At the ER. Not a comforting way to pass the time. #boston So sad. – GanderHeroDog (@veterantraveler) April 15, 2013

• An eyewitness during the explosions at the #Boston #Marathon says “it sounded like a cannon blast.” Video: on.cnn.com/1399A40 – CNN Video (@CNNVideo) April 15, 2013
Automated Activity on Web

- Wikipedia bot
- Twitter bot
- Spam bot
- IRC bot
- Automated tasks on the web
- Bots can perform tasks that are simple and repetitive, at a higher rate than would be possible for a human.
Twitter Bots

• Posts updates to Twitter automatically.

• Used in:
  • Regularly updating users about information,
  • Spam campaigns,
  • Directing Twitter users to outside webpages.
Research Motivation

- Bots are becoming popular on Twitter.
- Will impact discussions, information flow, credibility, information security, etc.
- People rely on Twitter for important updates, crucial information.
- More during high impact events.

- How do bots impact discussion, information diffusion on Twitter during high impact events?
Related Work

• Work on High Impact Events
• Work on Automated Activity
### Related Work - Work on High Impact Events

<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kwak et. al. 2010</td>
<td>Twitter is a social network or a news media?</td>
</tr>
<tr>
<td>Palen et. al. 2008</td>
<td>Analyzed Twitter adoption and use during emergencies.</td>
</tr>
<tr>
<td>Zhao et. al. 2011</td>
<td>Compared Twitter with traditional media (New York Times).</td>
</tr>
<tr>
<td>Mendoza et. al. 2010</td>
<td>Use of Twitter during the 2010 earthquake in Chile.</td>
</tr>
<tr>
<td>Castillo et. al. 2011</td>
<td>Information credibility on Twitter.</td>
</tr>
<tr>
<td>Gupta et. al. 2012</td>
<td>A mechanism to rank credible information on Twitter.</td>
</tr>
<tr>
<td>Longueville et. al. 2009</td>
<td>Location based mining to analyze forest fires in France.</td>
</tr>
<tr>
<td>Sakaki et. al. 2010</td>
<td>Used Twitter to locate epicenter &amp; impact of earthquakes.</td>
</tr>
<tr>
<td>Agrawal et. al. 2011</td>
<td>Tracked the Mumbai terrorist attack on Twitter.</td>
</tr>
<tr>
<td>Verma et. al. 2011</td>
<td>Used NLP, extract “situational awareness tweets”.</td>
</tr>
<tr>
<td>Vieweg et. al. 2010</td>
<td>Use of Twitter during two natural hazards events.</td>
</tr>
<tr>
<td>Gupta et. al. 2013</td>
<td>Analyzed fake content during the Boston marathon blast.</td>
</tr>
</tbody>
</table>
Related Work - Work on Automated Activity

Zhang et. al. 2011
Analyzing Time Trends in bot activity on Twitter.

Chu et. al. 2010
Distinguish between automated & human accounts on Twitter.

Roure et. al. 2013
Observed bots in the “wild” on Twitter.

Boshmaf et. al. 2011
Showed that online social networks are vulnerable to infiltration by bots on a large scale.

Messias et. al. 2013
Studied social bots and their influence over the network in which they were active.

Tavares et. al. 2013
Temporal analysis to compare activity of bots and humans.

Wald et. al. 2013
Studied the users that were susceptible to bots on Twitter.
Research Findings

• Bots in high impact events spread information obtained from “trusted and verified” sources.

• In the Boston marathon blasts we show that bots push updates to at least 9.53% new users.

• We show that bots do not propagate rumors and even if they do; they do it after some time.

• Bots are moving away from Twitter API based approach to web automation softwares/systems.

• We create a classifier based on user based features with an accuracy of 85.10% in classifying bot and non-bot accounts.
Methodology

- Event Selection
- Event Data Collection
- Data Annotation
- Data Enrichment
Event Selection

Boston Marathon Blast:
2 bombs exploded, 3 people killed. Bombing was followed by shootings, manhunt and firefight.
(April 15, 2013)

Oklahoma Tornado:
Tornado stuck Moore, 24 people were killed. Damages of about $2 Billion.
(May 20, 2013)
Event Selection

**Navy Yard Shooting:**
Lone gunman shot 12 people and injured 3 at NAVSEA, Washington Navy Yard.
(September 16, 2013)

**Cyclone Phailin:**
Cyclonic storm hit Bay of Bengal causing damages of $696 Million and 45 deaths.
(October 4 - 14, 2013)
Event Selection

**US Ice-storm 2014:**
Fierce storm hit East and South Coast of USA. 22 people were killed damages of about $15 Million.
(February 11 - 17, 2014)
Event Data Collection

<table>
<thead>
<tr>
<th>Name</th>
<th>Hashtags</th>
<th>Data Collection Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Blasts</td>
<td>Dzhokhar Tsarnaev, #Watertown, #manhunt, Sean Collier, #BostonStrong, #bostonbombing, #oneboston, bostonmarathon, #prayforboston, boston marathon, #bostonblasts, boston blasts, bostonterrorist, boston explosions, bostonhelp, boston suspect</td>
<td>April 15 - April 21, 2013</td>
</tr>
<tr>
<td>Oklahoma Tornado</td>
<td>Oklahoma, tornado, PrayForOklahoma, care4kidsOK</td>
<td>May 20 - May 22, 2013</td>
</tr>
<tr>
<td>Cyclone Phailin</td>
<td>phailin, cyclonephailin</td>
<td>October 4 - October 16, 2013</td>
</tr>
<tr>
<td>Ice-storm</td>
<td>Icestorm, #USIcestorm, #SnowInUS</td>
<td>February 11 - February 19, 2014</td>
</tr>
</tbody>
</table>

Data collected using Twitter API endpoints.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Boston Blast Marathon</th>
<th>Ice-storm</th>
<th>Navy Yard Shooting</th>
<th>Oklahoma Tornado</th>
<th>Cyclone Phailin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total tweets</td>
<td>7,888,374</td>
<td>433,880</td>
<td>484,609</td>
<td>809,154</td>
<td>76,136</td>
</tr>
<tr>
<td>Total unique users</td>
<td>3,677,531</td>
<td>198,391</td>
<td>257,682</td>
<td>542,049</td>
<td>34,776</td>
</tr>
<tr>
<td>Tweets with URLs</td>
<td>3,420,228</td>
<td>233,576</td>
<td>290,887</td>
<td>388,541</td>
<td>44,990</td>
</tr>
<tr>
<td>Number of re-tweets</td>
<td>5,217,769</td>
<td>209,556</td>
<td>262,362</td>
<td>509,732</td>
<td>41,718</td>
</tr>
</tbody>
</table>
Data Annotation

- Top 200 “tweeters” from each event
- 5 Events
- 1,000 users manually annotated
- Each user annotated by 3 annotators
- All annotators were Masters (CS) and PhD (CS) students
- Criteria: Must be using Twitter for 1 year at least
Data Annotation

• Please see the document for annotation instructions. (Cyborgs discouraged, different from bots)

• **Strict approach:**
  All 3 annotators must have labelled an account as either bot or nonbot. (100% agreement - for quality)

<table>
<thead>
<tr>
<th>Event</th>
<th>Bots</th>
<th>Non-Bots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Blasts</td>
<td>97</td>
<td>17</td>
</tr>
<tr>
<td>Icestorm</td>
<td>87</td>
<td>7</td>
</tr>
<tr>
<td>Navy Yard Shooting</td>
<td>90</td>
<td>11</td>
</tr>
<tr>
<td>Oklahoma Tornado</td>
<td>47</td>
<td>42</td>
</tr>
<tr>
<td>Cyclone Phailin</td>
<td>56</td>
<td>38</td>
</tr>
<tr>
<td>Total (Includes accounts active in multiple events)</td>
<td>377</td>
<td>115</td>
</tr>
</tbody>
</table>
Data Enrichment

• “Enrich” for further analysis.

• We collected their Timeline, Friends and Followers

• “Enriched” by collecting profile meta-data.
Analysis

- Twitter bot creation methodology
- Some characteristics
- Bot Friends
- Bot Followers
- Tweet Analysis
- Impact on Information Diffusion
- Features and Classification
- Detailed analysis of few bots
- Changes (2011 vs 2013)
Exploratory Numbers

• 377 bots, 5 events

• 309 distinct bots (26 active during multiple events)

• 115 distinct nonbots

• 5 events from 2013, 7 events from 2011 (later)
Bot Creation Methodology

• Twitter API rules:
  
  • Don't post *duplicate* content.
  
  • Don't post content *unrelated to the topic*.
  
  • Don't *repeatedly follow or unfollow*.
  
  • Don't *send large number of @replies*.
  
  • Don't *Retweet* aggressively.
  
  • 1,000 tweets per day limit: 41 tweets per hour, 0.68 tweet per min.
Bot Creation Methodology

- 4 major creation methodologies:
  - **Popular tweet based**: “listen” for popular tweets (via API, Trends, etc.) and repost
  - **Keyword based**: look for certain keywords tweets and repost.
  - **Source based**: Content posted/reposted from a particular other Twitter users
  - **Outside content based**: look for updates outside Twitter like RSS feeds, web feeds, blog updates, dedicated databases, etc.
Characteristics - Friends and Followers

- Data points clustered around the origin for bots, more spaced out for nonbots. (similar to Chu et. al.)
Characteristics - Profile

Description

Bots

Nonbots
Bot Friends and Data Sources

• One way: frequency of the “@mention” used to cite source, hold conversations on Twitter. (Retweets, @replies, etc.)

• Retweets in API: “RT @mention tweet text” or “tweet text via @mention”
Bots in high impact events spread information obtained from “trusted and verified” sources.
Bot Followers - Profile Description

Total # of Bot Followers collected: 623,198.
Bot Followers - Network (Boston Marathon Blasts)

- Total # of Bot Followers collected: 623,198.
- Network created using Gephi (open graph viz platform).
- RAM issues!!! Not possible to create the whole network!
- We created 2 random subsets of 40,000 & 20,000 users.
- Created network for the same.
Bot Followers - Network (Boston Marathon Blasts)

Average graph degree: 1.01545
Bot Followers - Network
Bot Followers - Network (Boston Marathon Blasts)

- Most users follow 1 bot account, very few follow more than 1.

- Following more than one bot accounts will “flood” their Timeline; coz bots tweet a lot!
Impact on Information Diffusion during High Impact Events

• Focused on Boston blasts dataset.

• In the Boston marathon blasts we show that bots push updates to at least 9.53% new users.

• These bots can be seen as “brokering” information to other users.
Role in Rumor Propagation (Boston Blasts)

- High impact events: spread of misinformation, incredible information, fake images, hoaxes, malicious content, etc.

- Castillo et. al. & Gupta et. al.: Information credibility on Twitter.

- Gupta et. al. (separate paper) Analyzed spread of fake images during Boston bombings, working on the exact same dataset.

- Look at some rumors in the Boston event and correlate the Timeline with annotated 97 bots from our dataset.
Role in Rumor Propagation (Boston Blasts)

• Rumor 1: Suspect became citizen on 9/11

• “RT @pspoole: Dzhokhar Tsarnaev received US citizenship on Sept 11, 2012 – 11 years to the day after 9/11 attacks http://t.co/kHLL7mkjnn”

• 1st tweet was by @pspoole on Friday April 19, 2013, at 15:34:56 (+0000).

• Only 2 bots Retweet this rumor on Friday, April 19, 2013, at 15:41:35 and 15:47:58 (+0000).
Role in Rumor Propagation (Boston Blasts)

- **Rumor 2:** Sandy hook child killed in bombing
  - “RT @CommonGrandma: She ran for the Sandy Hook children and was 8 years old. #prayforboston http://t.co/cLir6nI7tB”
  - 1st tweet on Friday, April 19, 2013, at 09:56:45 (+0000).
  - This rumor was never picked up by any of the 97 bots in our dataset.
Role in Rumor Propagation (Boston Blasts)

- **Rumor 3:** Donating 1$ Tweet.

- A Tweet by a fake account @_BostonMarathon, “For each RT this gets, $1 will be donated to the victims of the Boston Marathon Explosions. #DonateTo- Boston”

- Tweet time: Monday, April 15, 2013, at 11:29:23 (+0000).

- Only 1 bot in our dataset picked up this Tweet that too on Wednesday, April 17, 2013 at 00:50:24 (+0000)

- Gupta et. al. claimed **28,350 Retweets**. Working on the same dataset.

- Follow bots you will be safe! **We show that bots do not propagate rumors and even if they do; they do it after some time.**

- Bot Source: Verified accounts.
Tweet Analysis - URL Analysis

• 44,071 tweets by bots.
• 7,099 tweets by nonbots.
• 36,672 (83%) bot tweets and 4,849 (68%) nonbot tweets have URLs
• 188 bot and 27 nonbot URLs were marked “malicious” by “Google Safe Browsing API”
• Bots don't spread malicious content.
### Tweet Analysis - Source Analysis

Twitter API gives the “source” of a tweet.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Boston Bots Source</th>
<th>Count</th>
<th>Boston Non Bots Source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>twitterfeed</td>
<td>11405</td>
<td>Web</td>
<td>2603</td>
</tr>
<tr>
<td>2</td>
<td>web</td>
<td>5844</td>
<td>twitterfeed</td>
<td>1969</td>
</tr>
<tr>
<td>3</td>
<td>Tweet Old Post</td>
<td>4052</td>
<td>Tweet Button</td>
<td>1042</td>
</tr>
<tr>
<td>4</td>
<td>dlvr.it</td>
<td>3962</td>
<td>Twitter for iPhone</td>
<td>453</td>
</tr>
<tr>
<td>5</td>
<td>IFTTT</td>
<td>2049</td>
<td>TweetDeck</td>
<td>298</td>
</tr>
<tr>
<td>6</td>
<td>TweetDeck</td>
<td>1515</td>
<td>Botize</td>
<td>246</td>
</tr>
<tr>
<td>7</td>
<td>Crime News Updates</td>
<td>950</td>
<td>Twitter for iPad</td>
<td>220</td>
</tr>
<tr>
<td>8</td>
<td>VenturaCounty.Retweets</td>
<td>609</td>
<td>Echofon</td>
<td>216</td>
</tr>
<tr>
<td>9</td>
<td>WordPress.com</td>
<td>530</td>
<td>Twitterfall</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>Strictly Tweetbot for Wordpress</td>
<td>353</td>
<td>Instagram</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>Bitly Composer</td>
<td>186</td>
<td>HootSuite</td>
<td>3</td>
</tr>
</tbody>
</table>
Tweet Analysis - Source Analysis

- Use of automation services.
- IFTTT (“IF This Then That”) and dlvd.it
- Makes creating bots easy.
- Triggered via outside/inside Twitter activity and posted on Twitter.
Tweet Analysis - Time Analysis

1,000 tweets per day limit: 41 tweets per hour, 0.68 tweet per min.

- Zhang et. al.: automated accounts will exhibit timing patterns that are not found in non-automated users.

- Automated activity is invoked by job schedulers that execute tasks at specific times or intervals.

![Bot Distribution](image1)

![Nonbot Distribution](image2)
Tweet Analysis - Time Analysis

During High impact event

Bot

Nonbot
Tweet Analysis - Time Analysis

During High impact event - Inter-tweet time mean vs s.d.

![Graph](image_url)

- **Bot**
- **Nonbot**
Tweet Analysis - Time Analysis

During High impact event - Average tweet time

Temporal features do not help us in differentiating bots and nonbots during high impact events as proposed by Zhang et. al.
Features

• Chu et. al.’s Twitter features:
  • Tweet Count: Bots > Humans
  • Long term hibernation: Bots show long periods of no activity.
  • Ratio of Friends vs Followers: Bots have more friends; humans balanced.
  • Temporal tweeting pattern: Bots are more active during specific days of the week.
  • Account Creation Date: Bots are more “recent”.
  • Tweet source: Bots use API, Humans use mobiles, web, etc.
  • Presence of URLs: Bots have more.
  • Time Trends: Zang et. al.’s argument on Twitter API.
Features

• Temporal based features: Time trends, hibernation, temporal patterns.

• User based features: Tweet count, ratio of friends vs followers, account creation date, tweet source, presence of URLs.
Features

• We argue that some features are not there during high impact events.
  • Long term hibernation: Bots are active, participate.
  • Temporal tweeting pattern: Event can occur any day of the week.
  • Presence of URLs: 83% bots and 68% nonbot tweets had URLs. Small difference.
  • Time Trends: We compared them in a previous section.
Classification

- Decision Trees (J48 using WEKA) to classify if an account is bot or not.

- 2 sets:
  One with only user based features (F1), and one with user + temporal features (F2).

- Results show that user based features are better in predicting bots.

- Temporal based features are not so helpful.

- Similar result was obtained by Chu et.al.

- We get these results when we remove Time based features.
Classification

• Best knowledge gain features:
  • Tweet source.
  • Presence of URLs.
  • Ratio of Followers vs. Friends.
  • Account creation date.
  • Tweet count.

• Similar results were obtained by Chu et. al.
Detailed Analysis of few Bots

• How bots behave in normal condition?

• We took 5 bots and monitored them for a period of : 5 March 2014 to 9 April 2014.

• We took daily snapshots of their Timeline, Friends, Followers, Mentions, Retweets and @replies. (except Direct Messages - unavailable through API)
## Detailed Analysis of few Bots

<table>
<thead>
<tr>
<th>User ID</th>
<th>Screen Name</th>
<th>Tweets</th>
<th>Following</th>
<th>Followers</th>
</tr>
</thead>
<tbody>
<tr>
<td>21287212</td>
<td>FintechBot</td>
<td>22.2K</td>
<td>297</td>
<td>1,137</td>
</tr>
<tr>
<td>219913533</td>
<td>DTNUSA</td>
<td>565K</td>
<td>83</td>
<td>3,108</td>
</tr>
<tr>
<td>591713080</td>
<td>tipdoge</td>
<td>23.9K</td>
<td>1,333</td>
<td>5,622</td>
</tr>
<tr>
<td>1348277670</td>
<td>Warframe_BOT</td>
<td>3,397</td>
<td>1</td>
<td>4,251</td>
</tr>
<tr>
<td>606204776</td>
<td>BBCWeatherBot</td>
<td>535 (Deletes many tweets)</td>
<td>29</td>
<td>307</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Screen Name</th>
<th>Profile Description</th>
<th>Location Mentioned</th>
<th>External Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>@FintechBot</td>
<td>A little twitter bot that curates financial services tech news. Also some curating by @Aden.76.</td>
<td>England</td>
<td>adendavies.com</td>
</tr>
<tr>
<td>@DTNUSA</td>
<td>Comprehensive Daily News on United States of America Today ~ Copyright (c) DTN News Defense-Technology News <a href="http://defense-technologynews.blogspot.ca/">http://defense-technologynews.blogspot.ca/</a></td>
<td>Canada</td>
<td></td>
</tr>
<tr>
<td>@tipdoge</td>
<td>available commands: balance, deposit, withdraw, tip on the way to moon</td>
<td></td>
<td>tipdoge.info</td>
</tr>
<tr>
<td>@Warframe_BOT</td>
<td>This bot retweets ONLY Warframe ?Alerts from @WarframeAlerts Sorry, I know multiple RT bug. I’ll fix it.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>@BBCWeatherBot</td>
<td>Ask us about the weather! Start a tweet @BBCWeatherbot and tell us where and when you want to know about. Trial service by @NixonMcInnes @BBCRDat #BBCconnected</td>
<td>Everywhere</td>
<td>bbc.co.uk/weather</td>
</tr>
</tbody>
</table>
Detailed Analysis of few Bots - @Warframe_BOT

Changed itself after the event.

News aggregator to game updates.

All previous tweets deleted!
Detailed Analysis of few Bots - @FintechBot & @DTNUSA

Spread financial and general news.

Users tend to Retweet sometimes @reply

These bots never reply back. They do not hold conversations. (Frequently observed pattern)
Detailed Analysis of few Bots - @BBCWeatherBot & @tipdoge

Engage in interaction with users.

Computer programs that require certain input in particular format.

@BBCWeatherBot: requires place and day of the week. Used to search weather info in events.

@tipdoge: different commands like “balance”, “tip”, etc. Used to donate money for victims.
Detailed Analysis of few Bots -
@BBCWeatherBot & @tipdodge

Just Skywalker @fisherir777 · 23m
@tipdoge balance

Doge Tip Bot @tipdoge · 23m
@fisherir777 wuff! your balance is:673.10956668 DOGE

Fred Saunderson @fredsaunderson · 14h
@BBCWeatherBot Edinburgh this afternoon

BBC Weather Bot @BBCWeatherBot · 14h
@fredsaunderson Sorry, we need a place and a time. Try including a day. For a 9-day forecast for that location visit: bbc.co.uk/weather/265022...

Paul Sorrell @paul_sorrell · 8h
@bbcwatherbot Nottingham Thursday

BBC Weather Bot @BBCWeatherBot · 8h
@paul_sorrell We’re expecting temperatures up to 18°C, sunny intervals and a little wind. bbc.co.uk/weather/264117...

from Nottingham

BBC Weather

BBC Weather - Nottingham
By BBC Weather @bbcwather
Hourly weather for Nottingham with a 5 to 10 day forecast, giving a look further ahead.

View on web
Detailed Analysis of few Bots - @BBCWeatherBot & @tipdoge pattern in tweets

Average Levenshtein distance for @BBCWeatherBot: 0.678

Highly repetitive and tweet similarity.

Programs that have a limited output.
Changes 2011 vs 2013

• How did bot participation in high impact events change with time?
• Used the 2011 crisis dataset. (Gupta et. al.)
• Similar annotation scheme (previously discussed)
• Note: Data annotation done in March 2014. Behavior in 2014 does not mean bot/nonbot activity in 2011.

<table>
<thead>
<tr>
<th>Event</th>
<th>Hashtag/Trending Topic</th>
<th>Tweets</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virginia Earthquake</td>
<td>#earthquake, Earthquake SF</td>
<td>277,604</td>
<td>Magnitude 5.8 earthquake</td>
</tr>
<tr>
<td>Indiana Fair Tragedy</td>
<td>Indiana State Fair</td>
<td>49,924</td>
<td>Stage accident at Fair</td>
</tr>
<tr>
<td>Hurricane Irene</td>
<td>Hurricane Irene</td>
<td>90,237</td>
<td>Caused 55 deaths</td>
</tr>
<tr>
<td>Libya Crisis</td>
<td>Libya, tripoli</td>
<td>389,506</td>
<td>Rebel against Qaddafi</td>
</tr>
<tr>
<td>Mumbai Blast</td>
<td>#mumbaiblast,#needhelp</td>
<td>32,156</td>
<td>3 bomb blasts</td>
</tr>
<tr>
<td>UK Riots</td>
<td>#ukriots, #londonriots</td>
<td>542,685</td>
<td>Riots in United Kingdom</td>
</tr>
</tbody>
</table>
Changes 2011 vs 2013

• Annotation results:

<table>
<thead>
<tr>
<th></th>
<th>Bots</th>
<th>Non-Bots</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 2011 Events</td>
<td>183</td>
<td>95</td>
</tr>
</tbody>
</table>

• Major difference was observed in tweet source. Comparing it to 2013 gives us new methods used by bots to post data to Twitter.

<table>
<thead>
<tr>
<th>Rank</th>
<th>2011 Bot Sources</th>
<th>2011 Non Bots Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>twitterfeed</td>
<td>web</td>
</tr>
<tr>
<td>2</td>
<td>TweetDeck</td>
<td>Twitter for iPhone</td>
</tr>
<tr>
<td>3</td>
<td>web</td>
<td>Twitter for BlackBerry</td>
</tr>
<tr>
<td>4</td>
<td>Twitter for iPhone</td>
<td>Twitter for Android</td>
</tr>
<tr>
<td>5</td>
<td>PageBase.Net</td>
<td>Tweet Button</td>
</tr>
<tr>
<td>6</td>
<td>Echofon</td>
<td>Visibili</td>
</tr>
<tr>
<td>7</td>
<td>Tweet Button</td>
<td>Twitter for iPad</td>
</tr>
<tr>
<td>8</td>
<td>Resonancers</td>
<td>UberSocial for BlackBerry</td>
</tr>
<tr>
<td>9</td>
<td>Mobile Web</td>
<td>HootSuite</td>
</tr>
<tr>
<td>10</td>
<td>Butting In</td>
<td>Mobile Web</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Boston Bots Source</th>
<th>Count</th>
<th>Boston Non Bots Source</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>twitterfeed</td>
<td>11405</td>
<td>Web</td>
<td>2603</td>
</tr>
<tr>
<td>2</td>
<td>web</td>
<td>5844</td>
<td>twitterfeed</td>
<td>1969</td>
</tr>
<tr>
<td>3</td>
<td>Tweet Old Post</td>
<td>4052</td>
<td>Tweet Button</td>
<td>1042</td>
</tr>
<tr>
<td>4</td>
<td>dlvr.it</td>
<td>3962</td>
<td>Twitter for iPhone</td>
<td>453</td>
</tr>
<tr>
<td>5</td>
<td>IFTTT</td>
<td>2049</td>
<td>TweetDeck</td>
<td>298</td>
</tr>
<tr>
<td>6</td>
<td>TweetDeck</td>
<td>1515</td>
<td>Botize</td>
<td>246</td>
</tr>
<tr>
<td>7</td>
<td>Crime News Updates</td>
<td>950</td>
<td>Twitter for iPad</td>
<td>220</td>
</tr>
<tr>
<td>8</td>
<td>VenturaCounty_Retweets</td>
<td>609</td>
<td>Echofon</td>
<td>216</td>
</tr>
<tr>
<td>9</td>
<td>Wordpress.com</td>
<td>530</td>
<td>Twitterfall</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>Strictly Tweetbot for Wordpress</td>
<td>353</td>
<td>Instagram</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>Bitly Composer</td>
<td>186</td>
<td>HootSuite</td>
<td>3</td>
</tr>
</tbody>
</table>
Take Aways

• Key Points
• Future Work
• Limitations
Take Aways

• We show that bots in high impact events spread information from “trusted and verified” sources.

• We show that bots aid in information distribution. In the Boston marathon blasts bots push updates to at least 9.53% users.

• We show that bots do not spread rumors and even if they do; they do it after some time.
Take Aways

- Bots are moving away from Twitter API based approach to web automation softwares.
- We analyze the network of bots and bot followers.
- We show temporal based features don’t add value in differentiating bot and non-bot accounts on Twitter during high impact events.
Take Aways

• We create a classifier based on “user based features” with an accuracy of 85.10%.

• We analyze 5 bots and give insights on working and behavior.

• We analyze growth and changes in bot activity in 2011 vs 2013.
Future Work

• Analyze more events.

• Check and verify these results for other high impact events.

• Analyze bot impact on malicious and fake content on Twitter.

• Develop tool to differentiate b/w bots and nonbots.
Limitations

• Annotations were done after few months.

• We can’t claim that collected data is representative.
References


. De Longueville, B., Smith, R. S., and Luraschi, G. ”omg, from here, i can see the flames!”: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In Proceedings of the 2009 International Workshop on Location Based Social Networks (New York, NY, USA, 2009), LBSN ’09, ACM, pp. 73–80.


Acknowledgement

- IIITD
- Dr. PK
- AG
- PJ & PJ
- AA
- NS
- AB
- MG
- SG
- To all my annotators
Thank You

Bye bye IIITD!!! 5 years were fun!!!
Thank you for everything you have taught me!!

After July 2014, you can find me at UMBC ebiquity.

sudipmittal@gmail.com | sudipmittal@umbc.edu
sudip09068@iiitd.ac.in