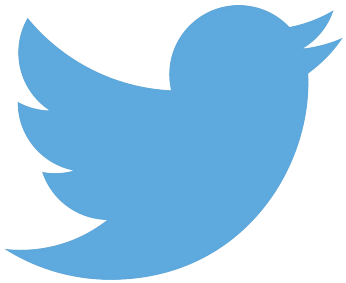
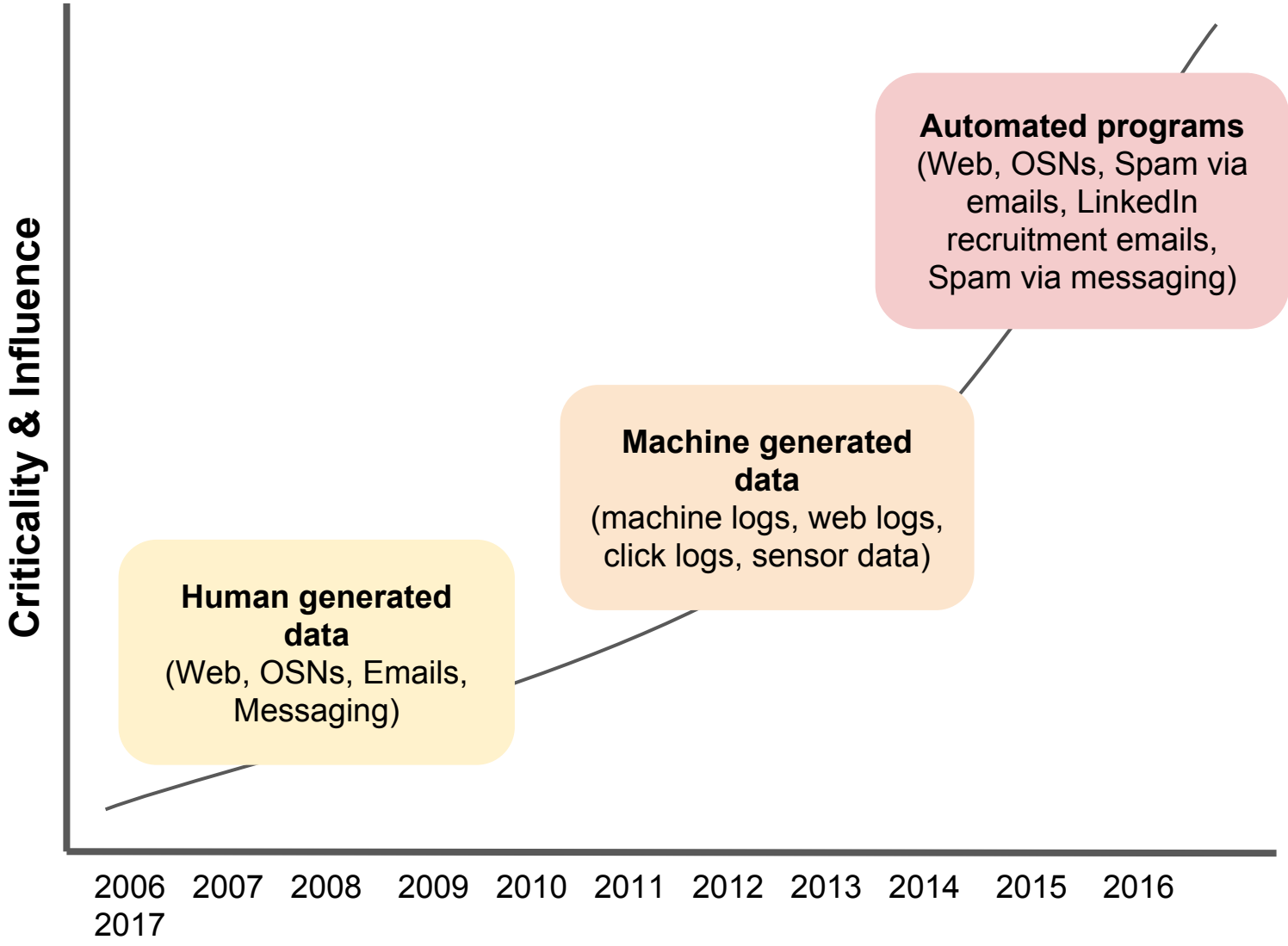
UNIVERSITY OF CAMBRIDGE

Bots on 

Zafar Gilani
zafar.gilani@cl.cam.ac.uk
www.cl.cam.ac.uk/~szuhg2/

Supervisor: Jon Crowcroft

The age of automation: resources & information



Taters-Calder-BoA bot interaction

http://www.theatlantic.com/technology/archive/2014/07/that-time-2-bots-were-talking-and-bank-of-america-butted-in/374023/

Taters-Calder-BoA bot interaction

http://www.theatlantic.com/technology/archive/2014/07/that-time-2-bots-were-talking-and-bank-of-america-butted-in/374023/

Engineering & Automating social interaction

- Bots may artificially inflate support for politicians or launch smear campaigns against opponents, 2010 U.S. midterm elections. [Ratkiewicz et al. 2011]
- Social botnets designed to expose private information, phone numbers, addresses. [Boshmaf et al. 2013]
- Bots contributing to strong polarisation of political discussion by creating an impression of grassroots movement of contrarians. [Conover et al. 2011]
- Bots can alter social media perception by artificially enlarging audience of some people [Edwards et al. 2014]
- Or ruin a reputation of a company [Messias et al. 2013]
- Calling volunteers to action using online bots [Savage et al. 2016]
- Propel TV ratings [Wu et al. 2013]
- … and many more.

Challenges: Bot detection

- Taxonomy:

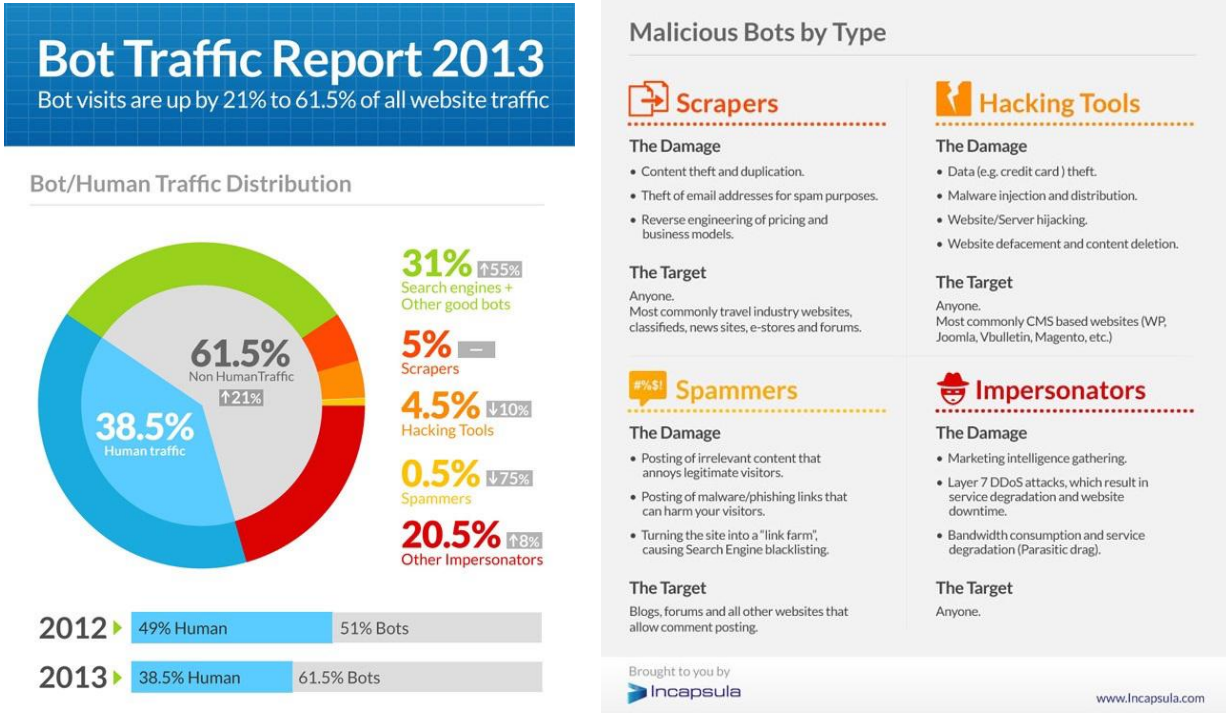
Type	Classes of approaches
Automated	Detection based on social network information
Organic/Manual	Crowd-sourcing and leveraging human intelligence
Automated	Machine learning methods based on identification of features
Automated	Text classification / NLP (*)

- Publicly available APIs: ***BotOrNot***[1]
  - Evaluation: Weak
- Other works: ***SybilRank***[2] (but focused on malicious botnets)

(1) http://truthy.indiana.edu/botornot/
(2) https://users.cs.duke.edu/~qiangcao/sybilrank_project/index.html

Challenges: Bot impact

- Some work on studying discrete cases.
- But, hardly any work on measuring impact on information propagation in OSNs.
- Despite the claims:



https://www.incapsula.com/blog/bot-traffic-report-2013.html

World from ***Sarah's*** eyes

TalentWise liked your Tweet — Feb 23

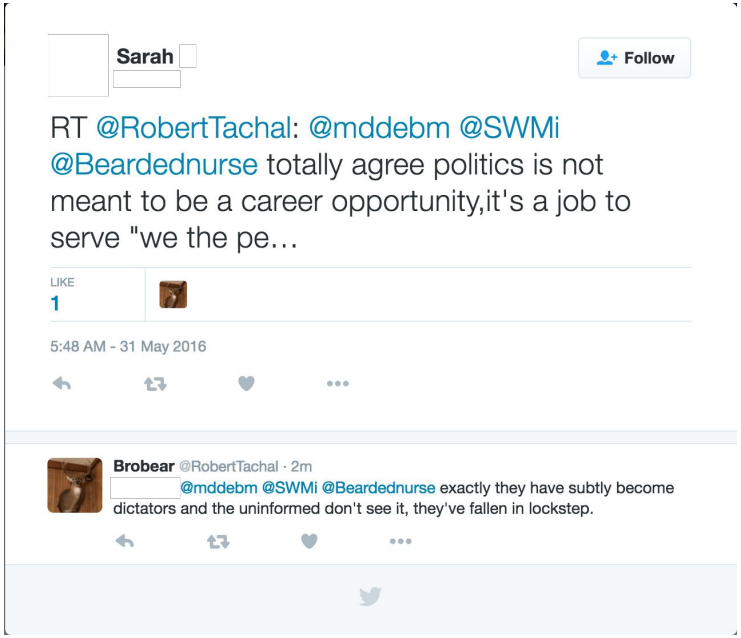
Feb 17: #Career #Advice #Job Get Paid for Consulting Work: Summary: Providing consultations is an opportunity to get p... tnyurl.uk/zy8ia

Badii Bouteraa liked your Retweet — Feb 23

Feb 18: Job Opportunity: Web Developer . Are you a fit? in Charlotte, NC lnkplg.co/XMxP5 #angularjs #javascript

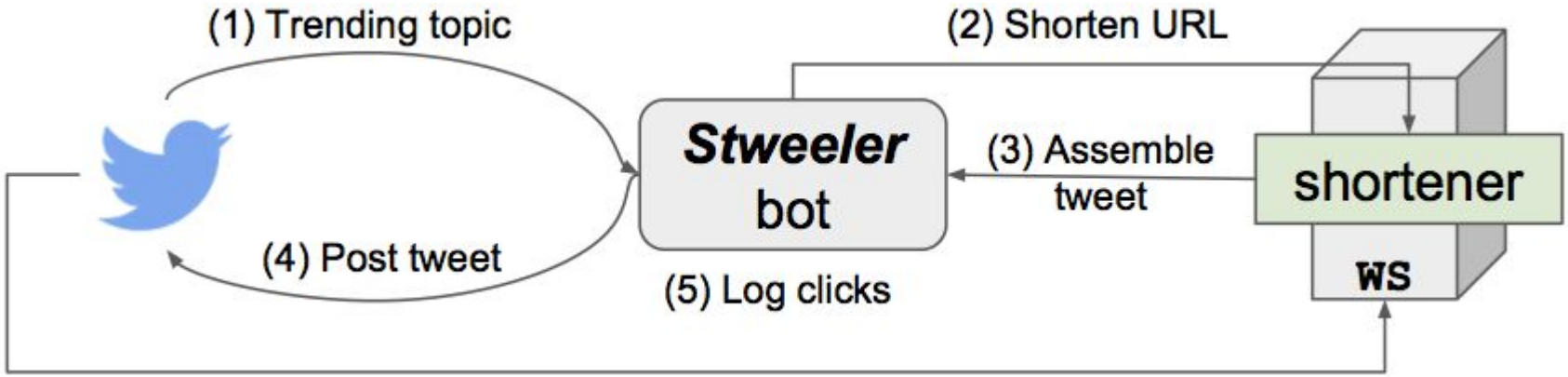
Jojo Retweeted you — 5h

5h: #jobforitaly New job: Software Engineer at Sendo.vn, Vietnam(International Opportunity in VIETNAM) Milano, tnyurl.uk/yfb6q



Stweeler: A toolkit for bot analysis

- Tweet collection system
- A bot (***Sarah***)
- A URL shortener
- A web server / database to collect clicks on bot's tweets / URLs



Clicks on bot's URLs (361 days, as of 16/10)	Clicks by bots
208,844	**97,908 (> 45%)**

https://github.com/zafargilani/stcs

What do we collect?

- Tweets from the Streaming API (tweets in JSON):
  - Bot impact measurement and modelling (primary goal of this research)
  - Bot detection algorithm (secondary goal of this research)
  - Characterisation (future)
- Click data (timestamp, hashed IP, User Agent string):
  - Supplementary to measuring bot impact
  - Supplementary to detecting bots

Note: we do not collect any personal information except whatever is made public by the Streaming API or by User Agent strings of human operated browsers or automated web crawling tools (bots).

On going activities and what's next …

- Bot detection using:

  (1) Human annotated dataset divided into four popularity bands

  (2) Various supervised classifiers, such as Random Forest classifier, and prioritising recall (completeness) over precision (usefulness)

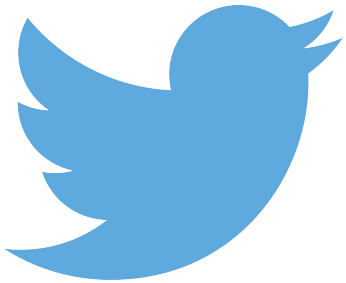
- Define and measure impact in terms of:

  (1) User engagement (clicks, retweets, replies, follows and likes)

  (2) Content dissemination (tweet graph, retweet graph, URL measure)

  (3) Activity and content generated on the Twitter CDN (activity sources, size of content)

UNIVERSITY OF CAMBRIDGE

Bots on 

Zafar Gilani
zafar.gilani@cl.cam.ac.uk
www.cl.cam.ac.uk/~szuhg2/

Supervisor: Jon Crowcroft