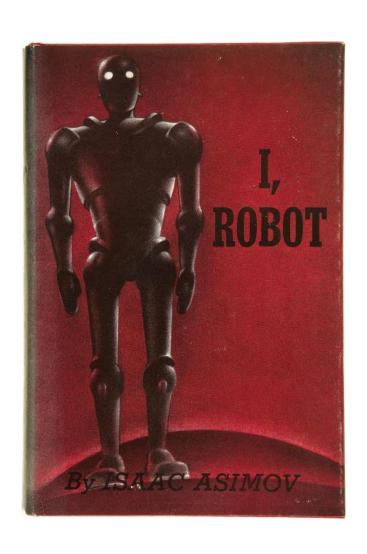
10 years after "I, Robot"



Is that a bot running the social media feed?

C. Edwards, A. Edwards, P. R. Spence, A. K. Shelton

Detecting Automation of Twitter Accounts

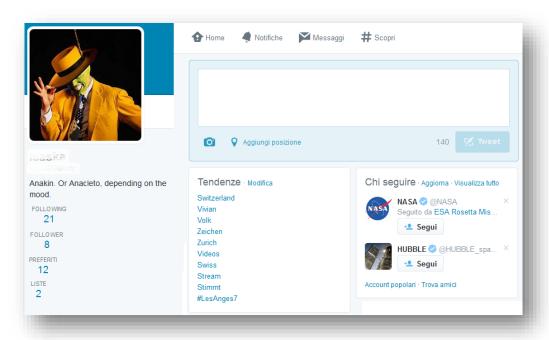
Z. Chu, S. Gianvecchio, H. Wang, S. Jajodia

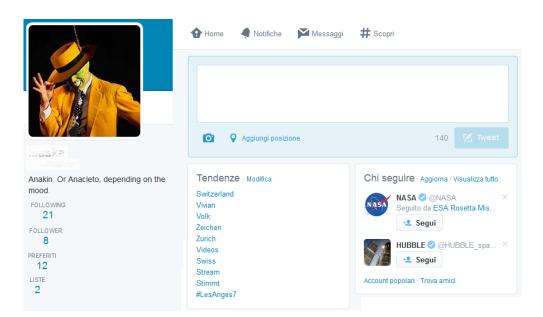
What is Twitter



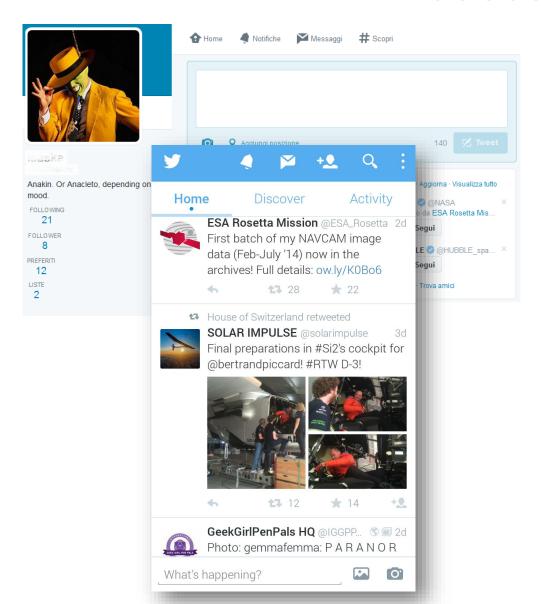
What is Twitter



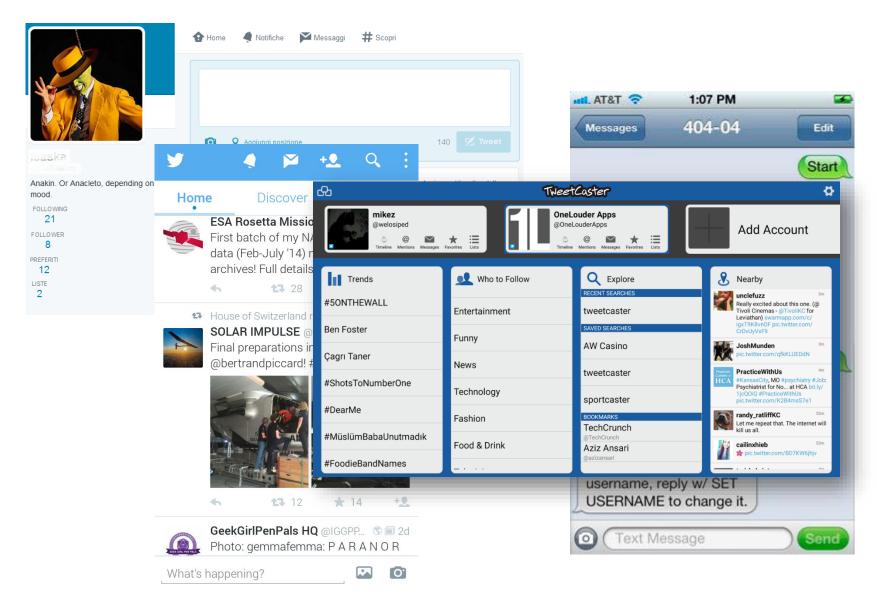


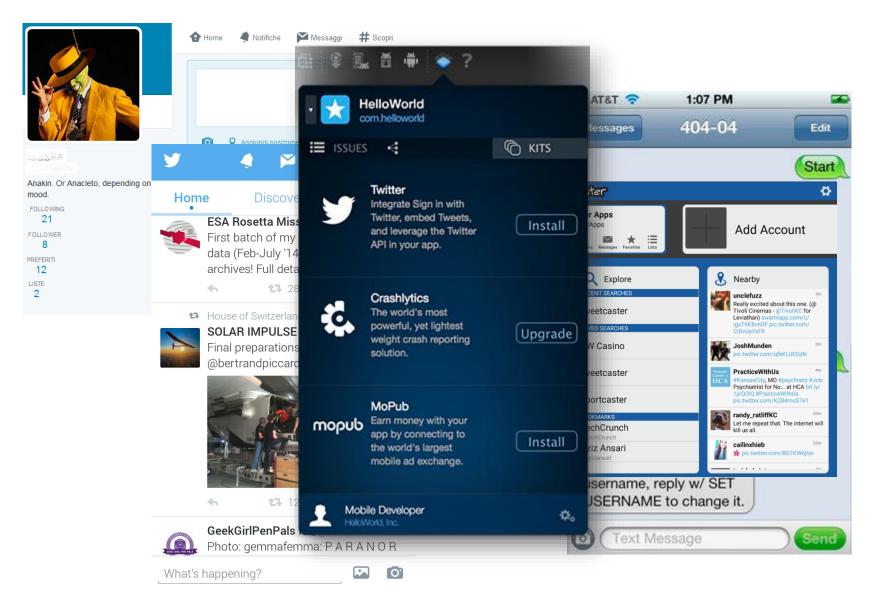












Twitter Evolution



i can't decide if i should pursue my hopes and dreams or start watching Downton Abbey from the start of season one.

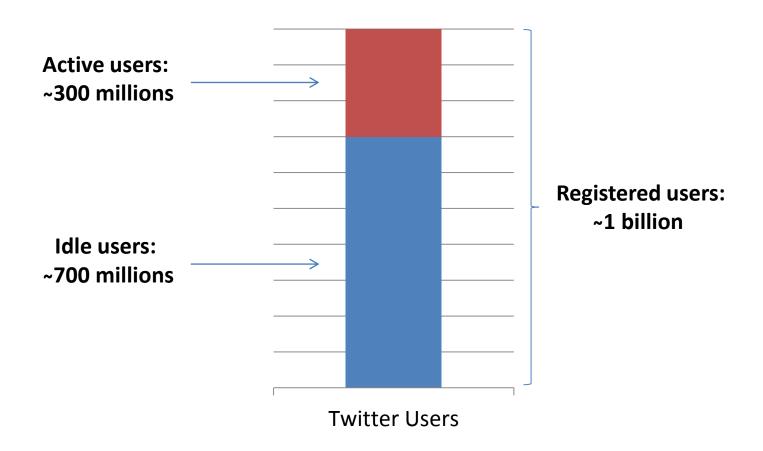


Continues heavy air strikes and counter strikes from all around Sanaa . My god . Hell is on the door. #yemen

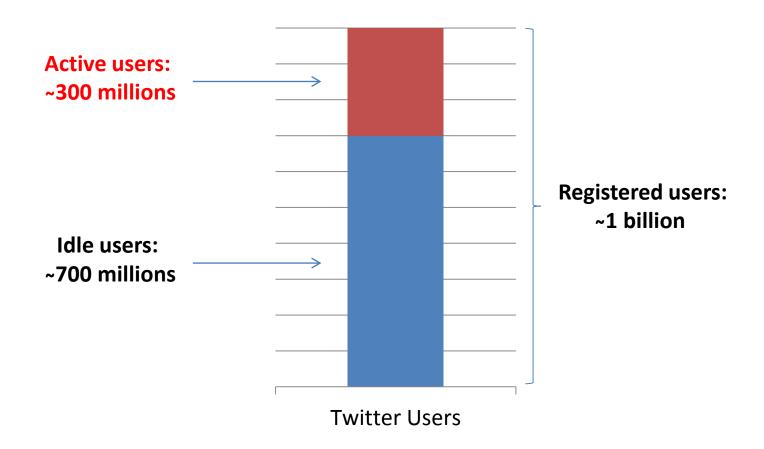


Tunisian authorities say they've killed militant linked to #bardomuseum attack, as march against terrorism begins bbc.in/1EkjpC2

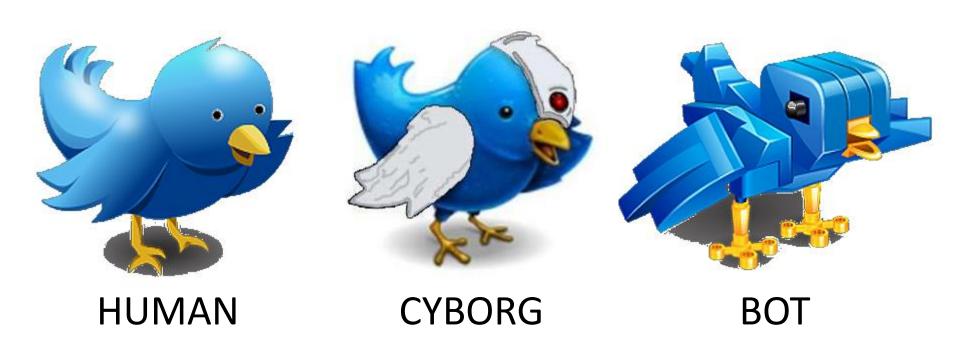
Who are the users?

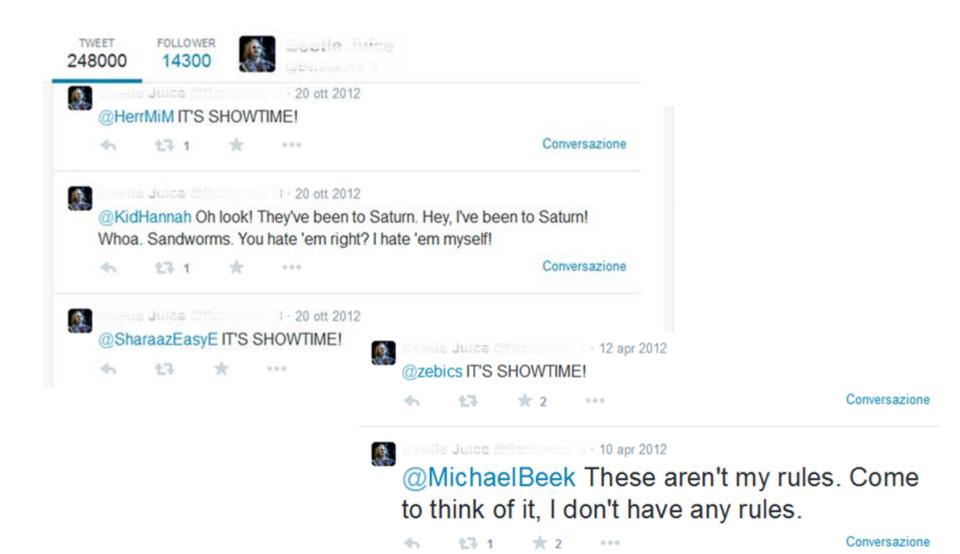


Who are the users?



Who are the users?





@a_dilEmma Let's turn on the juice and see what shakes loose.





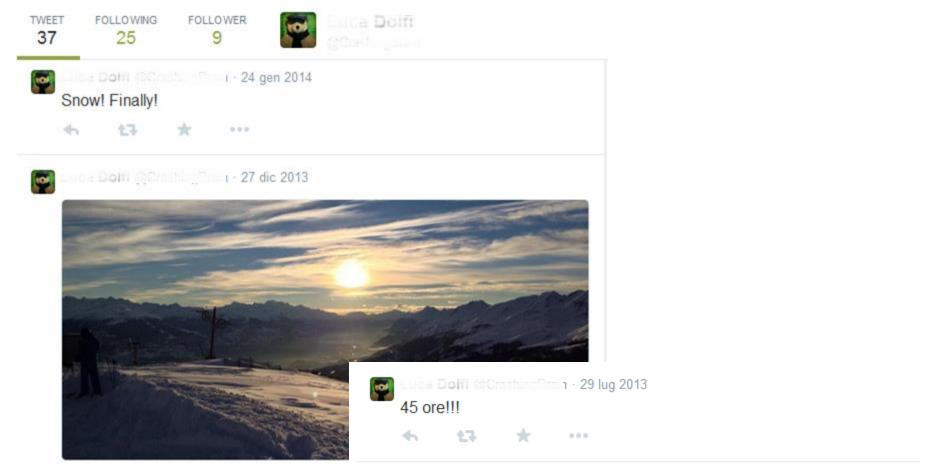
Beetle Juice @Betelgeuss 3 · 10 apr 2012

@MichaelBeek These aren't my rules. Come to think of it, I don't have any rules.

← t+ 1 ★ 2 ··· Conversazione

Beetle Juice @Betalgeuse 3 · 6 apr 2012

@a_dilEmma Let's turn on the juice and see what shakes loose.



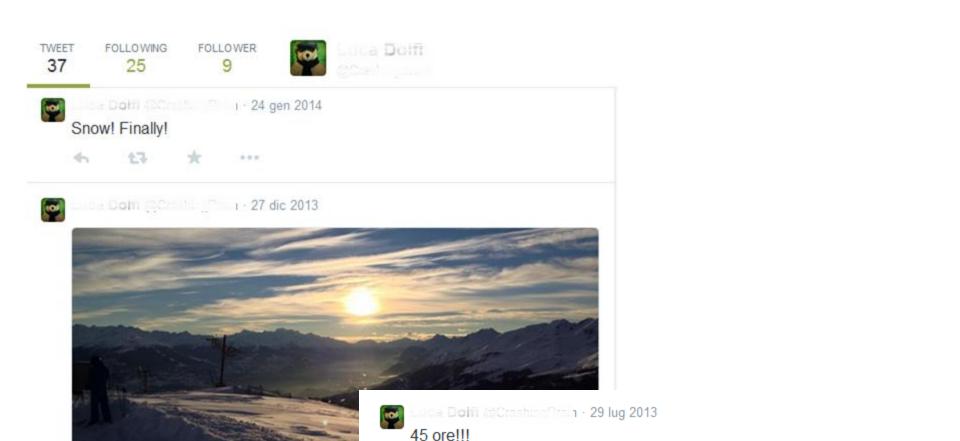


I estimated where Felix Baumgartner will land when he attempts the @redbullstratos record breaking freefall jump win.gs/LaMakNW



Luca Dolfi @CrashingBrain · 2 ott 2012

I'm in a group video chat with 5 others - Click to watch or join: tinychat.com





Luca Dolfi @CrashingBrain - 9 ott 2012

I estimated where Felix Baumgartner will land when he attempts the @redbullstratos record breaking freefall jump win.gs/www.skinw

h 17 * •

Luca Dolfi @CrashingBrain · 2 ott 2012

I'm in a group video chat with 5 others - Click to watch or join: tinychat.com

/lucapizziment



FOLLOWER 677





Stock Market Status @StokklinktStatus - 5 mar

Analysts suspect that Wal-Mart (WMT) was down -0.95% today because there loyalty usually resides else where.









Stock Market Status @StckMiddStatus - 5 mar

Paychex (PAYX) lost -1.09% today amid reports that of the light pollution) but it isn't impossible in dark suburbs.









Stock Market Status @StokklidStatus - 4 mar

Investors traded 2,218,708 shares of Viacom (VIAB) because I'm a miss my people's going to Florida ..









Stock Market Status @Stckl/lrktStatus · 4 mar

Investors traded 8,025,449 shares of Coca-Cola (KO) because MY COMPUTER'S STOPPED WORKING AND I CAN'T WATCH







6 位 * …



Stock Market Status @StckWrktStatus · 4 mar

MARKET REPORT: The S&P 500 lost -0.45% today because tomorrow it could be you.









Stock Market Status @StokklinktStatus · 5 mar

Analysts suspect that Wal-Mart (WMT) was down -0.95% today because there loyalty usually resides else where.









Stock Market Status @StckWirktStatus - 5 mar

Paychex (PAYX) lost -1.09% today amid reports that of the light pollution) but it isn't impossible in dark suburbs.









Stock Market Status @Stokklid Status - 4 mar

Investors traded 2,218,708 shares of Viacom (VIAB) because I'm a miss my people's going to Florida ..











Stock Market Status @StckWirktStatus · 4 mar

Investors traded 8,025,449 shares of Coca-Cola (KO) because MY COMPUTER'S STOPPED WORKING AND I CAN'T WATCH





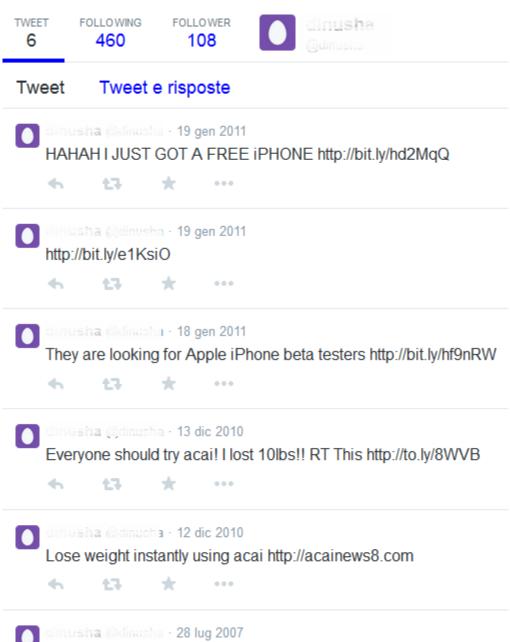




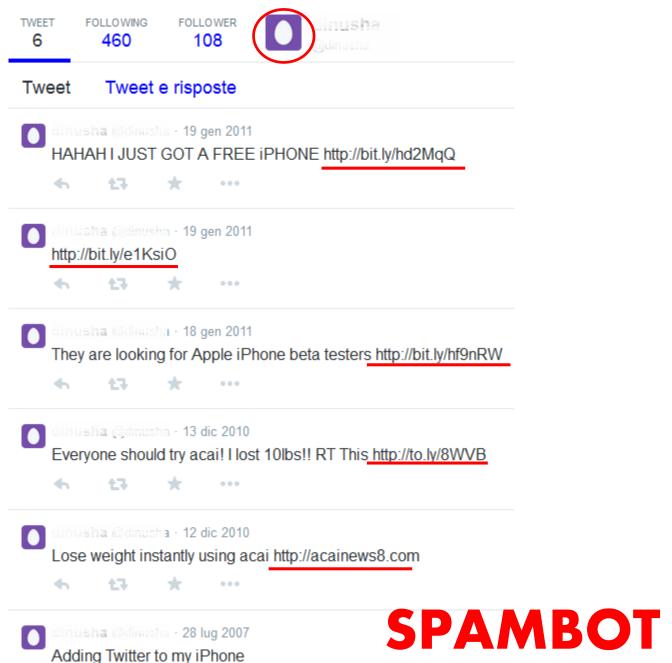


Stock Market Status @StckMirktStatus · 4 mar

MARKET REPORT: The S&P 500 lost -0.45% today because tomorrow it could be you.



Adding Twitter to my iPhone





FOLLOWING. 2.008

FOLLOWER 2.315





GeekGirlPenPals HQ @IGGPPCHQ ⋅ 5 mar

Poll Time! In honour of #WorldBookDay we want to know what your favourite book was as a child....









GeekGirlPenPals HQ @IGGPPCHQ · 5 mar

Photo: tearyplant: tearyplant: littlealienproducts: Cat Pot Plant small // \$45 plant blog plant blog tmblr.co/Zutles1f1-7v9









GeekGirlPenPals HQ @IGGPPCHQ - 5 mar

.@adventgeekgirl this is awesome, we love @smrtgrls! There are loads of science women in #IGGPPC so hopefully some will take part!



GeekGirlPenPals HQ @IGGPPCHQ · 5 mar

It's #WorldBookDay! Show us your costumes!



£3 1









Geel. Girl Pen Pals HQ ha ritwittato



AdventGeekGirl @adventgeekgirl · 5 mar

@IGGPPCHQ Love Science and Women? Check this out! amysmartgirls.com/2015/02/scienc... #iggppc #iggles









GeekGirlPenPals HQ @IGGPPCHQ - 4 mar

Photo: tmblr.co/Zutles1e-J9oc











FOLLOWING. 2.008

FOLLOWER 2.315





GeekGirlPenPals HQ @IGGPPCHQ - 5 mar

Poll Time! In honour of #WorldBookDay we want to know what your favourite book was as a child....









GeekGirlPenPals HQ @IGGPPCHQ · 5 mar

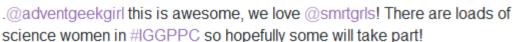
Photo: tearyplant: tearyplant: littlealienproducts: Cat Pot Plant small // \$45 plant blog plant blog tmblr.co/Zutles1f1-7v9







GeekGirlPenPals HQ @IGGPPCHQ - 5 mar





GeekGirlPenPals HQ @IGGPPCHQ · 5 mar

It's #WorldBookDay! Show us your costumes!







GirlPenPals HQ ha ritwittato



AdventGeekGirl @adventgeekgirl · 5 mar

@IGGPPCHQ Love Science and Women? Check this out! amysmartgirls.com/2015/02/scienc... #iggppc #iggles









GeekGirlPenPals HQ @IGGPPCHQ - 4 mar

Photo: tmblr.co/Zutles1e-J9oc









How to distinguish users?



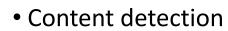
• Friend-to-Follower ratio







Access patterns

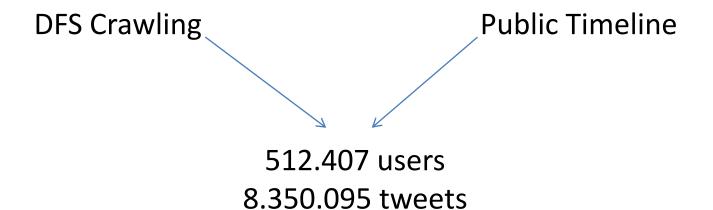






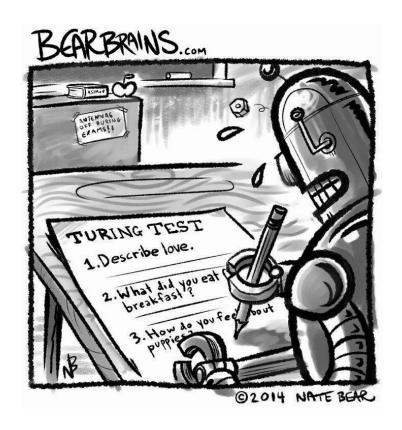
Account heuristics

Data Collection



Ground Truth Creation

Manual check of logs and homepages to collect 2000 users per category

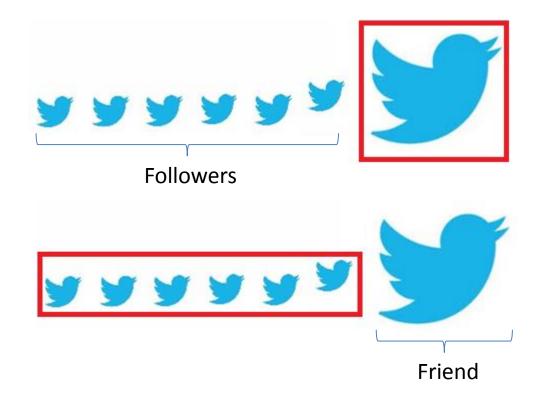


Ground Truth Creation

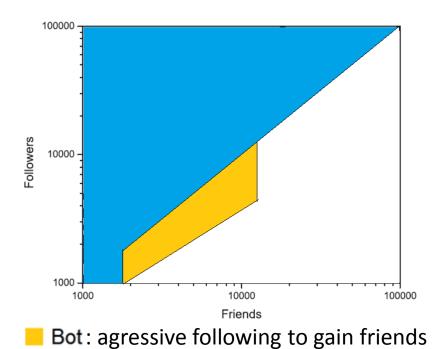
Intelligence Originality Consciousness Unintelligent Unoriginal Automation



Followers, Friends



Friend-to-Follower Ratio













Human: reciprocal relationships

Friend-to-Follower Ratio

Taste =
$$\frac{1}{n}$$
 sum_{friends}(Reputation(f))



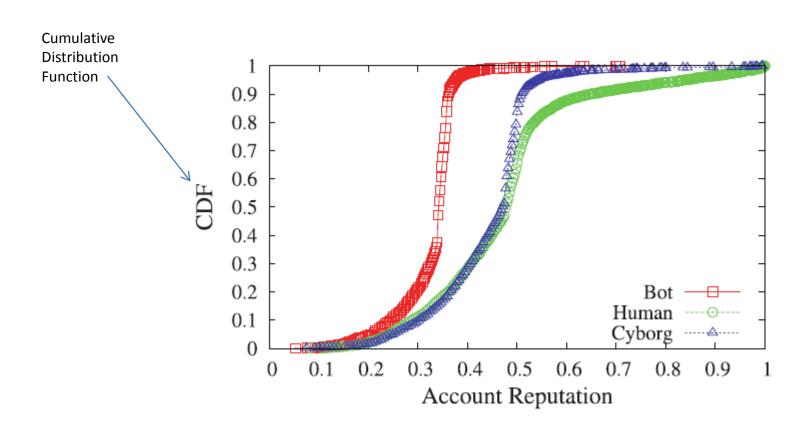








Friend-to-Follower Ratio





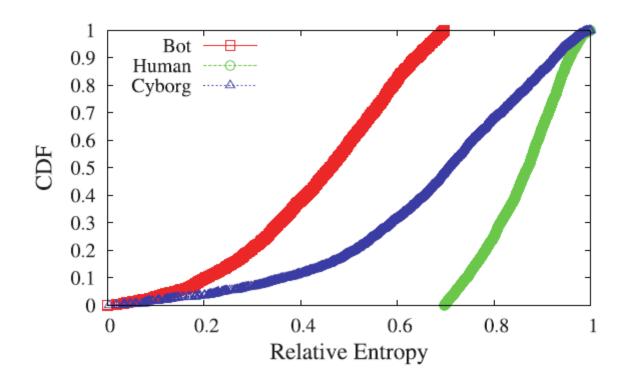








Tweeting Timing













Tweeting Timing

Human Behavior



Irregularities



Larger Coding



High Entropy











Tweeting Timing

Human Behavior



Irregularities



Larger Coding



High Entropy



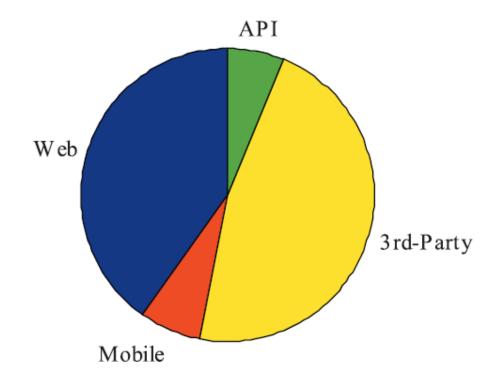








Access Patterns













Content Detection







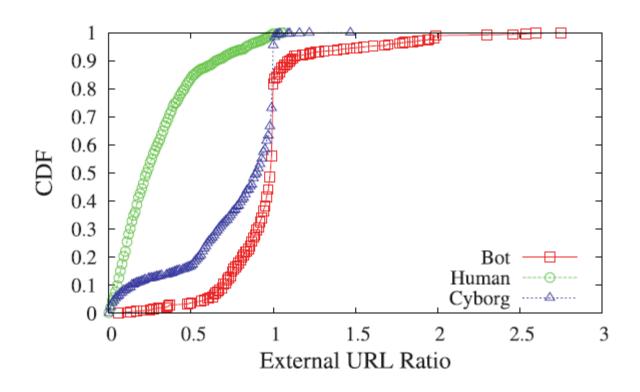








Content Detection













Malicious URLs: decide whether a message is spam based on blacklists















Spam detection:

Bayesian classifier decide whether a message is spam based on its content

$$P(spam|M) = \frac{P(M|spam)P(spam)}{P(M)}$$











Message is Spam

Author is a Bot











Message is Spam





What about non-malicious bots?





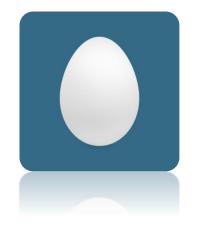






Account Heuristics







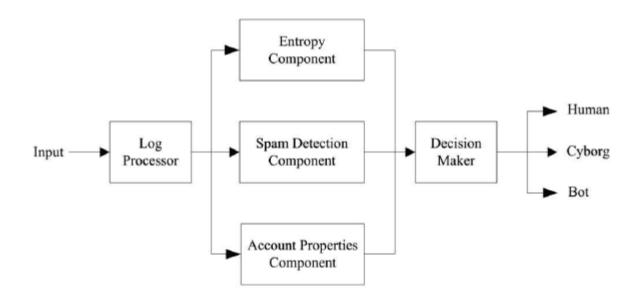














FOLLO WING 2.008

FOLLOWER 2.315





GeekGirlPenPals HQ @IGGPPCHQ ⋅ 5 mar

Poll Time! In honour of #WorldBookDay we want to know what your favourite book was as a child....







GeekGirlPenPals HQ @IGGPPCHQ · 5 mar

Photo: tearyplant: tearyplant: littlealienproducts: Cat Pot Plant small // \$45 plant blog plant blog tmblr.co/Zutles1f1-7v9







GeekGirlPenPals HQ @IGGPPCHQ - 5 mar

.@adventgeekgirl this is awesome, we love @smrtgrls! There are loads of science women in #IGGPPC so hopefully some will take part!



GeekGirlPenPals HQ @IGGPPCHQ · 5 mar

It's #WorldBookDay! Show us your costumes!



£3 1





43





Geel. Girl Pen Pals HQ ha ritwittato



AdventGeekGirl @adventgeekgirl · 5 mar

@IGGPPCHQ Love Science and Women? Check this out! amysmartgirls.com/2015/02/scienc... #iggppc #iggles









GeekGirlPenPals HQ @IGGPPCHQ - 4 mar

Photo: tmblr.co/Zutles1e-J9oc









Feature	Accuracy (%)
Entropy	82.8
URL Ratio	74.9
Automated Device %	71.0
Bayesian Spam Detection	69.5
Manual Device %	69.2
Registration Date	62.9
Mention Ratio	56.2
Link Safety	49.3
Hashtag Ratio	47.0
Followers to Friends Ratio	45.3
Account Verification	35.0

No single tree works well, a forest gives the best accuracy

		Classified				
		Human	Cyborg	Bot	Total	True Pos
	Human	1972	27	1	2000	98.6%
Actual	Cyborg	65	1833	102	2000	91.7%
	Bot	2	46	1952	2000	97.6%
				Avg	96.0%	

Minimal misclassification between human and bot

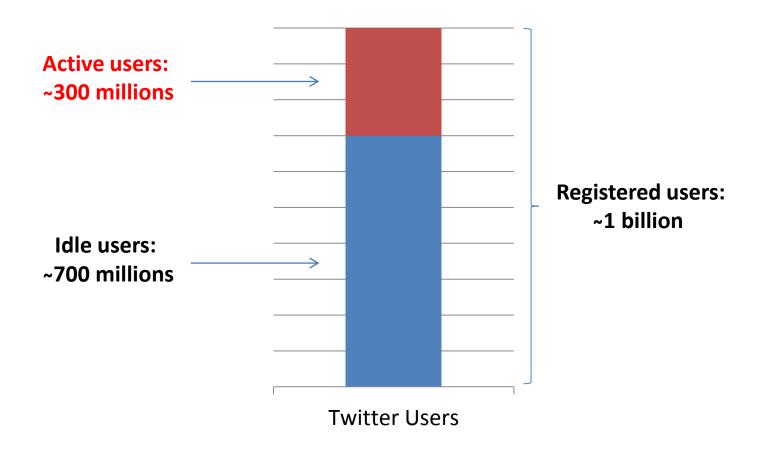
		Classified				
		Human	Cyborg	Bot	Total	True Pos
	Human	1972	27	1	2000	98.6%
Actual	Cyborg	65	1833	102	2000	91.7%
	Bot	2	46	1952	2000	97.6%
				Avg	96.0%	



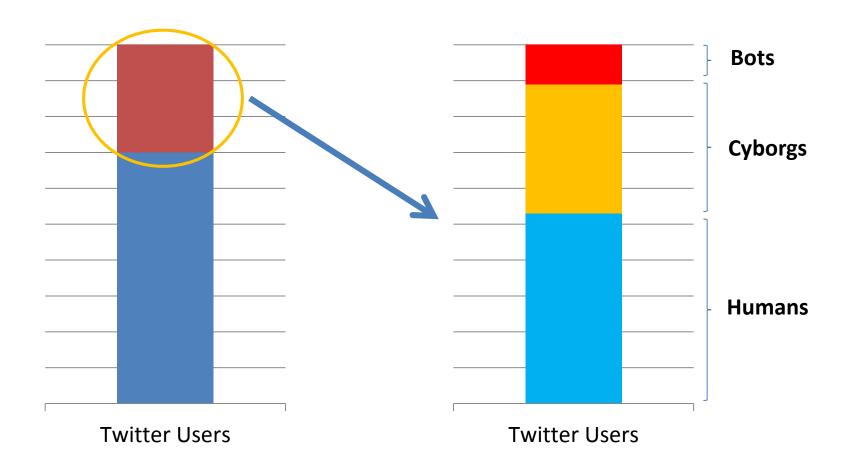
Training dataset (spam/nonspam) missing inputs:

- Bot tweets with no URLs left out
- Human and cyborg tweets with URLs left out

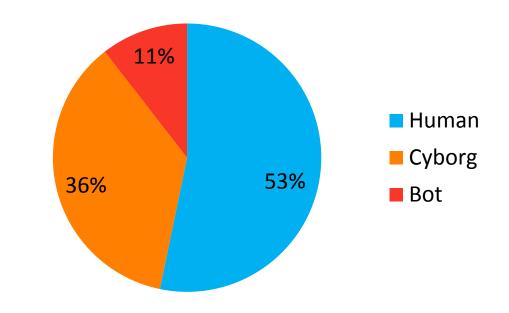
Who are the users?



Twitter composition



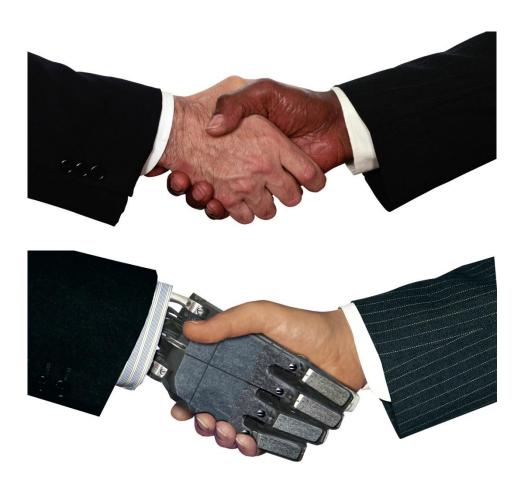
Twitter composition



Only one half is human!!!

CASA Paradigm

Social scripts are the same in both human-human and human-computer interaction



CASA Paradigm



Is the *communication quality* percieved differently between a bot and a human?

Communication Quality













Communication Quality







- Source Credibility
- Interpersonal Attraction
- Communication Competence
- Intent to Interact







Testing Perception



Communication Quality - Results

Means and Standard Deviations for the Two Conditions on the Dependent Variables.

Variable	Twitter	bot	Twitter human	
	M	(SD)	M	(SD)
Competence	5.12 _a	(1.07)	5.31 _a	(1.05)
Character	5.01 _a	(1.13)	5.27 _a	(1.03)
Caring	4.72 _a	(1.26)	4.95_{a}	(1.05)
Task attraction	3.40_{a}	(0.63)	3.58_{b}	(0.62)
Social Attraction	2.61 _a	(0.74)	$2.87_{\rm b}$	(0.63)
Communication competence	2.98 _a	(0.72)	3.09 _a	(0.73)
Intent to interact	2.62 _a	(1.08)	2.55 _a	(1.00)

Twitterbot got generally lower scores than the human, but it is still percieved as credible source

Communication Quality - Results

Means and Standard Deviations for the Two Conditions on the Dependent Variables.

Variable	Twitter	bot	Twitter human	
	M	(SD)	M	(SD)
Competence	5.12 _a	(1.07)	5.31 _a	(1.05)
Character	5.01 _a	(1.13)	5.27 _a	(1.03)
Caring	4.72_{a}	(1.26)	4.95_{a}	(1.05)
Task attraction	3.40_{a}	(0.63)	3.58_{b}	(0.62)
Social Attraction	2.61_{a}	(0.74)	2.87_{b}	(0.63)
Communication competence	2.98_{a}	(0.72)	3.09_{a}	(0.73)
Intent to interact	2.62_{a}	(1.08)	2.55_{a}	(1.00)

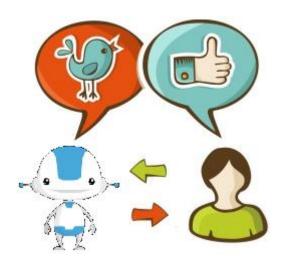


What if they rated the source and not to the author?

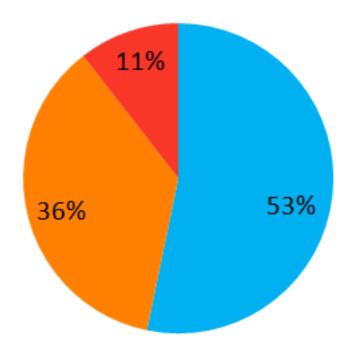
Communication Quality - Results

Followers can perceive twitterbots as credible attractive competent in communcation

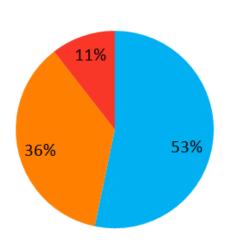
and can have <u>intention to interact</u>

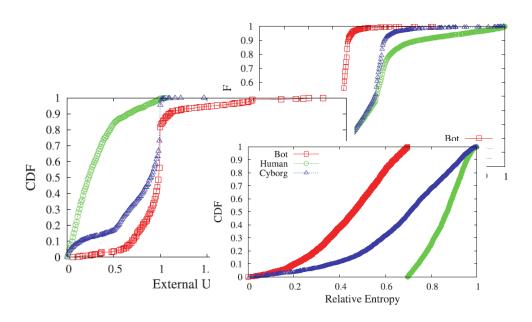


Summary

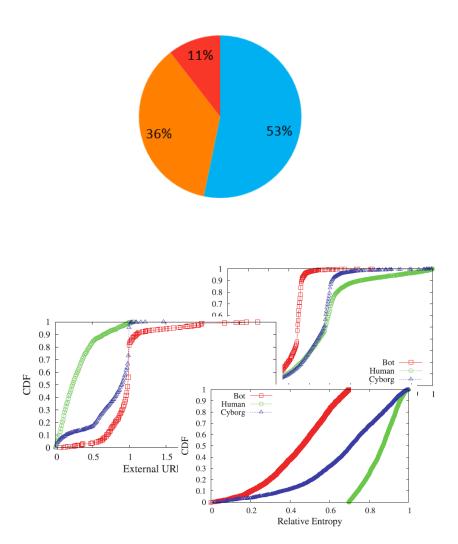


Summary





Summary





Thank you

