

CSC 574

Computer and Network Security

Social Networks Security

Alexandros Kapravelos
kapravelos@ncsu.edu

(Derived from slides by Gianluca Stringhini)

Social Networks Rely on Trust

Social network users share content from sources they trust.

Trust can come from two sources:



A screenshot of a Twitter profile for Steven Murdoch (@sjmurdoch). The profile picture shows a man with short brown hair. The header is green. Below the profile picture is a blue rounded rectangle with the word "Connections" in white. The profile statistics show 483 tweets, 120 following, and 1,613 followers. The bio is empty. A tweet from Steven Murdoch is visible at the bottom.

Connections

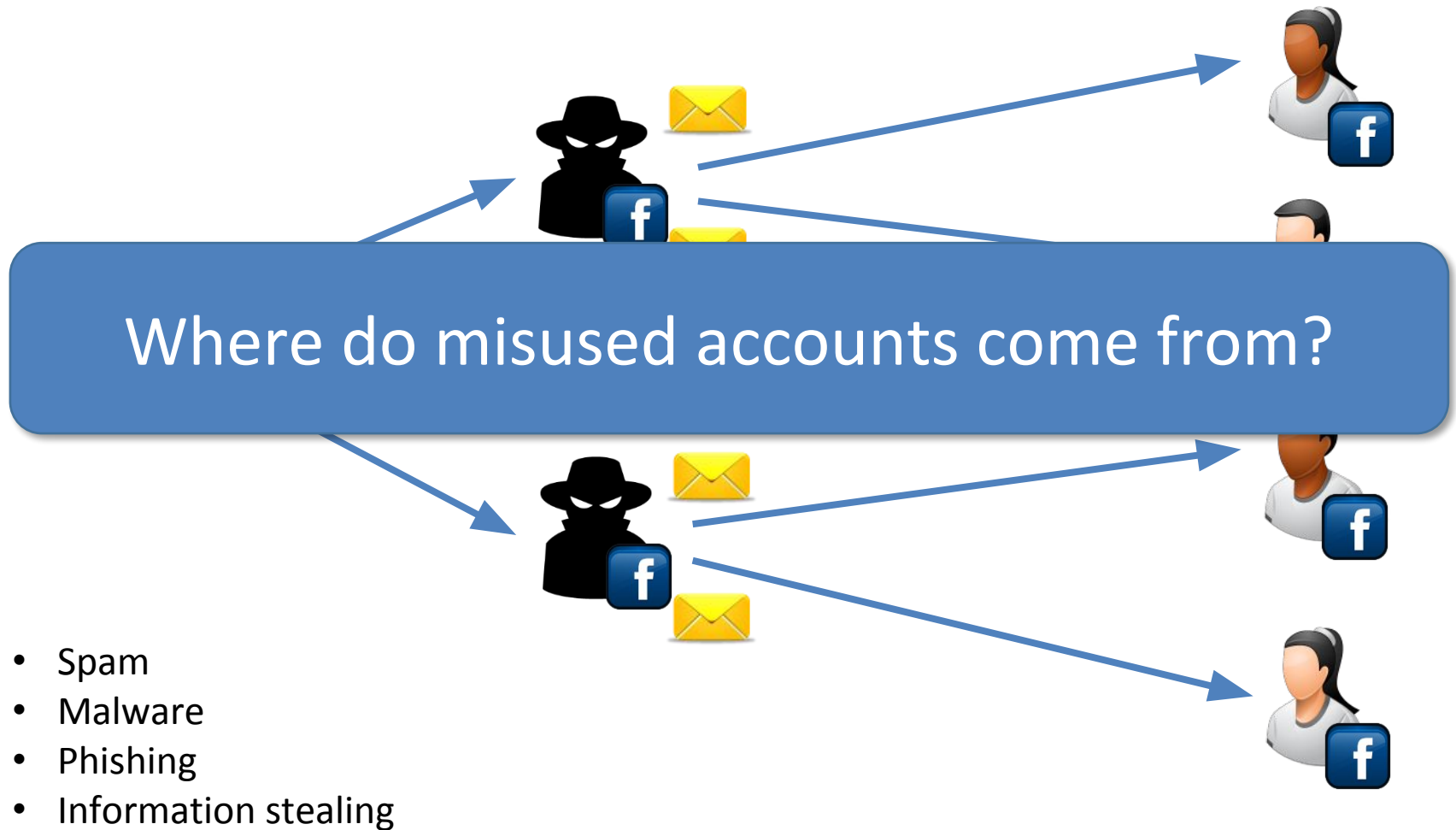


A screenshot showing two Twitter profiles. The top profile is for The Associated Press (@AP), with a header image of a man in a white shirt and red tie. The bottom profile is for Cambridge University (@Cambridge_Uni), with a header image of purple wisteria. A blue rounded rectangle with the word "Popularity" in white is overlaid on the Cambridge University profile. The Cambridge University profile statistics show 4,803 tweets, 651 following, and 178K followers. The bio mentions "News and Events from the University of Cambridge" and includes the website "cam.ac.uk".

Popularity

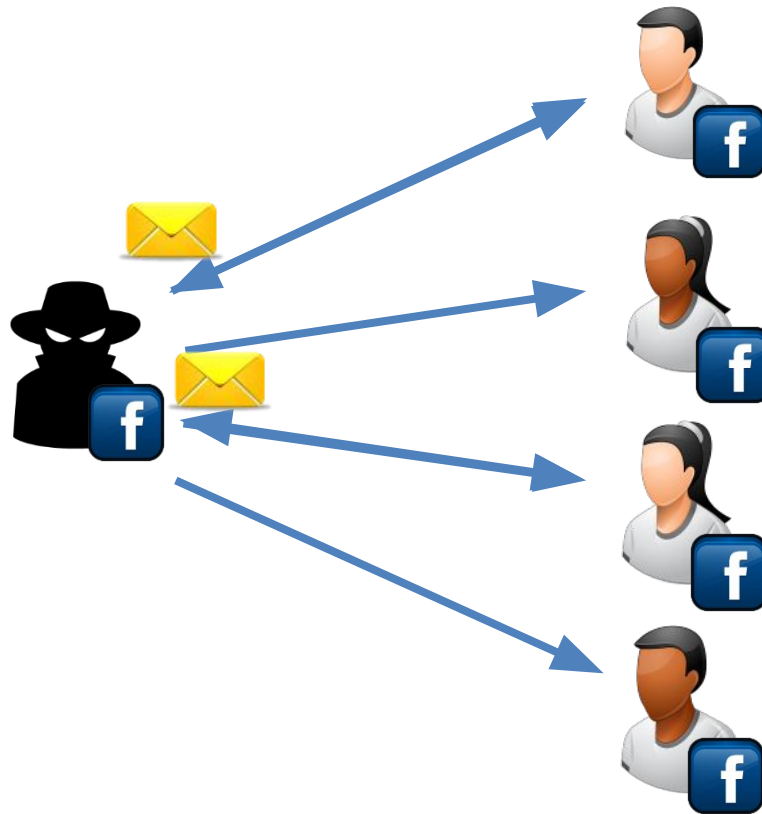
Cybercriminals use social networks too

Anatomy of Social Network Abuse



Fake Accounts

Created by cybercriminals or purchased on the black market



Leverage the *modus operandi* and the characteristics of fake accounts for detection

“Detecting Spammers on Social Networks”
[ACSAC 2010]

Very high accuracy, low false positives

Disadvantages of using fake accounts

- Need to build connections
- Can be deleted at no cost
- Are expensive to create

Without the victim's
trust, social network
abuse is ineffective

Manipulating User Trust

- Hijacking trust
- Building trust

Twitter CEO: 'We suck at dealing with abuse'

Dick Costolo says trolls are costing Twitter users

By [Nitasha Tiku](#) and [Casey Newton](#) on February 4, 2015 09:25 pm [✉ Email](#)

Hijacking someone
else's reputation

Compromised Accounts

 **McDonalds** 
@BurgerKing

We just got sold to McDonalds... already built

← Reply ↻ Retweet ★ Favorite ⋮ More

3,642 RETWEETS 572 FAVORITES

9:01 AM - 18 Feb 13

 **The Associated Press** 
@AP

Breaking: Two Explosions in the White House and Barack Obama is injured

← Reply ↻ Retweet ★ Favorite ⋮ More

3,063 RETWEETS 144 FAVORITES

12:07 PM - 23 Apr 13

Tweets



Giovanni Vigna @giovanni_v
@vineapp 2014 New Year c
coupon code WINTERSALE
@http://www.parka-ca.com

Expand



Giovanni Vigna @giovanni_vigna
@parkacat 2014 New Year discount \$20 of car
coupon code WINTERSALE at
@http://www.parka-ca.com

Expand

3 hours ago via web ☆ Favorite ↻ Retweet ← Reply

Mentioned in this Tweet



BarackObama Barack Obama 

This account is run by #Obama2012 campaign staff. Tweets from the President are signed -BO.

The AP Twitter Hack

Not only humans read tweets



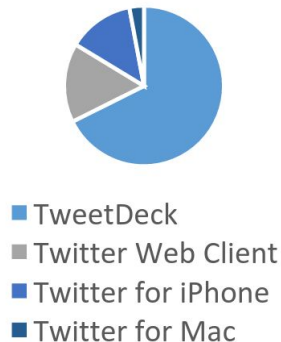
Insight

People develop habits when using social networks

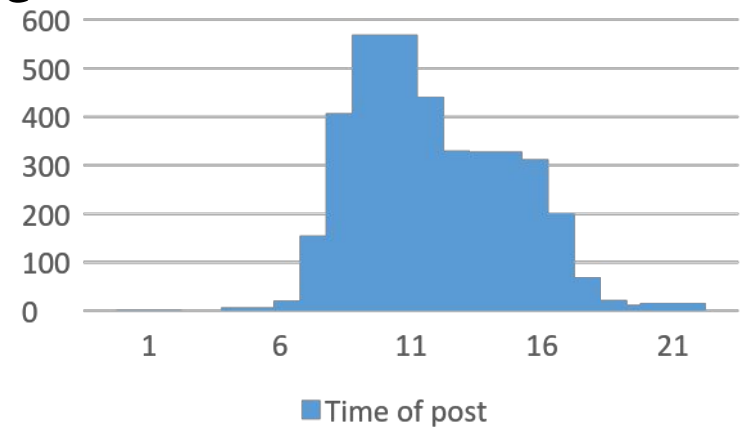
These habits hardly change over time



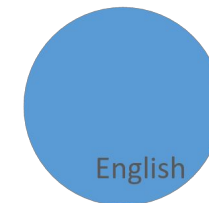
Client Used



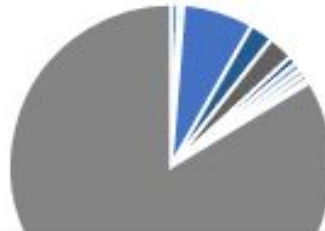
URL domain



Language



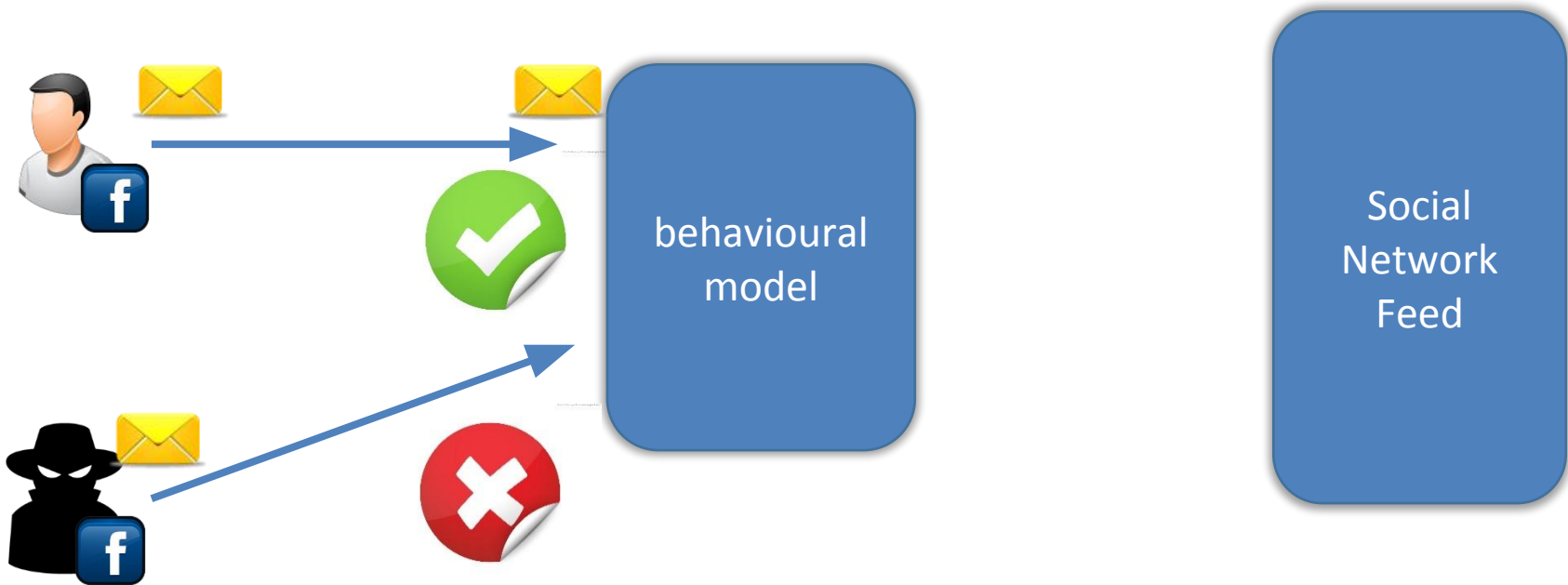
Hashtags



A message sent by an attacker will show differences from the typical behaviour

- #camresearch
- #extremesleepover
- #fresheradvice1660
- #robottakeover
- #underthemicroscope
- #picturethis

COMPA



COMPA maintains a behavioural model for Online Social Network accounts (using SVMs)

Any time a new message is posted, it is checked against the model

- Messages that match the profile are used to update the model
- Messages that do not match the profiles are discarded and flagged as an attack

COMPA in Action



COMPA would have blocked the Associated Press compromise

- Time: 0.00
- **Source: 0.99** (Web, usually SocialFlow)
- Hashtag: 0.06
- **Domain: 0.88** (No URL present)
- Mentions: 0.07
- Language: 0.00

Does it work in all cases?

 **Arturo Toledo**
@arturot



My first tweet from METROTWIT bit.ly/bazdFI
- I LIKE it already way more than TweetDeck.

Much cl... [Send Message](#)

← Reply

2
RETWEETS

10:26 PM



Robin Evans
@RobinEvansSC



My first message sent from HootSuite!

Compromises against regular users are large scale



Gianni V.
@RetroGee



We are excited to post our first picture to Instagram! Welcome to the world of Rich Kim NYC. [#RichLife...](#) instagram.com/p/jfvZt3yRgb/

← Reply ↻ Retweet ★ Favorite 📌 Pocket ... More 📎 HootSuite

7:16 PM - 22 Jan 2014



'Sent'
@Sentiient



Official first tweet from Tweetdeck

[@Joeyoe31](#) tweetdeck tutorial tl;dr version pls

← Reply ↻ Retweet ★ Favorite ... More 📎 HootSuite

4:37 AM - 23 Jan 2014

Detecting Large-scale Compromises



In COMPA, we detect large-scale campaigns by:

- Grouping “similar” messages
- Checking messages against their behavioural profile
- Detecting as a compromise groups with high rates of anomalous messages

COMPA: Evaluation

Twitter

- 1.4 billion tweets (10% of the total)
- 343,229 compromised accounts

Facebook

- 106 million status updates
- 11,499 compromised accounts

False positives between 3.6% and 0.5%, depending on the history of the user's activity

COMPA: Discussion

Large-scale compromises

An attacker would have to learn the behaviour of each of his victims → **UNFEASIBLE**

To achieve their goal, attackers' messages have to be different from what legitimate users usually post

Behavioural Modelling to Fight Spearphishing

Email users develop habits too

We can use these habits to detect attackers who compromise email accounts and send forged emails (**spearphishing**)

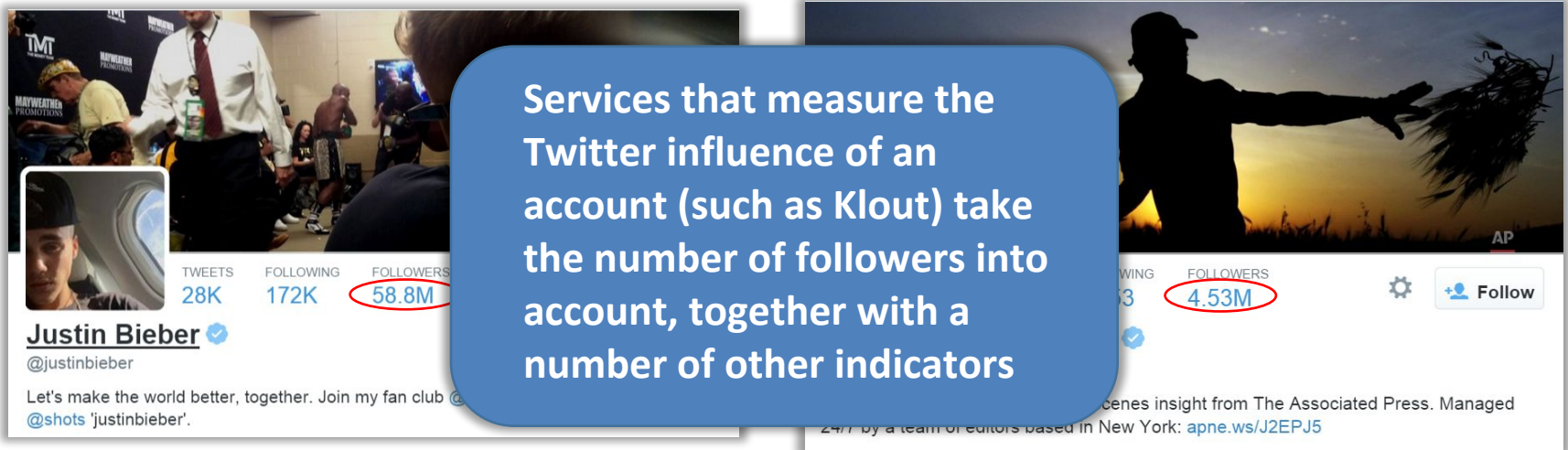
- Writing habits (stylometry)
- Composition habits
- Interaction habits

Able to detect between 90 and 98% attack emails

False positives between 8% and 1% depending on the length of the email history

Quickly building an online reputation

Twitter Followers = Perceived Reputation



Building a network of followers is difficult!

Shortcuts to Success

News > World news > Italy

Hot or bot? Italian professor casts doubt on politician's Twitter popularity

Comedian turned politician Beppe Grillo says academic is wrong to suggest more than half of his online followers are fake

Andrea Vogt in Milan

The Guardian, Sunday 22 July 2012 12:59 EDT

Does Louise Mensch have 80,000 fake followers on Twitter?

A Twitter stats website suggests that 97% of her followers are fake. (UPDATE: Someone's messing with Mensch.)

BY ALEX HERN | PUBLISHED 24 JULY 2012 15:04

Library 12 0 159



Recommend Share 0

Tweet 3

+1 0

Share

counts, 23% never

LATEST UPDATE: 14/08/2013 - AUSTRALIA - PRESIDENTIAL ELECTIONS - SOCIAL

Australia's opposition leader accused of fake followers

Fake Twitter account for newly appointed minister receives thousands of followers

ISTANBUL - Hürriyet Daily News

Beğen 0 Tweet 1 +1 1 Pinit Share



04 Jul 08 Jul

Can One Really Buy Followers?



The screenshot shows the NewFollow.info website. At the top, there is a blue header with a Twitter bird logo and the text "Enjoy The Unique Holidays !". The main logo "NewFollow.info" is in large, colorful letters. To the right of the logo are three buttons: "Terms Of Use", "Privacy Policy", and "Remove Account". Below the header, there is a section titled "Latest Members Get More Twitter Followers" with a row of small profile pictures. The main content area has a light blue background. It features a large, colorful banner with the text "MORE FOLLOWERS = MORE FRIENDS + MORE VISITERS" in purple. Below this, the banner says "Buy VIP NOW ONLY £ 9.99" in large, bold letters, with a "New Plan !" button. At the bottom of the banner, it says "Get Max Of 30,000 Followers". Below the banner, there is a text box with the following content:

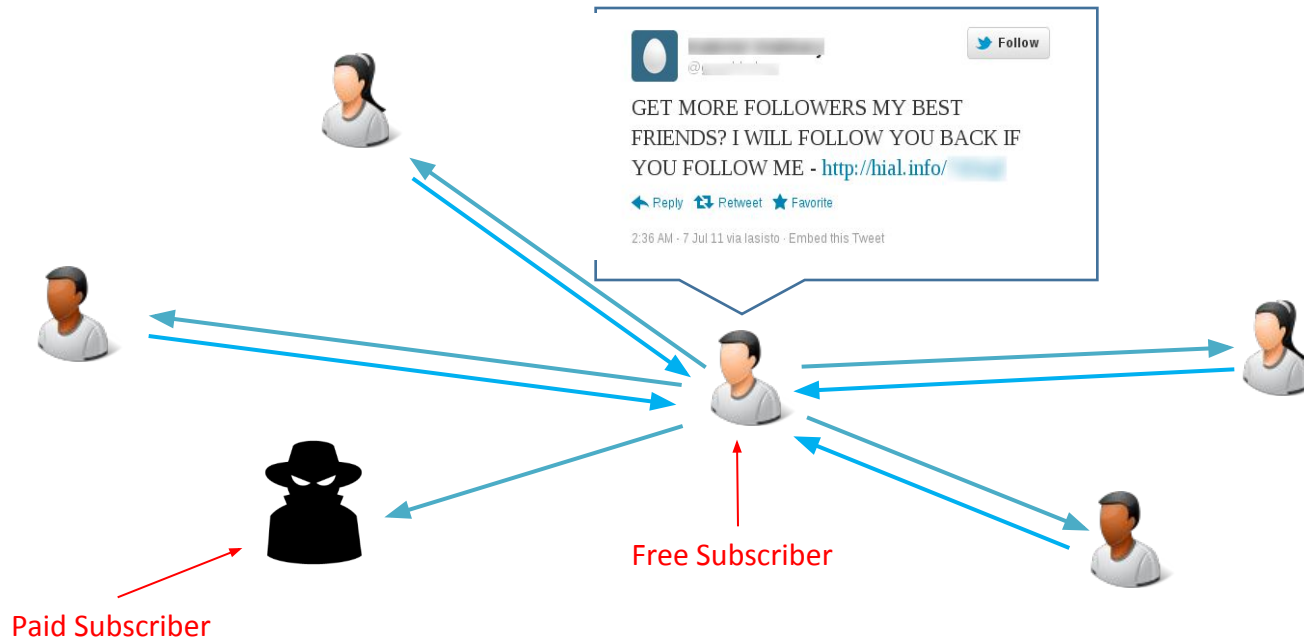
Get Followers on Twitter : Twitter is a really powerful tool for online social networking, promoting your online work and getting to know people in the web world around you. To maximize your use of Twitter, you want to gain followers on the site. The more people that you have following you there, the more likely it is that they will see what you're doing and be interested in your work around the web. **NewFollow.info** is simple way to help you can do to increase the number of people who follow you on Twitter.

Twitter Follower Markets

Different types of followers for sale

- Fake accounts (Sybils)
- Compromised accounts
- Pyramid schemes

Pyramid Markets



- Free subscribers → Victims
- Paid subscribers → Customers

Twitter's ToS forbids users to participate in Twitter Follower Markets

Active Twitter Follower Markets

Market (sorted by order of returned results)	\$ for 10K Followers	Pyramid?
Newfollow.info	\$216	YES
Bigfologo.com	\$91.99	YES
Bigfollow.net	\$70	YES
Intertwitter.com	\$65	NO (fake accounts)
Justfollowers.in	\$95	YES
Twriends.com	\$169	NO (fake accounts)
Socialwombat.com	\$49	NO (fake accounts)
Devumi.com	\$64	NO (fake accounts)
Hitfollow.info	\$214	YES
Plusfollower.info	\$214	YES
Buyactivefans.com	\$40	NO (fake accounts)

Market Sizes

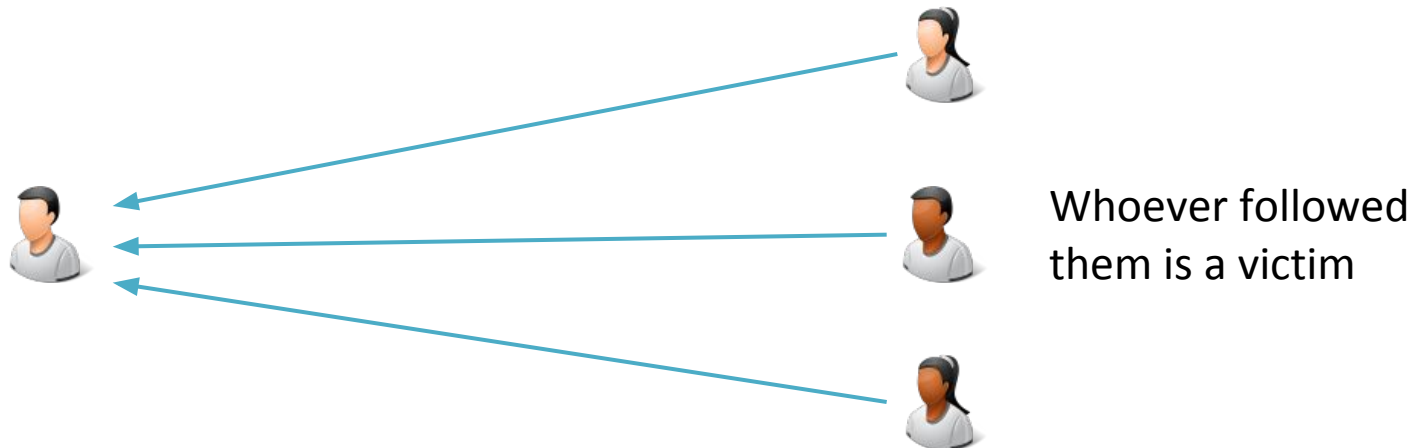
Let's look at tweets advertising the top five markets

10% of the all public tweets (3.3 billion tweets), collected over a period of four months

Market	Tweets	Victims
BigFollow	662,858	90,083
BigFolo	4,732,016	611,825
JustFollowers	302	257
NewFollow	77,865	38,341
InterTwitter	0	0
Total	5,473,041	740,506

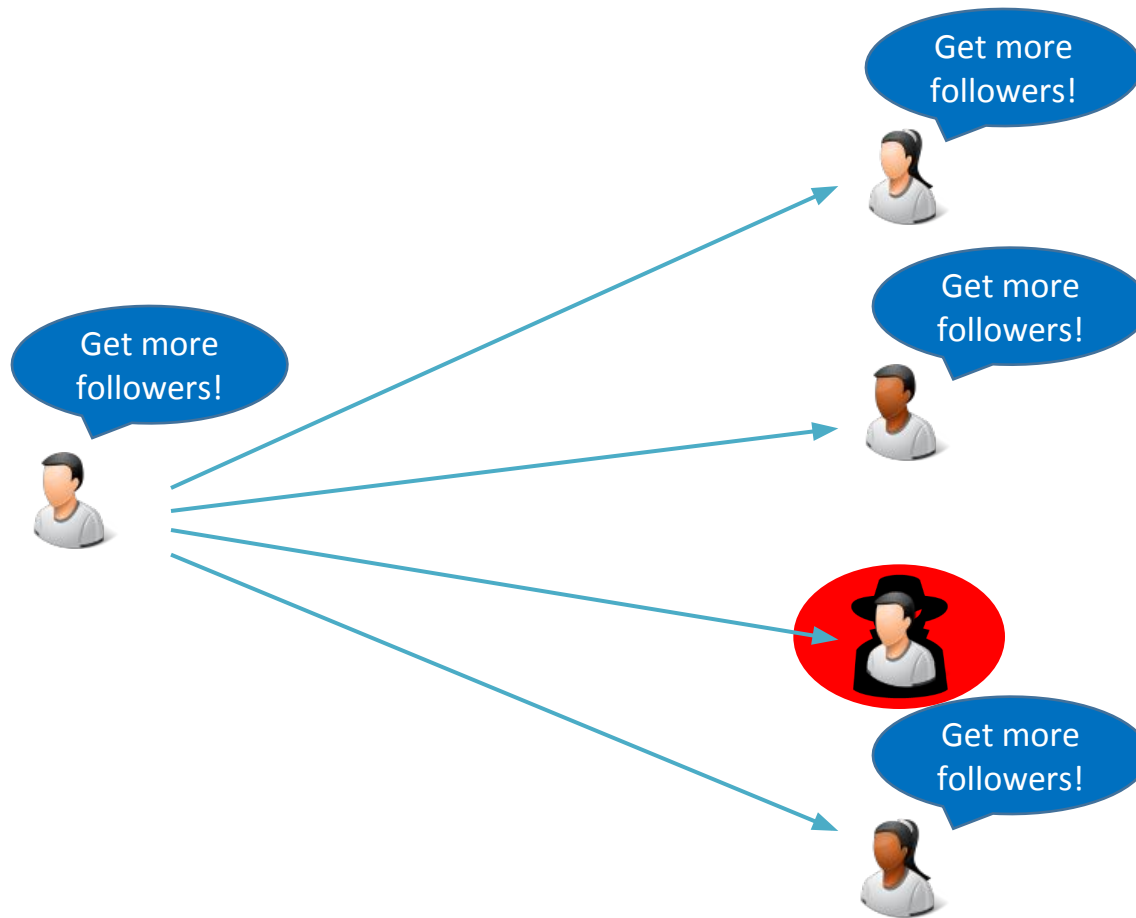
Detecting Market Victims

Purchased followers from the most popular five markets



In total, the authors identified 69,222 victims

Detecting Market Customers



Detecting Market Customers

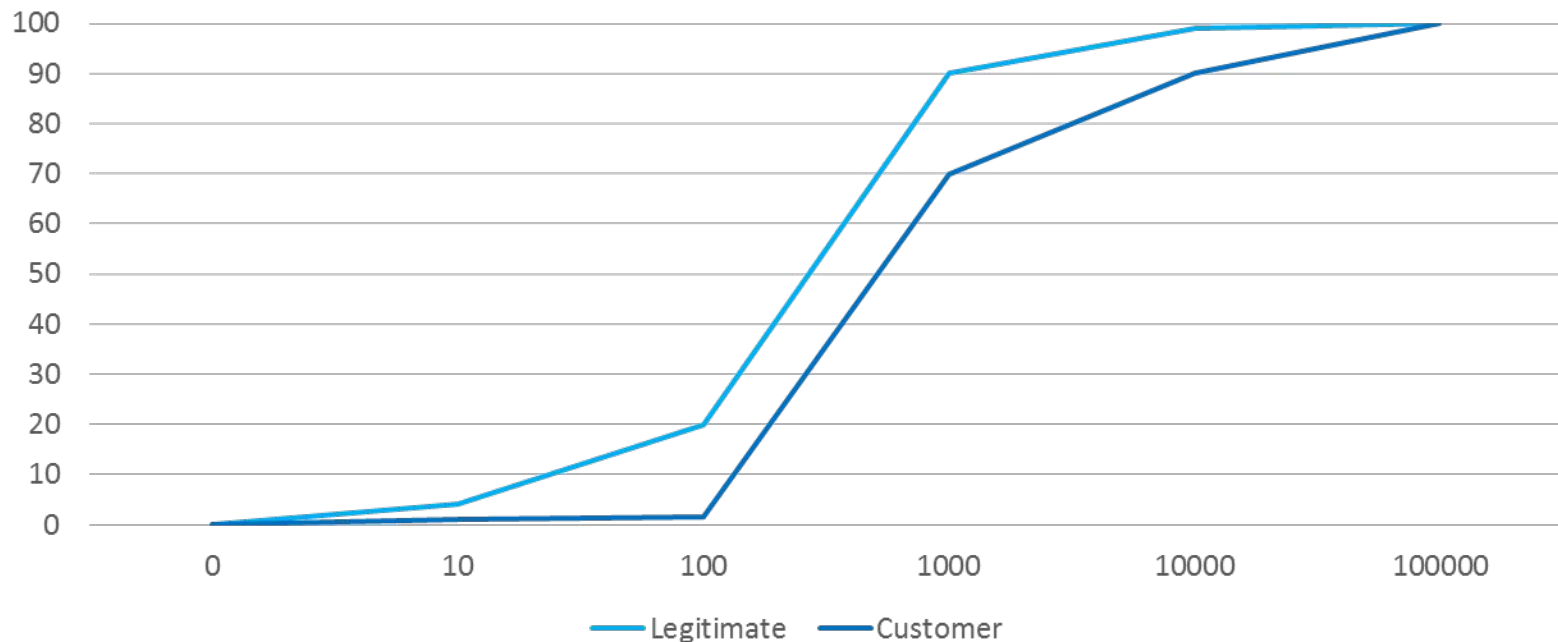
Signed up 180 newly-created accounts as market victims

Identified 2,909 market customers

Customer Characteristics

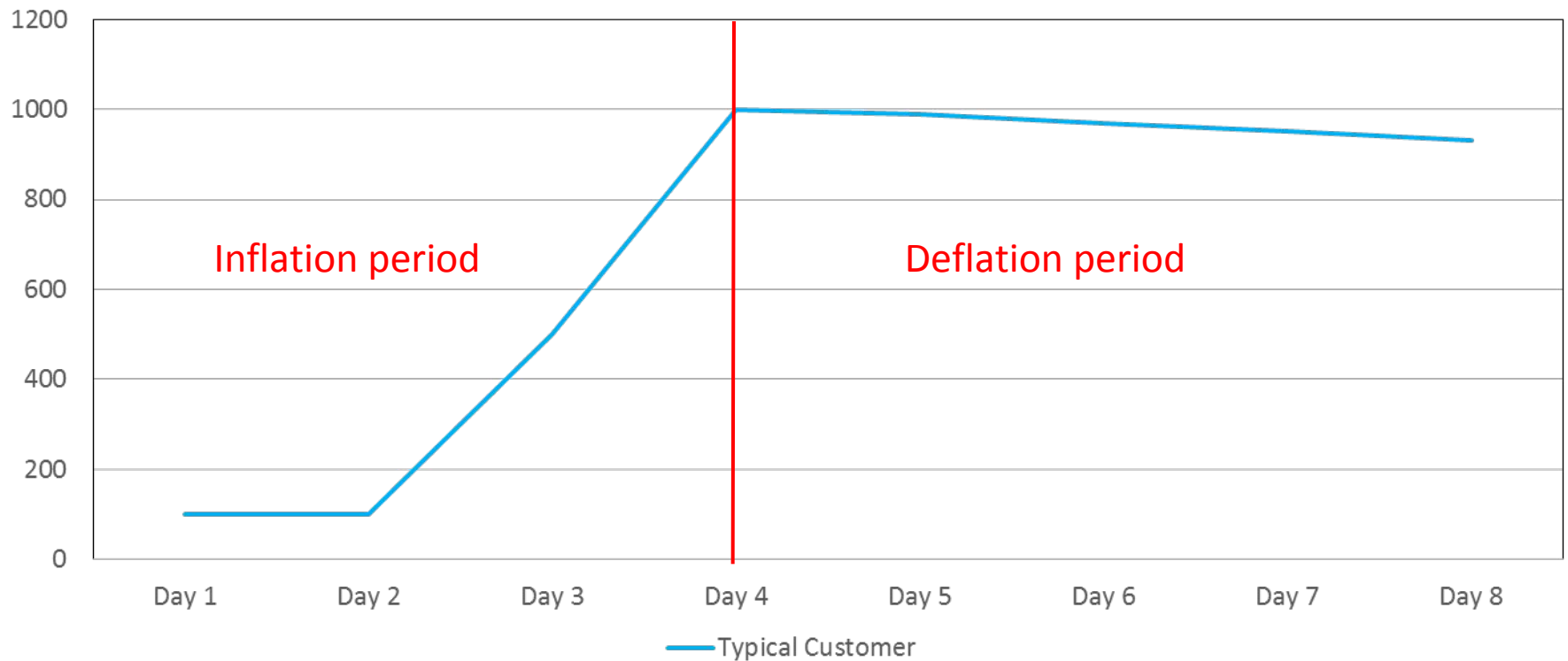
Compared our set of customers to a set of two million regular users picked at random

Customers have more followers than regular users



Customer Follower Dynamics

Followers over time



Customer Follower Dynamics

During an observation period of one week:

- Spike in Followers ≥ 50 over an hour:
50% Customers, 0.4% Regular
- Steady decrease of followers for ≥ 10 consecutive hours:
60% Customers, 0.05% Regular

Follower Dynamics Detection

Developed a classifier to detect customers in the wild

Ground truth: Set of 2,909 customers and 10,000 regular accounts (monitored for a week)

Classifier: Random Forests

10-fold cross validation: 98.4% true positive rate
0.02% false positive rate

Detecting Customers in the Wild

Monitored a set of two million regular accounts for two weeks

Detected 684 customers

- Observed only two million accounts
- Purchase needs to happen during our observation

Analysis of the Identified Customers

The detected accounts have the expected characteristics of customers

- They belong to wanna-be celebrities and small businesses
- They do not post interesting content

Buying followers does not help in becoming influential (median Klout 45)

- A customer with 103,000 followers → Klout score of 57

Twitter fails in detecting customers: 2 out of 684 were suspended