Twitter

Twitter Twitter: Twitter

Twitter twitter - Twitter Twitter

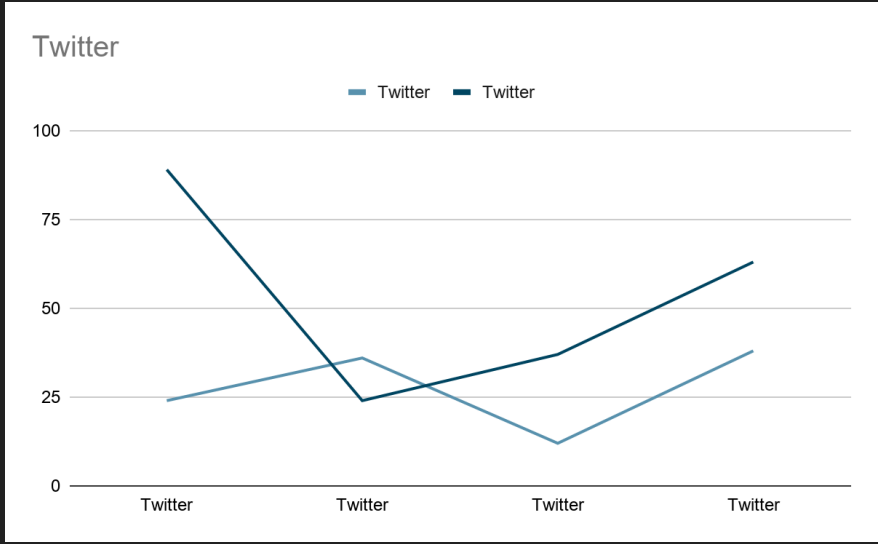
Twitter twitter twitter, twitter… twitter twitter

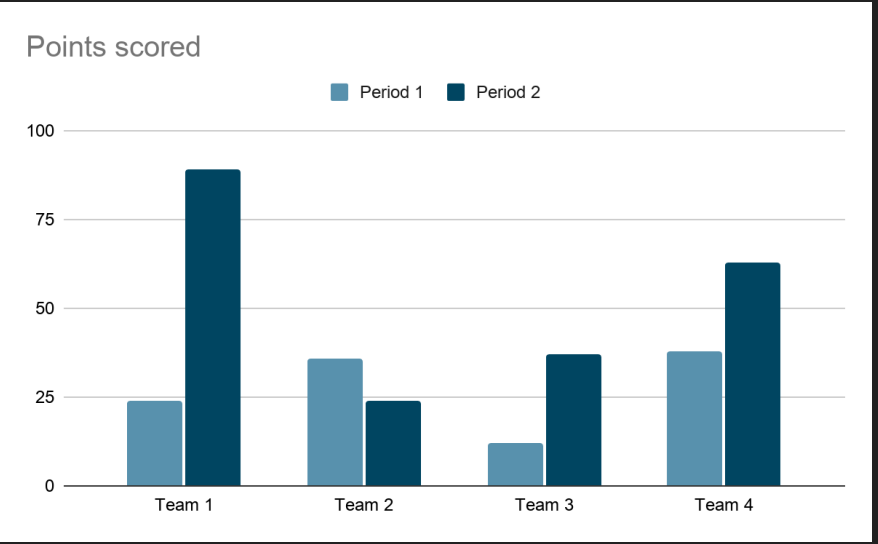
32.4% Twitter twitter twitter, twitter twitter twitter twitter

Twitter twitter! @twitter twitter -> twitter*

* Twitter twitter twitter, twitter

Twitter twitter twitter twitter





April Fools

Twitter

A Taste of Tweets: Reverse Engineering Twitter Spammers

Paper By Yang et al.
Presentation by Thomas Quig

Have you ever seen this before?

Figure 1: Shortened links with the #mcdonalds gift card hash tag

Elon Musk @patheuk

I'm giving 10 000 Bitcoic (BTC) to all community!

I left the post of director of Tesla, thank you all for your suppoot!

I decided to make the biggest crypto-giveaway in the world, for all my readers who use Bitcoin.

Participate in giveaway - spacex.plus

200 339 1,284

Promoted


twitter
Search
Home
Profile
Messages
Heidi Illems
@heidiillms
Business and Marketing Intelligence Specialist
Follow
Tweets
Favorites
Following
Followers
Lists
heidiillms Heidi Illems
White Paper: Social Media and the Future of Selling
http://bit.ly/lVUw9J
6 minutes ago
heidiillms Heidi Illems
White Paper: How Social Media Creates Empowered Customers http://bit.ly/mswd7k
22 minutes ago
heidiillms Heidi Illems
White Paper - Market Research and Social Media in the 21st Century: http://bit.ly/hudDTd
43 minutes ago
heidiillms Heidi Illems
White Paper - Using Social Media to Build Client Trust and Involvment http://bit.ly/eAS8fe
1 hour ago
About
1,338
Tweets
Following
About Help
Businesses

Ben Collins @oneunderscore__ · Aug 22

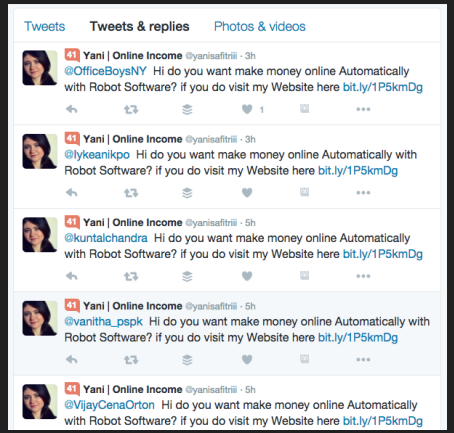
Simply incredible.

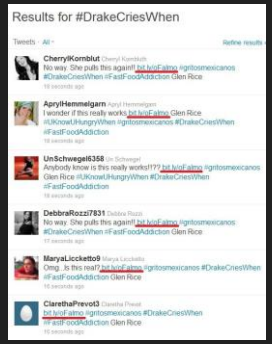
(h/t @josh_emerson)

> **Dr. Michael Stuart** @mstuartdds · 23h
> And NONE of it is RUSSIAN COLLUSION.
> 41 6 336
>
> **Debra Hirsch**
> @DebraHirsch1
> Follow
>
> Replying to @mstuartdds @SpryUte and 2 others
>
> Dr. Stuart, you've been dead for 3 years.
>
> 
>
> **Dr Michael Stuart's Obituary on Dallas Morning News**
> Read the Obituary and view the Guest Book, leave condolences or send flowers. | STUART, Dr. Michael Went home to meet his Lord and Savior Wednesday, Feb. 18, 2015, in Dallas, Texas. D...
> obits.dallasnews.com
>
> 5:30 PM - 21 Aug 2018
>
> **25** Retweets **258** Likes
>
> 24 25 258

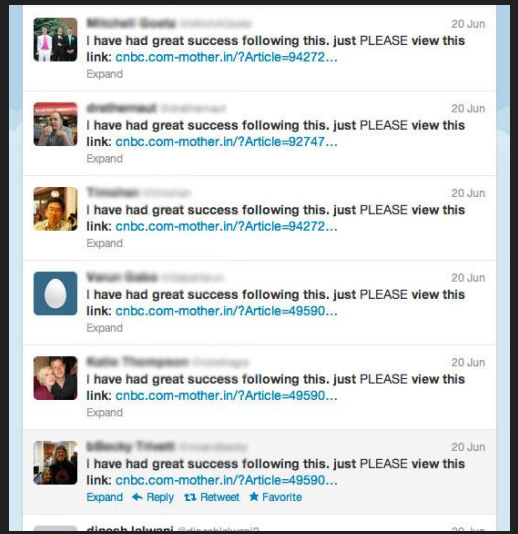
509 9.3K 28K


Results for #priiconf12
Tweets Top / All
Berenice Alegria @Bereniceptkff 23m
WTF? bit.ly/ld8LQN?=l9 #priiconf12 Charles Taylor Southwest is the epitome of excellence!
Denice Vaugeois @Deniceeigd 45m
Wow..wow..wow Simply f'n great #priiconf12 bit.ly/lu8DYL?7
Manda Amoroso @Mandaxiny 47m
#priiconf12 WTF? bit.ly/ld8LQN?=m2 is frickin awesome! 'Nough Said!
Neville Hobson @jangles 52m
Deck I used in my presentation at #priiconf12 in Dublin yesterday now on Slideshare slidesha.re/IusXct
View media
Manda Amoroso @Mandaxiny 1h
#priiconf12 #GoodThingsInTheWorld Moscow Wtf? - Where did this come from? bit.ly/ld8LQN?=v4
Rachel Honors @RachelHfb 1h
bit.ly/lu8DYL?3 Harry or Eleanor #priiconf12 Kidrauhl is Justincredible is hotsss!
Markham @markhamnolan 1h
@shanehegarty @sineadgleeson @lukeoneill We had a nice ol' chat about that yesterday on the panel at #priiconf12

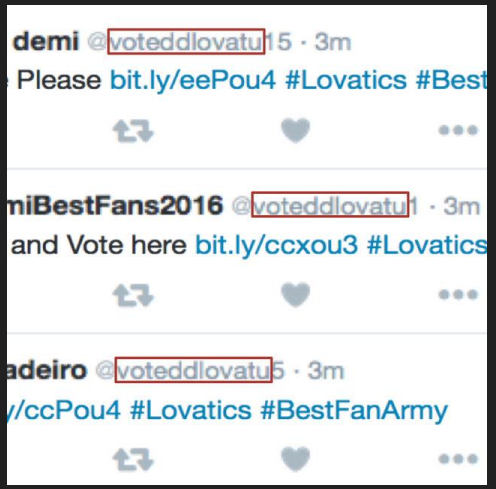

Tweets
Tweets & replies
Photos & videos
41 Yani | Online Income @yanisafitrili · 3h
@OfficeBoysNY Hi do you want make money online Automatically with Robot Software? if you do visit my Website here bit.ly/1P5kmDg
41 Yani | Online Income @yanisafitrili · 3h
@lykeanikpo Hi do you want make money online Automatically with Robot Software? if you do visit my Website here bit.ly/1P5kmDg
41 Yani | Online Income @yanisafitrili · 5h
@kuntalchandra Hi do you want make money online Automatically with Robot Software? if you do visit my Website here bit.ly/1P5kmDg
41 Yani | Online Income @yanisafitrili · 5h
@vanitha_pspk Hi do you want make money online Automatically with Robot Software? if you do visit my Website here bit.ly/1P5kmDg
41 Yani | Online Income @yanisafitrili · 5h
@VijayCenaOrton Hi do you want make money online Automatically with Robot Software? if you do visit my Website here bit.ly/1P5kmDg


Results for #DrakeCriesWhen


Acne Free Ebook @AcneFREEebook 1m
Hi! New Unique Weight Loss Method - Blue Heron Health News
dlvr.it/1vfKqm
Expand Reply Retweet Favorite
Body Wrap Tips @BodyWrapTips 1m
Hi! New Unique Weight Loss Method - Blue Heron Health News
dlvr.it/1vfKnW
Expand
healthtips_a2z @healthtips_a2z 1m
Rich Perry speech story diet myths headache treatments: Sometimes when we become so self-conscious about our...
is.gd/82fkyP
Expand

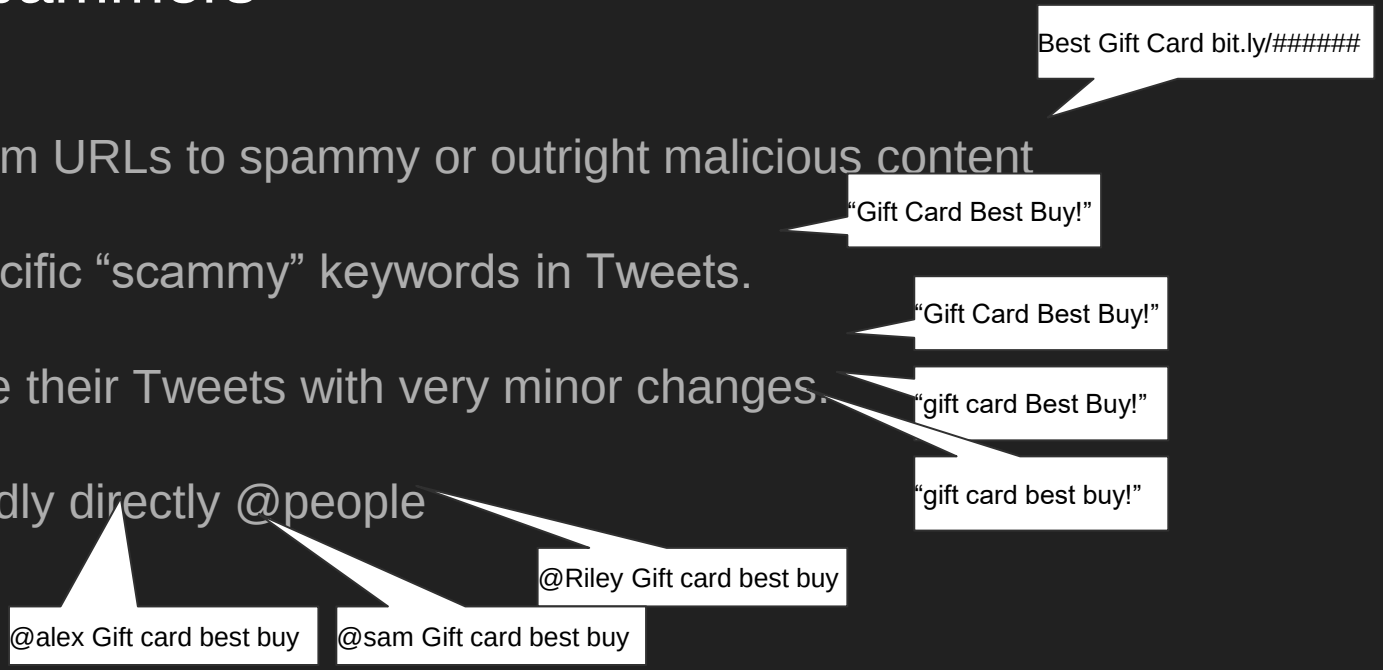

20 Jun
I have had great success following this. just PLEASE view this link: cnbc.com-mother.in/?Article=94272...
Expand
20 Jun
I have had great success following this. just PLEASE view this link: cnbc.com-mother.in/?Article=92747...
Expand
20 Jun
I have had great success following this. just PLEASE view this link: cnbc.com-mother.in/?Article=94272...
Expand
20 Jun
I have had great success following this. just PLEASE view this link: cnbc.com-mother.in/?Article=49590...
Expand
20 Jun
I have had great success following this. just PLEASE view this link: cnbc.com-mother.in/?Article=49590...
Expand
20 Jun
I have had great success following this. just PLEASE view this link: cnbc.com-mother.in/?Article=49590...
Expand Reply Retweet Favorite


twitter
aiww
Home
Profile
← Back to Home
Delete this saved search
Results for aiww
Tip: use operators for advanced search.
Tweets
Tweets with links
Tweets near you
People
superonlyone xizihupandedengdai
艺术界的精英意识和权威意识是任何行业里最重的， #aiww #jubao5mao #cn501 #cnjasmine #5mao
31 seconds ago
eiuielk 灯火阑珊
艺术界的精英意识和权威意识是任何行业里最重的， #aiww #jubao5mao #cn501 #cnjasmine #5mao
58 seconds ago
365t 室
@noooo0000 抱歉 我错了 窃以为有些不妥 抱歉 向爱帅#aiww 欠
1 minute ago
lidamink 李大明
再回头来看，流氓燕不急，这也是钱的原因。 #aiww #cn501 #cnjasmine #5mao #jubao5mao
2 minutes ago
rachmarsden Rachel Marsden
The final day of @aiww show at Tate Modern...it's been emotional...saw a naked protester try to make her mark
http://twitpic.com/4sdhz1
2 minutes ago
eiuielk 灯火阑珊
上了长安街，反而什么都不是，那是吸引媒体作秀的地方，是否被媒体围住的感觉很飘飘然？ #aiww #jubao5mao #cn501 #cnjasmine #5mao
5 minutes ago
eiuielk 灯火阑珊
不是见风使舵的人，你们已经是英雄。 #aiww #jubao5mao #cn501 #cnjasmine #5mao
6 minutes ago


demi @voteddlovatu15 · 3m
Please bit.ly/eePou4 #Lovatics #Best
miBestFans2016 @voteddlovatu1 · 3m
and Vote here bit.ly/ccxou3 #Lovatics
adeiro @voteddlovatu5 · 3m
/ccPou4 #Lovatics #BestFanArmy

Twitter Spammers

1. Post spam URLs to spammy or outright malicious content
2. Post specific “scammy” keywords in Tweets.
3. Duplicate their Tweets with very minor changes.
4. Repeatedly directly @people



Solution

- Twitter Policy

Platform manipulation and spam policy

Overview

September 2020

You may not use Twitter's services in a manner intended to artificially amplify or suppress information or engage in behavior that manipulates or disrupts people's experience on Twitter.

We want Twitter to be a place where people can make human connections, find reliable information, and express themselves freely and safely. To make that possible, we do not allow spam or other types of platform manipulation. We define platform manipulation as using Twitter to engage in bulk, aggressive, or deceptive activity that misleads others and/or disrupts their experience.

Platform manipulation can take many forms and our rules are intended to address a wide range of prohibited behavior, including:

Misuse of Twitter product features

You can't misuse Twitter product features to disrupt others' experience. This includes:

Tweets and Direct Messages

- sending bulk, aggressive, high-volume unsolicited replies, mentions, or Direct Messages;
- posting and deleting the same content repeatedly;
- repeatedly posting identical or nearly identical Tweets, or repeatedly sending identical Direct Messages; and
- repeatedly posting Tweets or sending Direct Messages consisting of links shared without commentary, so that this comprises the bulk of your Tweet/Direct Message activity.

Following

- "follow churn" – following and then unfollowing large numbers of accounts in an effort to inflate one's own follower count;
- indiscriminate following – following and/or unfollowing a large number of unrelated accounts in a short time period, particularly by automated means; and
- duplicating another account's followers, particularly using automation.

What is in violation of this policy?

Under this policy we prohibit a range of behaviors in the following areas:

Accounts and identity

You can't mislead others on Twitter by operating fake accounts. This includes using misleading account information to engage in spamming, abusive, or disruptive behavior. Some of the factors that we take into consideration include:

- use of stock or stolen profile photos, particularly those depicting other people;
- use of stolen or copied profile bios; and
- use of intentionally misleading profile information, including profile location.

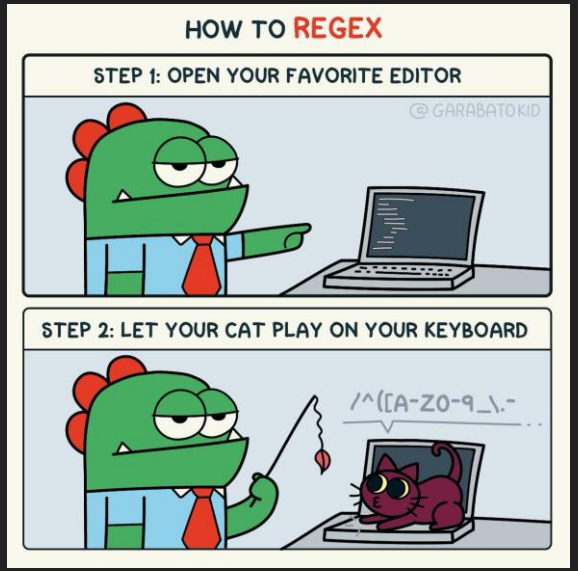
You can't artificially amplify or disrupt conversations through the use of multiple accounts or by coordinating with others to violate the Twitter Rules. This includes:

Q: Do Spammers Care ?

A: lol no

How do we enforce policy?

- Spam Filters!
  - Regex
  - NLP
    - Machine Learning
  - Manual Review



- “How Twitter is fighting spam and malicious automation” (June 2018)
  - https://blog.twitter.com/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html

- Not great, a very fragile system

Discussion Question

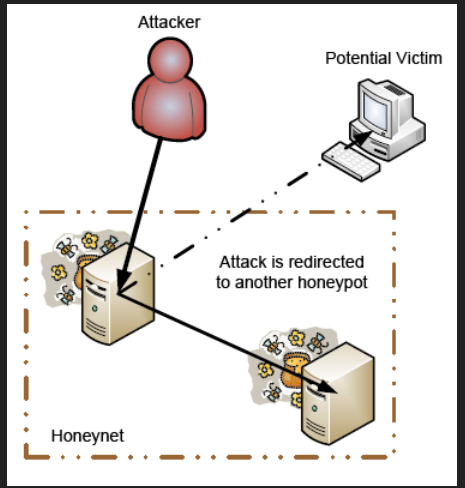
How can spammers get around spam filters? Can they avoid traps?

How to improve the spam filter?

Honeypots

- Fake accounts meant to catch attract attackers
- Run by software
- Isolated and Non-impactful
  - If they get pwned nothing really is lost

- In the context of OSN’s (Online Social Networks)
  - Have spammers contact you.
  - Scam attempts
  - Success measured in CR (Capture Rate)



Taste of Tweets - Goals

- Problem Statement
  - Gain insight into spammer's attack tastes and general characteristics
  - Deepen an understanding of Social Honeypot measurement
  - Develop new guidelines of sampling to get more spam accounts.

- Discover Spammer's Tastes
  - Many... many honeypots
- Active scanning of network to find spammers (based on Spammer targeting)
  - Utilize characteristics of the spammers to find them before they spam
  - Who they follow
  - What they tweet about
  - Get them banned before they can do much harm

Social Honeypots - Parameters

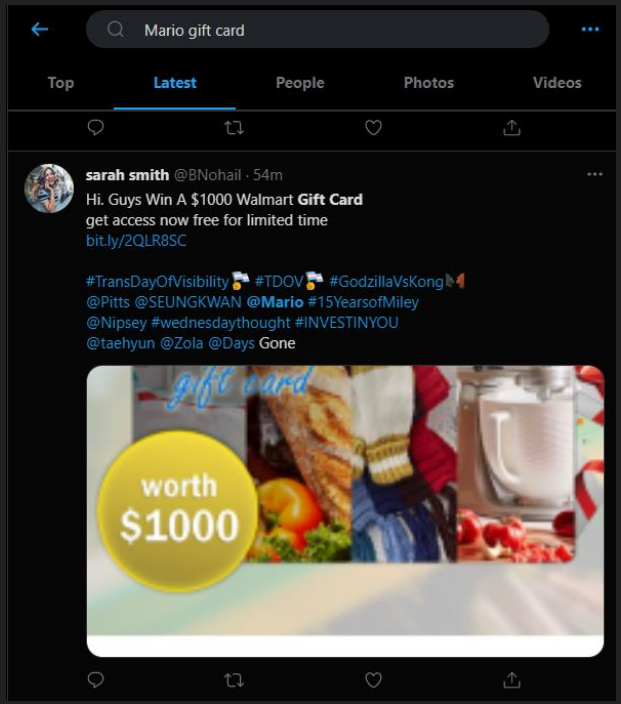
- Tweet Behavior
  - Keywords
  - Frequency
  - Topics
- Follow Behavior
  - Field
  - Famous
  - Verified *
- App Behavior
  - What third party apps are used
    - TweetDeck
    - Instagram

Table 1: 96 "benchmark" social honeypots with 24 fine-grained social behavior patterns

Index	Category	Sub-Category	Pattern	Index	Category	Sub-Category	Pattern
1-5	Tweet	Frequency	Once per day	6-10	Tweet	Frequency	Twice per day
11-15	Tweet	Frequency	Once per hour	16-20	Tweet	Keywords	Trending Topics
21-25	Tweet	Keywords	Arbitrary Hashtags	26-30	Tweet	Keywords	Current Affairs
31-35	Tweet	Keywords	Bait Words	36-40	Tweet	Keywords	No Hashtags
41-45	Tweet	Topic (Twice per day)	Entertainment	46-50	Tweet	Topic (Twice per day)	Expertise
51-55	Tweet	Topic (Twice per day)	Sports	56-60	Tweet	Topic (Twice per day)	Economics
61-62	Tweet	Topic (Once per hour)	Entertainment	63-64	Tweet	Topic (Once per hour)	Expertise
65-66	Tweet	Topic (Once per hour)	Sports	67-68	Tweet	Topic (Once per hour)	Economics
69-70	Follow	Two accounts per day	Entertainment	71-72	Follow	Two accounts per day	Expertise
73-74	Follow	Two accounts per day	Sports	75-76	Follow	Two accounts per day	Economics
77-81	App	NA	Twitpic	82-86	App	NA	Instagr
87-91	App	NA	Twiends	92-96	Default	NA	NA

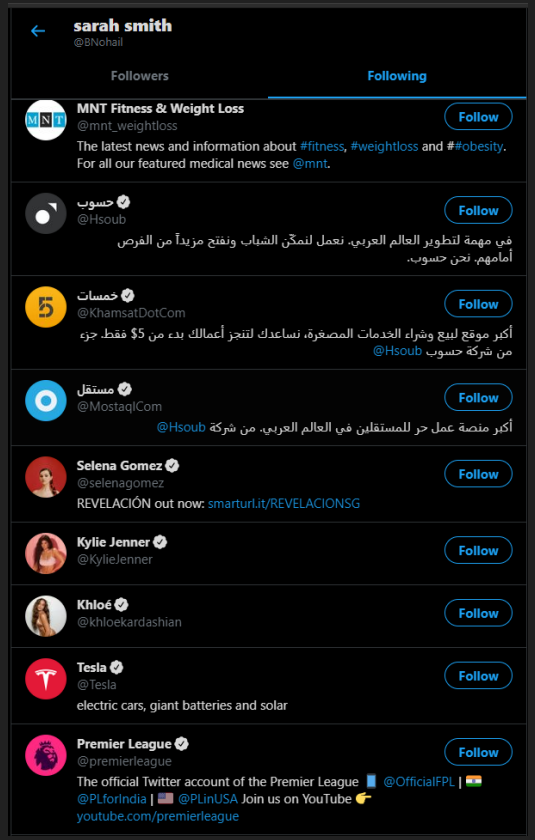
Social Honeypots - Tweet Behavior

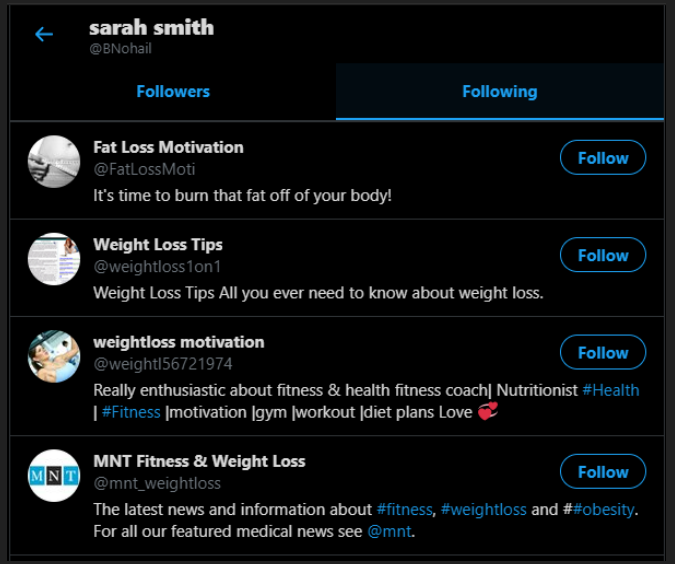
- Keywords
  - Trending Topics
  - Hashtags (Arbitrary and None)
  - Bait Words
- Topics
  - Entertainment
    - TV, Movies, Games, Books Etc
  - Expertise
    - IT, Tech, Science, Fashion… Household…
  - Sports
    - Sports…
  - Economics
    - Business, finance, charity



Social Honeypots - Follow Behavior

- Follow famous people
- Follow based on topics
- Verified accounts
- Spammers do this too!





Social Honeypots - App Behavior

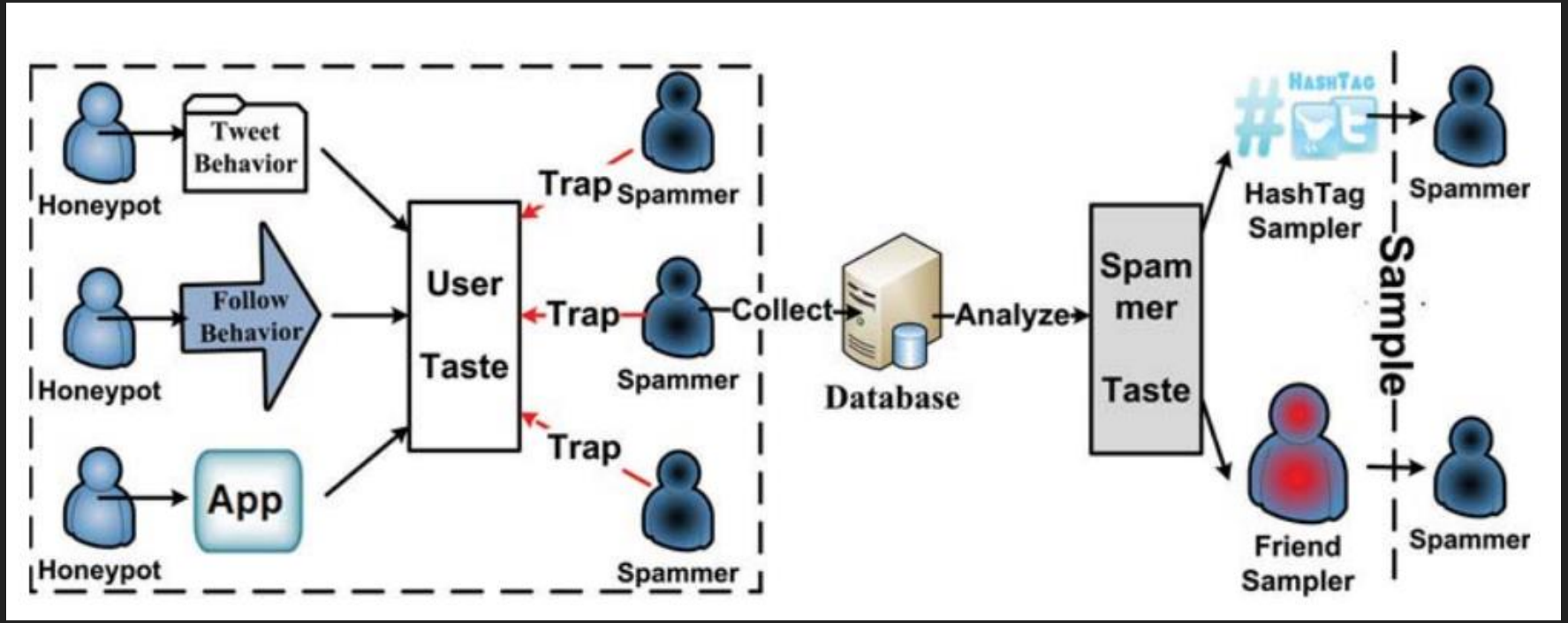
- Instagram / Facebook integration,
- 3rd party Twitter support apps (tweetdeck)
- Games

Questions?

Discussion Q: Part of bot spam relies on publicly visible following / follower lists. Should there be a privacy setting for this on the personal level? Should twitter change this overall to prevent bots follow sampling?

Social Honeypots

Social Honeypots - System

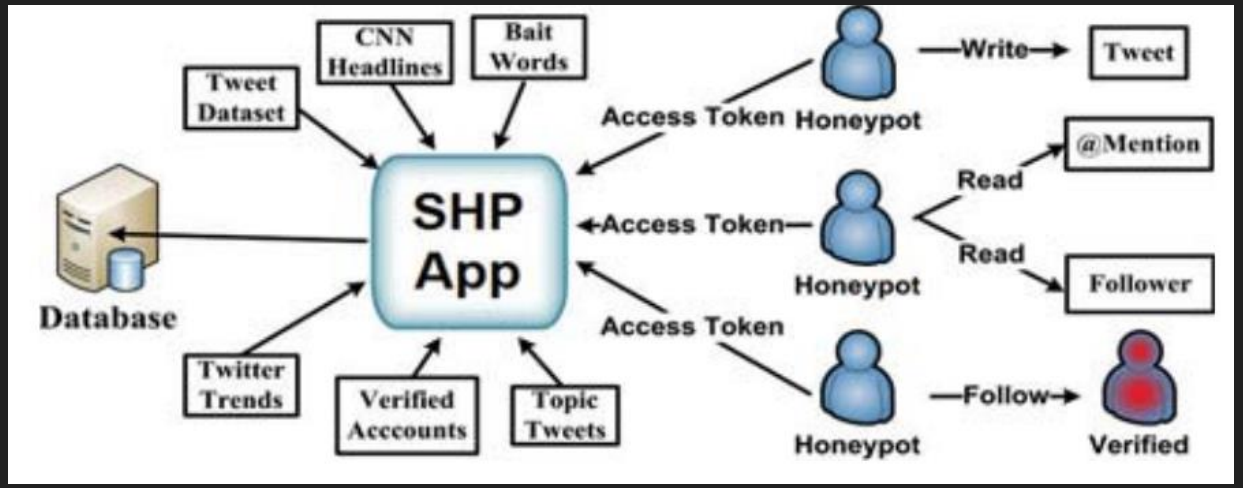


Social Honeypots - System Design

- 96 Honeypots
  - Vary each parameter
- Centralized application
  - Honeypots send their access tokens to the SHP app

Table 1: 96 "benchmark" social honeypots with 24 fine-grained social behavior patterns

Index	Category	Sub-Category	Pattern	Index	Category	Sub-Category	Pattern
1-5	Tweet	Frequency	Once per day	6-10	Tweet	Frequency	Twice per day
11-15	Tweet	Frequency	Once per hour	16-20	Tweet	Keywords	Trending Topics
21-25	Tweet	Keywords	Arbitrary Hashtags	26-30	Tweet	Keywords	Current Affairs
31-35	Tweet	Keywords	Bait Words	36-40	Tweet	Keywords	No Hashtags
41-45	Tweet	Topic (Twice per day)	Entertainment	46-50	Tweet	Topic (Twice per day)	Expertise
51-55	Tweet	Topic (Twice per day)	Sports	56-60	Tweet	Topic (Twice per day)	Economics
61-62	Tweet	Topic (Once per hour)	Entertainment	63-64	Tweet	Topic (Once per hour)	Expertise
65-66	Tweet	Topic (Once per hour)	Sports	67-68	Tweet	Topic (Once per hour)	Economics
69-70	Follow	Two accounts per day	Entertainment	71-72	Follow	Two accounts per day	Expertise
73-74	Follow	Two accounts per day	Sports	75-76	Follow	Two accounts per day	Economics
77-81	App	NA	Twitpic	82-86	App	NA	Instagr
87-91	App	NA	Twiends	92-96	Default	NA	NA



Social Honeypot Campaign - Aggregate Results

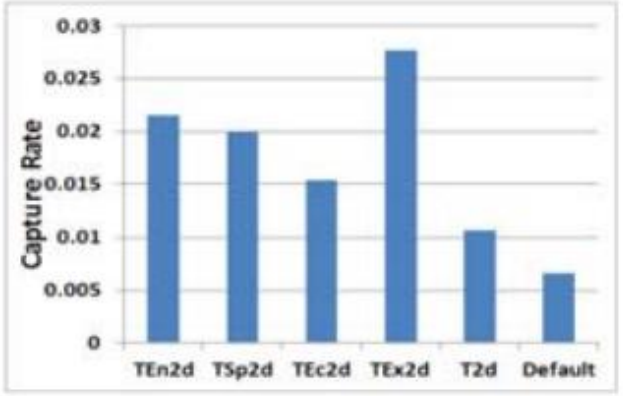
- 96 Benchmark Honepots, 5 Months
  - 1,077 unique accounts that follow one of the honeypots.
  - 440 @mentions
  - 1,512 Unique Accounts (That literally does not add up… April fools I guess?*)
- 1,512 Accounts
  - 303 Suspended by Twitter
    - Presumed spammers
  - 278 Identified by manual examination
  - 578 total spam accounts (Again… this math does not add up)

- Ground Truth
  - Difficult, there is no 100% way to identify a spammer.

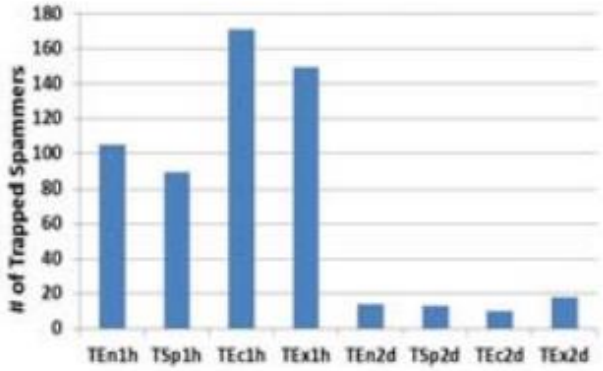
Original Presentation date: April 1, 2021

Social Honeypot Campaign - Specific Results

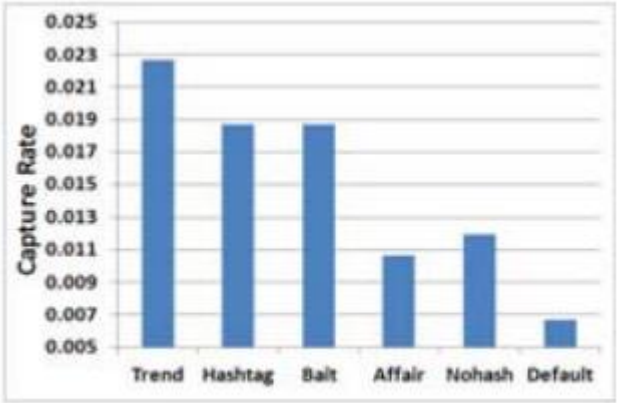
- CR (Capture Rate), number of trapped spammers / day.
  - Exact metric not specified, would suspect it is either a follow or a mention.
  - Higher CR = Better Honeypot
- Topic Expertise at 2 Tweets/Day was highest for just tweeting topics
  - .0275 vs 0.021 (Second highest)
- Following a specific topic = Higher CR
- Bait words
  - Very high CR
- “Advanced” Honeypots
  - Deployed 10 more honeypots for a week.
    - Specific optimized parameters (6 steps)
  - **Significantly** higher CR
    - 2.17!! (25.5 times starter GU)



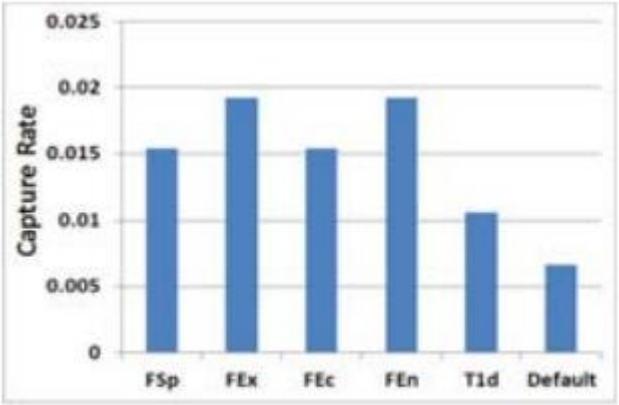
(a) Tweet Topics



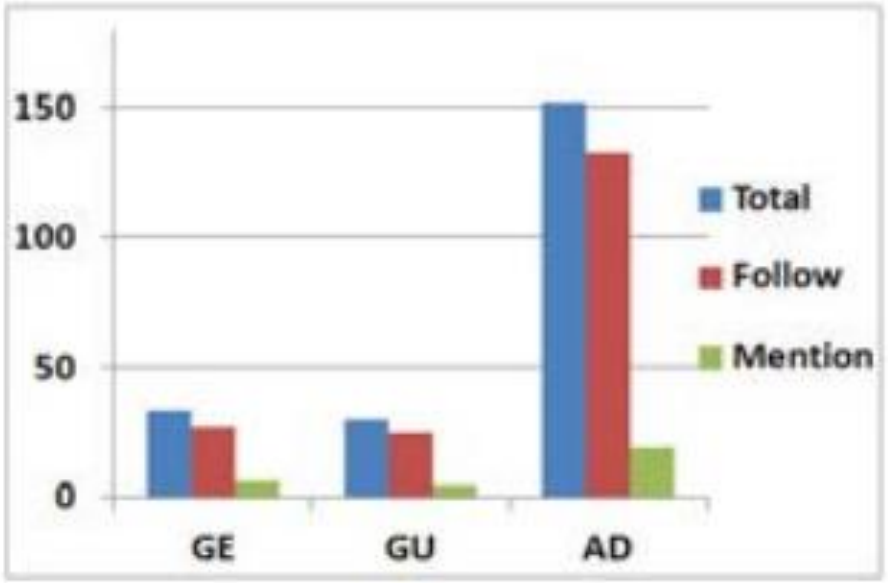
(b) # of Trapped Spammers



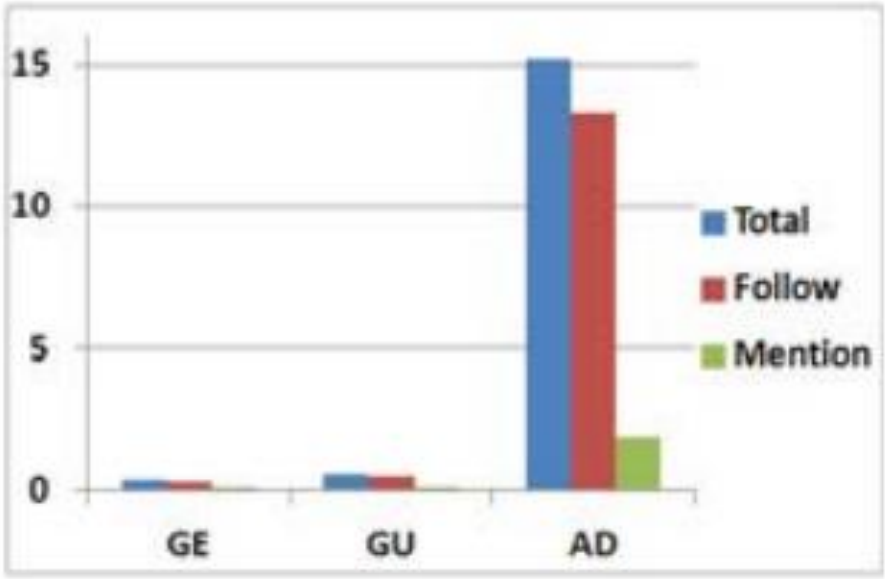
(a) Tweet Keywords



(b) Follow Behavior



(a) # of Spammers



(b) per Honeypot

Figure 6: The effectiveness of advanced honeypots.

Spammer’s Tastes

If accounts post more specific semantic topics (Economics), do they attract spammers?

YES

Do accounts with more specific terms attract more spammers?

YES

Do user’s following behaviors attract to spammers?

YES

Do accounts with different app behaviors attract more spammers?

NO

Questions?

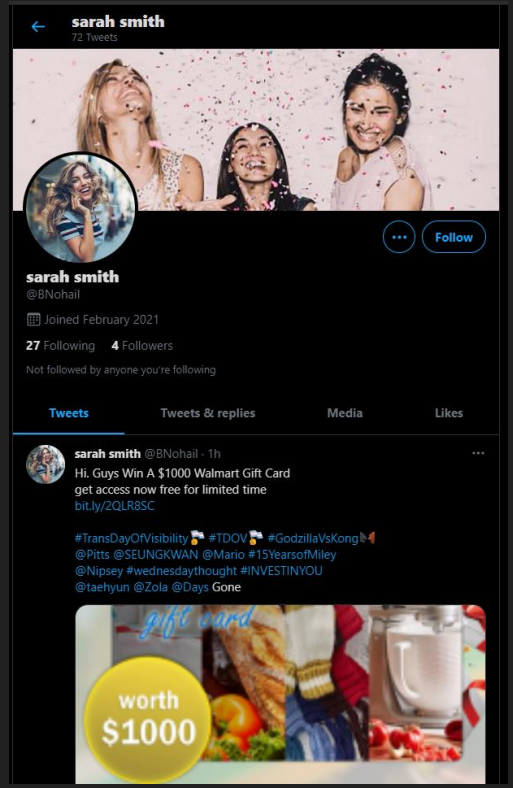
Discussion Q: How effective can a honeypot truly be? Methods used to detect spammers could also be used by spammers to avoid honeypots.

Active Sampling

Active Sampling - Overview

- Don't sit around and wait for attacks
  - Find the spammers yourself
- Spammers need to look legitimate
  - Follow their targets

- Active Data Collection
  - Sample Various data groups to find likely spammers
- Samplers
  - Friend Sampler
  - Hashtag Sampler



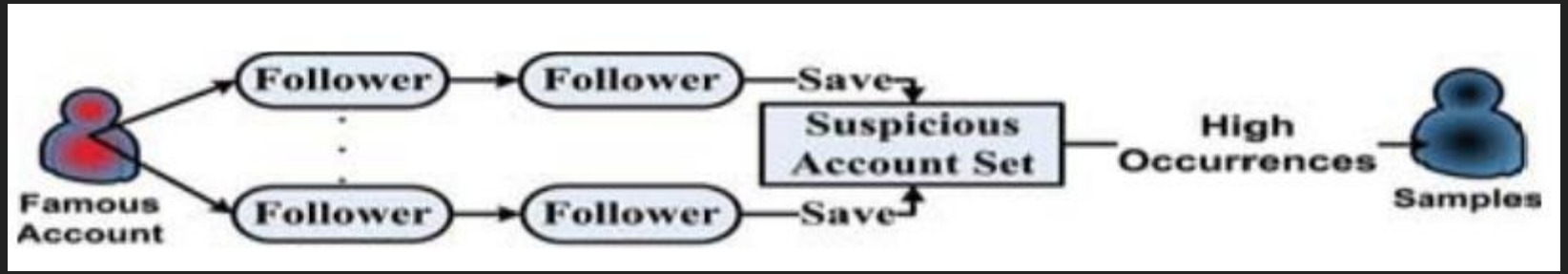
Active Sampling - Motivation

- Collecting spam accounts is first step to analyze behavior
- Honeypots require lots of time and luck (Passive)
  - Methods may evolve
  - People don’t want to wait
  - There is a cost to every day
- Manual Identification is Tedious and Costly

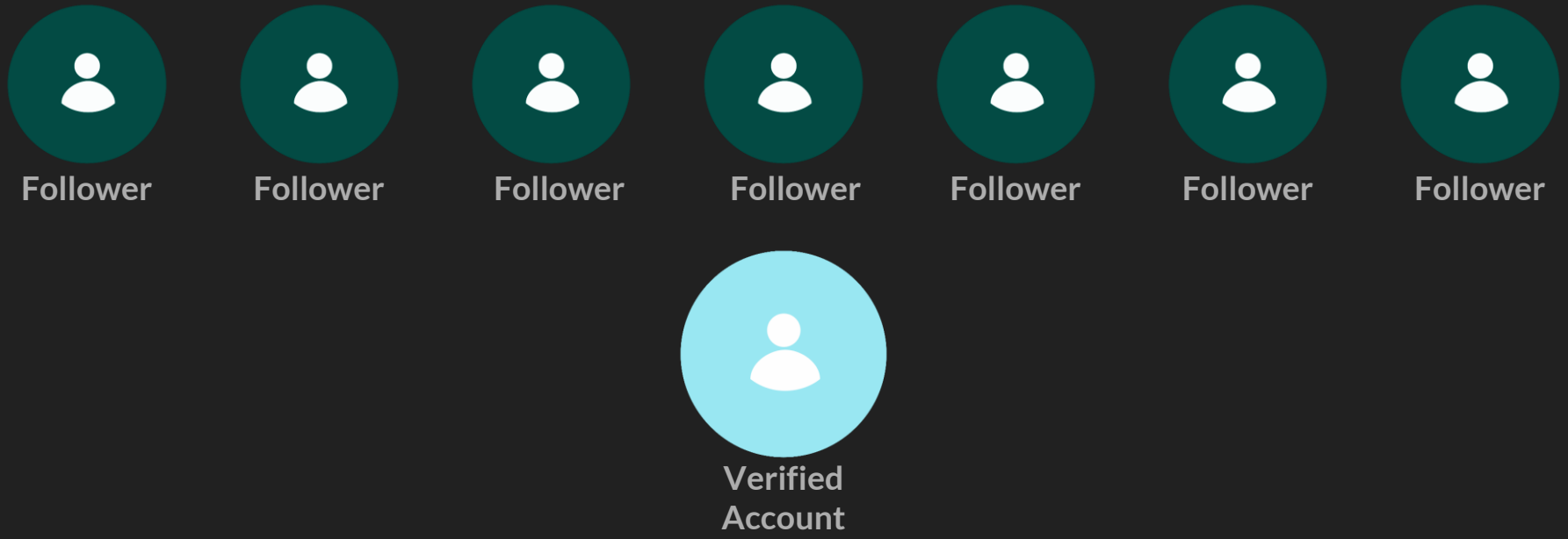
- Find an **efficient** and **accurate** algorithm to find spam accounts
  - Accuracy is important as false positives are bad!

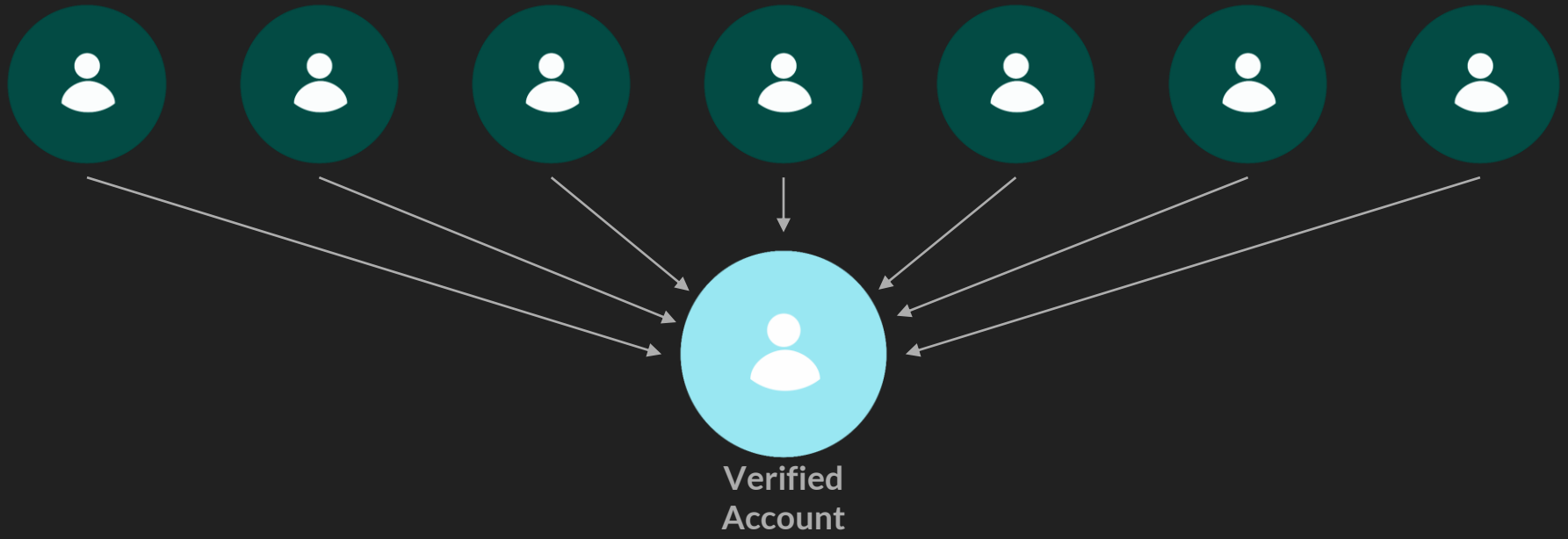
Friend Sampler

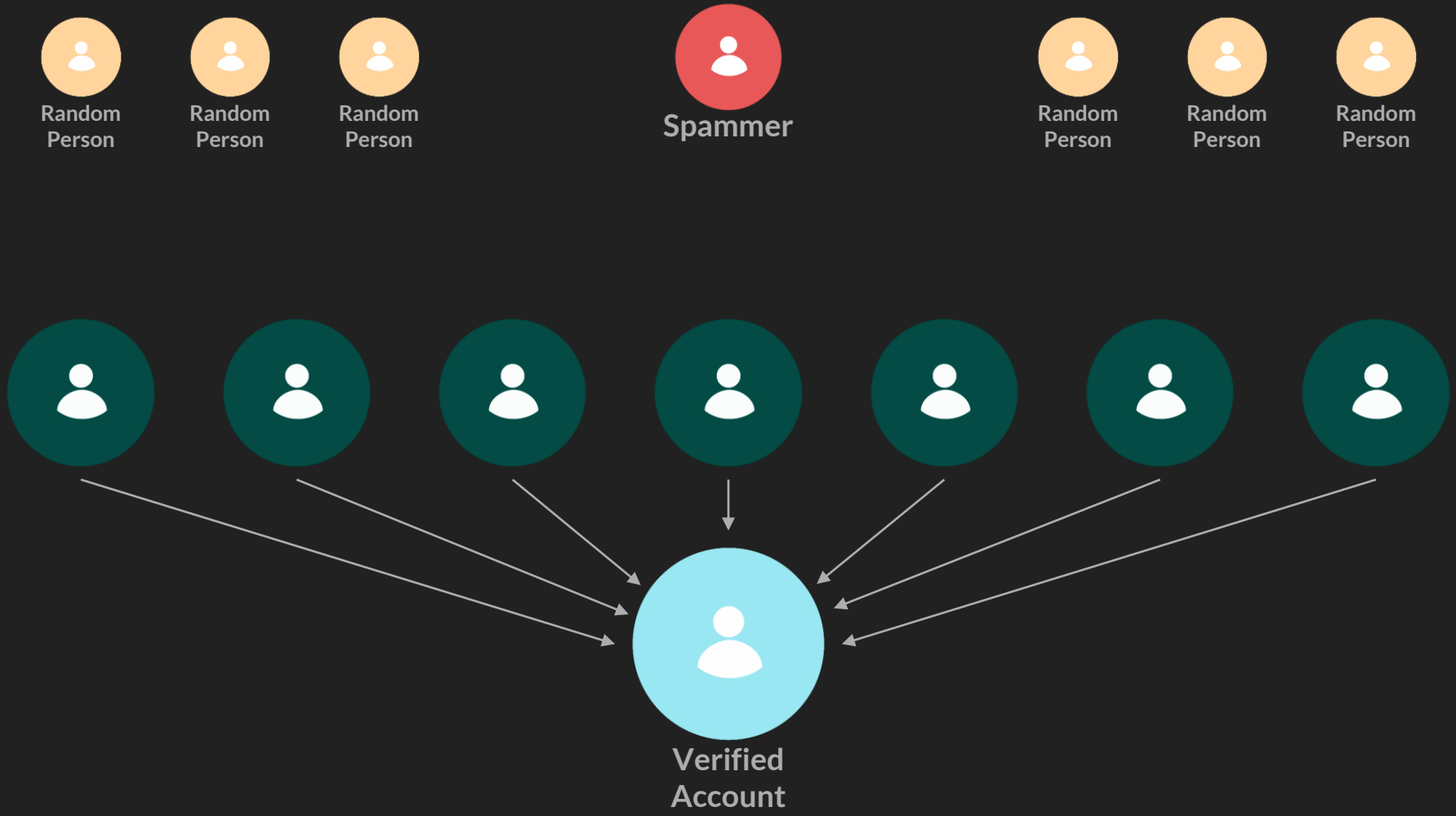
Friend Sampler

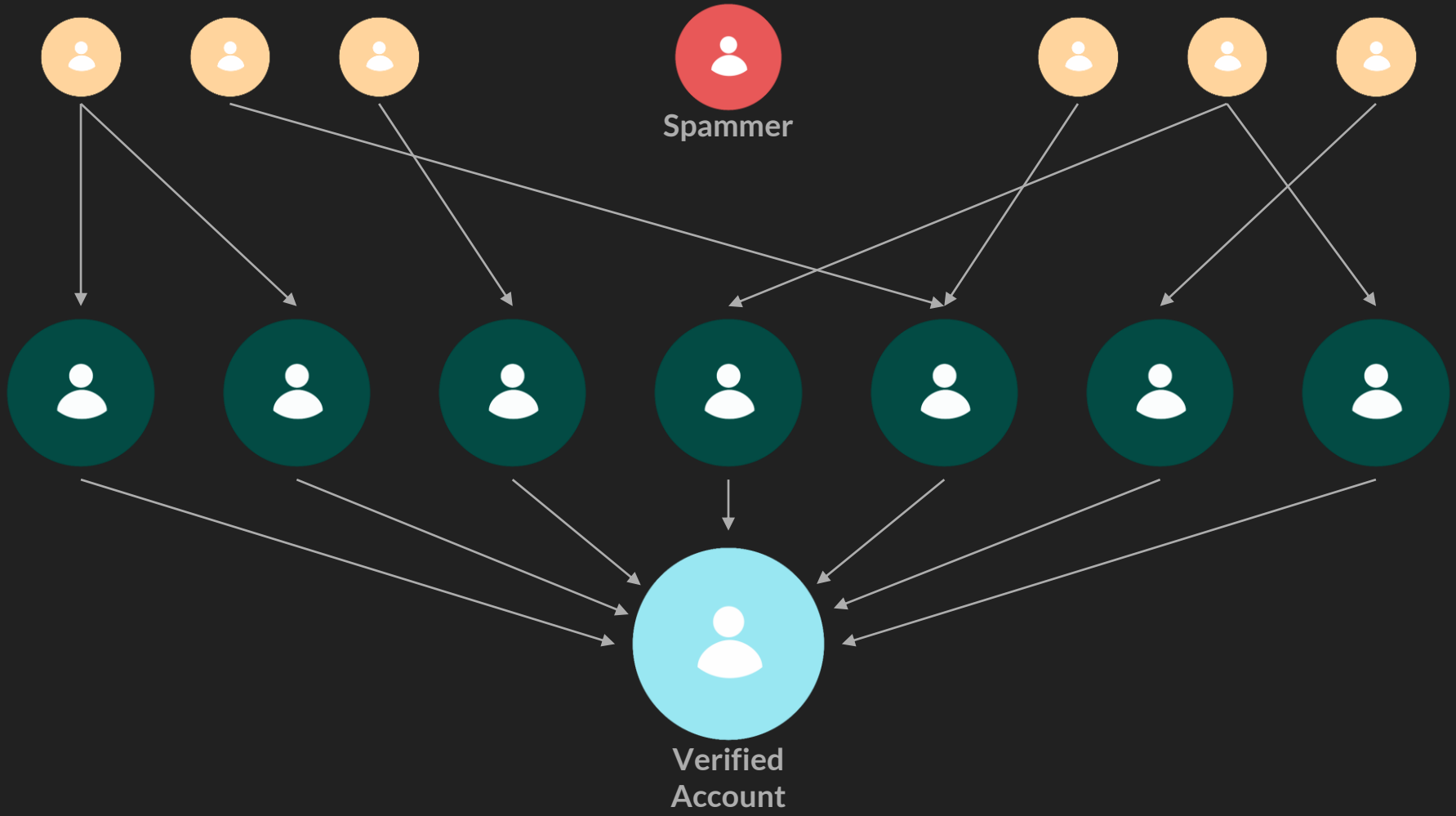


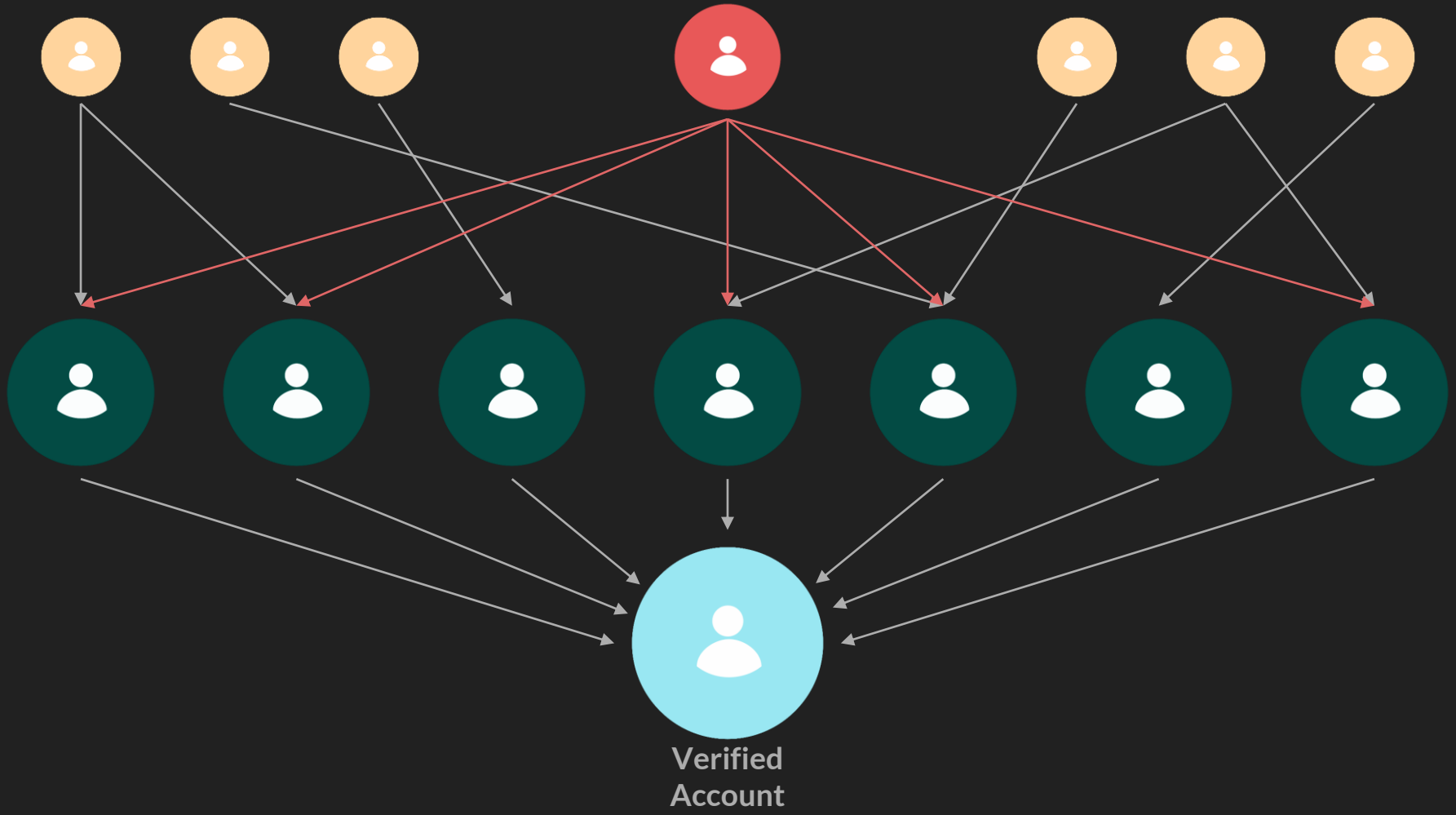
Verified
Account


Follower
Follower
Follower
Follower
Follower
Follower
Follower
Verified Account


Verified
Account


Random Person
Random Person
Random Person
Spammer
Random Person
Random Person
Random Person
Verified Account


Spammer
Verified
Account


Verified
Account

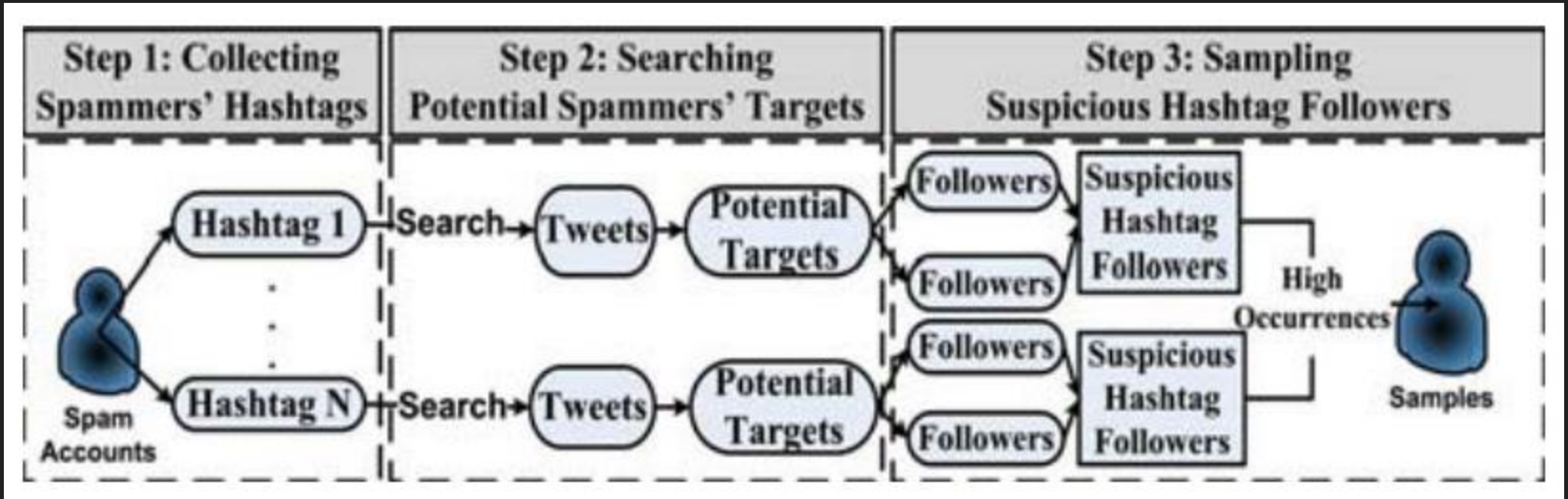
Friend Sampler

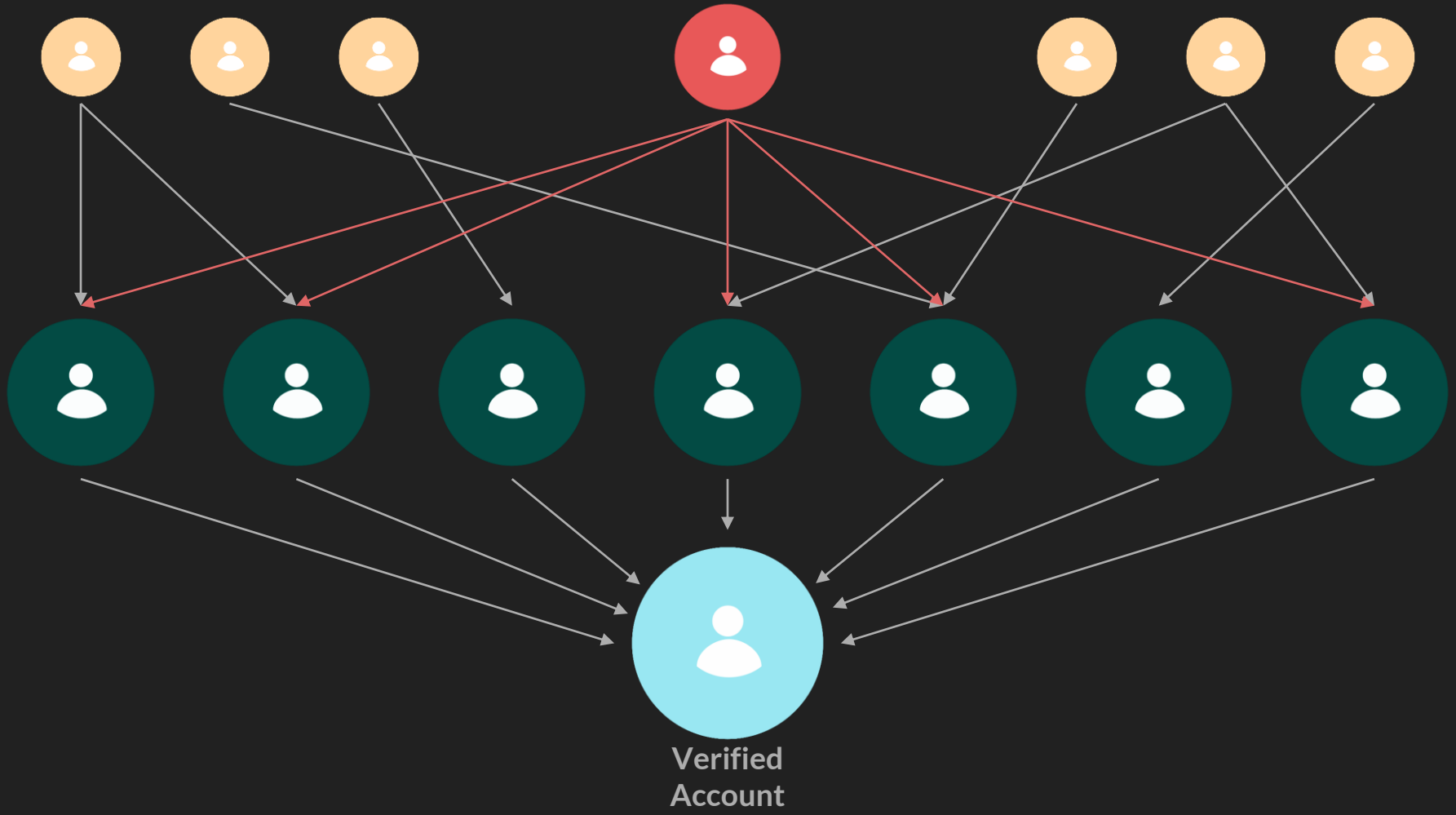
- Spammers select famous accounts' followers as their targets
- Steps
  1. Select M verified accounts
  2. Collect the N followers for each account
  3. Examine the followers of the N followers.
  4. Look for high overlap

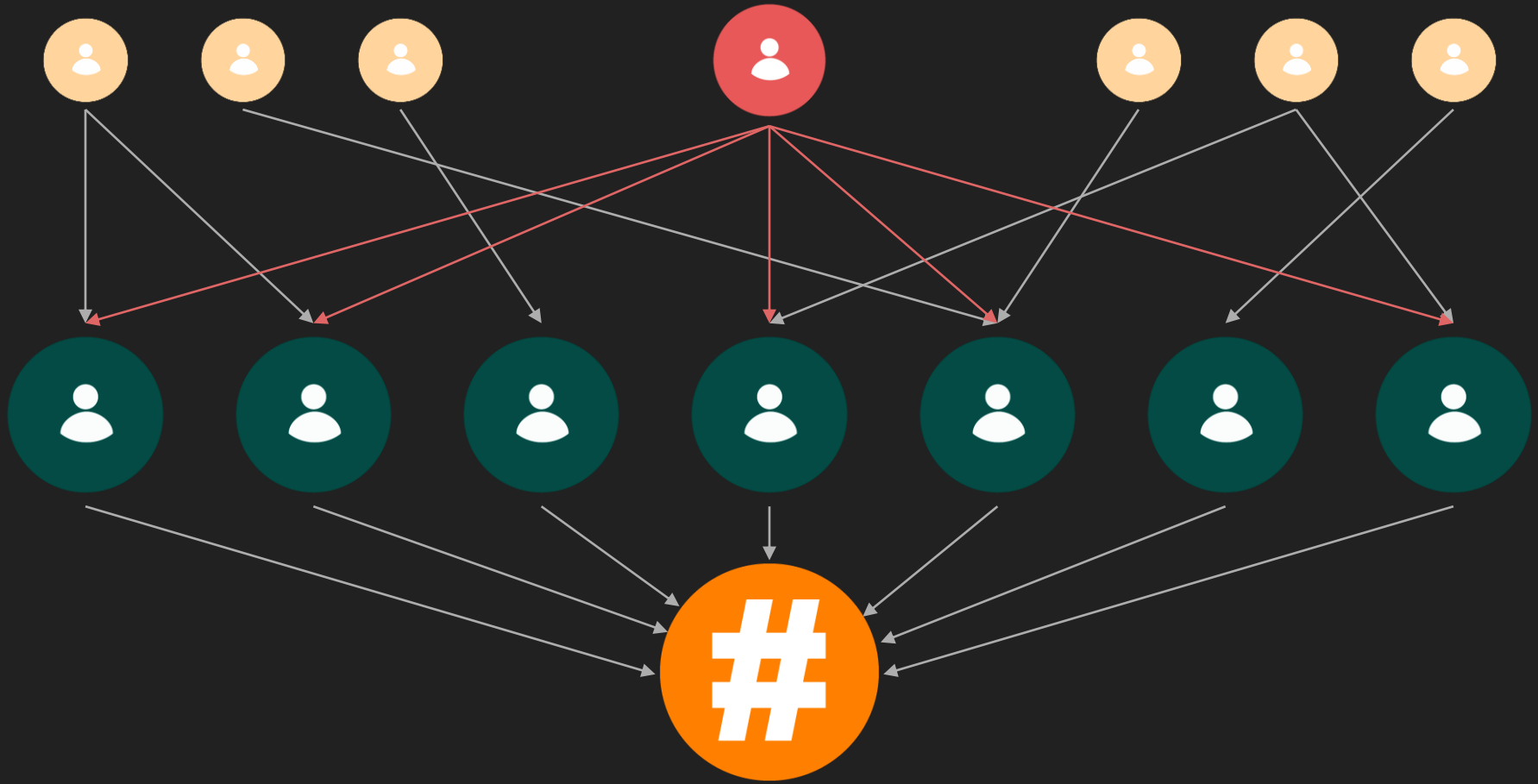
- High overlap = Sus
  - Scammer is following many of the followers

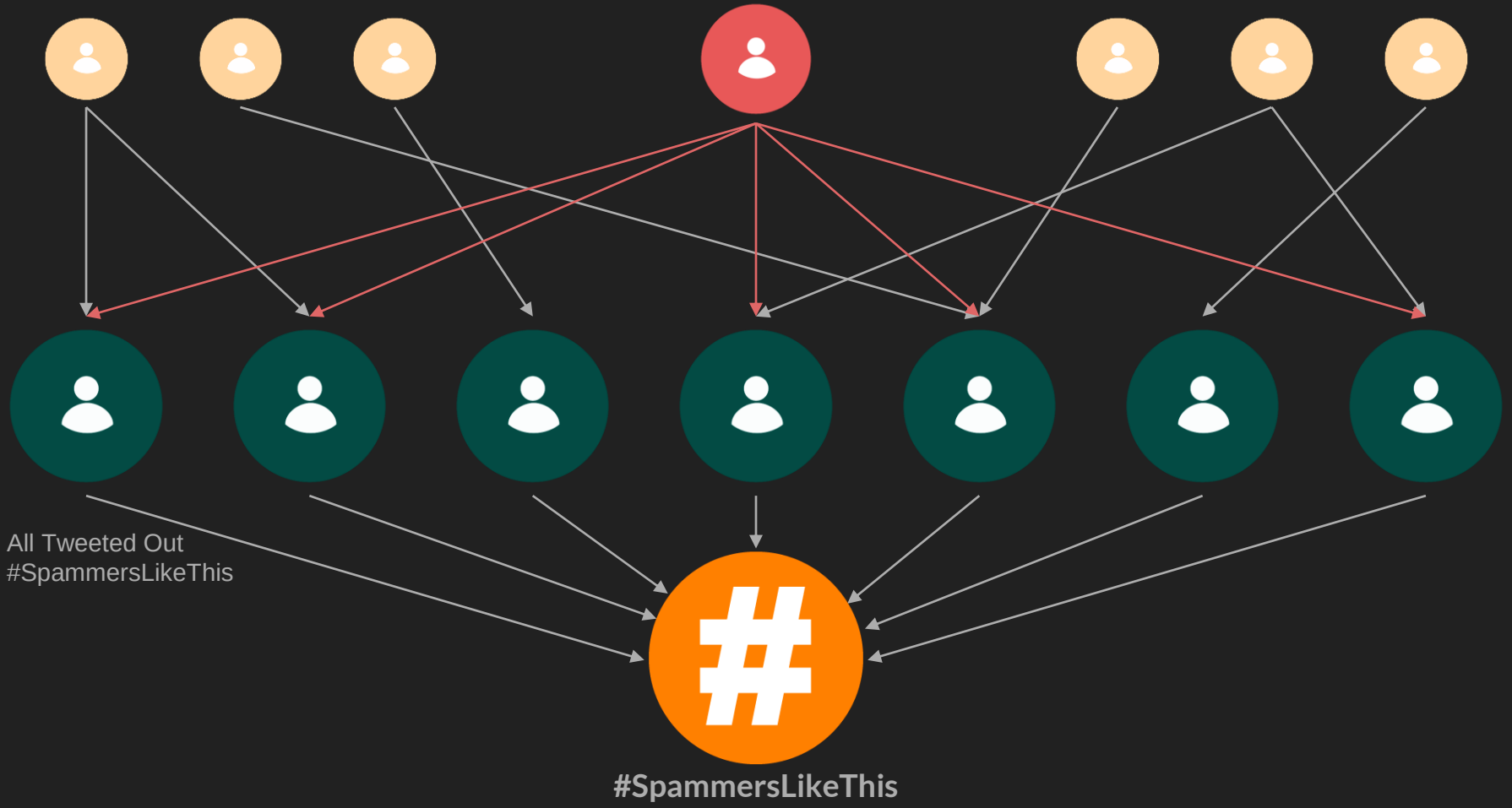
Hashtag Sampler

Hashtag Sampler



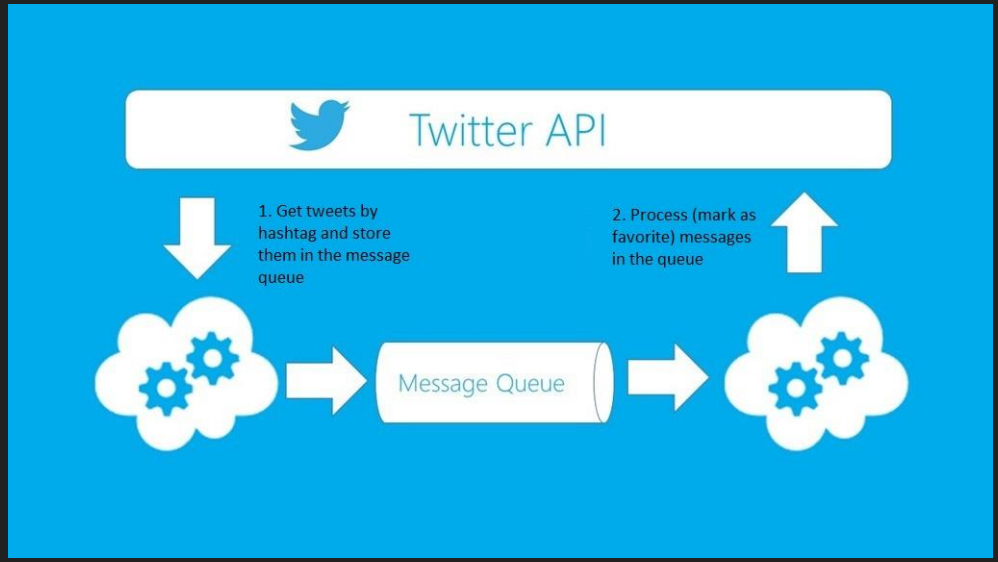

Verified
Account


#SpammersLikeThis


All Tweeted Out
#SpammersLikeThis
#SpammersLikeThis

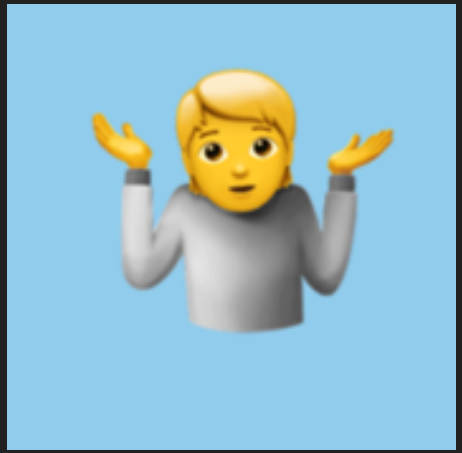
Implementation Details

- Twitter search API
  - Strange arbitrary limitations
- Friends
  - 400 Verified accounts
- Hashtag
  - 1500 Tweets
  - 3246 Hashtags by 278 spammers
  - 500 Sus # Followers
  - Limitations...
- Ground Truth
  - Hard to get…
  - Suspended = Spammer
  - Machine Learning on classifier (for non-suspended accounts)



Active Sampling Results - Confirmation

- Closest to ground truth = account getting suspended
- Ground truth is difficult
  - Non Suspended accounts
  - Machine learning 10-fold cross validation
- False positives
  - Exist
  - Really hard to tell...



Active Scanning Results - Raw Results

- Hashtag Sampler
  - 8,983 unique accounts flagged as likely spam
    - 262 suspended
    - 4,665 identified by machine learning classifier
  - Hit ratio of 54.89%
- Friend Sampler
  - 21,686 accounts flagged
    - 4,000 suspended
    - 9,781 identified by filter
  - Hit ratio of 63.55%
- Combined
  - Combining the algorithms results in 62% accuracy
- Complementary
  - High Exclusivity ratio, should be run in conjunction with each other.

Takeaways

- Spammers have specific tastes
- Active sampling wildly increases the rate of spammer detection
- These algorithms and methods can be implemented into Twitter’s spam policy and filters, and should be.

Discussion Questions

Has this field changed since 2014?

What were the accounts that are not spam, were they just confused people? Bots? Etc.

How could this be applied to other social networks (Facebook, Reddit, Youtube)

Is this an ongoing battle? How can spammers get around spam filters?

The social honeypots collect information from individuals without their explicit consent. Is this ethical?