Badly Evolved?

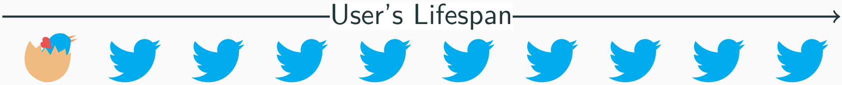
Exploring Long-Surviving Suspicious Users on Twitter

Majid Alfifi and James Caverlee

Department of Computer Science and Engineering
Texas A&M University

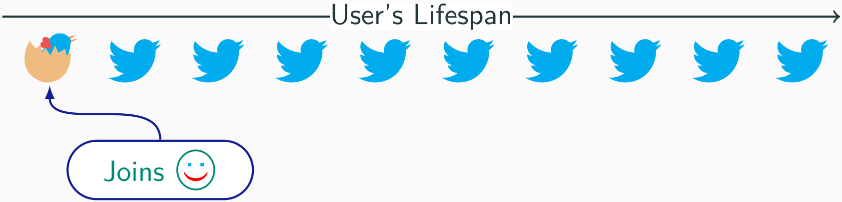
Lifecycle of a social account

An ideal account:



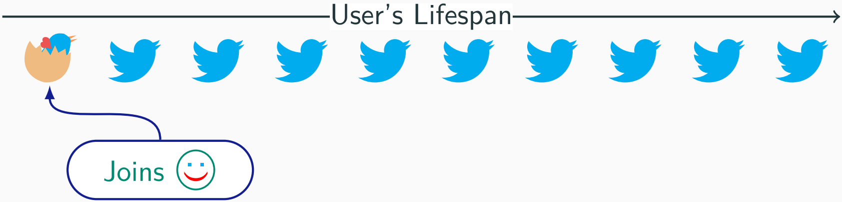
Lifecycle of a social account

An ideal account:

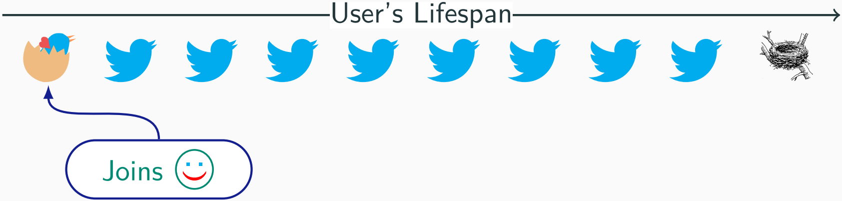


Lifecycle of a social account

An ideal account:

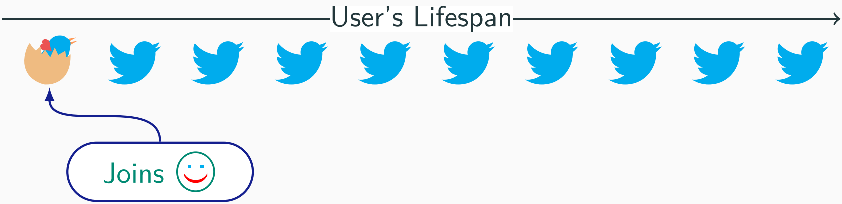


An OK account:

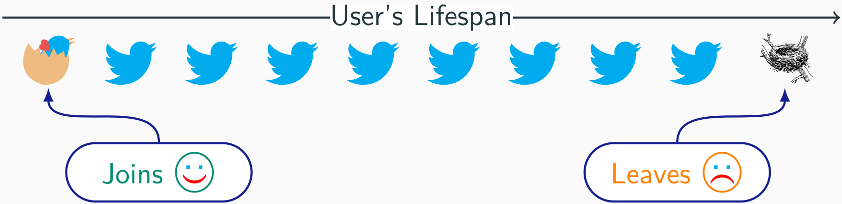


Lifecycle of a social account

An ideal account:

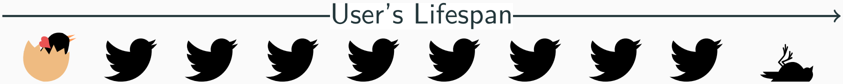


An OK account:



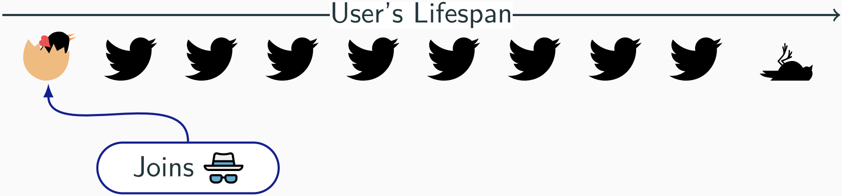
Lifecycle of a social account

A bad account:



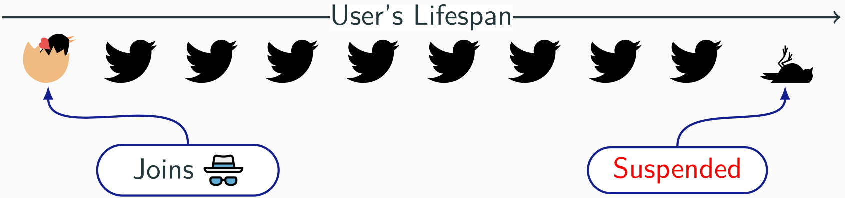
Lifecycle of a social account

A bad account:

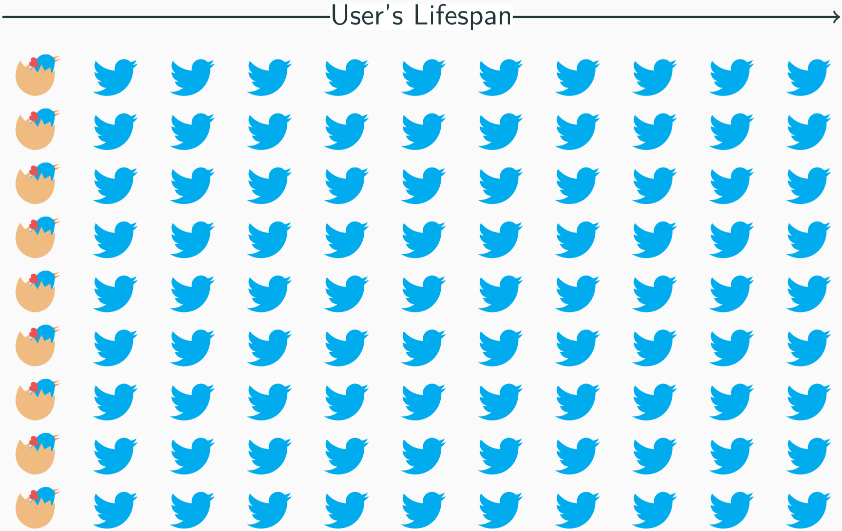


Lifecycle of a social account

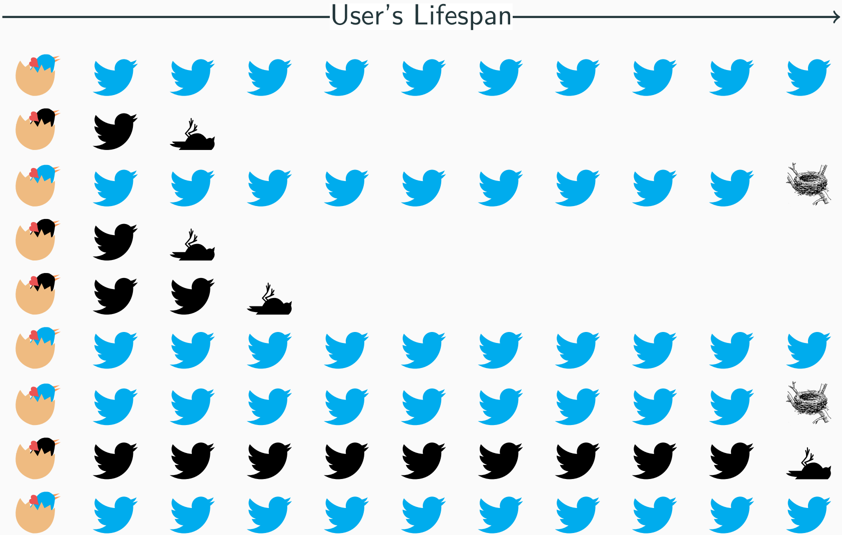
A bad account:



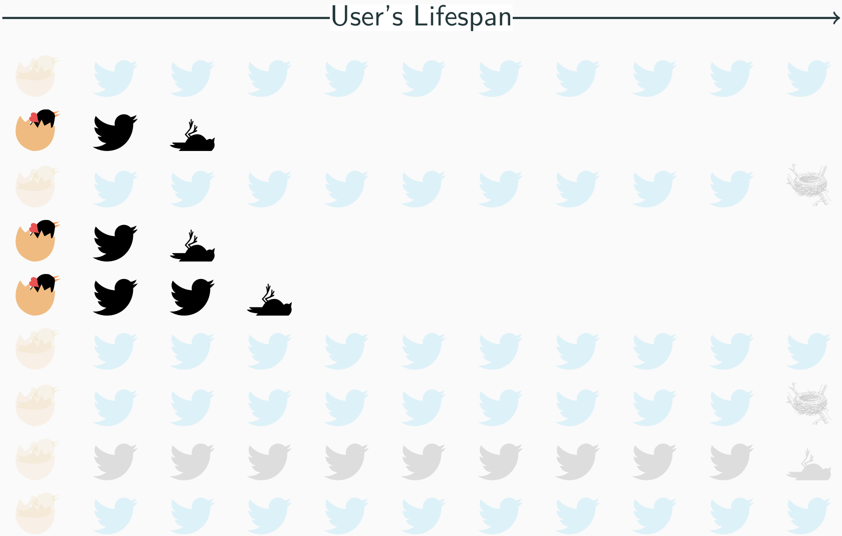
Ideally



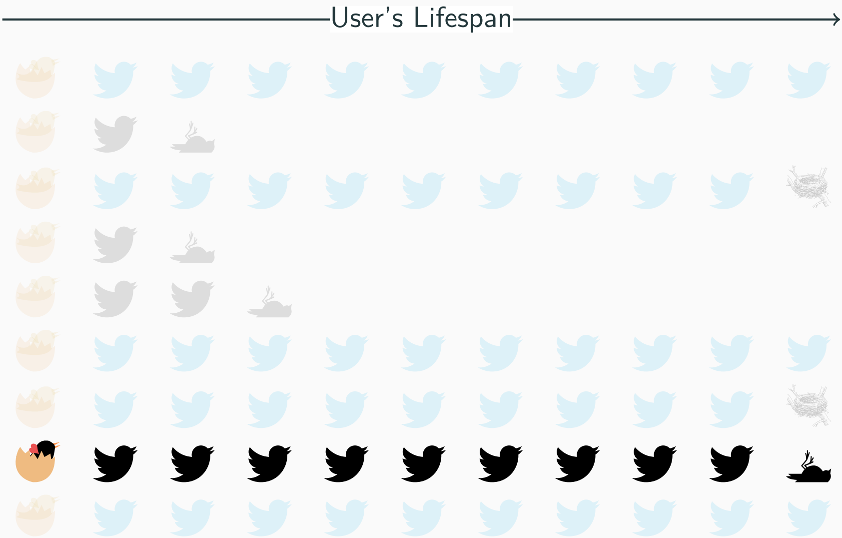
Reality



Well researched area



Our focus



Research goals

Do long-lived suspended accounts always engage in bad behaviors?

Research goals

Do long-lived suspended accounts always engage in bad behaviors?

Do they abruptly become bad accounts?

Research goals

Do long-lived suspended accounts always engage in bad behaviors?

Do they abruptly become bad accounts?

Or do they gradually evolve into bad accounts?

Do long-lived suspended accounts always engage in bad behaviors?

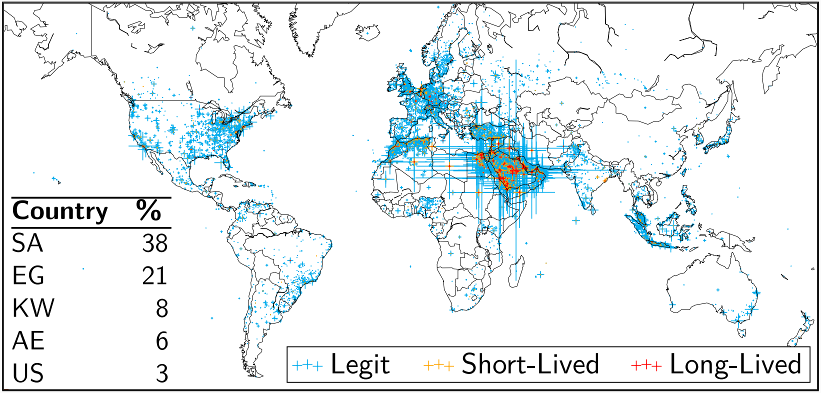
Do they abruptly become bad accounts?

Or do they gradually evolve into bad accounts?

How are they different from short-lived suspended accounts?

Dataset

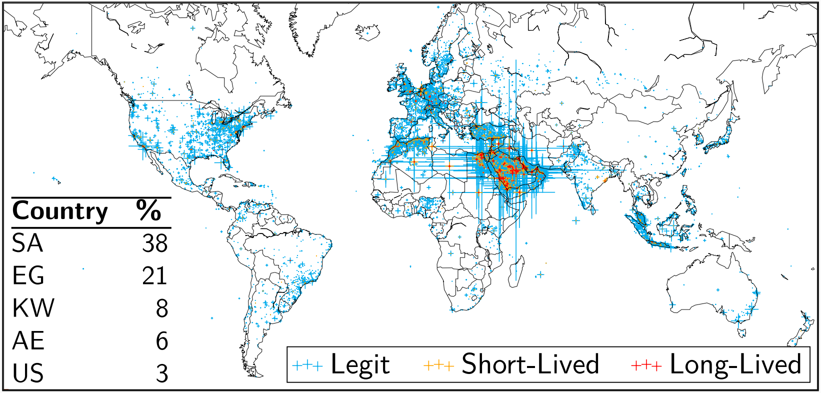
All Arabic tweets in 2015



Country	%
SA	38
EG	21
KW	8
AE	6
US	3

Dataset	Size
Tweets	9,285,246,636
Accounts	26,711,275
Tweets from Suspended Accounts	1,960,160,536
Suspended Accounts	6,175,113

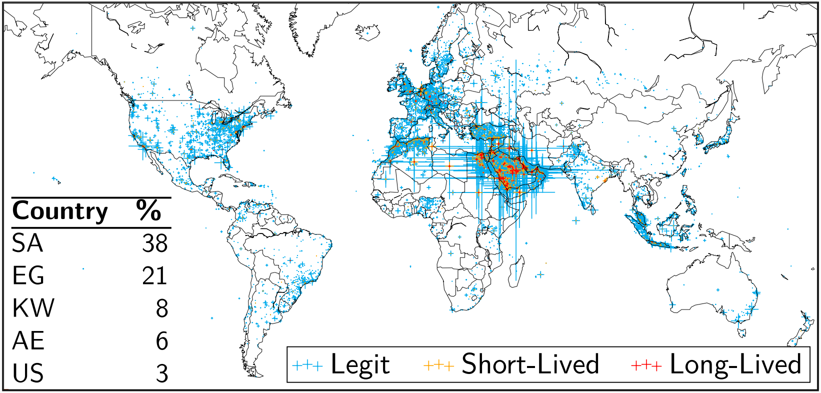
All Arabic tweets in 2015



Country	%
SA	38
EG	21
KW	8
AE	6
US	3

Dataset	Size	
Tweets	9,285,246,636	
Accounts	26,711,275	
Tweets from Suspended Accounts	1,960,160,536	
Suspended Accounts	6,175,113	23%

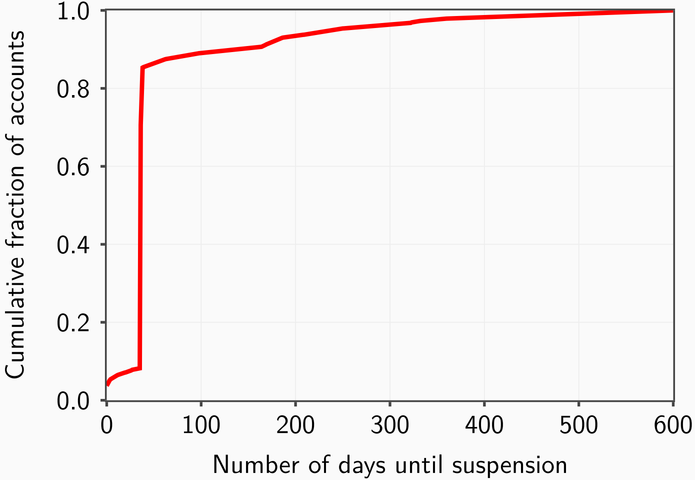
All Arabic tweets in 2015



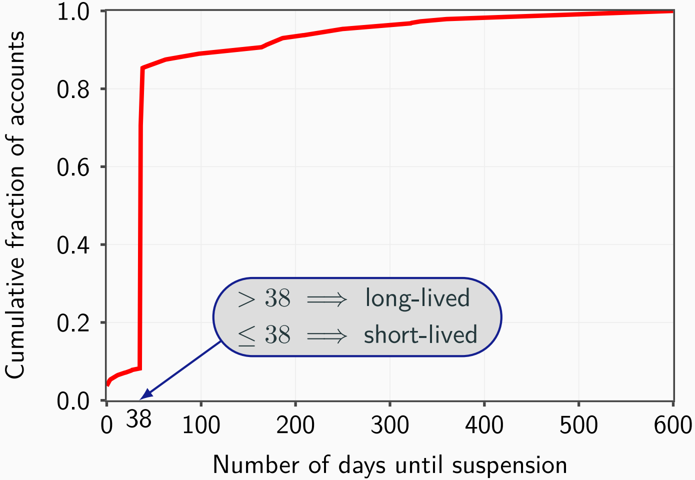
Country	%
SA	38
EG	21
KW	8
AE	6
US	3

Dataset	Size	
Tweets	9,285,246,636	
Accounts	26,711,275	
Tweets from Suspended Accounts	1,960,160,536	21%
Suspended Accounts	6,175,113	23%

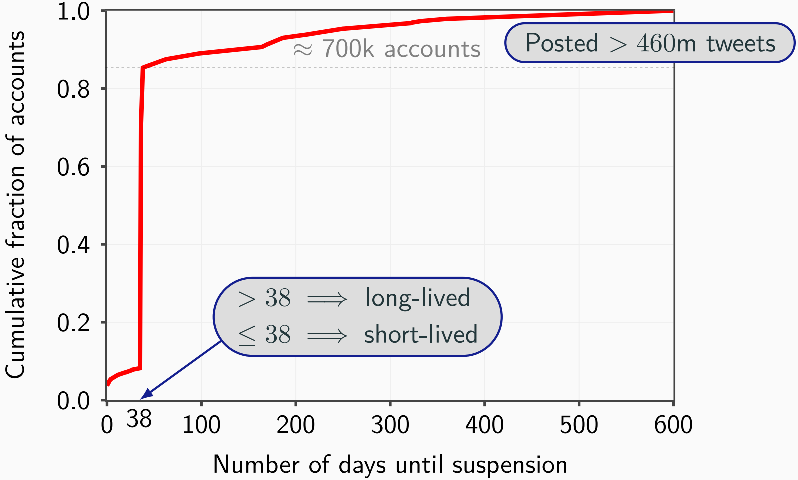
Suspended accounts: long-lived vs. short-lived



Suspended accounts: long-lived vs. short-lived



Suspended accounts: long-lived vs. short-lived



Four account groups

#	Group	Accounts	Tweets count
1	`long-lived`	17,909	42,630,795

All accounts that:

1. Were created in January 2015 or December 2014.
2. Were active on at least 6 different months.
3. Were eventually suspended by Twitter.

Four account groups

#	Group	Accounts	Tweets count
1	`long-lived`	17,909	42,630,795
2	`short-lived`	17,909	14,129,870

A random sample from accounts that:

1. Were suspended within 38 days of creation.
2. Posted at least 10 tweets.

Four account groups

#	Group	Accounts	Tweets count
1	`long-lived`	17,909	42,630,795
2	`short-lived`	17,909	14,129,870
3	`legit`	17,909	9,772,176

A random sample from accounts that:

1. Were created in January 2015 or December 2014.
2. Were Active on at least 6 different months.
3. Were still alive in November 2016.
4. Stopped tweeting in January/February 2016.

Four account groups

#	Group	Accounts	Tweets count
1	`long-lived`	17,909	42,630,795
2	`short-lived`	17,909	14,129,870
3	`legit`	17,909	9,772,176
4	`isis`	17,518	11,849,065

We exploit a list of ISIS accounts crowdsourced by the Anonymous group and recover their tweets.

We focus on accounts that:

1. Were actually suspended.
2. Were active in 2015 (>10 tweets).

`thehackernews.com`

Four account groups

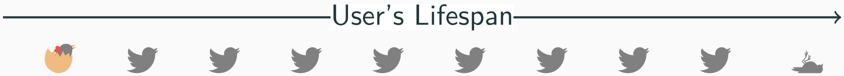
#	Group	Accounts	Tweets count
1	long-lived	17,909	42,630,795
2	short-lived	17,909	14,129,870
3	legit	17,909	9,772,176
4	isis	17,518	11,849,065

Methodology

Evolution modeling

To study users evolution, we split the lifespan of an account into 10 stages:

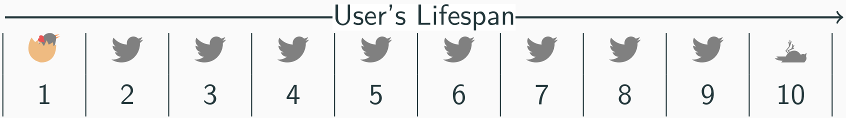
Longer lifespan...



Evolution modeling

To study users evolution, we split the lifespan of an account into 10 stages:

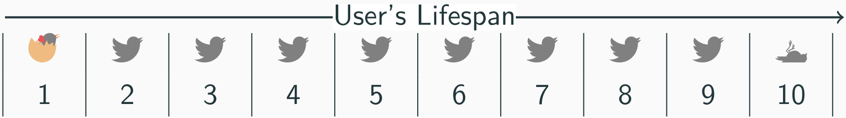
Longer lifespan...



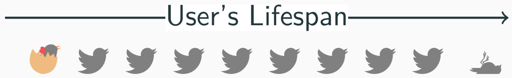
Evolution modeling

To study users evolution, we split the lifespan of an account into 10 stages:

Longer lifespan...



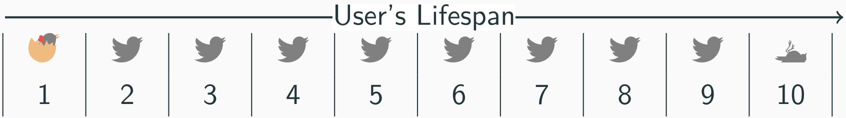
Shorter lifespan...



Evolution modeling

To study users evolution, we split the lifespan of an account into 10 stages:

Longer lifespan...



Shorter lifespan...

Stage-wise measures

At each stage, we measure several signals:

Behavioral

- Number of URLs
- Number of Hashtags
- Number of Mentions (in and out)

Linguistic

- Distance from the Twitter stream
- Self similarity

** See paper for all features **

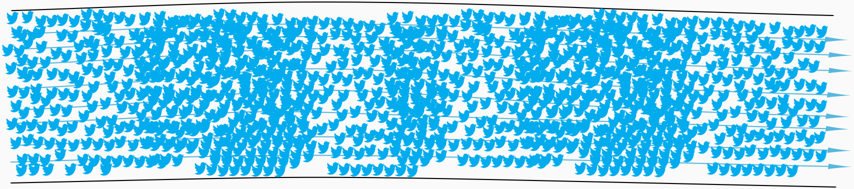
Linguistic distance from the Twitter stream

$$H_t(t, \mathrm{BLM}) = -\frac{1}{N}\sum_i \log P_{\mathrm{BLM}}(b_i)$$

Variable	**Meaning**
BLM	Background Language Model from the Twitter stream
t	Tweet
H_t	Cross-entropy of a tweet t according to the BLM
b_i	Bigram
N	Number of bigrams in a tweet t
$P_{\mathrm{BLM}}(b_i)$	Probability of a bigram b_i according to the BLM

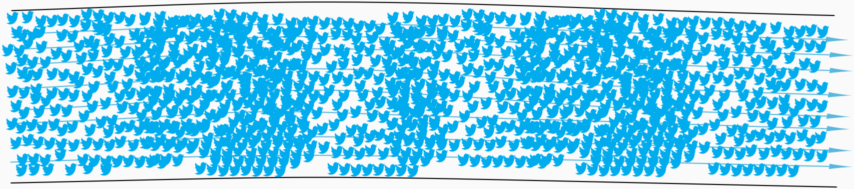
- Higher values indicate more sophisticated accounts. (e.g. humans)
- Repetitive low quality tweets get lower values

Linguistic distance from the Twitter stream: an example



Twitter stream

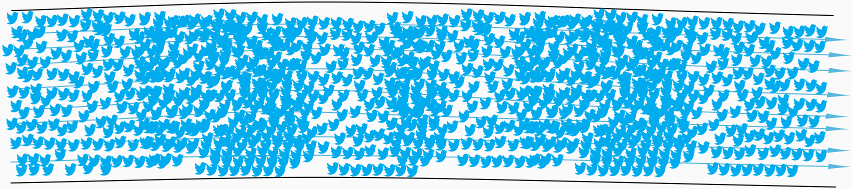
Linguistic distance from the Twitter stream: an example



Twitter stream

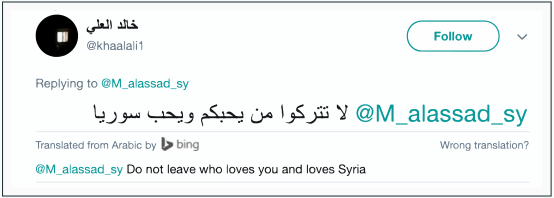
Linguistic distance from the Twitter stream: an example

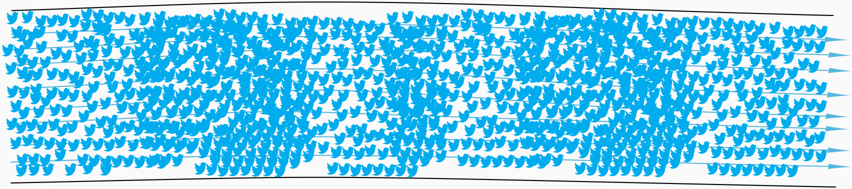
$$H_t = 1.28031$$



Twitter stream

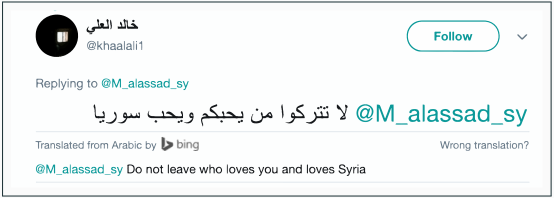
Linguistic distance from the Twitter stream: an example



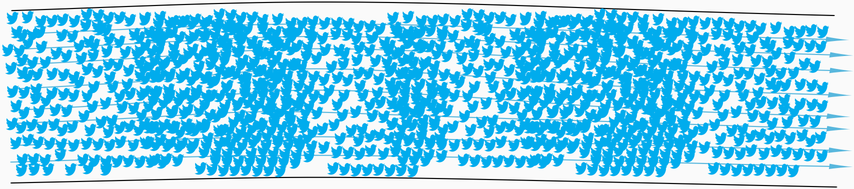


Twitter stream

Linguistic distance from the Twitter stream: an example



$H_t = 19.81468$

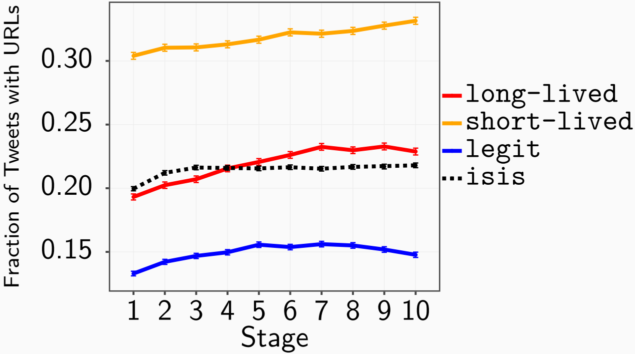


Twitter stream

Results & Conclusions

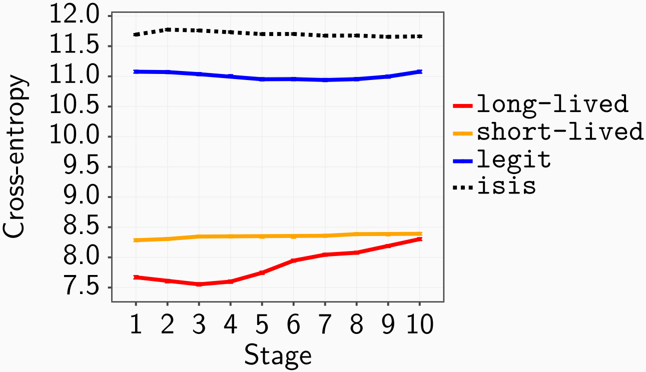
How do long-lived accounts evade detection?

They fine-tune their behavioral signals to remain under the radar.

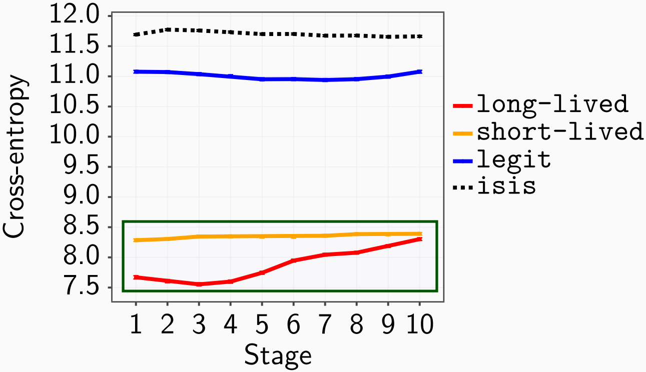


- long-lived may have evaded detection by limiting URL sharing among other signals.

What is the linguistic difference between groups?

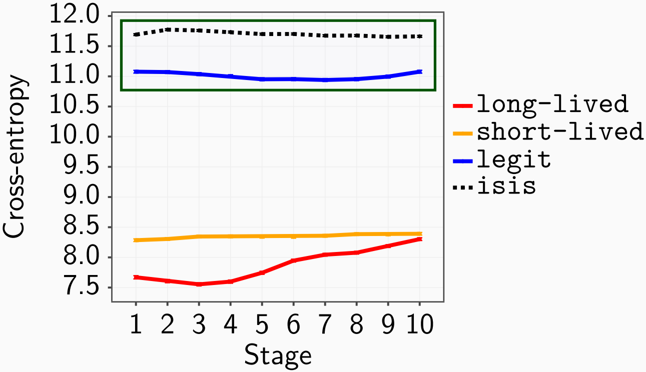


What is the linguistic difference between groups?



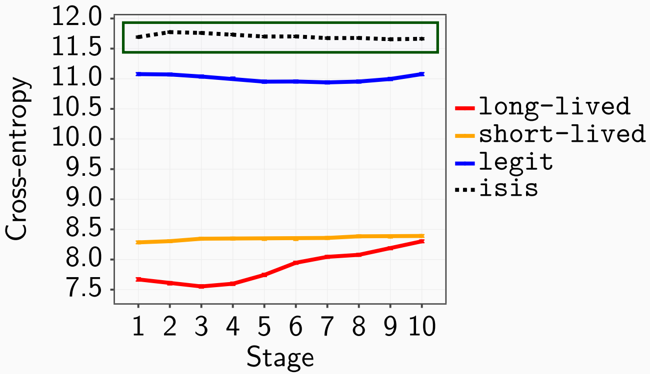
- long-lived fail to evade the linguistic distance measure.

What is the linguistic difference between groups?



- long-lived fail to evade the linguistic distance measure.
- isis and legit deviate the most hinting they may both represent real people.

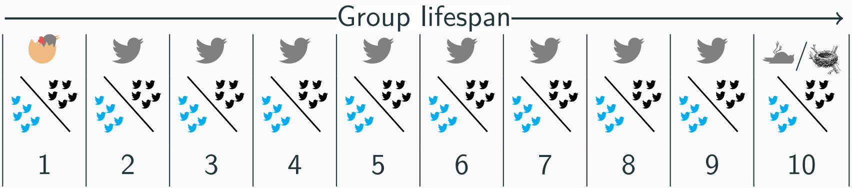
What is the linguistic difference between groups?



- `long-lived` fail to evade the linguistic distance measure.
- `isis` and `legit` deviate the most hinting they may both represent real people.
- `isis` deviates even more, potentially due to their extreme language.

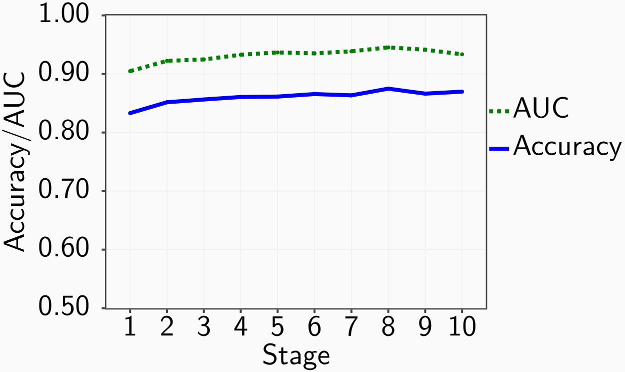
Can we detect long-lived accounts? How early?

We use a series of binary classifiers (Random Forest) one for each stage.
We use the signals measured at each stage as features.



Is an account long-lived?

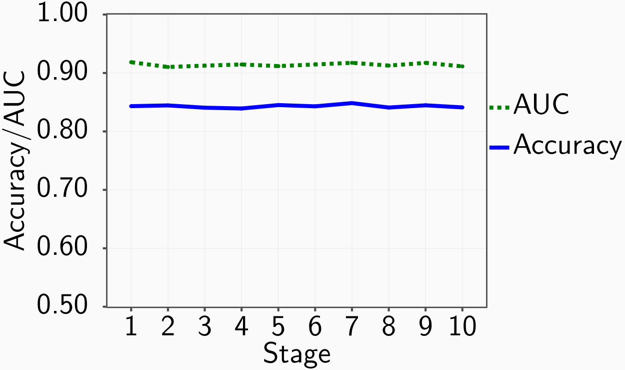
We train binary classifiers for long-lived and other groups:



- long-lived accounts can be detected very early.
- long-lived behavior slightly worsens over time resulting in better detection.

Is an account `isis`?

We train binary classifiers for `isis` and other groups:



- `isis` accounts are also detectable early.
- detection performance is consistent implying consistent behavior.

Conclusions

- The majority of long-lived suspicious accounts have most likely been born that way and didn't evolve into bad accounts.

- Long-lived suspicious accounts can be detected early greatly improving the quality of online social content.
- ISIS accounts are easily detectable regardless of their reportedly successful social media practice.

Thank You!

Slides available at:
`http://students.cs.tamu.edu/alfifima`