

Synthesized Social Signals: Computationally-Derived Social Signals from Account Histories

Jane Im⁺, Sonali Tandon⁺, Eshwar Chandrasekharan^{*},
Taylor Denby⁺, Eric Gilbert⁺

University of Michigan School of Information⁺

University of Illinois at Urbana-Champaign Computer Science^{*}



How do you decide whether to interact with a stranger online?

Social Signals on Social Platforms

The image shows a Facebook profile for Kyle Chung. The profile picture is a circular portrait of a man with long dark hair, labeled 'a'. The cover photo is a blue background with silhouettes of sharks and fish, labeled 'b'. The name 'Kyle Chung' is displayed below the profile picture. Navigation buttons include 'Add Friend', 'Follow', 'Message', and a three-dot menu. The main navigation tabs are 'Timeline', 'About', 'Friends', and 'Photos'. The 'Intro' section contains the text 'PhD student who loves research' (labeled 'c') and 'Followed by 77 people' (labeled 'd'). The 'Friends' section shows '819 (7 Mutual)' friends (labeled 'e') and three profile pictures: David Smith, Alicia Kim, and Jim G (labeled 'f'). The 'About' section is expanded, showing 'Works at School of Public Health' (labeled 'g'), 'Studied at School of Public Health' (labeled 'h'), 'Lives in United States' (labeled 'i'), and '1 family member' (labeled 'j'). A post by Kyle Chung is visible at the bottom, starting with 'My first conf' and 'Thank you so much everyone!' with a smiley face emoji (labeled 'k').

Social Signals F2F



Social Signals on Social Platforms

The image shows a Facebook profile for Kyle Chung. The profile picture is a circular portrait of a man with long dark hair, labeled 'a'. The cover photo is a blue background with silhouettes of sharks and fish, labeled 'b'. The name 'Kyle Chung' is displayed below the profile picture. Navigation buttons include 'Add Friend', 'Follow', 'Message', and a three-dot menu. The main navigation tabs are 'Timeline', 'About', 'Friends', and 'Photos'. The 'Intro' section contains the text 'PhD student who loves research' (labeled 'c') and 'Followed by 77 people' (labeled 'd'). The 'Friends' section shows '819 (7 Mutual)' friends (labeled 'e') and three profile pictures: David Smith, Alicia Kim, and Jim G (labeled 'f'). The 'About' section is expanded, showing 'Works at School of Public Health' (labeled 'g'), 'Studied at School of Public Health' (labeled 'h'), 'Lives in United States' (labeled 'i'), and '1 family member' (labeled 'j'). A post at the bottom of the 'About' section says 'My first conf... Thank you so much everyone! 😊' (labeled 'k').

Social Signals Matter

- ✱ People manage self-presentation using social signals
- ✱ Such self-presentation affect people's perception

Ellison, Nicole, Rebecca Heino, and Jennifer Gibbs. "Managing impressions online: Self- presentation processes in the online dating environment." *Journal of computer-mediated communication* 11.2 (2006): 415-441.

Munger, Kevin. "Tweetment effects on the tweeted: Experimentally reducing racist harassment." *Political Behavior* 39.3 (2017): 629-649.

Social Signals on Social Platforms



[Edit profile](#)

임제인 Jane Im
@im__jane

PhD student @UMich | Previously: Korea University, @MITAppInventor | Builds consentful & safe socio-technical systems #consent #feminism

📍 Ann Arbor, MI 🔗 imjane.net 📅 Joined August 2018

565 Following **250** Followers

Social Signals on Social Platforms



[Edit profile](#)

임제인 Jane Im
@im__jane

PhD student @UMich | Previously: Korea University, @MITAppInventor | Builds
consentful & safe socio-technical systems #consent #feminism

📍 Ann Arbor, MI 🔗 imjane.net 📅 Joined August 2018

565 Following **250** Followers

But social platforms provide an **important source of information** more easily compared to offline world.

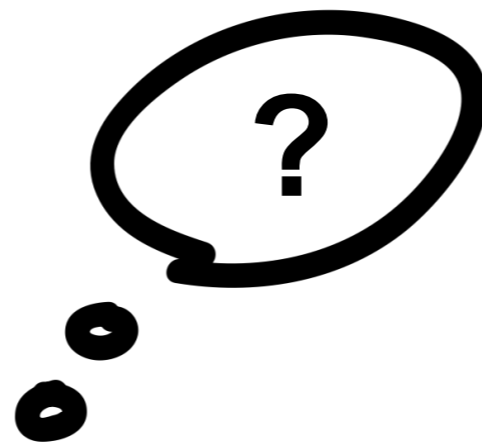
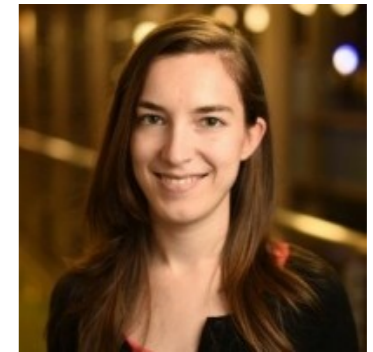


A history of a person's behavior

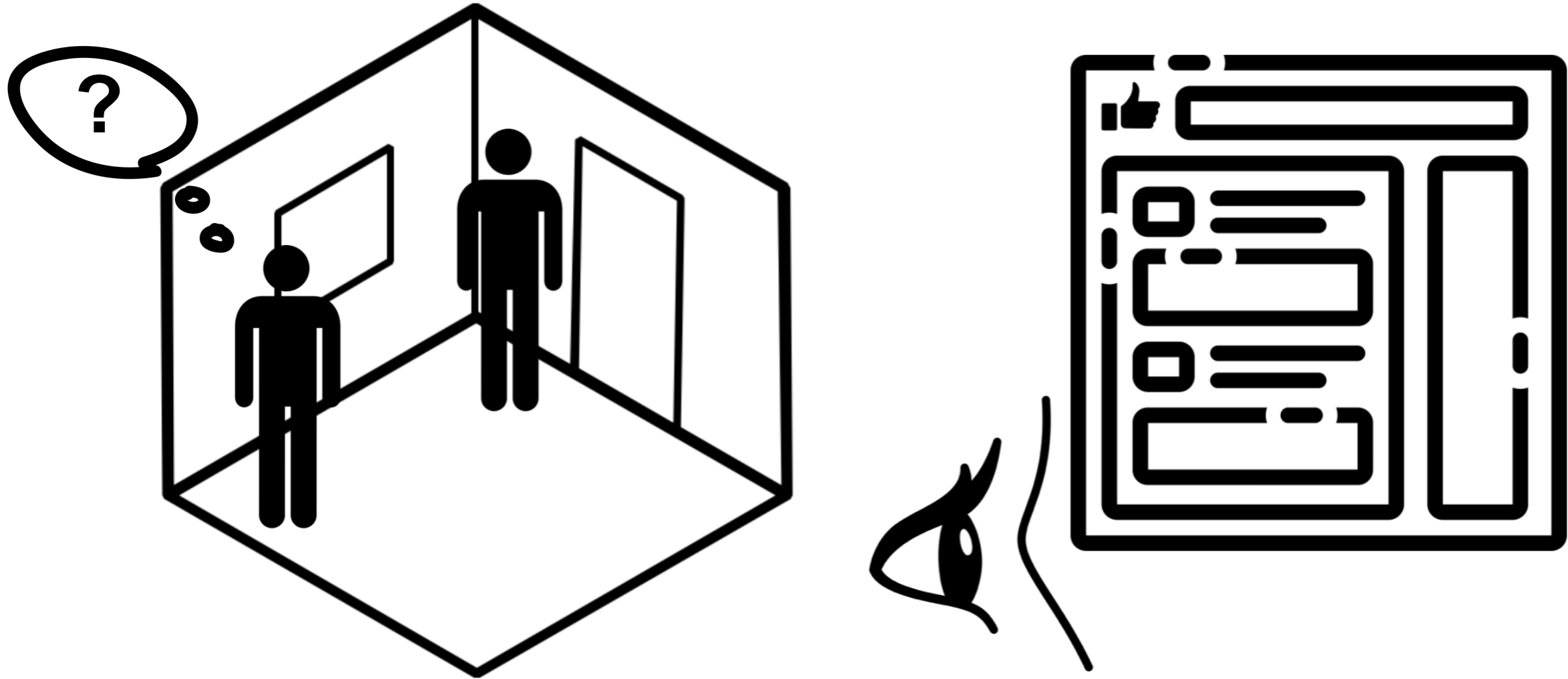


Eric, my advisor :-)

history of interactions



Eric's PhD students before me :)



A history of a person's behavior



Receiver cost is too high

Synthesized Social Signals (S3s)

social signals computationally derived from an account's history, and then rendered into the profile

Synthesized Social Signals (S3s)

social signals computationally derived from an account's history, and then **rendered into the profile**

Synthesized Social Signals (S3s)

- ✱ reduces cost to access important information
- ✱ harder to fake



Bob
@bob



Saw your posts [@im_jane](#) . Disgusting. You're just a foolish coward

10:07 AM · Jan 22, 2020 · [Twitter Web App](#)



Bob
@bob



You know you are an asshole [@strangerstran18?](#) Your tweets are garbage

12:50 AM · Jan 22, 2020 · [Twitter Web App](#)





Bob
@bob



Saw your posts @im iane . Disgusting. You're iust a foolish cov

10:07 AM · Jan 2



Bob
@bob



Hmm...there's more to the story?



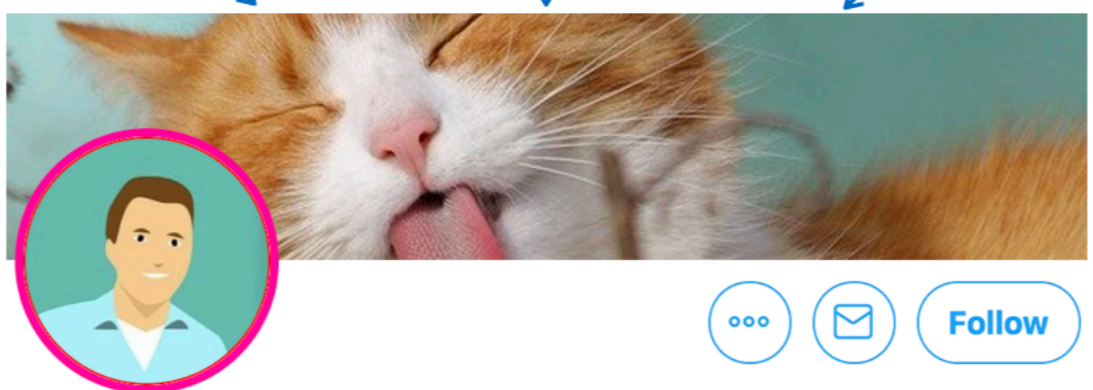
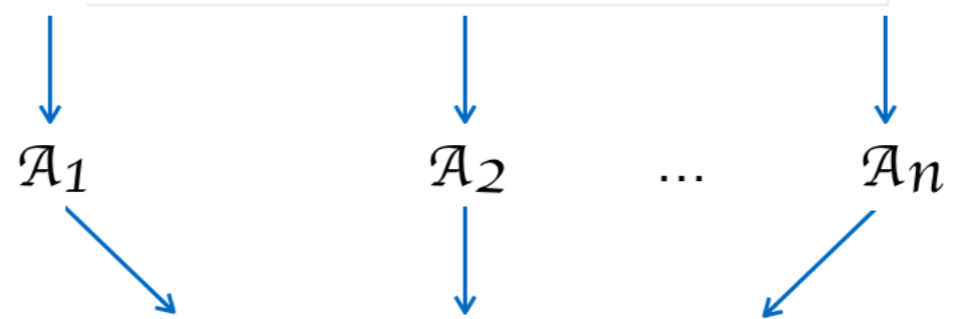
Who Targeted Ukraine Airlines Flight 752? Iran Shot It Down ...
The claim that Major General Qassem Soleimani was a "terrorist" on a mission to carry out an "imminent" attack tha...
globalresearch.ca

10:05 AM · Jan 22, 2020 · [Twitter Web App](#)



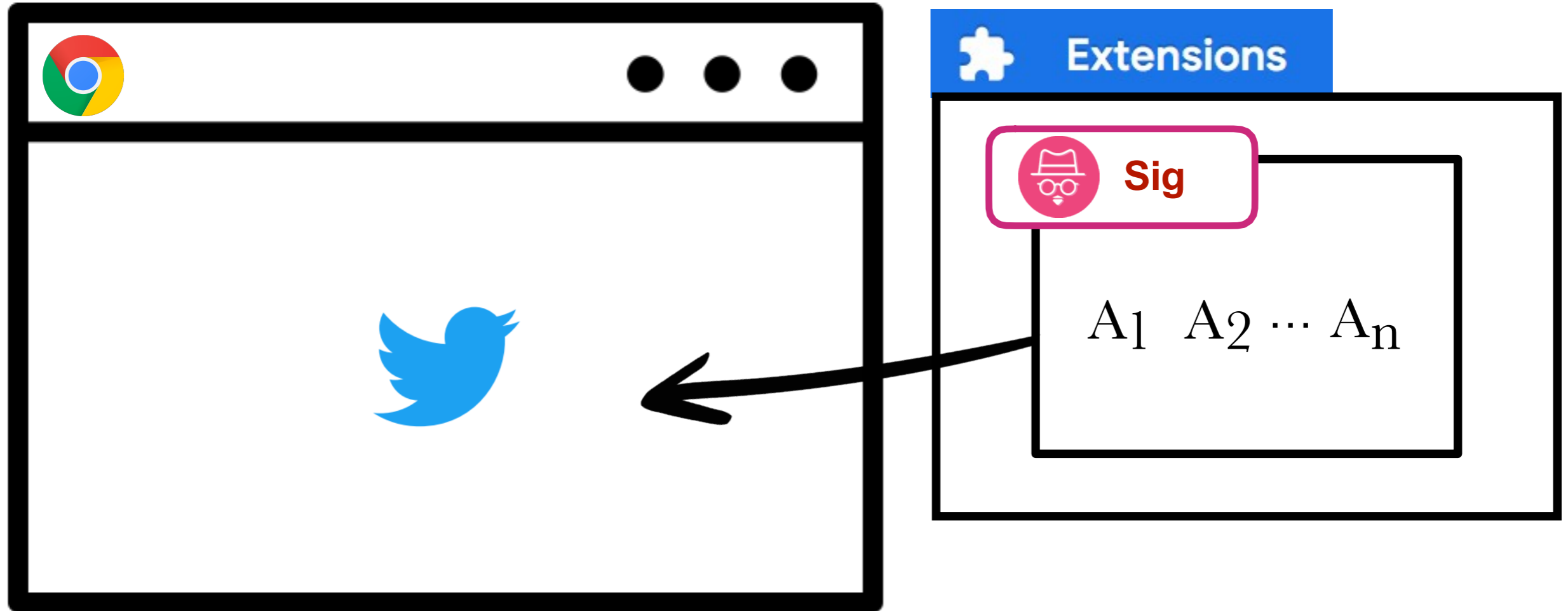
You
twe

12:50



Bob
@bob
toxicity misinfo
Photographer @ NY
Joined March 2011
203 Following 3,439 Followers

Sig



Sig

- ✦ For our field study we focused on **Twitter**
- ✦ Focused on **toxicity** and **misinformation**, two well-known problems on social media





Notifications



All

Mentions



Lucy liked your Retweet

The hardest part of a PhD and an academic job isn't intelligence, grants, or research chops..

My mom hugged me yesterday and started to cry, "I wish you lived closer" I was visiting for Christmas. She had brain surgery two weeks ago.
@AcademicChatter #phdchat #AcademicChatter



Lucy followed you



Stat World followed you



Lucy

@lucy_lim Follows you

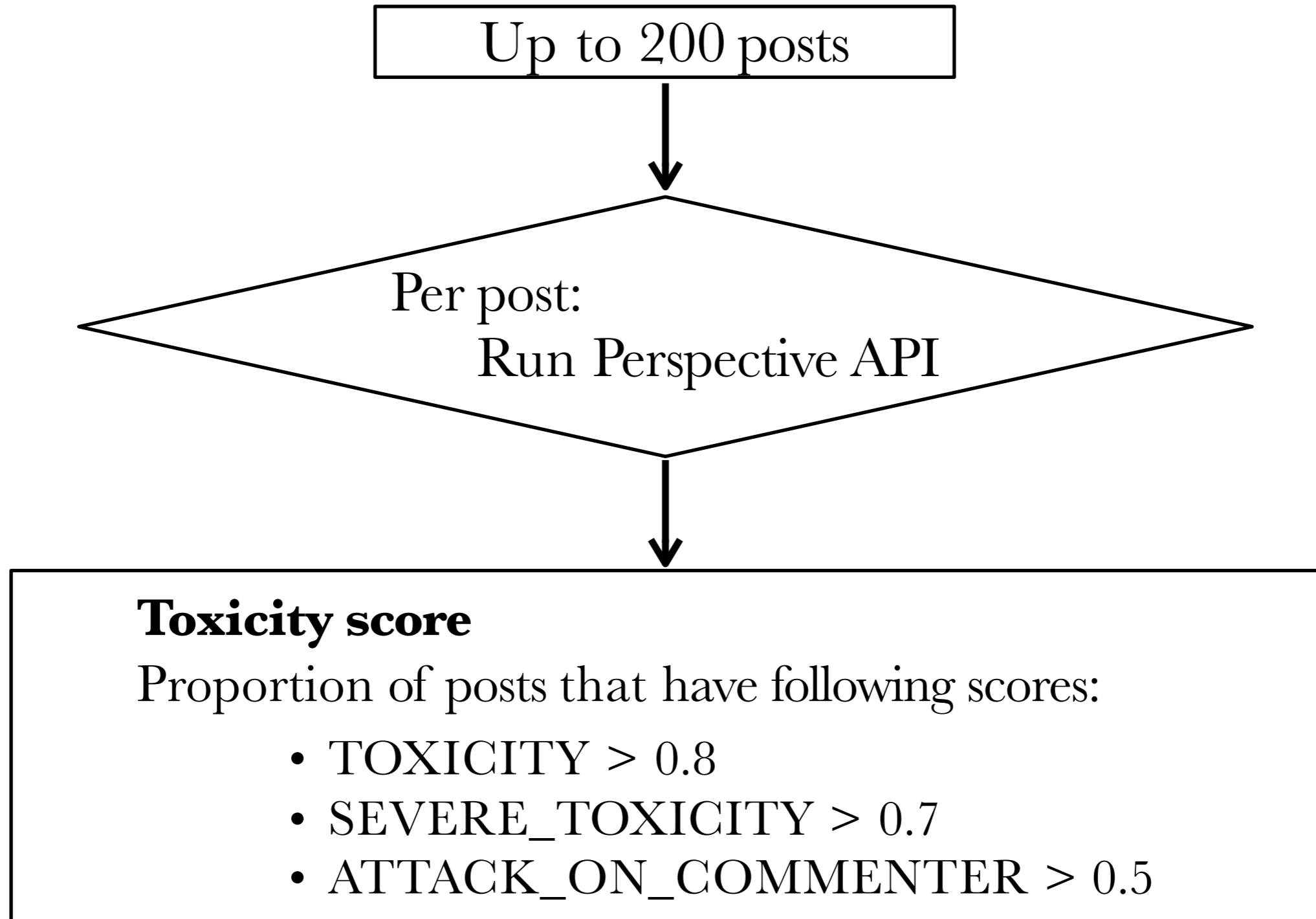
Toxicity

Misinformation

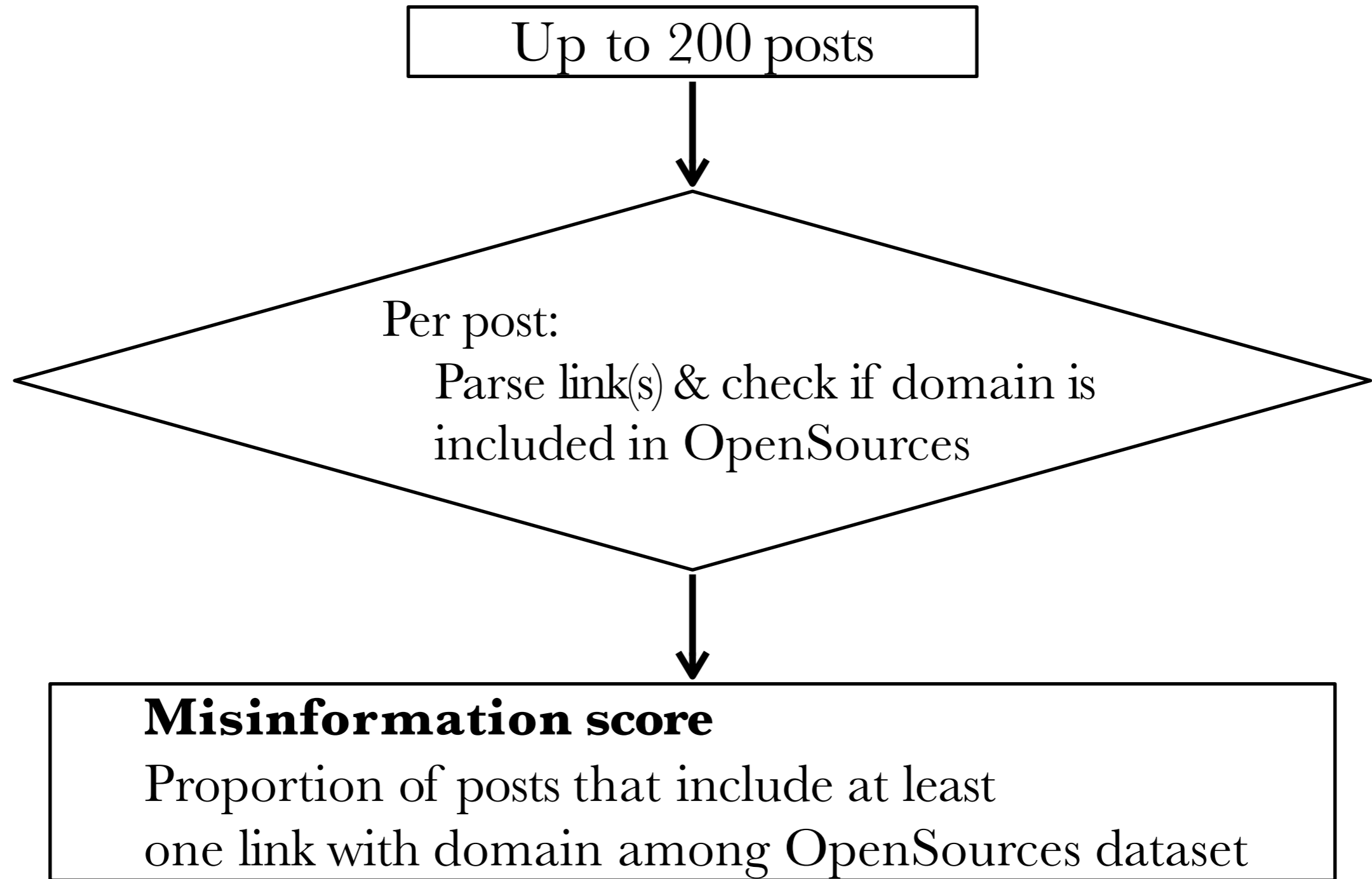
human, researcher, PhD student focusing on HCI

📅 Joined January 2020

Computing Toxicity S3



Computing Misinformation S3



Profile pages



Lucy

@lucy_lim Follows you

Toxicity Misinformation

human, researcher, PhD student focusing on HCI

Joined January 2020



Lucy

@lucy_lim Follows you

Misinformation

human, researcher, PhD student focusing on HCI

Joined January 2020

⊕ ☆ 🔔

Sig
Improving social signals on Twitter!

Maximum **toxic** tweet frequency you would allow : 50%

Maximum **misinfo.** tweet frequency you would allow: 1%

Log out

Profile pages

About **33%** of @jimmmmy605's recent tweets are likely to be toxic.
Below are the most recent toxic tweets of @jimmmmy605.

Here "toxic" is defined as "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion" according to [Perspective API](#).

You know you are an asshole @strangerstran18 ? Your tweets are garbage

You know you are an asshole @strangerstran18 ?

Joined March 2018

#ClearTheShelters
29.2K Tweets

Notification / Timeline

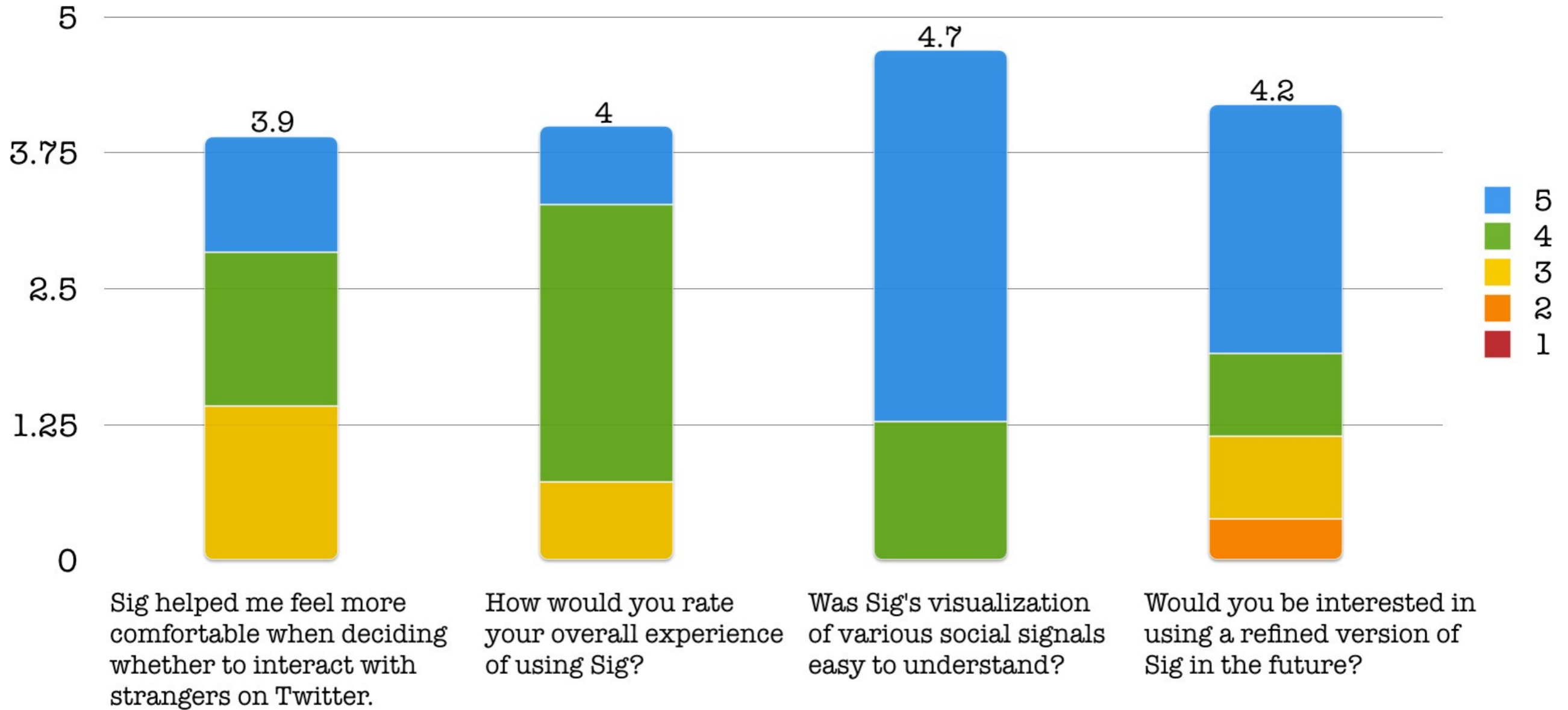
Alerts based on S3 were shown via the color of profile borders



Final Field Study

- ✱ Recruited **11 people from Twitter** by tweeting about our study and using Twitter Ads to promote our tweet.
- ✱ Participants used Sig for at least 30 minutes per day during 4 days.
- ✱ We used **semi-structured interviews** and **online survey**.

Findings



Conventional signals VS S3s

“It was like I figured, cause she had a **blue check mark** and everything, and she’s a politician so it’s kind of funny when you see that [flagged by Sig as misinformation spreading].” –P10



Conventional signals VS S3s

“...if someone retweeted something and I looked on their profile and it was someone with a lot of followers, my instinctual reaction to that is like, ‘Oh, they’re probably fairly legit if they’ve got **30,000 followers,**’ but this made me **rethink that, so those situations.**” –P5

Feeling safer

“I think I would **keep setting it low for general safety**, and then if something seemed like it was flagged incorrectly, then I could just note that to myself. I’d rather risk that than it not flagging people it should.” –P9

Social decision-making

The image shows a Twitter profile for a user named Lucy (@lucy_lim). The profile picture is a circular portrait of a young woman. The header features a scenic landscape background with mountains and a lake. The profile information includes the name 'Lucy', the handle '@lucy_lim', and a bio that reads 'human, researcher, PhD student'. It also shows 'Joined January 2020', '1 Following', and '0 Followers'. A context menu is open over the profile, listing various actions such as 'View Topics', 'Add/remove from Lists', 'View Lists', 'View Moments', 'Copy link to profile', 'Mute @lucy_lim', 'Block @lucy_lim', and 'Report @lucy_lim'. A hand cursor icon is pointing at the 'Block @lucy_lim' option. Below the profile, the 'Tweets' section is visible, showing a tweet from Lucy that says 'You know you are an as'. Another tweet from Lucy is partially visible below it.

Lucy
@lucy_lim Follows you

human, researcher, PhD student

Joined January 2020

1 Following 0 Followers

Not followed by anyone you're following

View Topics

Add/remove from Lists

View Lists

View Moments

Copy link to profile

Mute @lucy_lim

Block @lucy_lim

Report @lucy_lim

Follow

Tweets

Lucy @lucy_lim · 41m
You know you are an as

Lucy @lucy_lim · 46m

Social decision-making

“Double check to see if the flagged tweets matched up with what I thought was a problem, which it usually did...sometimes I’d **mute them or block them** preemptively.” –P9

Reducing receiver costs

“...the extension would be a useful way for me to **quickly get information** without having to scroll back a zillion pages...” –P11

What if S3s are implemented on Twitter?

S3s' downstream effect on communication

“I think people might be a **little more cautious talking to others** and that it would be **harder for misinformation to go viral...**” –P9

“...I wonder if people would **find ways to transverse the extension**. I think it's just me being a pessimist in a sense of finding ways to state things in a way that don't flag toxic for example.” –P2

Future work

Building tools like Sig that use human-in-the-loop to quickly address false positives/false negatives.

Chandrasekharan, Eshwar, et al. "Crossmod: A Cross-Community Learning-based System to Assist Reddit Moderators." *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019): 1-30.

Surfacing results of computation gives people the *agency* to make their own decisions.

Paper: imjane.net/papers/s3-chi2020.pdf

Jane Im  [@im__jane](https://twitter.com/im__jane)

Co-authors: Sonali Tandon, Eshwar Chandrasekharan,
Taylor Denby, Eric Gilbert