SICSS Edinburgh

THE UNIVERSITY of EDINBURGH

CSS in Practice

Predicting User's Attributes

Walid Magdy

29-May-2025

Task/Purpose

- Learn/Predict something about a given user
- Learn/Understand characteristics of certain user groups
- Consideration: ETHICS
- Examples RQs:
    - Can we predict users' actions in the future?
    - Can we predict users' hidden information?
    - What makes a given user have a given leaning?
- Example applications: predicting voting, extremism …

Outlines

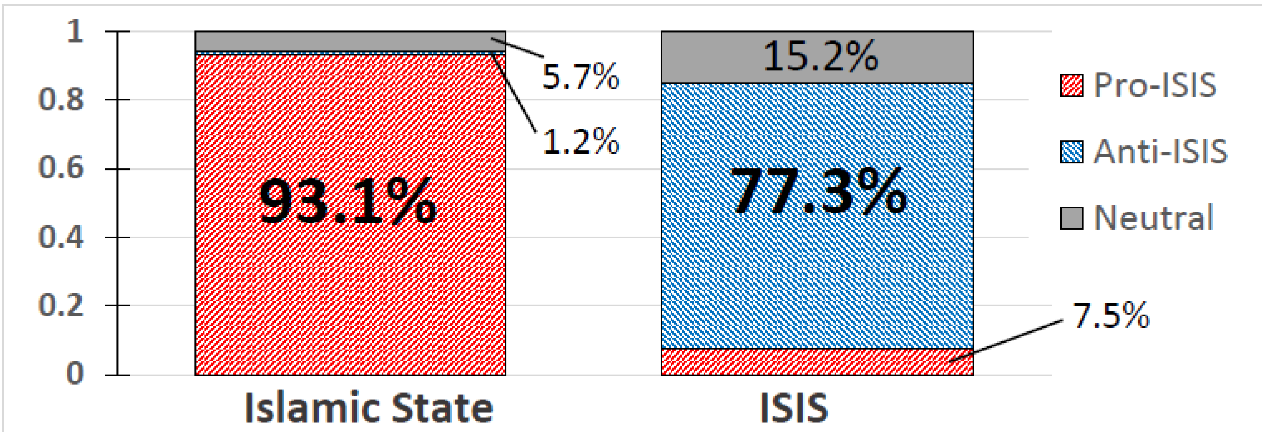
- Examples of using CSS for predicting user's information

- 6 Example Studies
- No technical details (ask if you need details)
- Sharing main methodology
- Topics: might be sensitive!!

Antecedent of Support?

Where ISIS supporters come from?

- Signals of ISIS support is frequently noticed on SM
- Collected 3 million tweets mentioning ISIS
- Labeling:

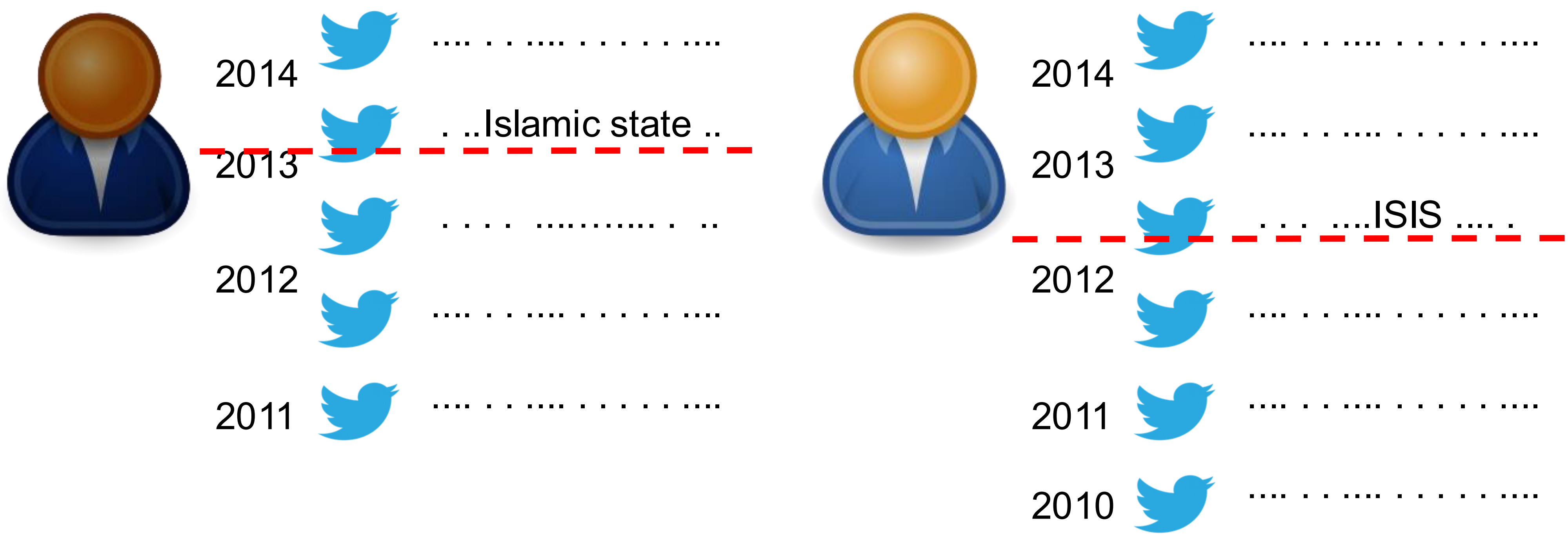


- 57K (11K + 46K) users talking about ISIS (10 tweets at least)

THE UNIVERSITY of EDINBURGH

Data Collection

- Collect tweets timeline for **57K** users → **123 million** tweets
- Identify tweets of users before even mention ISIS
- Filter-out accounts with no pre-ISIS data



THE UNIVERSITY of EDINBURGH

Classifier

- Pro-ISIS accounts with pre-ISIS tweets = **7,700** users
- Balance Data: random select **7,700** Anti-ISIS account with pre-ISIS tweets
- Train classifier with Pre-ISIS tweets
- Mission: Predict if the user will be in the future **Pro-** or **Anti-** ISIS
- Features: tweets content (BOW)

- Accuracy → **87%**
- **Analysis:**
  - Find most distinguishing feats for Pro-ISIS *(before being supporters to ISIS)*

Findings

- **Most distinguishing features:**
  - Related to Arab spring (Egypt, Syria, Libya)
  - Related to protesting against Arab regimes (SA, Kuwait, Iraq)
- **Qualitative**

Date	Tweet (translated)
May 25, 2012	Don't be surprised if it rains today ... martyrs are spitting on us
Nov. 9, 2014	Preliminary schizophrenia: I like ISIS, but I want to watch Chris Nolan's new movie
Nov. 17, 2014	Check the gazes of Bashar's soldiers before slaughter by #Islamic_State in #despite_the_disbelivers

- **Support of ISIS is <u>not ideological</u>, but for <u>revenge</u>**

Methodology

- Use old textual data on user's social media page to predict their potential future leanings

Ref:

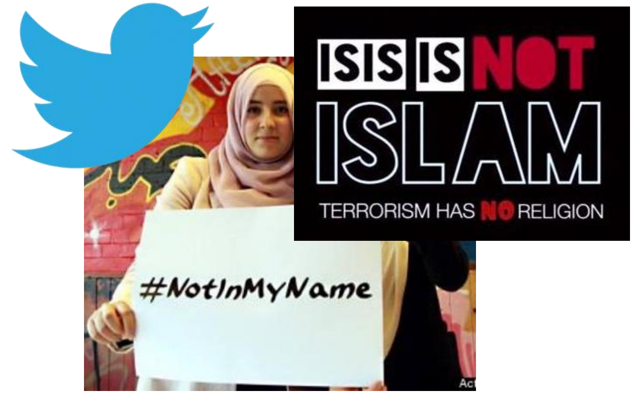
- Magdy W., K. Darwish, and I. Weber. "I like ISIS, but I want to watch Chris Nolan's new movie": Exploring ISIS Supporters on Twitter. *Hypertext 2015*
- Magdy W., K. Darwish, and I. Weber. #FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support. *First Monday, 2016*

Predicting Unseen Views!

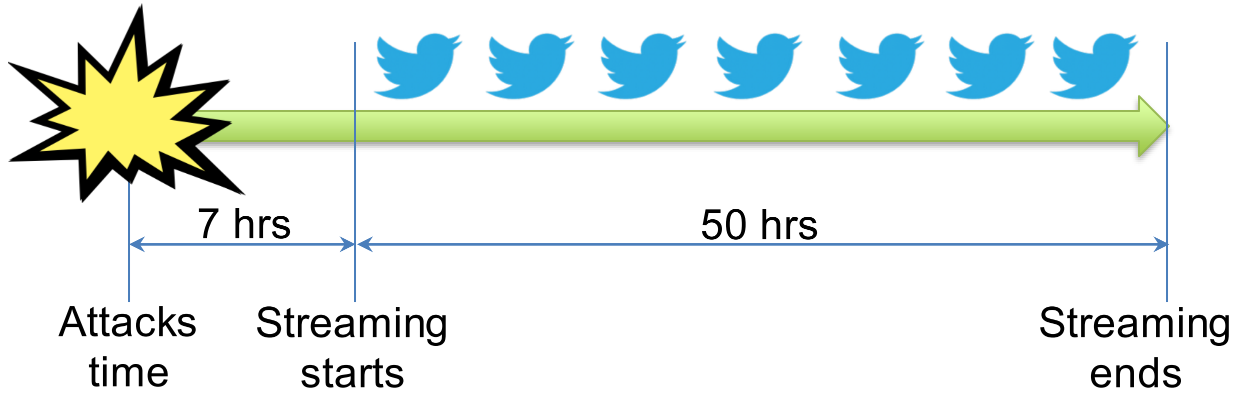
#ParisAttacks

#ParisAttacks – next few hours

#Pray4Paris



#ParisAttacks – Data Collection



- Collection: **8.4 million** tweets about #ParisAttacks
- **900K** tweets talking about Islam
- Sampling + label propagation → **336K** tweets *(in 10 languages)*
  Attacking Muslims / Defending Muslims / Neutral

#ParisAttacks – Top Hashtags

Positive	Count	Negative	Count
#MuslimsAreNotTerrorist	34,925	#IslamIsTheProblem	3,154
#MuslimAreNotTerrorist	17,759	#RadicalIslam	1,618
#NotInMyName	4,728	#StopIslam	1,598
#MuslimsStandWithParis	1,228	#BanIslam	460
#MuslimsAreNotTerrorists	1,106	#StopIslamicImmigration	333
#ThisisNotIslam	781	#IslamIsEvil	290
#NothingToDoWithIslam	619	#IslamAttacksParis	280
#ISISareNotMuslim	316	#ImpeachTheMuslim	215
#ExtremistsAreNotMuslim	306	#KillAllMuslims	206
#ISISisNotIslam	243	#DeportAllMuslims	186

Research Questions

- Can we predict user stances?
- What if the user never talked about the topic before?
- What are the key features?

- US-based polarized users → 44K users
- Latest 400 tweets/user before attacks + Profile info
- 12.6M tweets + Network interactions + Profile info

Predicting Stances

- **Features:**
  - **Content:** BOW, hashtags
  - **Profile:** name, desc., location
  - **Network:** retweets, replies, mentions
- **SVM + linear kernel**
- **10-fold cross-validation**
- **Divide to:**
  - **Mentioned Islam before** (10.5K → 6.6K+4K)
  - **Never mentioned Islam** (33.5K → 27.5K+6K)

Results

Features Set	Mentioned-before		Never mentioned-before	
	Accuracy	*F-score*	*Accuracy*	*F-score*
Content	0.83	0.82	0.84	0.73
Profile	0.73	0.70	0.79	0.62
Network	**0.86**	**0.85**	**0.88**	**0.77**
All	0.85	0.84	0.87	0.76

- **Network interactions are the most effective features**
- **Predictability is high even for users never mentioned the topic before**

Feature Analysis

Defending Muslims

THE HUFFINGTON POST

THE NEW YORKER

BLACK LIVES MATTER

LOVE WINS

NRA

EDM
ELECTRONIC DANCE MUSIC

Acabella

Attacking Muslims

FOX NEWS

NFL

NHL

b theblaze

NASCAR

DRUDGE REPORT

#TCOT

Lessons

- **People's unspoken views are predictable**
- **User's network is a key factor for future behavior**
- **Humans tend to group into homophily, even on SM**

Ref:

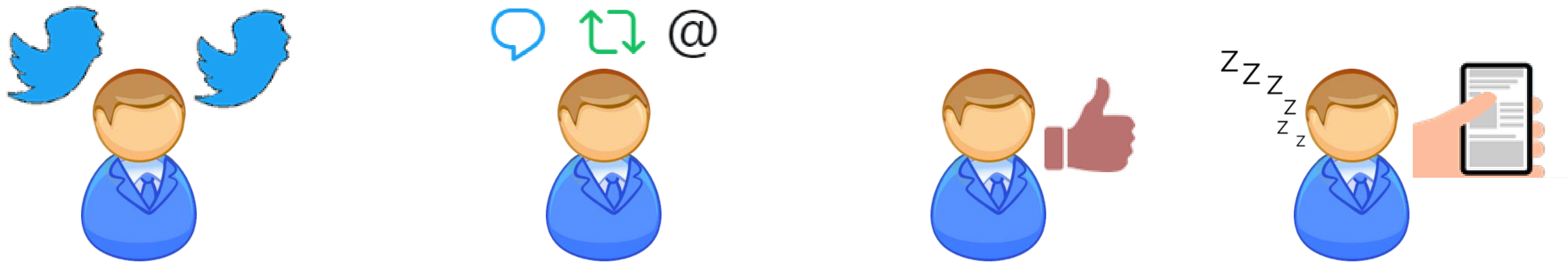
- Magdy W., K. Darwish, A. Rahimi, N. Abukhodair, T. Balswin. #ISISisNotIslam or #DeportAllMuslims? Predicting Unspoken Views. *Web Science 2016*
- Darwish K., W. Magdy, A. Rahimi, N. Abukhodair, T. Baldwin. Predicting Online Islamophopic Behavior after #ParisAttacks. *Journal of Web Science 2017*

What can reveal your Stance?

SemEval Stance Detection Task 2016

- **4K** tweets on **5 topics** labeled by stance {for, against, none}
- Topics: *Abortion, Atheism, Feminism, Clinton, & Climate Change*
- State-of-the-art:
  - SVM + n-gram features → F-score **69%**
  - Other approaches: deep learning → F-score < 69%
  - Focus on content features only! → user discussed the topic!
- RQ: How detecting stance could be done if:
  - User never discussed the topic!
  - User never tweeted, but has some online activity!
  - User has no content and no activity!

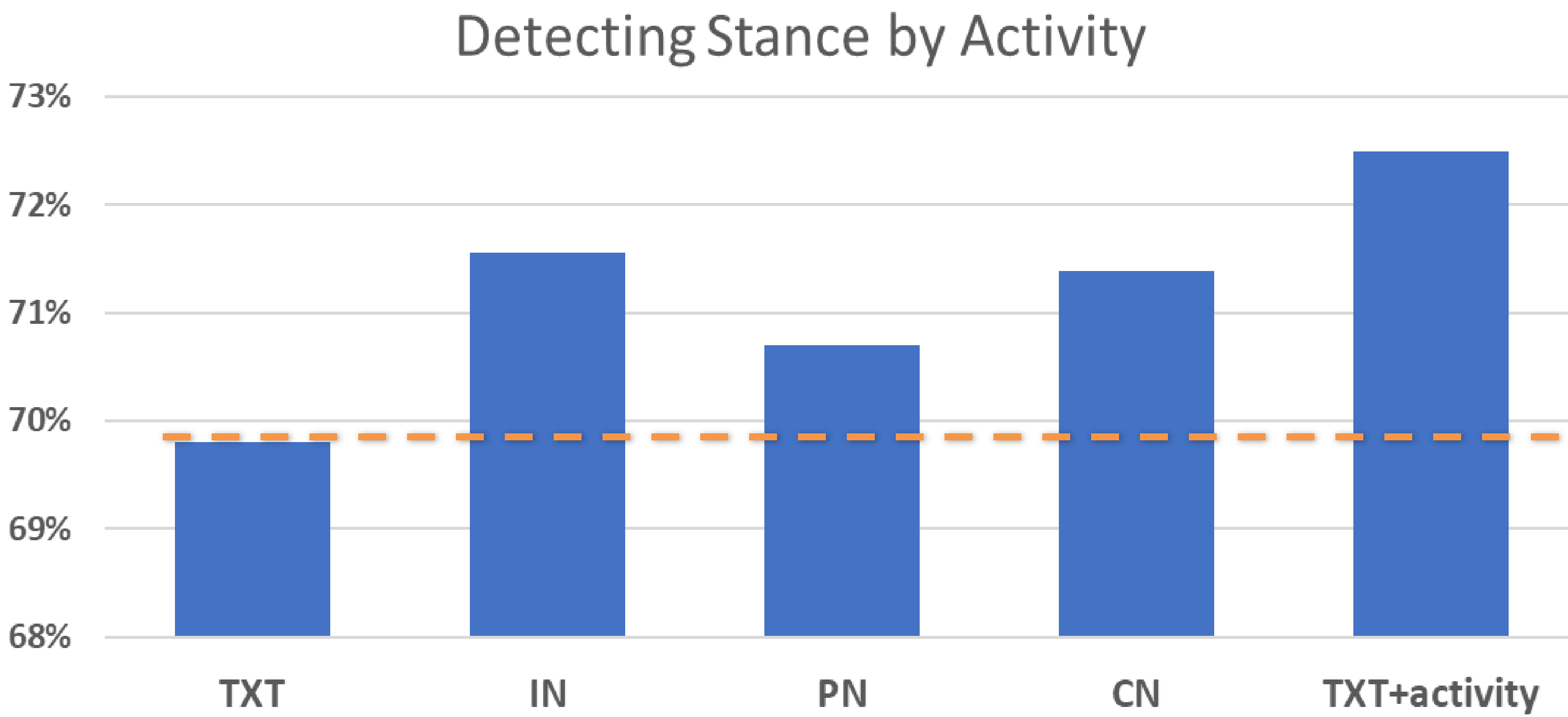
Detecting Stance in Four Situations



On-topic content | General activity | Silent User | Passive User

- TXT: tweet Text content
- IN: Interaction Network → network user retweet, reply, mention
- PN: Preference Network → network in tweets user like
- CN: Connection Network → network user follows

What can reveal your stance?



what you say | *your activity* | *your preferences* | *your existence* | *combine them*

Lesson

- **Every activity for us online can give indication about our stances and leanings, whether we express it or not!**

Ref:

- Aldayel A and W. Magdy. Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *CSCW 2019*

Fake Accounts / Catfishes
Can your style online show you are fake?

THE UNIVERSITY of EDINBURGH

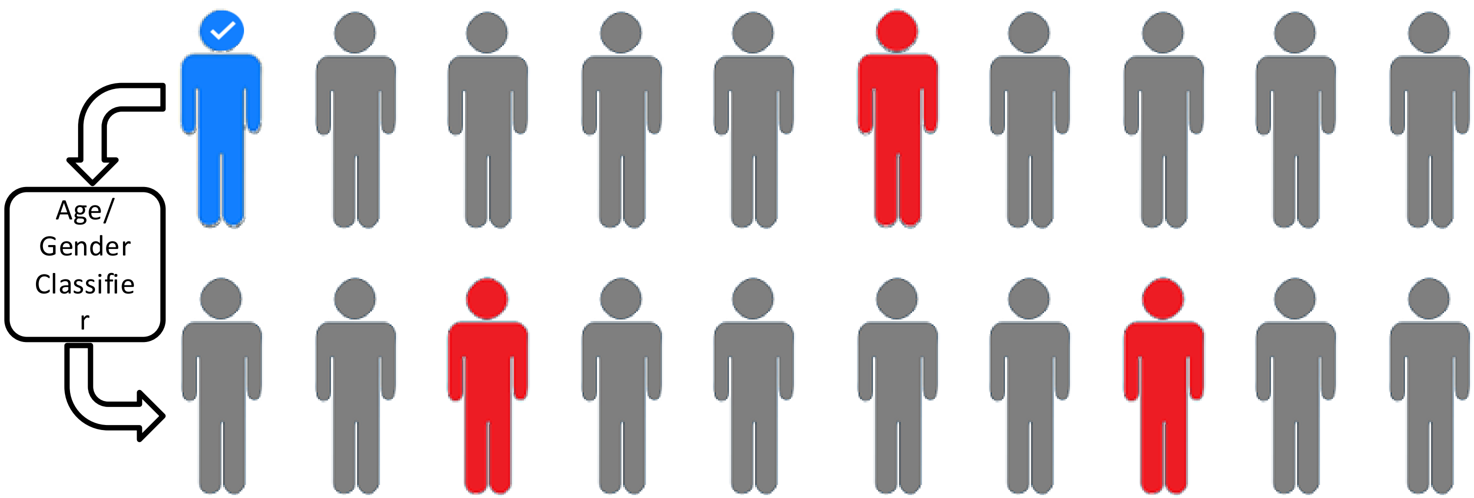
Catfishes

- **Catfish: pretends different person**

Data/Approach

Pornhub

- 100K user accounts
- 5% verified accounts



Data Sample


```
<USER>
    <ID>8423012381</ID>
    <TYPE>0</TYPE>  <!-- 1 for verified accounts -->
    <AGE>27</AGE>
    <PROFILE>Girl,Single,Girls,united%states</PROFILE> <!-- gender, status, interested in,
    country -->
    <ACTIVITY>7611,202</ACTIVITY> <!-- #videos watched, #posts -->
    <FRIENDS>1350,594,76,680</FRIENDS> <!-- #total friends, males, females, others -->
    <SUBSCRIBERS>605,246,25,334</SUBSCRIBERS> <!-- same as above -->
    <SUBSCRIBEDTO>106,40,7,59</SUBSCRIBEDTO> <!-- same as above -->
    <NCOMMENTS>11,10,87,54</NCOMMENTS> <!-- #comments, #unique comments, #words, #unique
    words -->
    <COMMENTS>
        <comment count=2>thanx sugar</comment>
        <comment count=1>I can definitely say the same! You are damn sexy!</comment>
        <comment count=1>Awwwww you guys are so sweet to this poor horny girl!!!</comment>
        <comment count=1>Thanx you all are real sweet</comment>
        <comment count=1>Mmmmmm PLEASE!!!!</comment>
        <comment count=1>Well thank you sugar!</comment>
        <comment count=1>Ohhh yeah I think I could</comment>
        <comment count=1>Heheheehehe thanx sugar!</comment>
        <comment count=1>No problem...my pleasure in fact!</comment>
        <comment count=1>Damn you are one fine woman!!!! *kiss kiss*</comment>
        <comment count=1>GRRRR I NEED SOMEONE XXXXXX MY XXXX AND XXX RIGHT NOW!!!</comment>
    </COMMENTS>
</USER>
```

Predicting Gender & Age

Gender

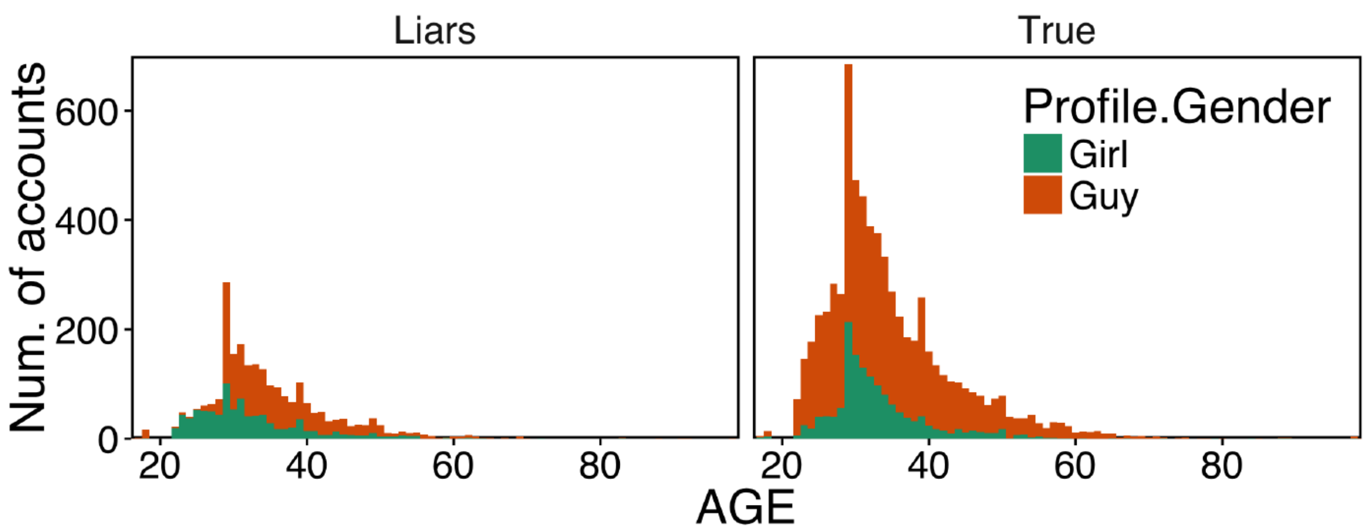
Features	Accuracy
Comments	85.9%
Network/Activity	88.7%
All	**92.0%**

Age

Features	Correlation	MAE (yrs)
Comments	**0.509**	**5.58**
Network/Activity	0.234	6.12
All	0.440	5.78

Who Catfishes?

- **25%** are likely lying about their age
- Males pretend to be young females
- Females pretend to be older males



Lesson

- **From your online public social activity, hidden personal information could be estimated**

Ref:

- Magdy W., Y. Elkhatib, G. Tyson, S. Joglekar, N. Sastry. Fake it till you Make it. Fishing for Catfishes. *ASONAM 2017*

Prediction is sometimes Scary!

Demographic Prediction

- Different research showed that demographic information can be predicted about users using their posts
- Experiment:
  - Collected timelines of **20,000** Twitter users (WW, NYC, London, Johannesburg)
  - Annotate <u>ethnicity</u> and <u>gender</u> of user based on profile pic.
  - Use their posts to predict their demographics

Classifier	Accuracy
Gender	80.6%
Ethnicity	85.2%

Male
Female

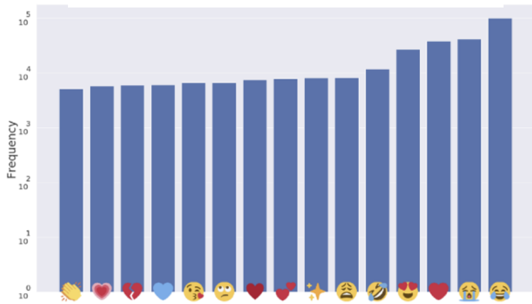
White
Black
Others

What about the Emoji they use?

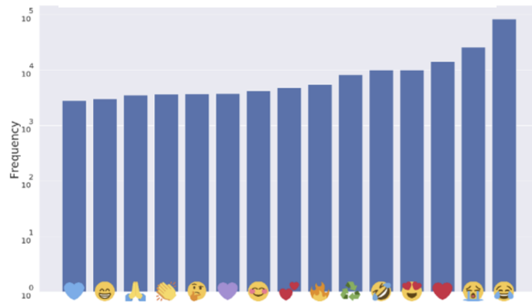
- There are **2600** emoji out there!
- **20%** of social media posts contain emoji
- Does general emoji usage differ by demographic?
- Can we use them for prediction instead of text?

Top Used Emoji

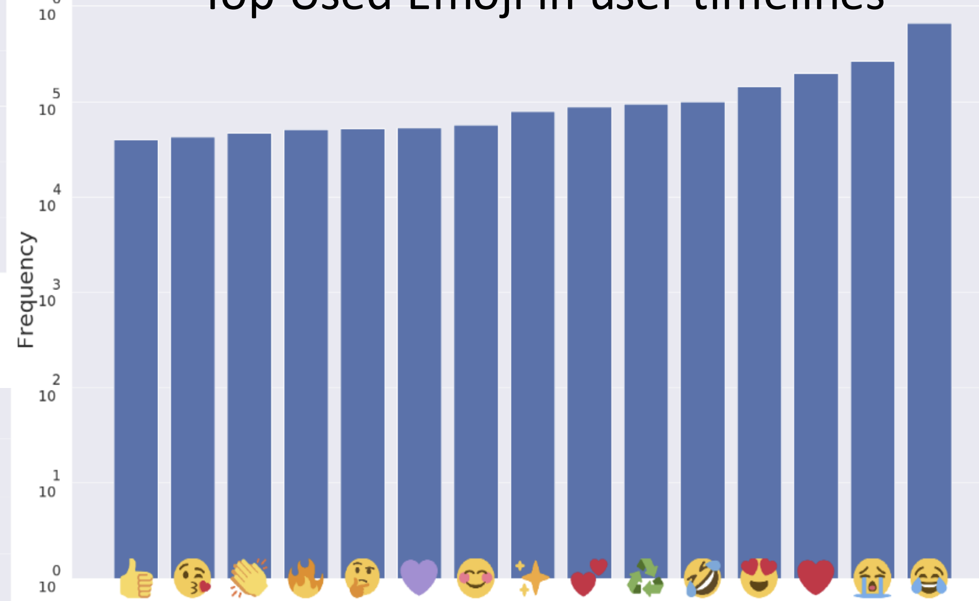
Top Emoji by **Female** users



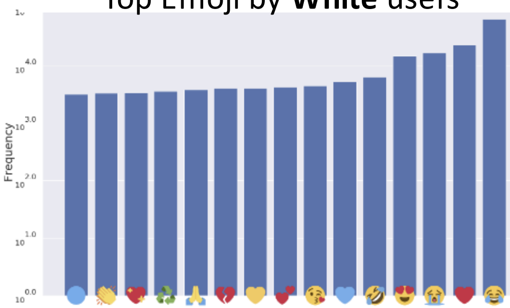
Top Emoji by **Male** users



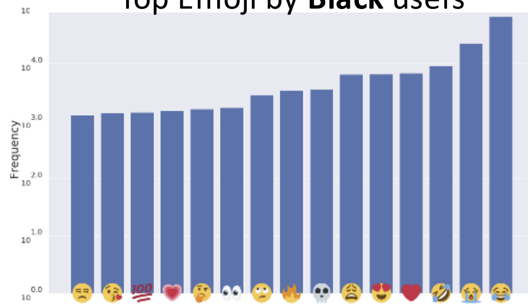
Top Used Emoji in user timelines



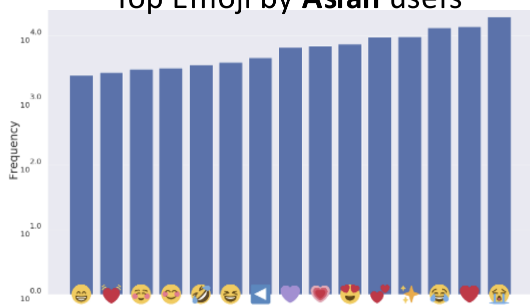
Top Emoji by **White** users



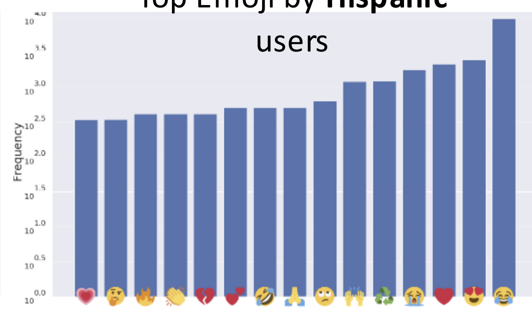
Top Emoji by **Black** users



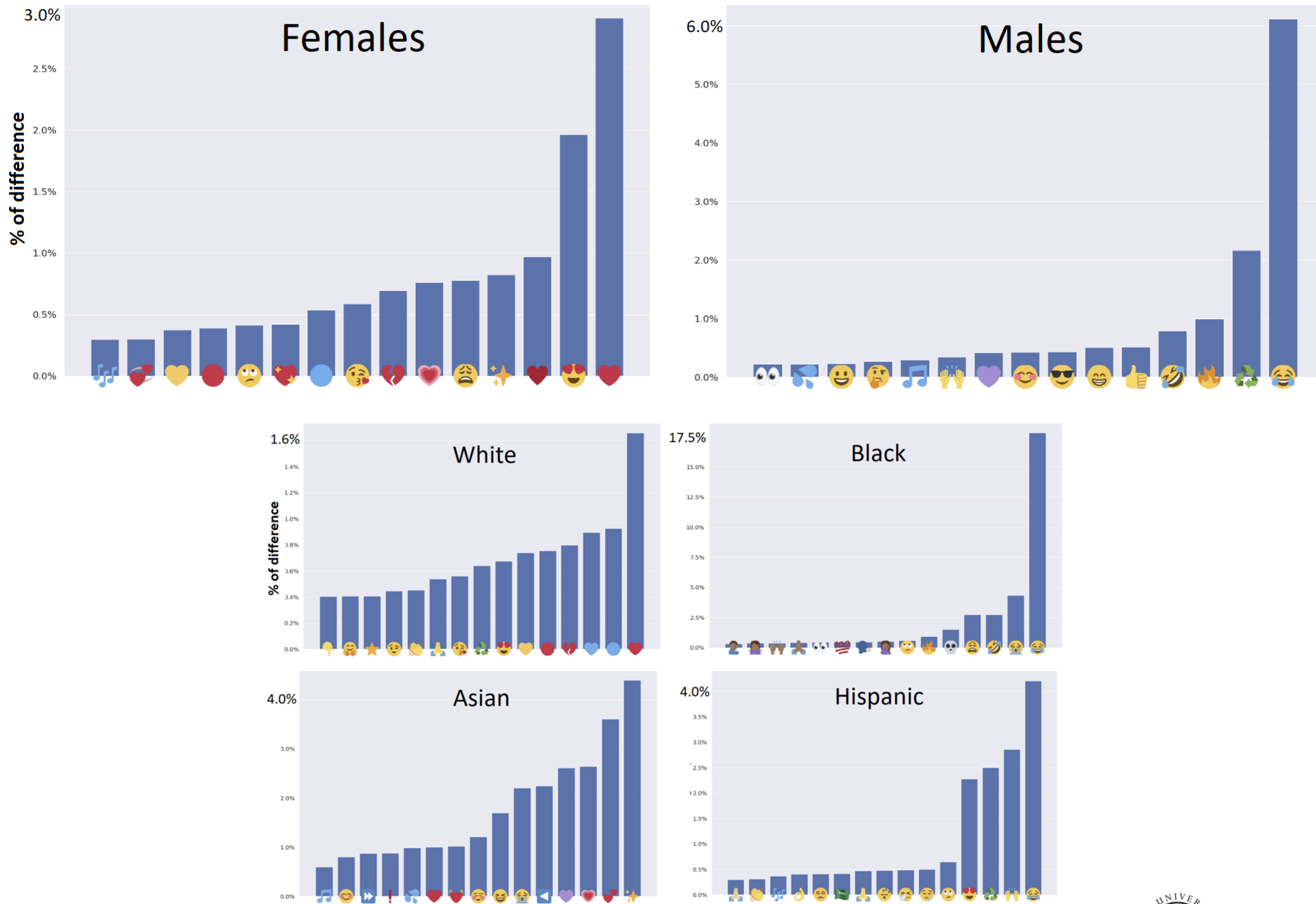
Top Emoji by **Asian** users



Top Emoji by **Hispanic** users



Difference in Usage by Demographics



Demographics Estimation by Emoji

- Are these differences enough to estimate user's demographic information?
- Train classifiers for gender/race using emoji in timeline only!

Classifier	Posts	Emoji
Gender	80.6%	80.4%
Ethnicity	85.2%	84.5%

Lesson

- Minor signals in our posts online, such as emoji, can tell a lot about our identity, including gender and ethnicity.

Ref:

- Benkhedda Y., P. Xiao, and W. Magdy. Emoji are Effective Predictors of User's Demographics. ASONAM 2023

Takeaways

- We have many signals and footprints online
- These signals can be used to infer information about us.
- We can predict user's leanings, biases, and demographics
- CSS is a method to make us learn about ourselves and our societies in a fast way → ETHICS
- Can be an excellent informative motivation for more in depth social studies.

- Presented work was using public social media data,
  → Think about the data that companies own!

Final Words

- **Computational Social Science**
- **Interdisciplinary field**
  - Social Science + Computer/Data Science
- **Conferences**:
  - CSCW, ICWSM, CHI, the WebConf, WebSci, SocInfo, ASONAM
- **Journals**:
  - Nature Human Behavior, ACM TSC, Springer SNAM
- **NLP**:
  - Models: Huggingface
  - Tasks: SemEval

Thank You

wmagdy@inf.ed.ac.uk

@walid_magdy

http://homepages.inf.ed.ac.uk/wmagdy/

SOCIAL MEDIA ANALYSIS AND SUPPORT FOR HUMANITY

@SMASH_Edin

https://smash.inf.ed.ac.uk/

THE UNIVERSITY of EDINBURGH