SICSS Edinburgh

THE UNIVERSITY of EDINBURGH

CSS in Practice
Global Analysis

Walid Magdy

16-Jun-2023

Task/Purpose

- Event, Phenomena, Topic
  → understand what is going on (social science RQ)
- Important aspects:
  Trends, change over time
- Interest → **Global Analysis**
- Method → Collect data, label (classify), analyse

- Examples:
  - Elections, polarization, responds after major event ... etc.

Outlines

- Examples of using CSS for global analysis

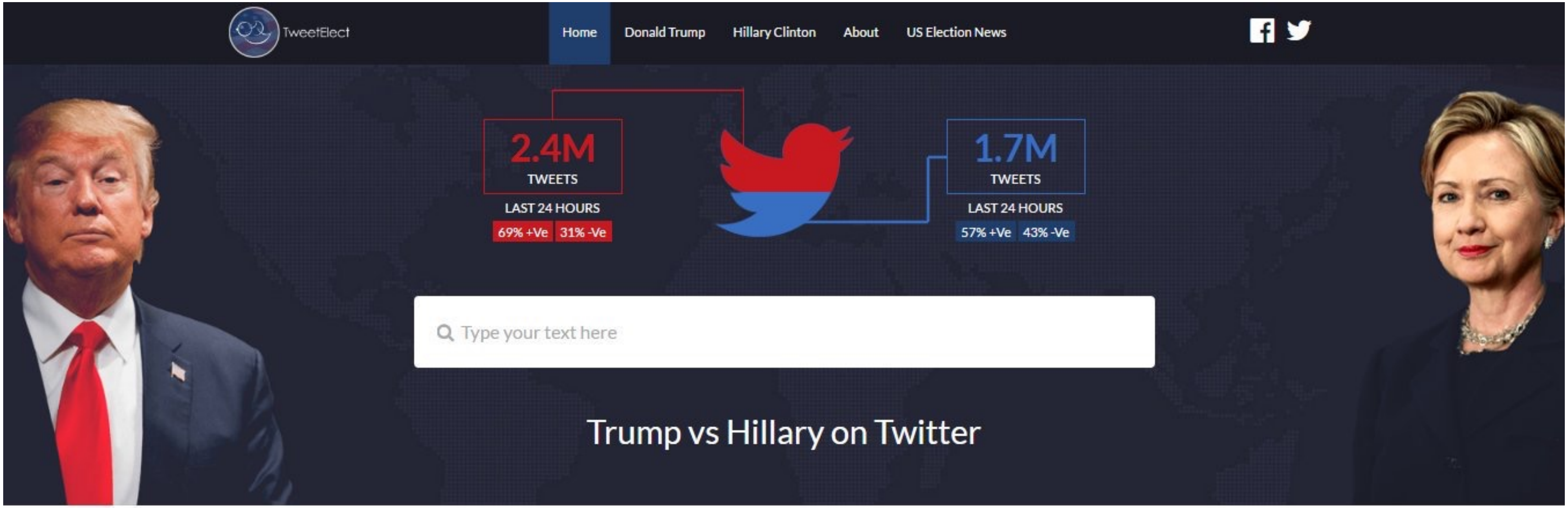
- 6 Example Studies
- No technical details (ask if you need details)
- Sharing main methodology
- Topics: might be sensitive!!

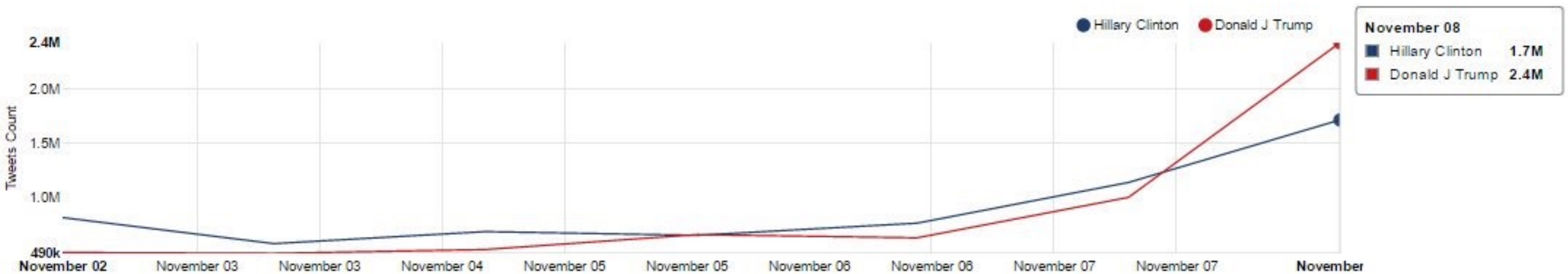
Manual Labeling

What Viral Posts can Tell us?

US Election

TweetElect.com



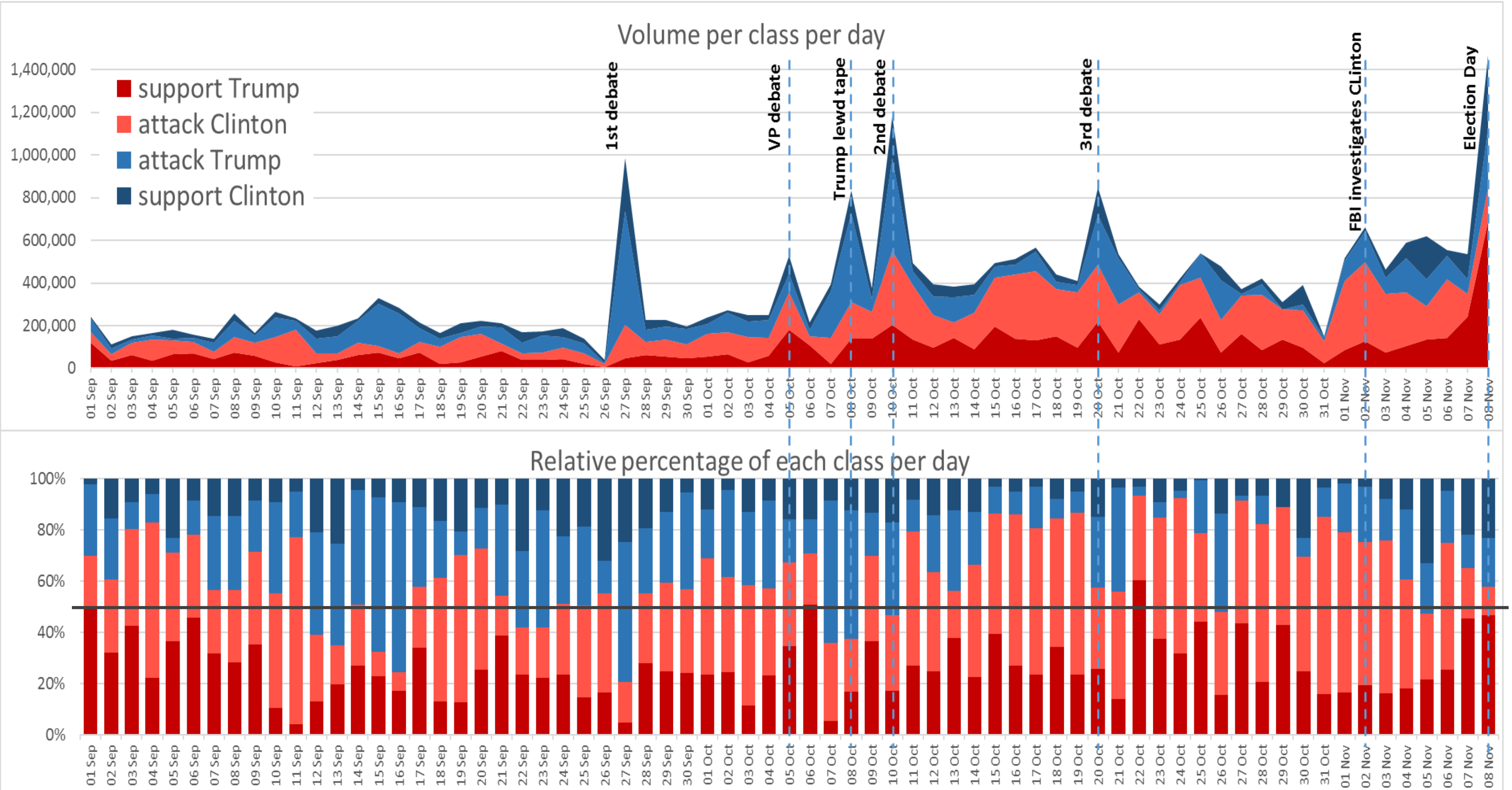


THE UNIVERSITY of EDINBURGH

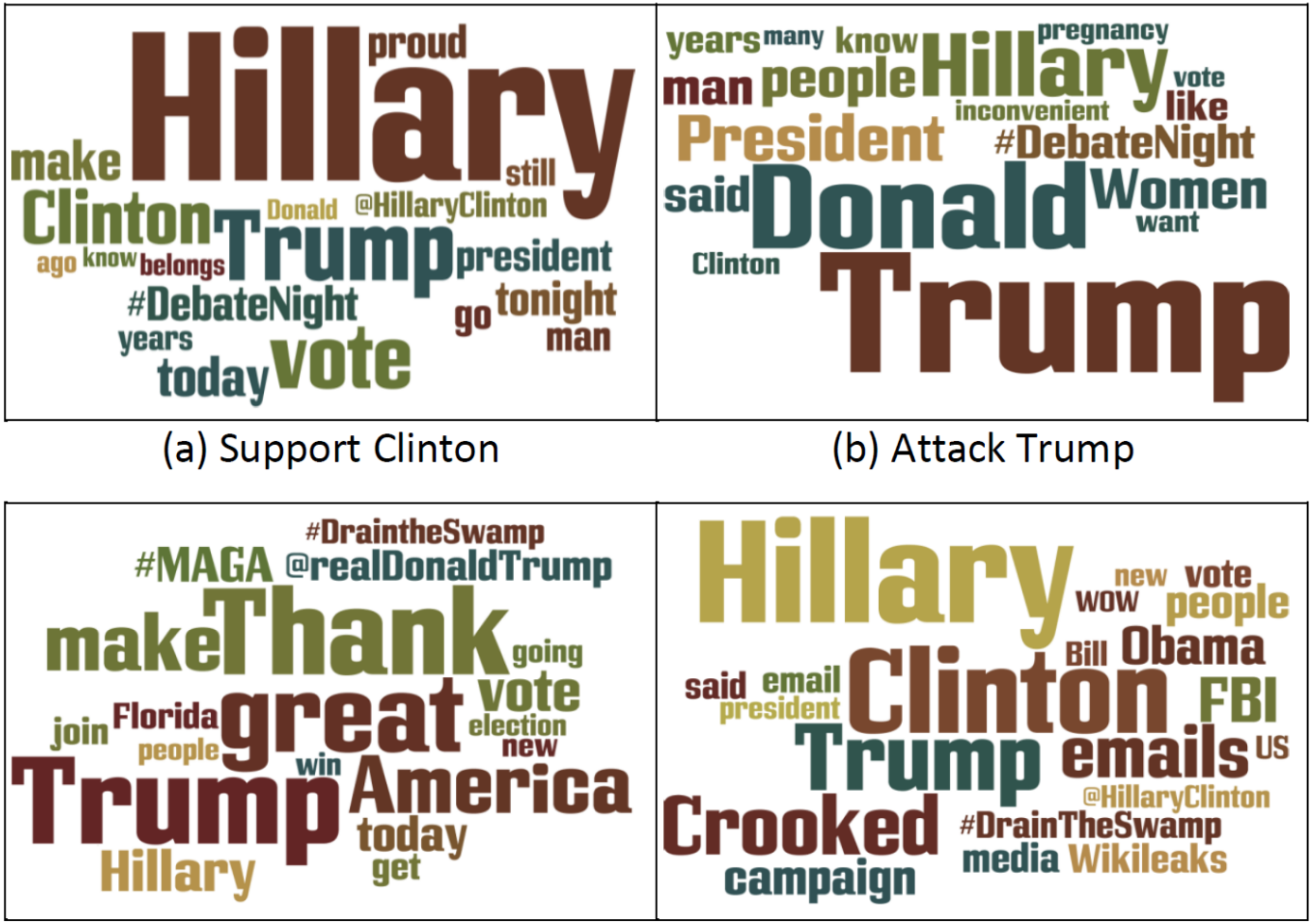
TweetElect.com (data collection)

- 38 keywords related to US Election
- Study period: 1 Sep 2016 – 8 Nov 2016 (election day)
- Total tweets: 6.8M tweets → 65.8M retweets
- Study:
  - Most 50 viral daily tweets
  - 3450 tweets → 26.6M retweets (40% of full volume)
- Label: support/attack Trump/Clinton or neither

Support/Attack Volume



Most Discussed Topics



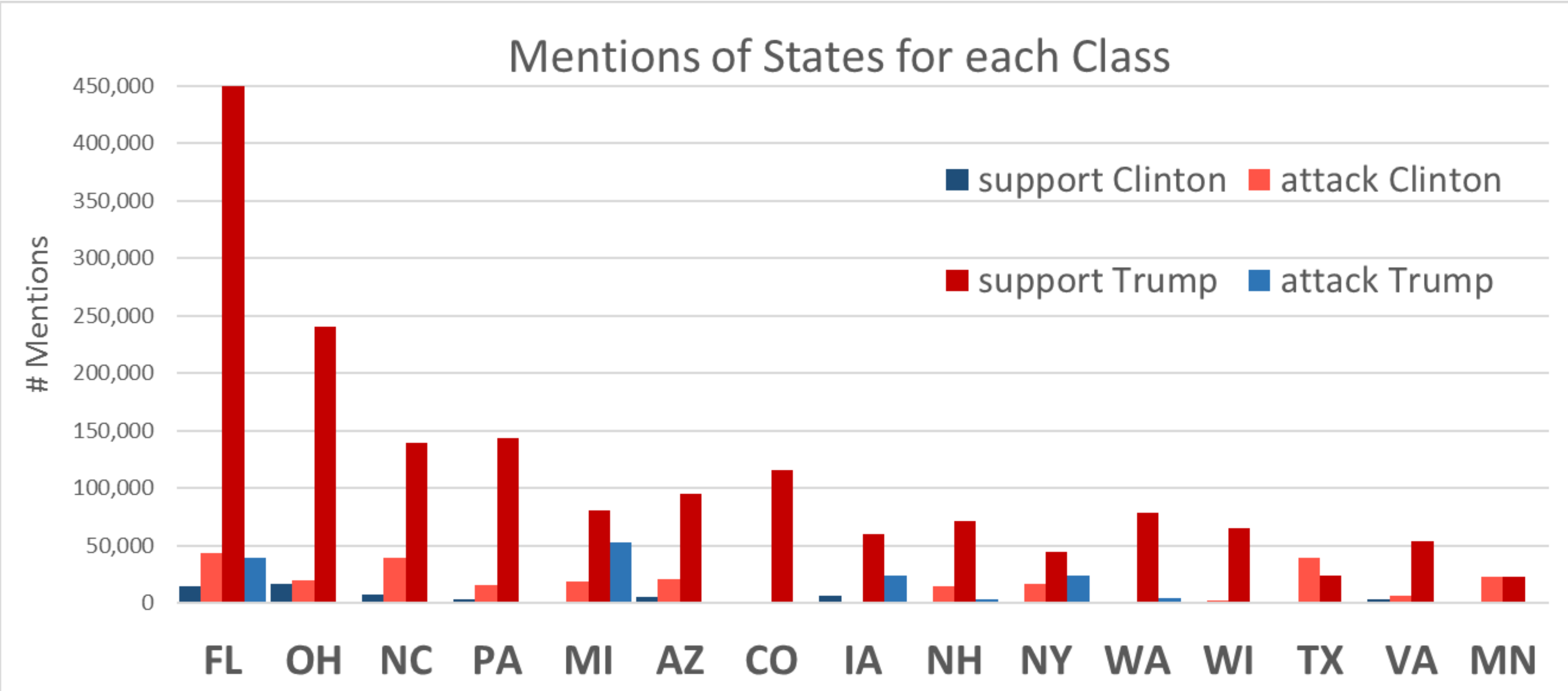
(a) Support Clinton

(b) Attack Trump

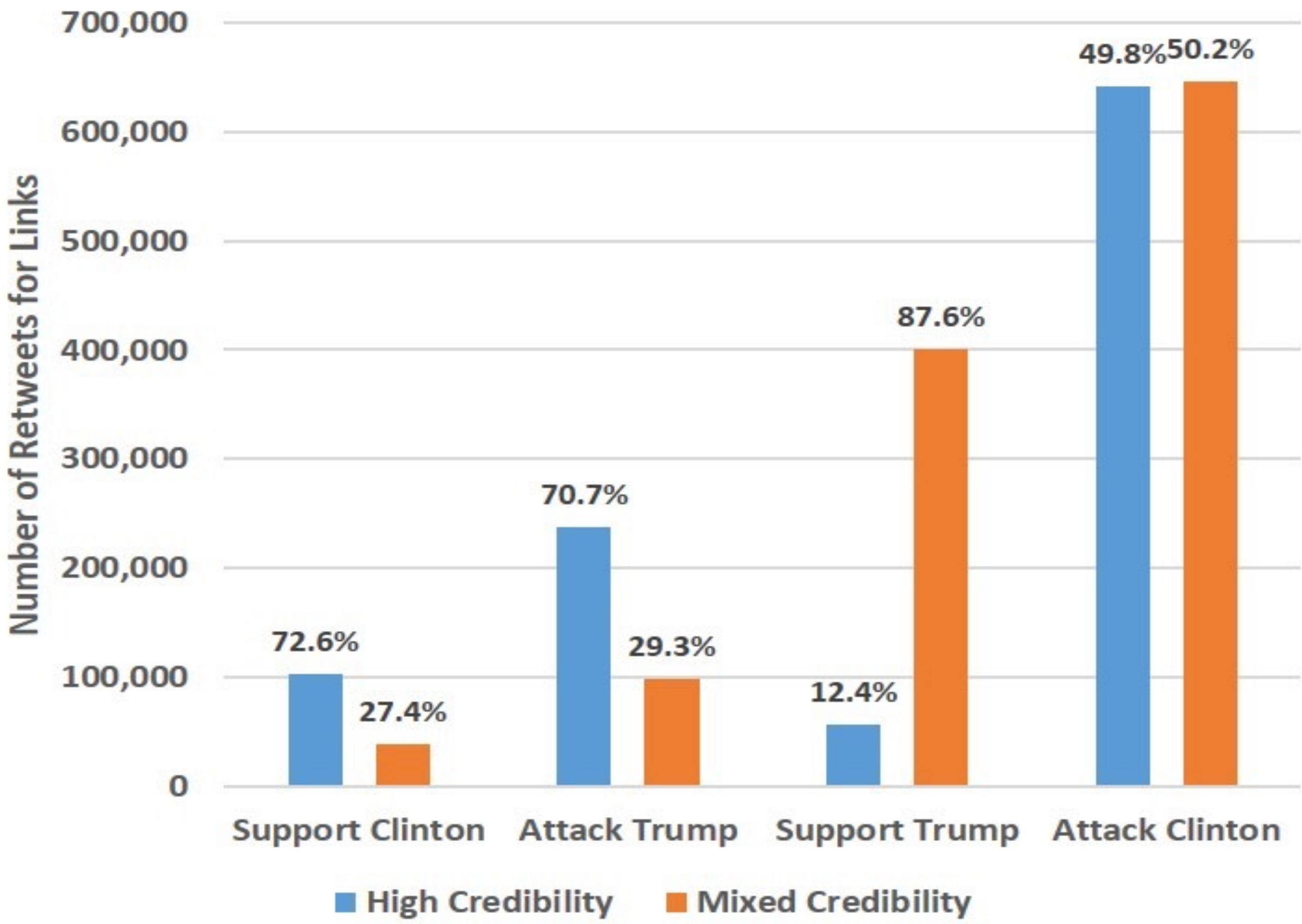
(c) Support Trump

(d) Attack Clinton

Mentions of States



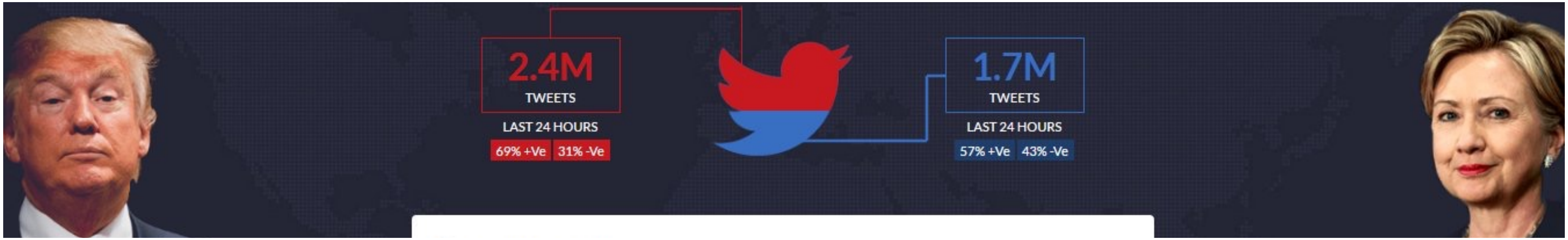
Fake News (Credibility)



Top Accounts

support Clinton			attack Trump		
Account	Count	Volume	Account	Count	Volume
Hillary Clinton	331	2,025,821	*Hillary Clinton*	363	2,698,209
President Obama	4	122,947	Bernie Sanders	11	304,860
Senator Tim Kaine	15	84,245	Ozzyonce	1	152,756
Jerry Springer	1	78,872	Bailey Disler	1	124,322
Erin Ruberry	1	72,167	Stephen King	2	121,635
Richard Hine	1	66,817	Kat Combs	1	105,118
Bernie Sanders	7	46,180	Es un racista	1	102,063
CNN	6	41,983	Rob Fee	1	99,401
Funny Or Die	1	27,909	Jerry Springer	1	78,872
Channel 4 News	1	27,409	Master of None	1	67,690
Support Trump			attack Clinton		
Donald J. Trump	446	4,992,845	*Donald J. Trump*	246	3,613,025
Kellyanne Conway	51	199,511	WikiLeaks	141	1,454,903
Mike Pence	36	195,824	*Kellyanne Conway*	92	349,025
Dan Scavino Jr.	40	181,601	Paul Joseph Watson	78	297,273
Official Team Trump	14	141,289	*Official Team Trump*	23	150,932
Donald Trump Jr.	20	112,835	*Donald Trump Jr.*	34	126,744
Eric Trump	8	79,387	Jared Wyand	19	85,742
Immigrants4Trump	4	52,256	Cloyd Rivers	7	84,063
Cloyd Rivers	2	45,493	Juanita Broaddrick	4	83,903
Paul Joseph Watson	10	38,474	James Woods	16	78,719

What viral tweets can tell us?



	Trump	Clinton
Volume in favour	63%	37%
Days in favour	85%	15% (with larger volumes)
Supporters	Continuously active	Active on specific events
Focus	42% support candidate 58% attack opponent	32% support candidate 68% attack opponent
Campaign slogan	Well spread	Low attention
Links Credibility	Mostly mixed credibility	50%-50% high-mixed credibility
Most attacked	After debate / leaked video tape	FBI / WikiLeaks
Swing states	Large focus	Almost no focus

Methodology

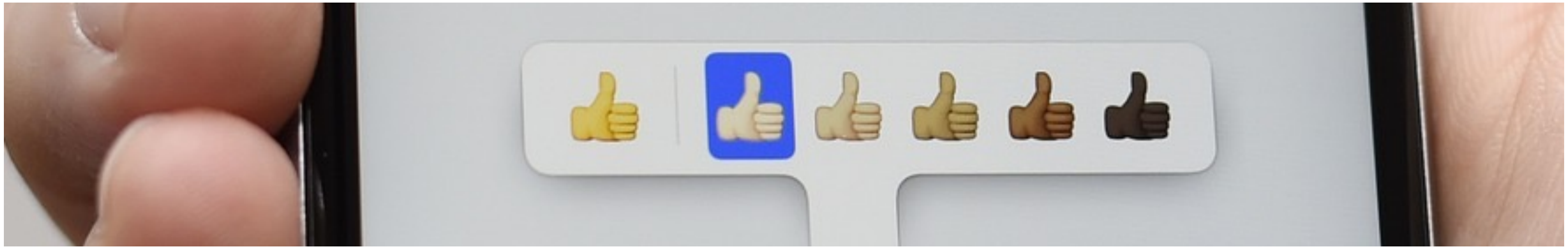
- Manual labeling of few trending tweets
- Expand to large number of retweets
- Focus on specific aspects (users, links, entities … etc)

Ref:

- Darwish K., W. Magdy, and T. Zanouda. Trump vs. Hillary: What went Viral during the 2016 US Presidential Election. *SocInfo 2017*

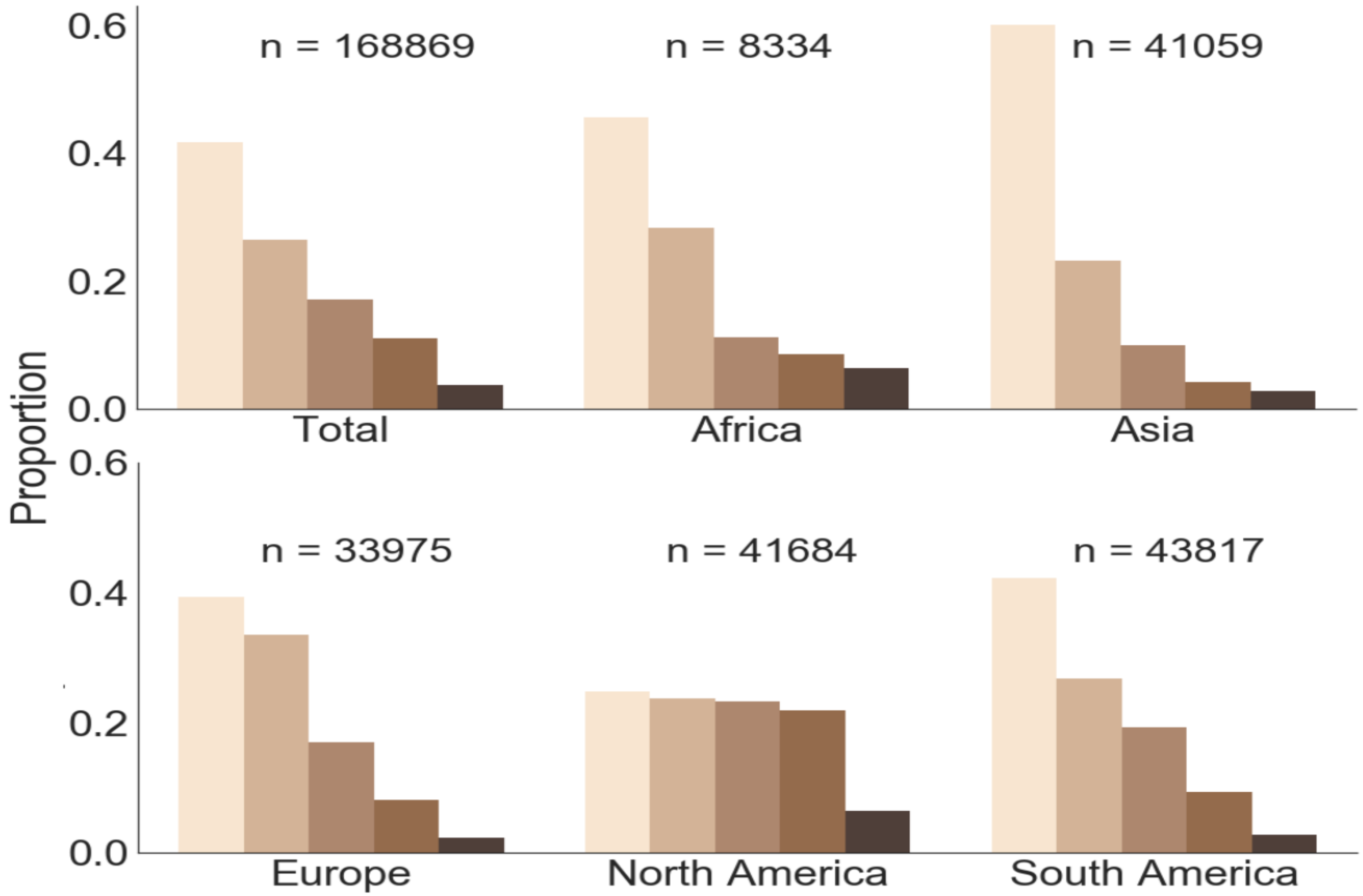
General Usage of New Features on Social Media

Usage of Emoji skin-modifier

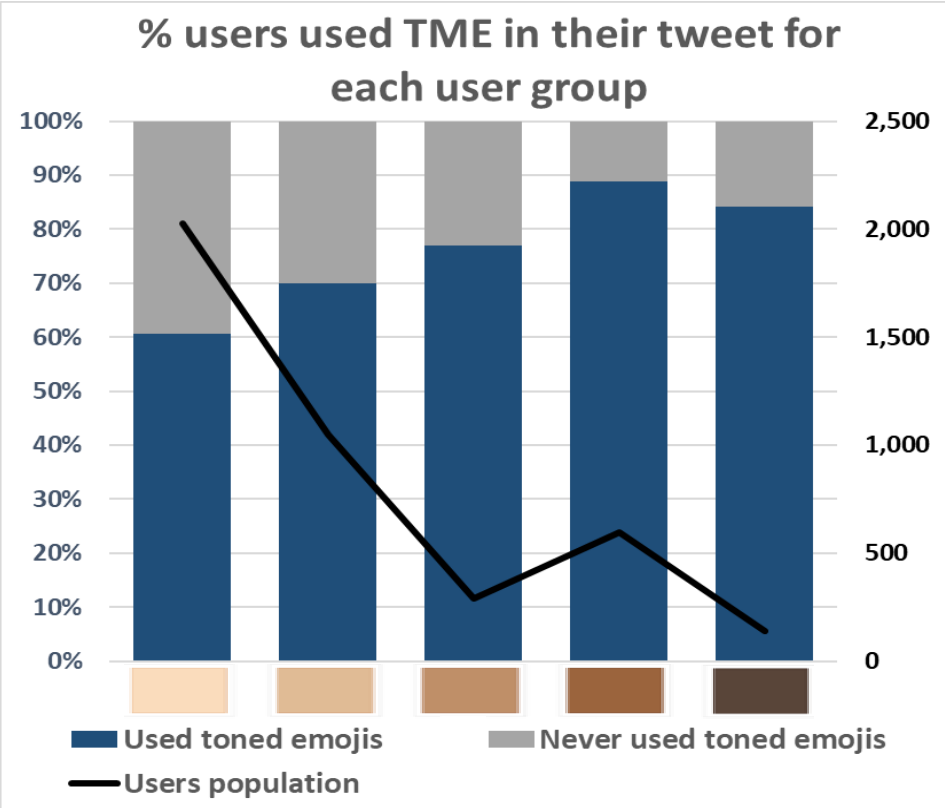


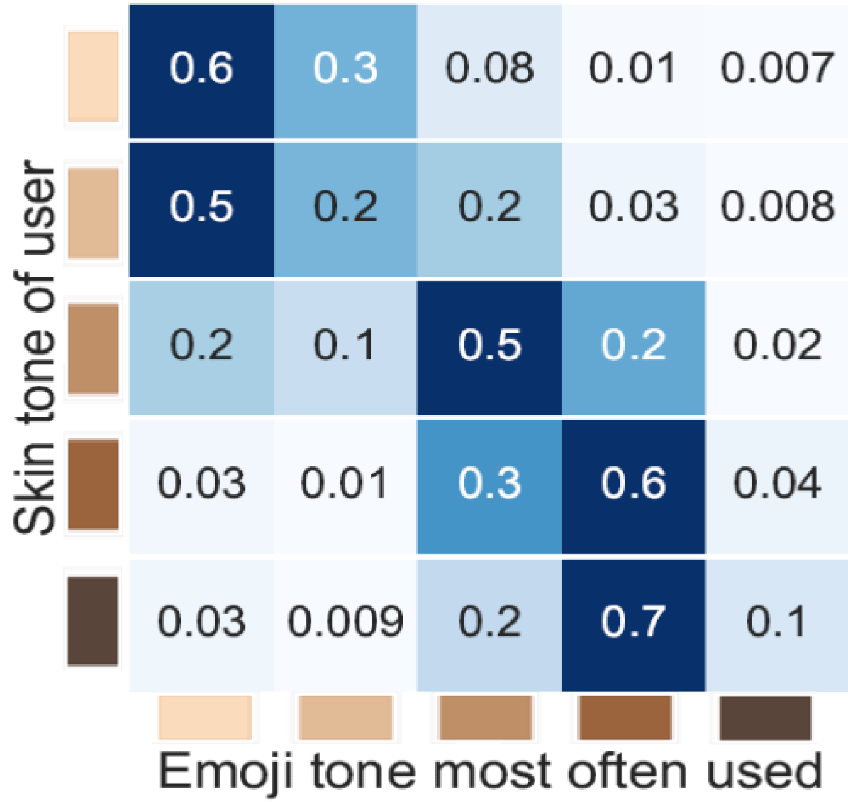
- Collection: **>1 billion** tweets from around the world
- Detect tweets with skin-modifiable emoji
- Check how skin color is modified by location
- Select **4,100** users (with their 7M tweets) who used these emoji
- Compare user skin color (manually annotated) to used color in emoji

Emoji Skin-Tone by Region



Who modifies the color? How?



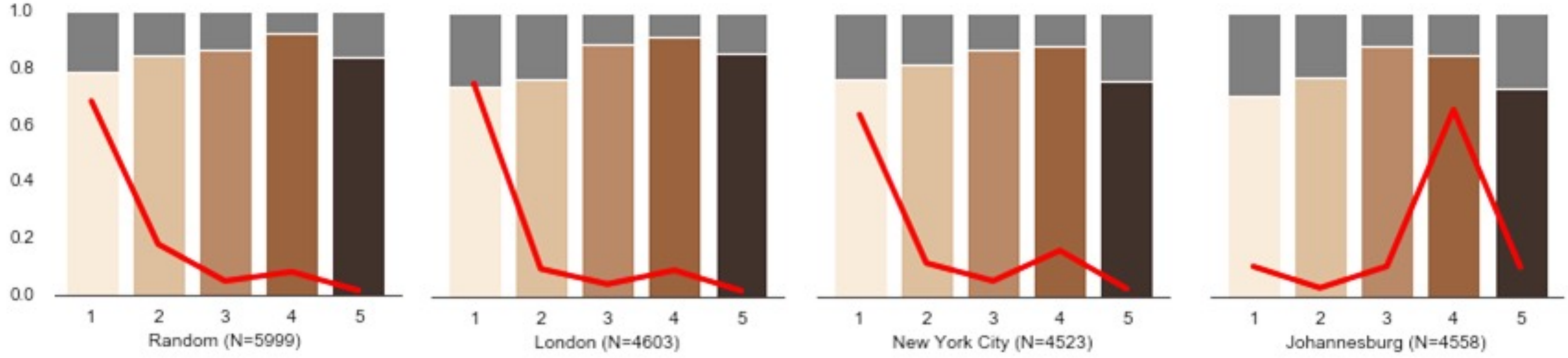


- Darker-skin users are more keen to modify skin-tone of emoji
- The majority of users use skin tones similar to their skin

Generalizability of Findings

- Our sample is random set of users from all over the world!
- Are these findings consistent over multiple samples/locations?
- Labeled a new random set of ~6000 accounts
- Labeled additional accounts, but from specific locations:
  - **London** (~4600 accounts)
  - **New York City** (~4500 accounts)
  - **Johannesburg** (~4500 accounts)
- Repeated the analysis

Generalizability of Findings



Random	1	2	3	4	5
1	0.6	0.3	0.08	0.01	0.01
2	0.5	0.2	0.2	0.02	0.02
3	0.3	0.1	0.4	0.2	0.02
4	0.04	0.04	0.3	0.6	0.06
5	0.04	0.04	0.1	0.6	0.2

London	1	2	3	4	5
1	0.6	0.3	0.09	0.01	0.01
2	0.3	0.3	0.4	0.08	0.004
3	0.05	0.03	0.4	0.4	0.05
4	0.02	0.006	0.2	0.7	0.1
5	0.02	0	0.08	0.7	0.2

New York City	1	2	3	4	5
1	0.5	0.4	0.1	0.01	0.01
2	0.2	0.3	0.4	0.07	0.01
3	0.06	0.1	0.5	0.3	0.02
4	0.02	0.03	0.2	0.6	0.09
5	0.01	0.01	0.1	0.7	0.2

Johannesburg	1	2	3	4	5
1	0.5	0.3	0.1	0.06	0.01
2	0.3	0.4	0.2	0.06	0.03
3	0.09	0.1	0.5	0.3	0.04
4	0.06	0.05	0.3	0.4	0.1
5	0.06	0.09	0.2	0.5	0.2

Methodology

- Large data collection
- Manual labeling for users (ethical consideration!)
- Comparative analysis of usage

Ref:

- Robertson A., W. Magdy, S. Goldwater. Self-Representation on Twitter Using Emoji Skin Color Modifiers*. ICWSM 2018*
- Robertson A., W. Magdy, S. Goldwater. Emoji Skin Tone Modifiers: Analyzing Variation in Usage on Social Media*. ACM Transaction of Social Computing*

Beliefs/Religious Discussions on Social Media

How Arabs Discuss Atheism on Twitter?

- Atheism in criminalised or shamed in the Arab world
- There are some Arabs who declare their negative stance towards religion on Twitter
- How these people are located in the Arab social network?
- How others interact/connect to them?

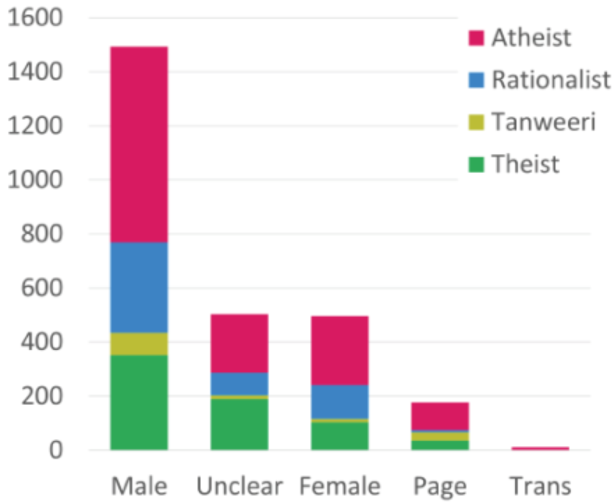
Data Collection

- Search Twitter accounts **bios** for those who explicitly define themselves as: atheists, anti-atheism (theists), Rationalists, or Tanweeri → 2,670 accounts
- For each account, collect all accounts they follow or interact with → 550K accounts
- Focus on accounts in their network that are:
  - Followed by at least 20 accounts in our seed list
  - Retweeted by at least 10 accounts in our seed list
  - Mentioned/replied by at least 10 accounts in our seed list
- Plot/Analyse networks

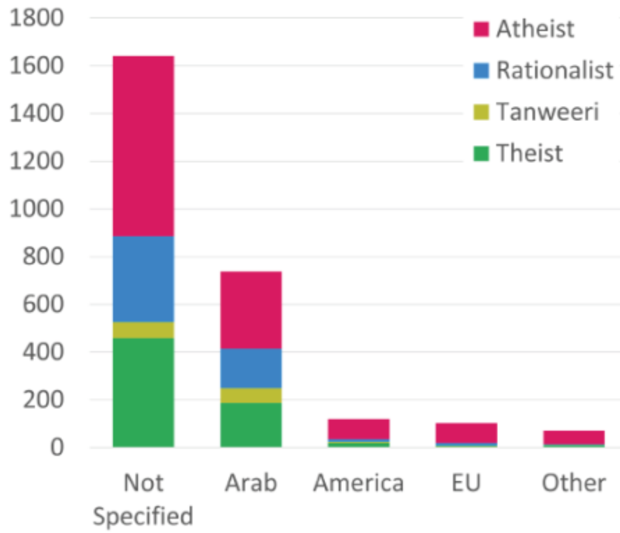
Seed List

Belief	Count
Atheist	1304
Theist	682
Rationalist	549
Tanweeri	138
Total	2673

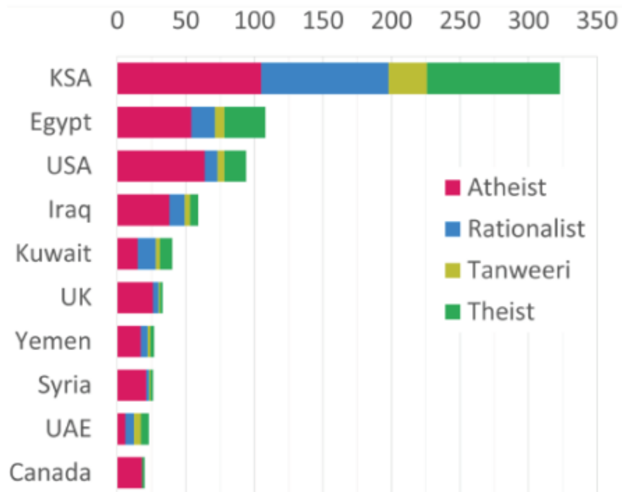
Religion	Count
Muslim	644
Christian	32
Hindu	3
jihadi	2
Jewish	1



(a) Gender/Type Distributions



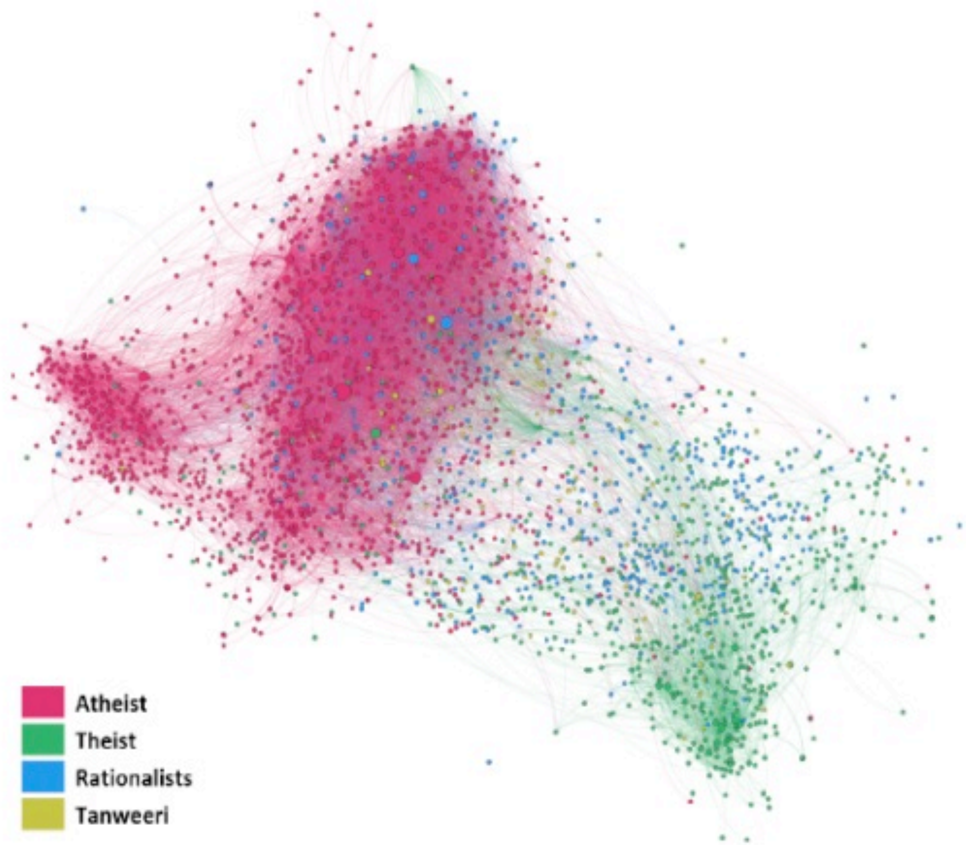
(b) Locations Distributions



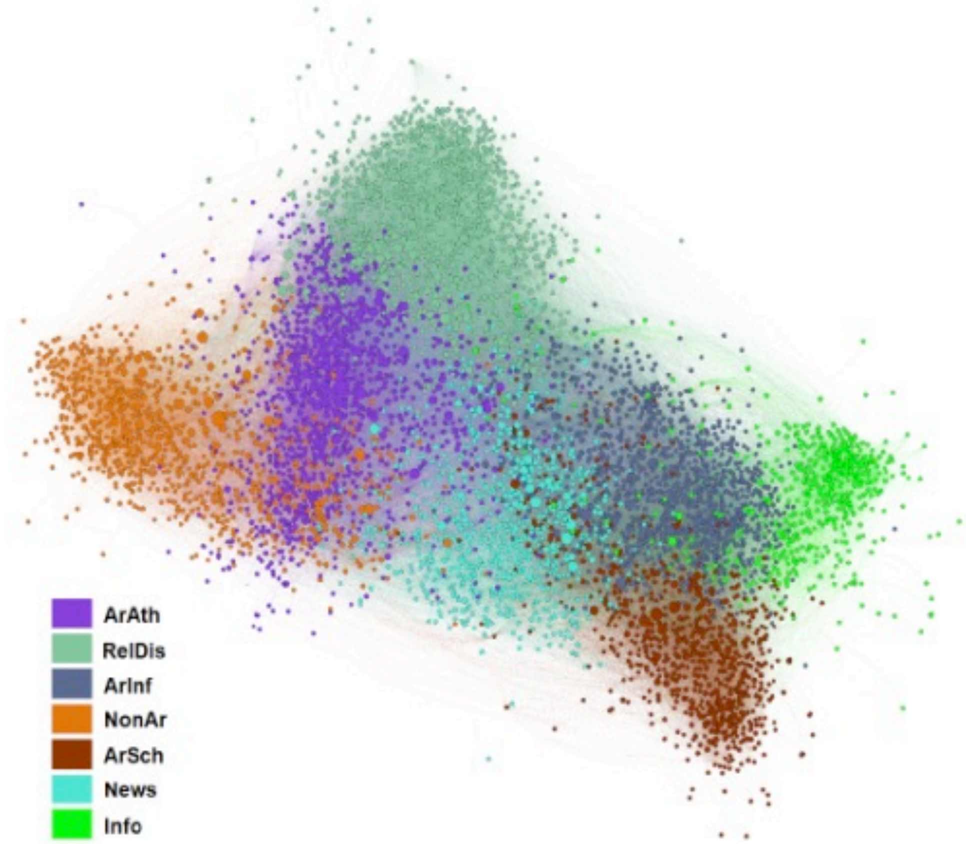
(c) Countries with Most Accounts

THE UNIVERSITY of EDINBURGH

Followers Network

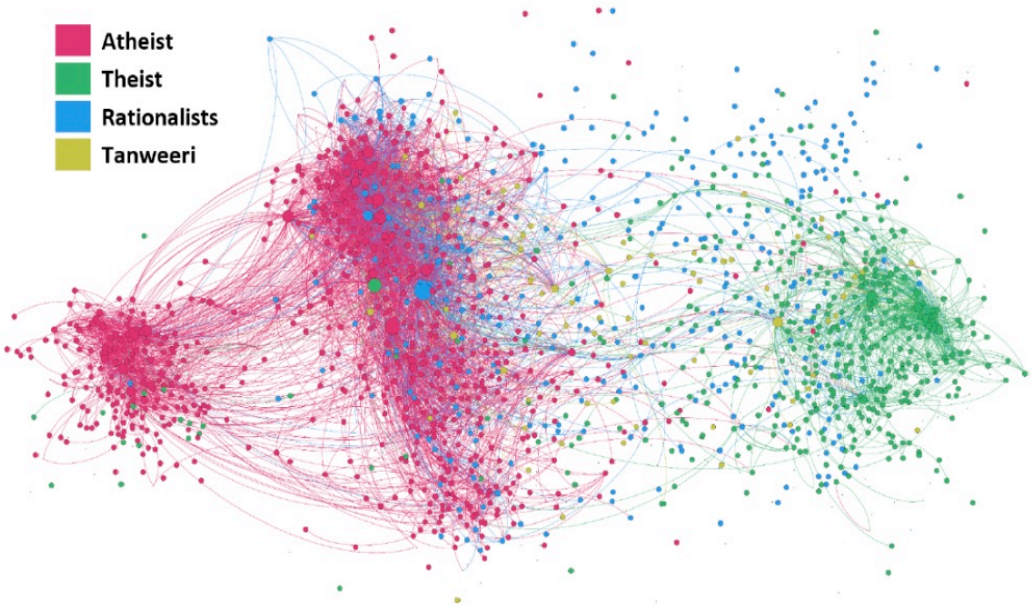


(a) The follow connections among our seed accounts. Nodes colors represent each group in our dataset.



(b) The follow connections among the seed accounts and their Follow network. Nodes colors represent different clusters obtained based on modularity

Retweets Network



(a) The retweeting interactions among our seed accounts. Nodes colors represent each group in our dataset.

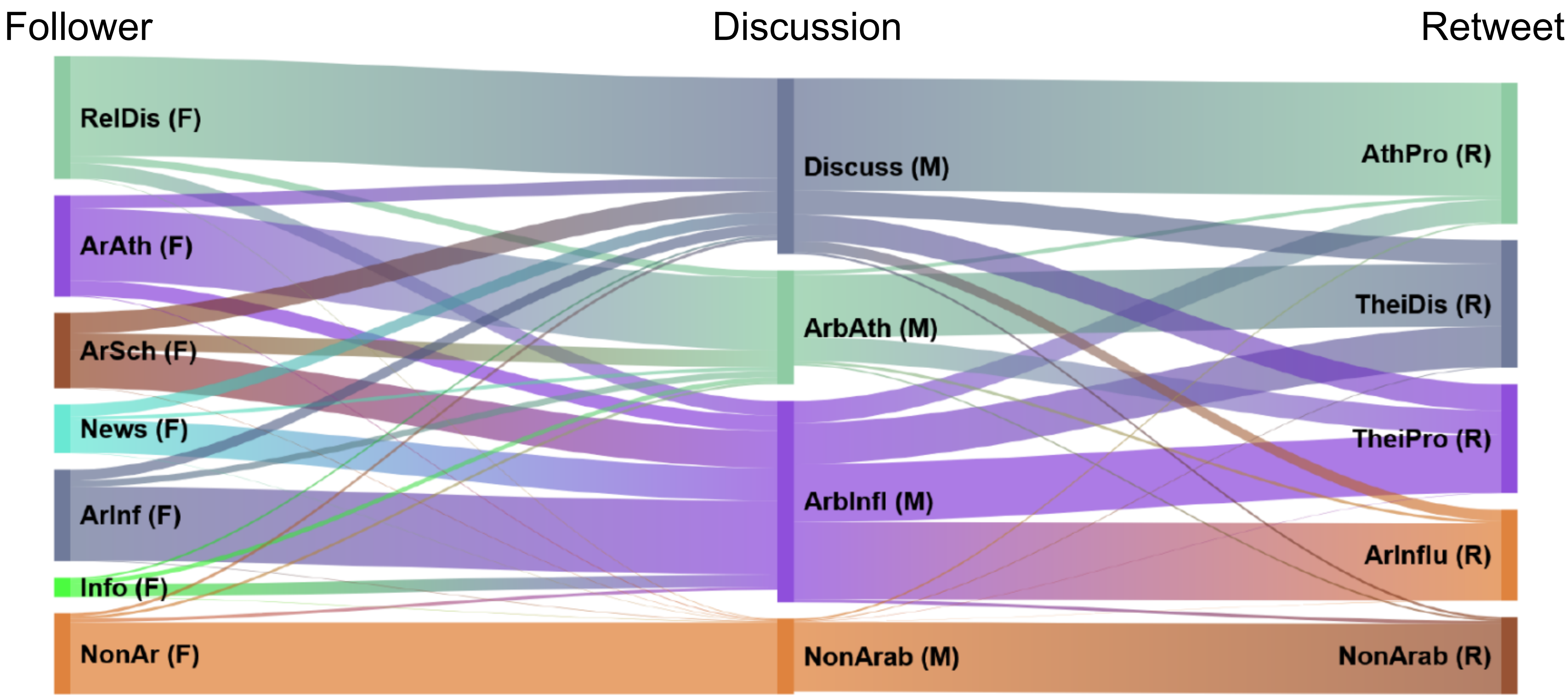
(b) The retweeting interactions among the seed accounts and their Retweet network. Nodes colors represent different clusters obtained based on modularity

Discussion Network

(a) The follow connections among our seed accounts. Nodes colors represent each group in our dataset.

(b) The follow connections among the seed accounts and their Follow network. Nodes colors represent different clusters obtained based on modularity

Intersection between networks



Observations (1/2)

- There is clear polarization on the atheism topic in the Arab social media, especially in the follower and retweets networks
- Discussion network shows large discussion among different groups
- In general, Tanweeri and Rationalist accounts act as bridges between atheist and theist accounts
- Theists accounts are tightly connected to famous Muslim scholars who refute atheism

Observations (2/2)

- Atheist are not one group, but multiple, including:
  - Religion refuters/arguers
  - Arab-connected atheists
  - Non-Arab connected atheists
- Arab atheists connected to western networks create an echo-chamber
  - Isolated from most Arab networks, including other atheists
  - Follow/Retweet/Discuss with each other only
  - Tweet mostly in English

Methodology

- **Initial search for Twitter account Bios**
- **Manual labeling → identify user groups**
- **User network collection (big data)**
- **In-depth network analysis → qualitative analysis**

Ref:

Al Hariri Y., W. Magdy and M. Wolters. Arabs and Atheism: Religious Discussions in the Arab Twittersphere. SocInfo 2019

Al Hariri Y., W. Magdy and M. Wolters. Atheists versus Theists: Religious Polarisation in Arab Online Communities. Under submission

Automatic Labeling

People Changing Opinion?

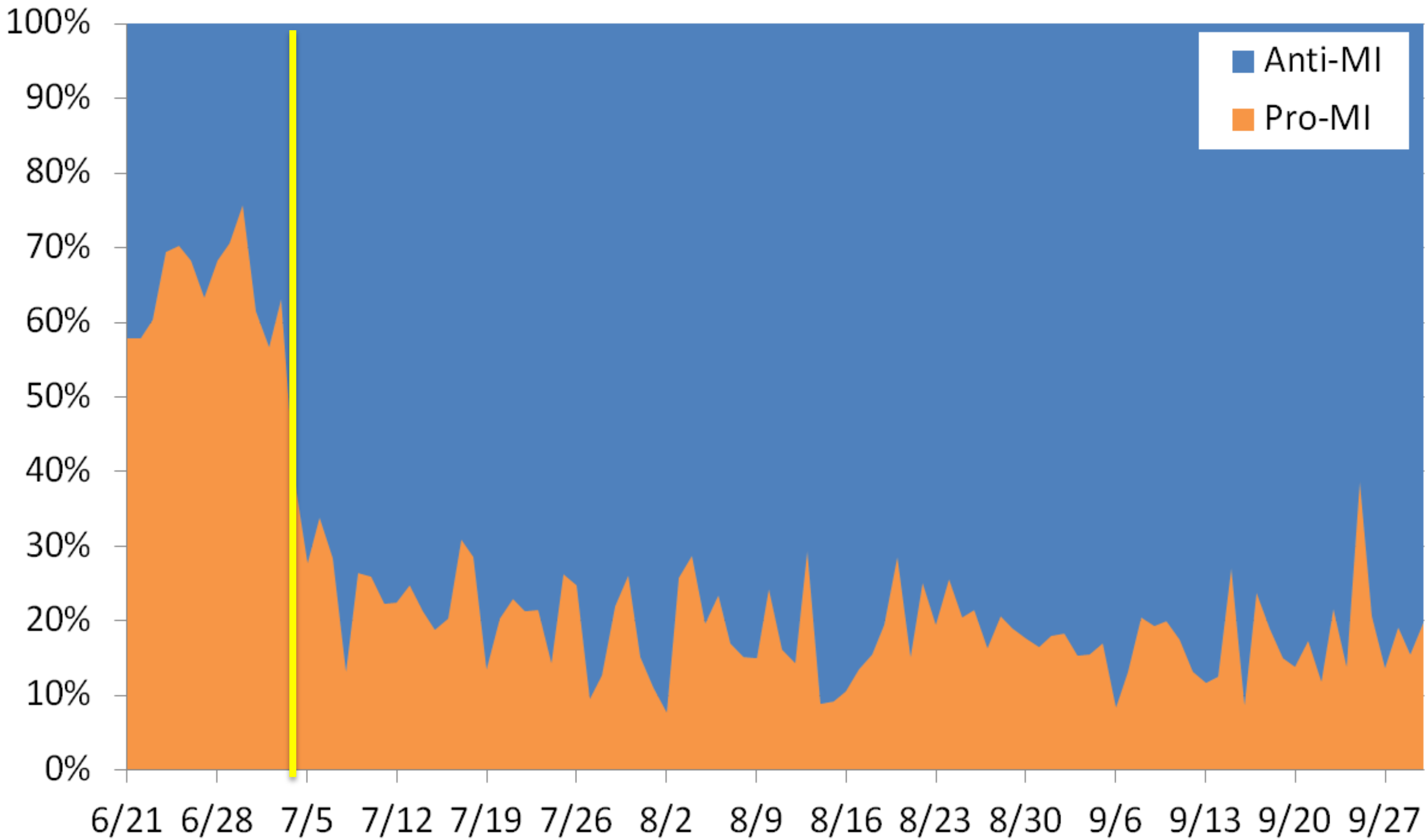
The Egyptian Military Intervention

- 30 June 2013: large demonstration in Egypt against Morsi
- 3 July 2013: Military ousted Morsi
- 5 July-13 Aug: Large Sit-in against military coup
- 14 Aug: Army ends Sit-in by force, while hundreds killed

Study

- Shift in trending tweets about Egypt was noticed since 3/7/2013
- **Data Collection:**
  - 6M tweets on Egypt → 21 July 2013 – 30 Sep 2013
- **Label tweets:**
  - Pro/Anti military intervention
  - Sample **1000** → train classifier → label the rest.
- **Classifier:**
  - Trained SVM binary classifier → Pro/Anti military intervention
  - Accuracy: 85% (on the tweet level)
  - Label all tweets on topic using the classifier

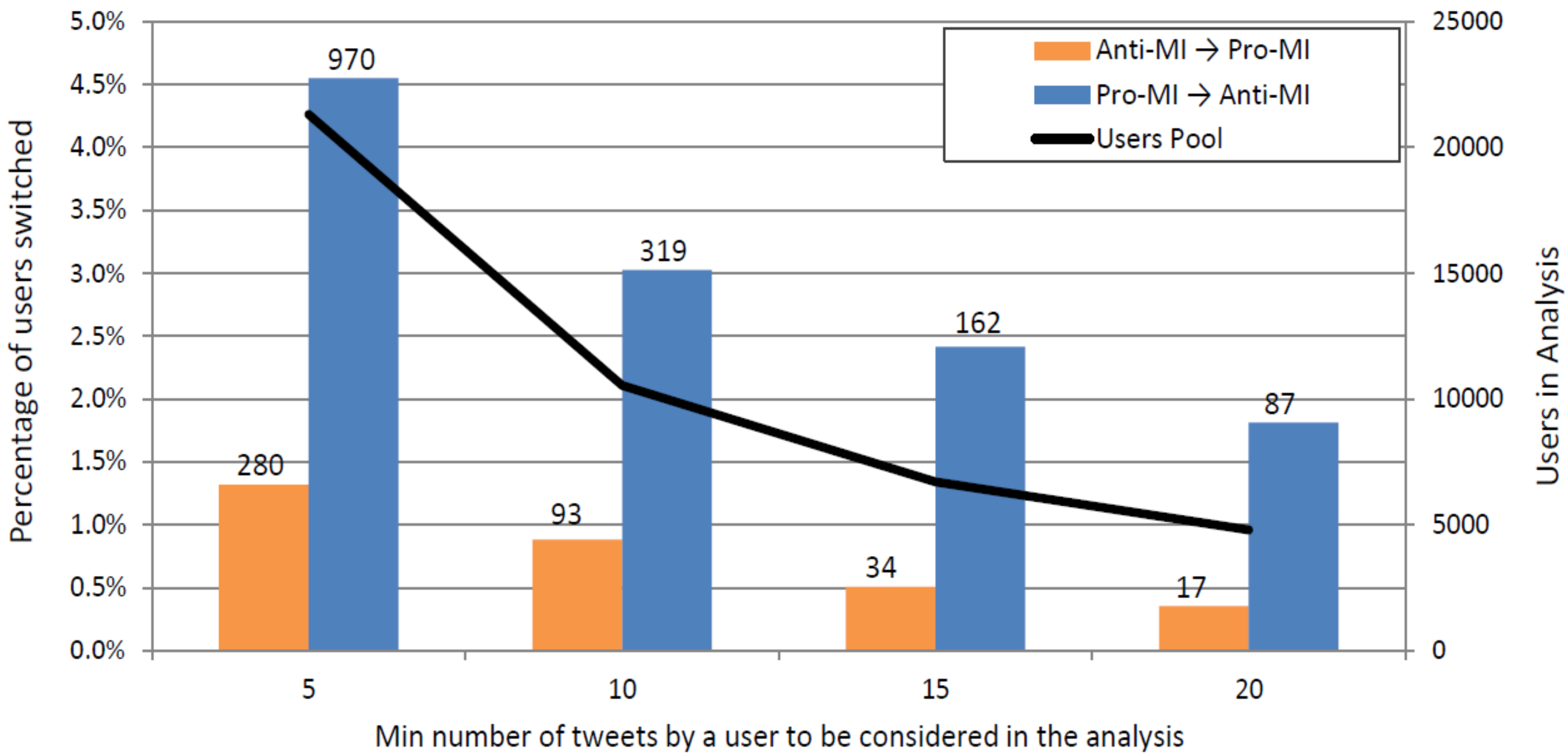
Volume of Tweets Pro/Anti-MI



On the User Level

- Identified 22K Twitter users with >5 tweets on topic
- Examine change in tweets stance over time
- Three confidence levels:
  - Users with 5+ tweets: 22K users
  - Users with 10+ tweets: 11K users
  - Users with 20+ tweets: 5K users
- Changes:
  - Pro-MI → Anti-MI
  - Anti-MI → Pro-MI

% of Users Changing Opinion



Methodology

- Data collection
- Sample labeling → classifier → label the rest
- Timeline analysis on:
  - Tweet level (trends) → can be misleading
  - User level → more indicative
- Lesson: It is not easy to have someone switching political belief

Ref:

Borge-Holthoefer J., W. Magdy, K. Darwish, and I. Weber. Content and Network Dynamics Behind Egyptian Political Polarization on Twitter. *CSCW 2015*

Summary

- **Many ways to analyse an event/topic globally using CSS**
- **Data:**
  - User → content
  - User → network
  - Content → users
  - Content → trends
- **Labeling strategy:**
  - Manual to trending samples
  - Manual to set of users
  - Automatic
- **Labeling strategy:**
  - Tweet-level vs user-level
  - Content vs Network

Next

- Analyse users' behavior (on the individual level)
- Prediction something about the user