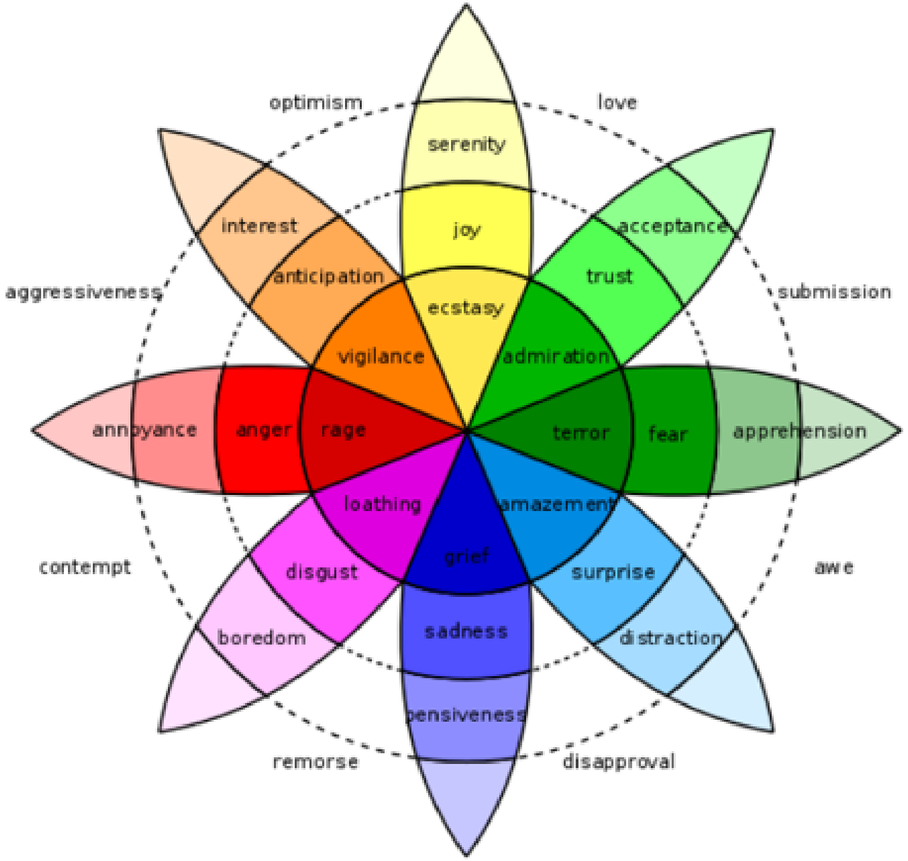
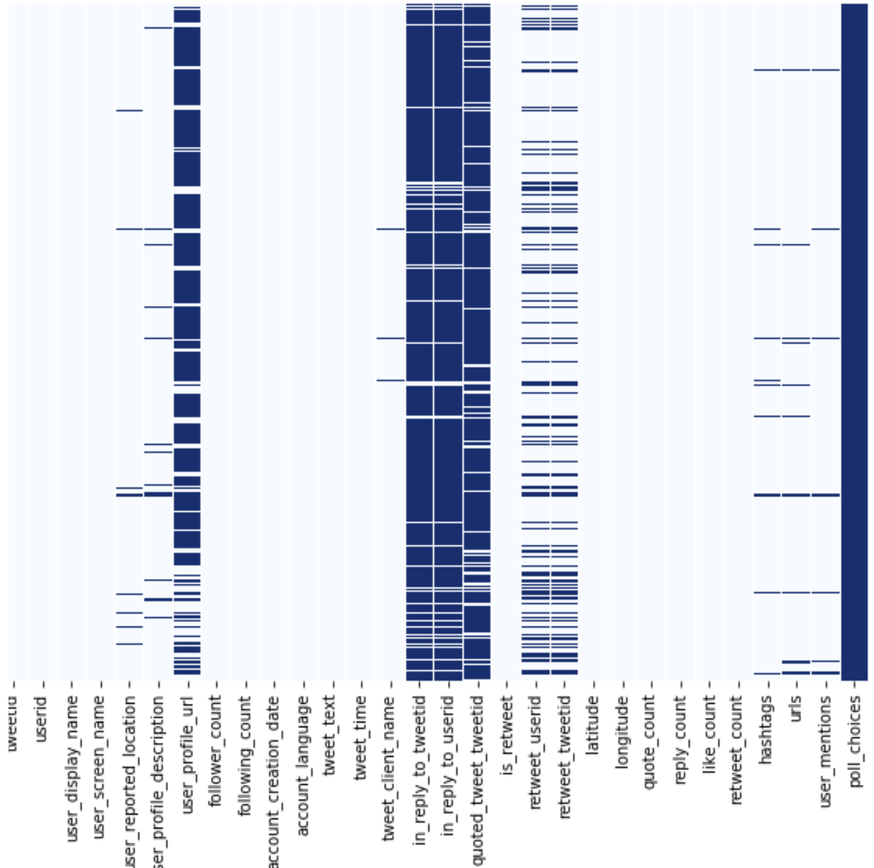


AI Emotion Detector and Topic Modeling in Suspended Elections-Related Twitter Accounts

nastyasimpson

Presented by Anastasia Simpson

Data



Twitter Election Integrity

Public Data Set, Russia

Released in January 2019

All public, non deleted tweets from accounts Twitter believes are **connected** to **state-backed information operations**.

416 accounts and their tweet content.

PROBLEM

Engineering Public Opinion is a significant concern to the public, officials and every one, really. It is on the forefront of any modern political and/or money divide.

Prevent the manipulation of news and information by potentially malicious actors.

TARGET

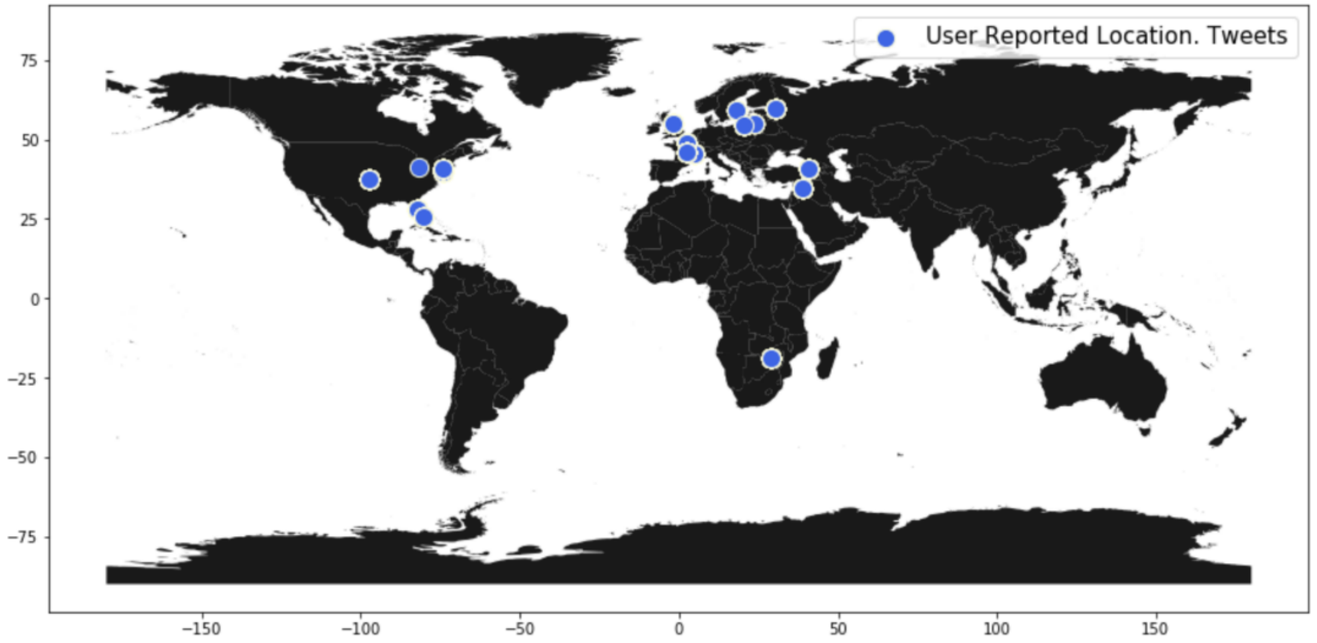
Natural Language Components of Tweets Text

Oh I love him! http://t.co/rXoZLhxNeR
#Conservatives #obama #ObamaHatesAmerica #obam...
@Salon @MMFlint #fatjokes #liberalbullshit #mu...
🗣🗣🗣🗣🗣\n\n🚨🚨🚨🚨🚨\n@SenSchumer hates America!\n@C...
OMG🤣🤣 https://t.co/5H2qVcopMT

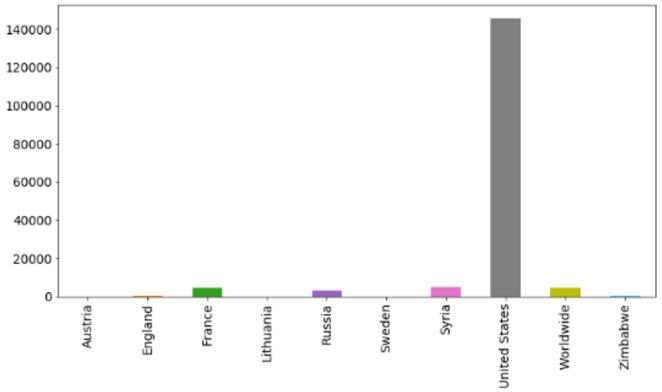
#SAA managed to liberate the village of Jinah .
الولايات المتحدة الأمريكية تقوض تحقيق السلام ف...
#حلب ويحبط هجوم# من مساحة 93% الجيش يسيطر على .
The third round of #SyrianTalks in #Astana wil.

"Democratic congressional candidate Alexandria Ocasio-Cortez
ally, that's the left's new talking point: they say everything
eason. Because they are racist themselves."
"@MylesMill @NYPDDetectives Eh.. I seen many kkkops choke-hol
kick them all around, plant drugs and so on and so forth.. I'm
its high time these killer cops were also arrested and JAILED!
"@JeepersMikkers @BrulloMaggie @ABC It's funny how the whole
tell a simple truth.\nWow... he is literally lying to all thes
'@thehill Its such a shame the POTUS behaves like a 2-y-o tod
'@cnni Hopefully he will address the UN General assembly as a

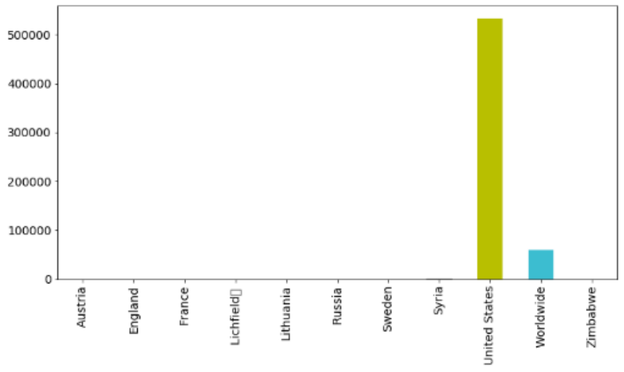
WHERE THE USERS SAY THEY ARE FROM?



Tweets



Retweets



REPORTED LOCATION

NLP

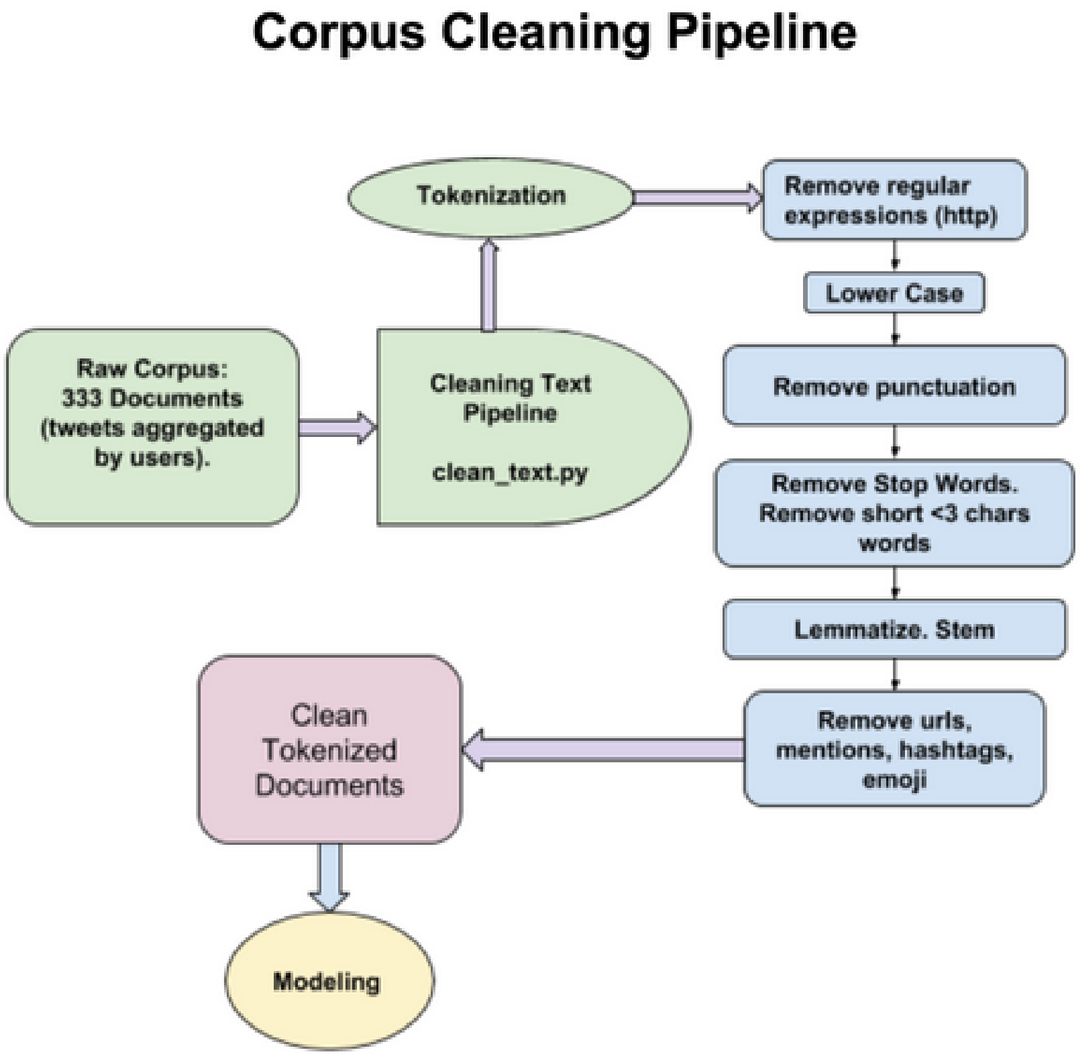
NATURAL LANGUAGE PROCESSING

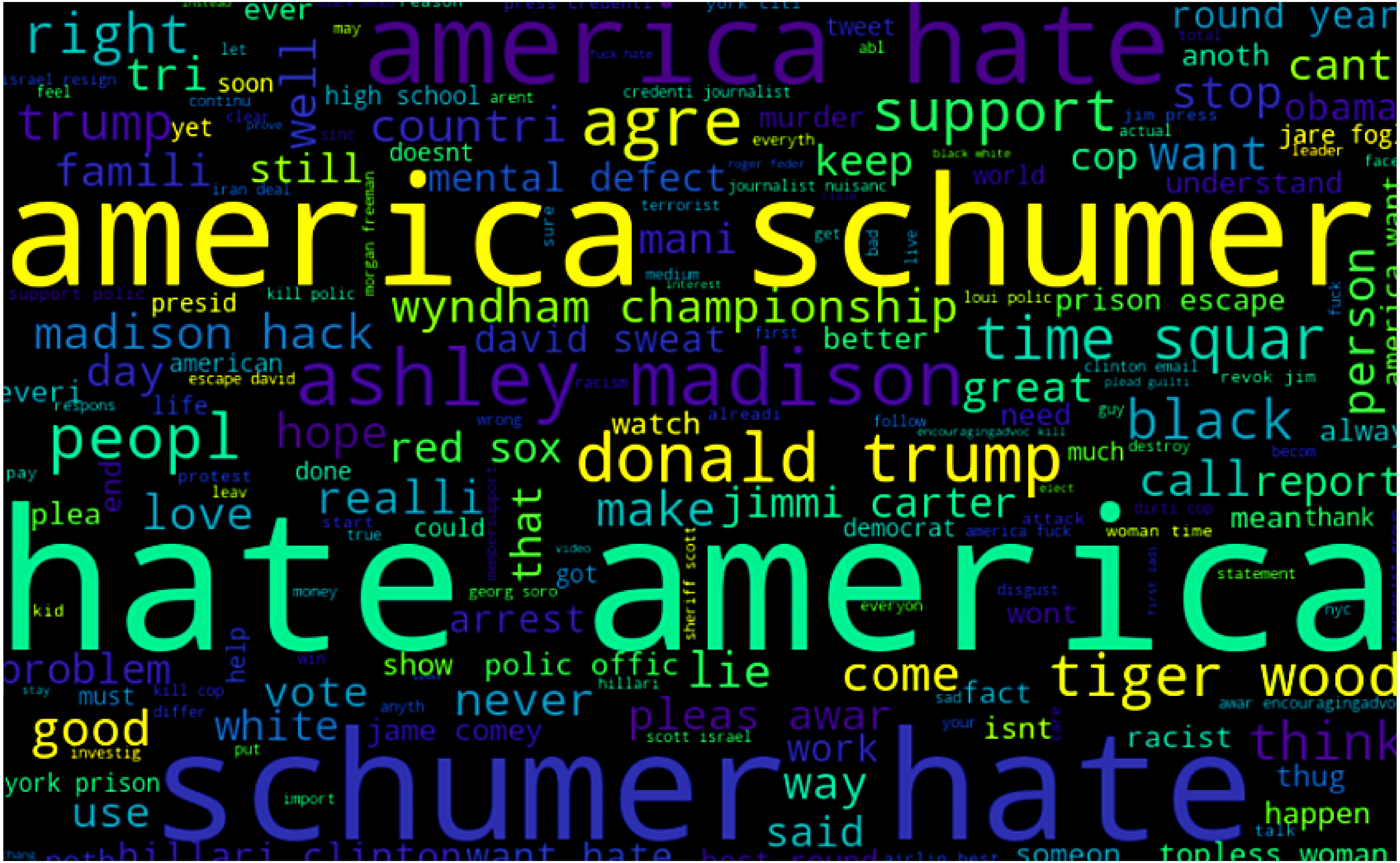
lemmer = WordNetLemmatizer()

stemmer = SnowballStemmer('english')

stop_words = set(nltk.corpus.stopwords.words('english'))

pip instal tweet-preprocessor



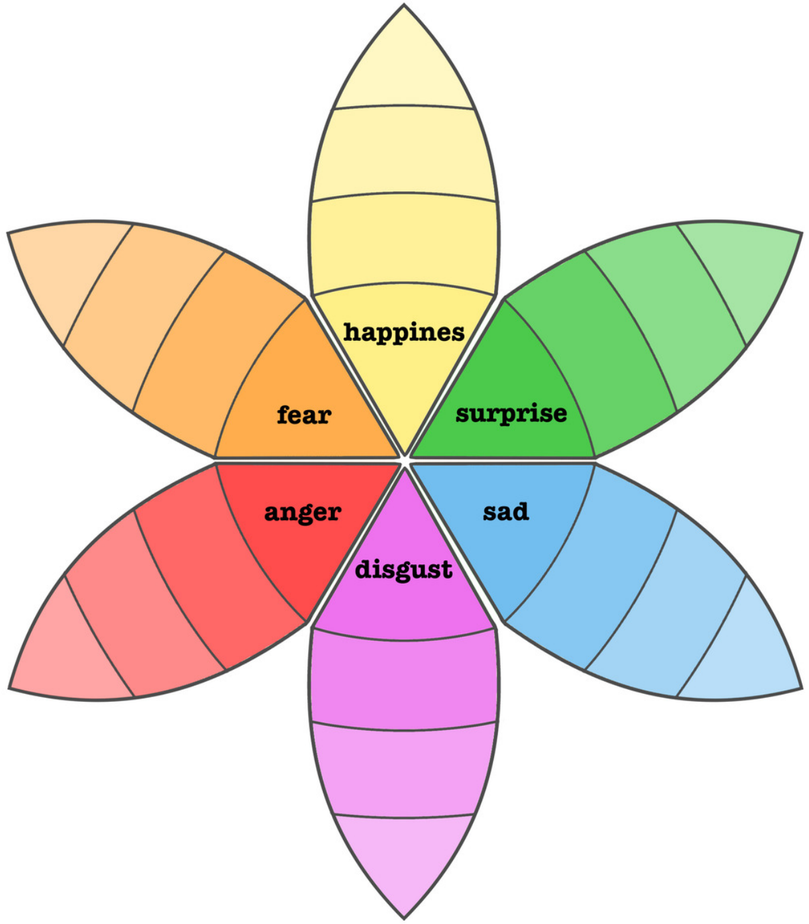

america schumer
hate america
schumer hate
america hate
ashley madison
donald trump
wyndham championship
time squar
tiger wood
madison hack
jimmi carter
mental defect
david sweat
red sox
prison escape
pleas awar
hillari clinton
want hate
topless woman
round year
right
tri
trump
famili
well
agre
support
countri
high school
stop
obama
cant
anoth
keep
cop
want
understand
still
peopl
hope
mani
black
great
person
call
report
make
realli
love
plea
day
everi
life
american
presid
better
need
watch
alway
much
that
could
done
democrat
meanthank
arrest
problem
vote
never
lie
come
fact
show
polic offic
white
good
use
work
way
said
think
racist
thug
happen
isnt
jame comey
someon
york prison
must
help
end
got
wont
soon
yet
ever
tweet
murder
world
doesnt
cop
kill polic
georg soro
sheriff scott
scott israel
journalist nuisanc
credenti journalist
clinton email
revok jim
statement
woman time
money
kid
investig
import
put
protest
terrorist
medium
kill cop

AI EMOTION DETECTOR

"The only thing we have to fear is fear itself - nameless, unreasoningm unjustified, terror which paralyzes needed efforts to convert retreat into Advance. ", smooth & tasty.

FDR -
FIRST INAUGURAL ADDRESS, MARCH 4TH, 1933

Paul Eckman Basic Emotions



emotion_detector.py

TWEETS

VOCABULARY

CORPUS COUNT VECTORS

word	disgust	surprise	neutral	anger	sad	happy	fear
magnetizing	0.023810	0.119048	0.023810	0.023810	0.023810	0.023810	0.023810
adequately	0.007143	0.035714	0.007143	0.035714	0.064286	0.021429	0.007143
belch	0.035714	0.035714	0.035714	0.107143	0.035714	0.035714	0.035714

```
(em_data.index == vectorizer.get_feature_names()).all()
```

```
True
```

```
em_data.values.shape
```

```
(23634, 7)
```

```
X.shape
```

```
(159272, 23634)
```

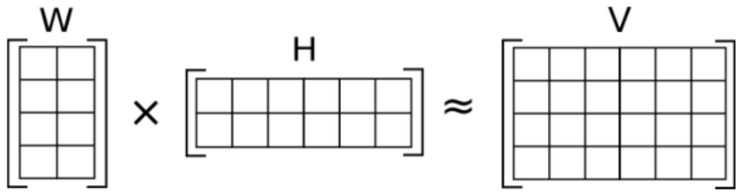
```
em_freq_np = X @ em_data.values
```

```
type(em_freq_np)
```

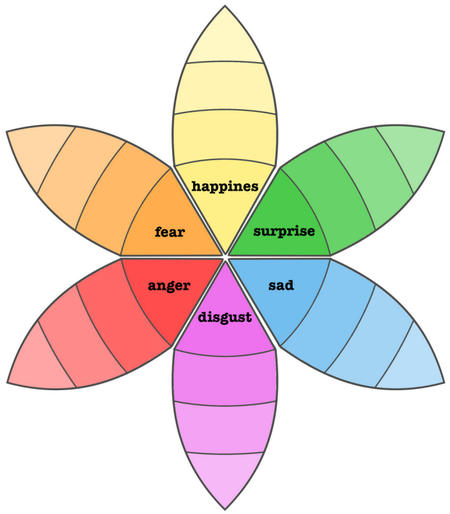
```
numpy.ndarray
```

```
em_freq_np.shape
```

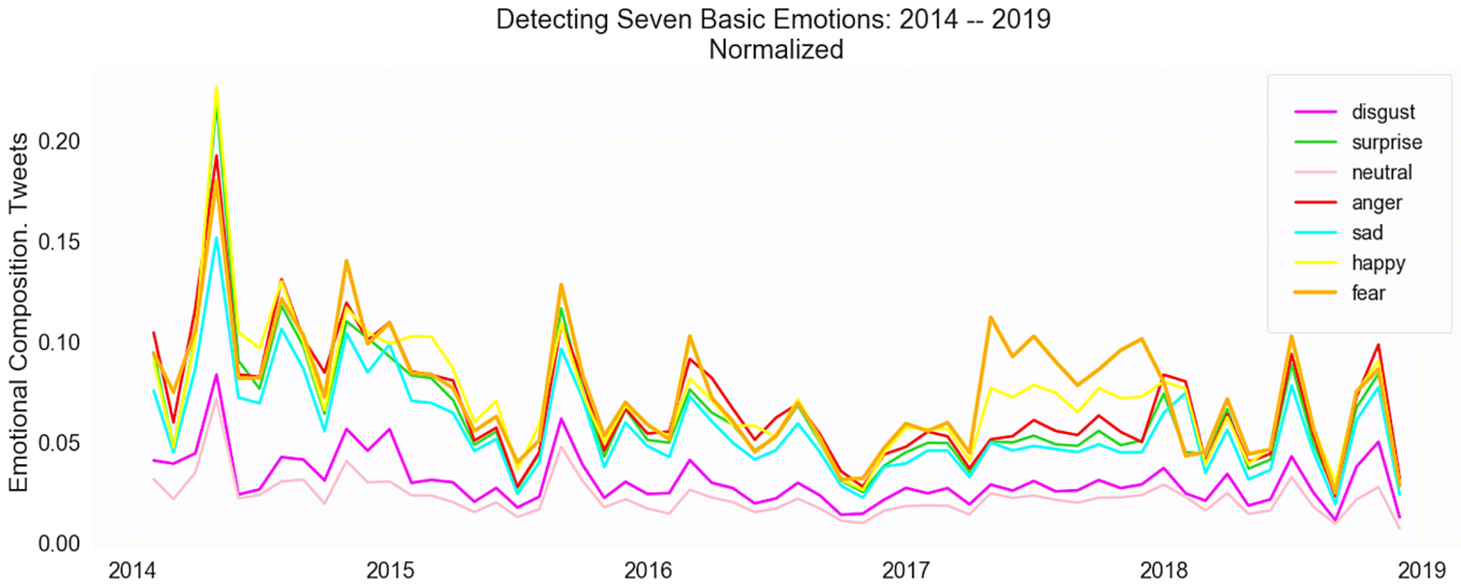
```
(159272, 7)
```

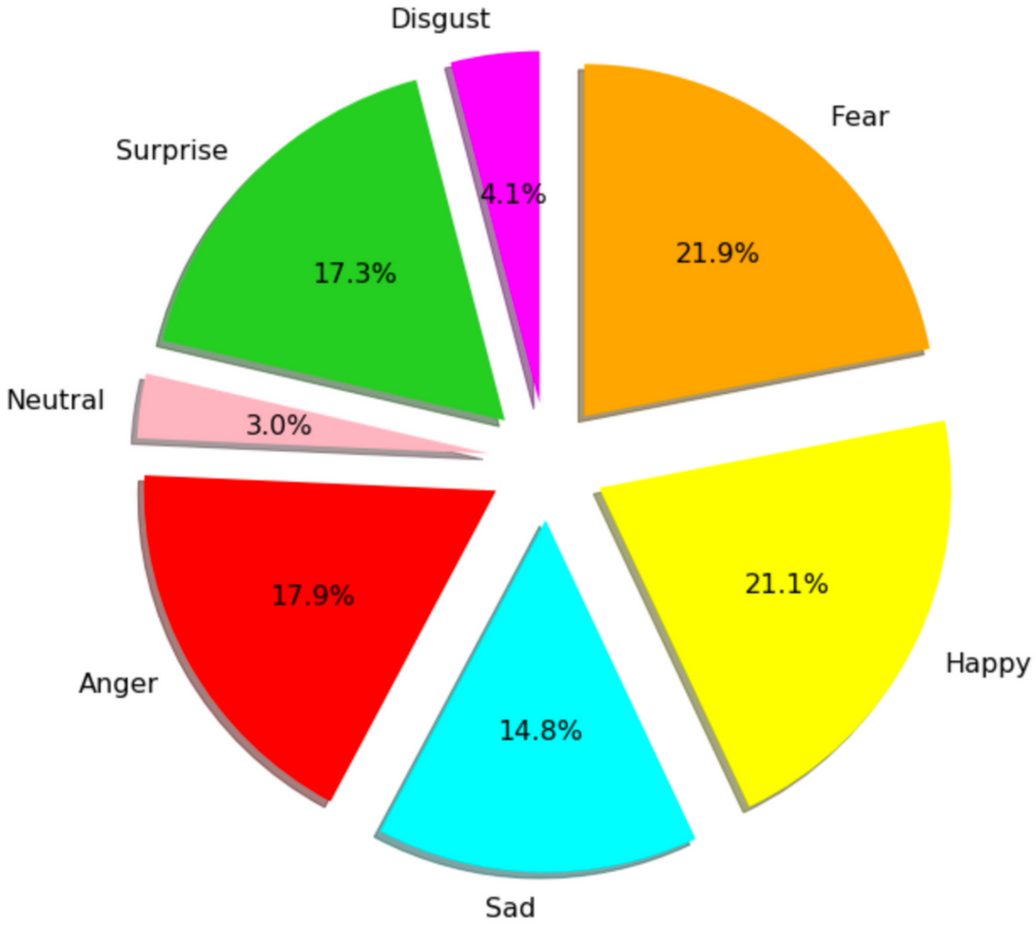


	tweet_text	tweet_time	disgust	surprise	neutral	anger	sad	happy	fear
0	ashley madison stats find new orleans among th...	2015-08-21 22:30:00	0.097226	0.271829	0.098647	0.144845	0.236115	0.097226	0.113099
1	httptcon1xwcjhpr8	2015-04-13 18:04:00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	@nytimesworld two face scumbag listen up obama...	2017-01-13 03:17:00	0.061261	0.114266	0.058528	0.102166	0.111456	0.262198	0.154196



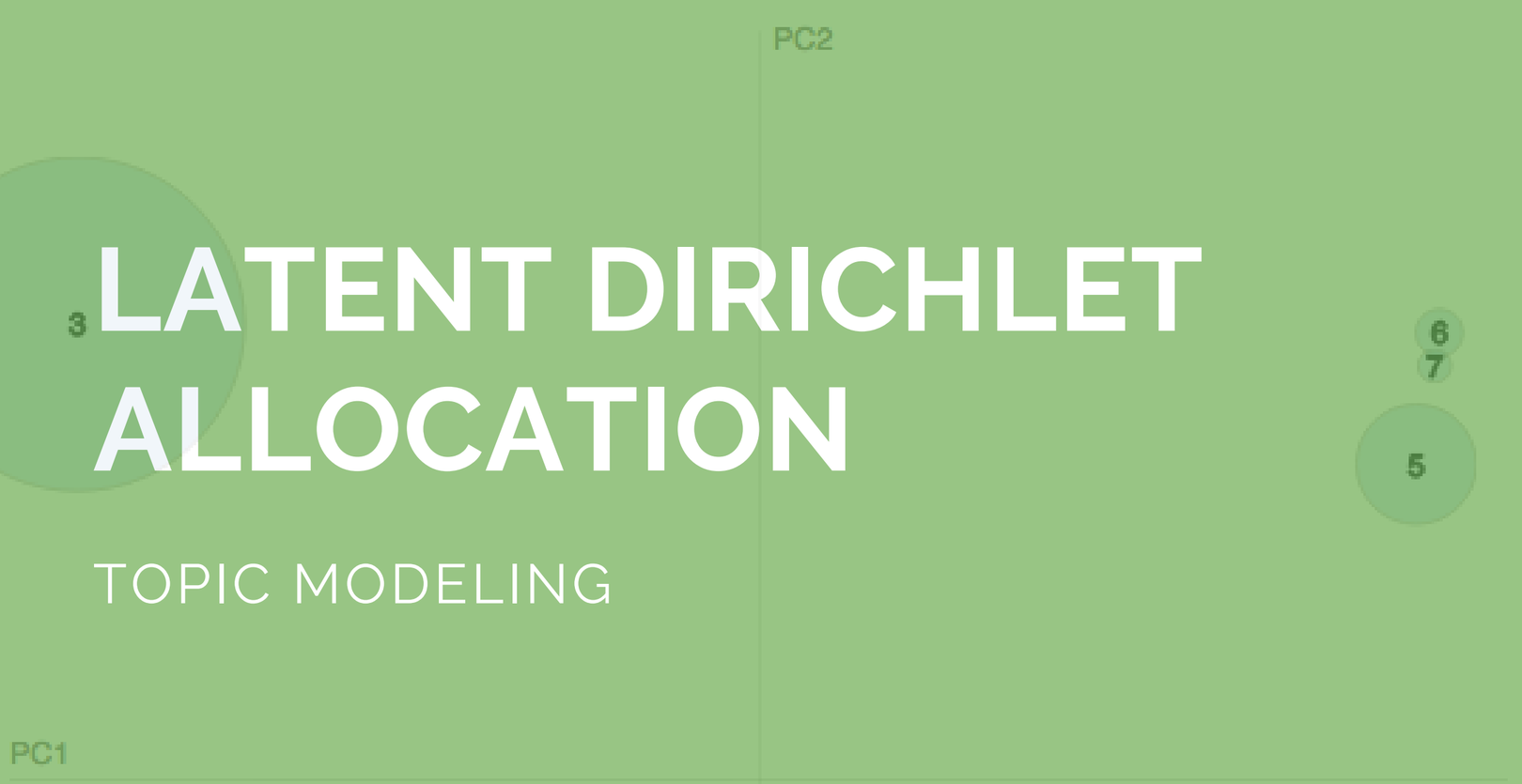
RESULTS

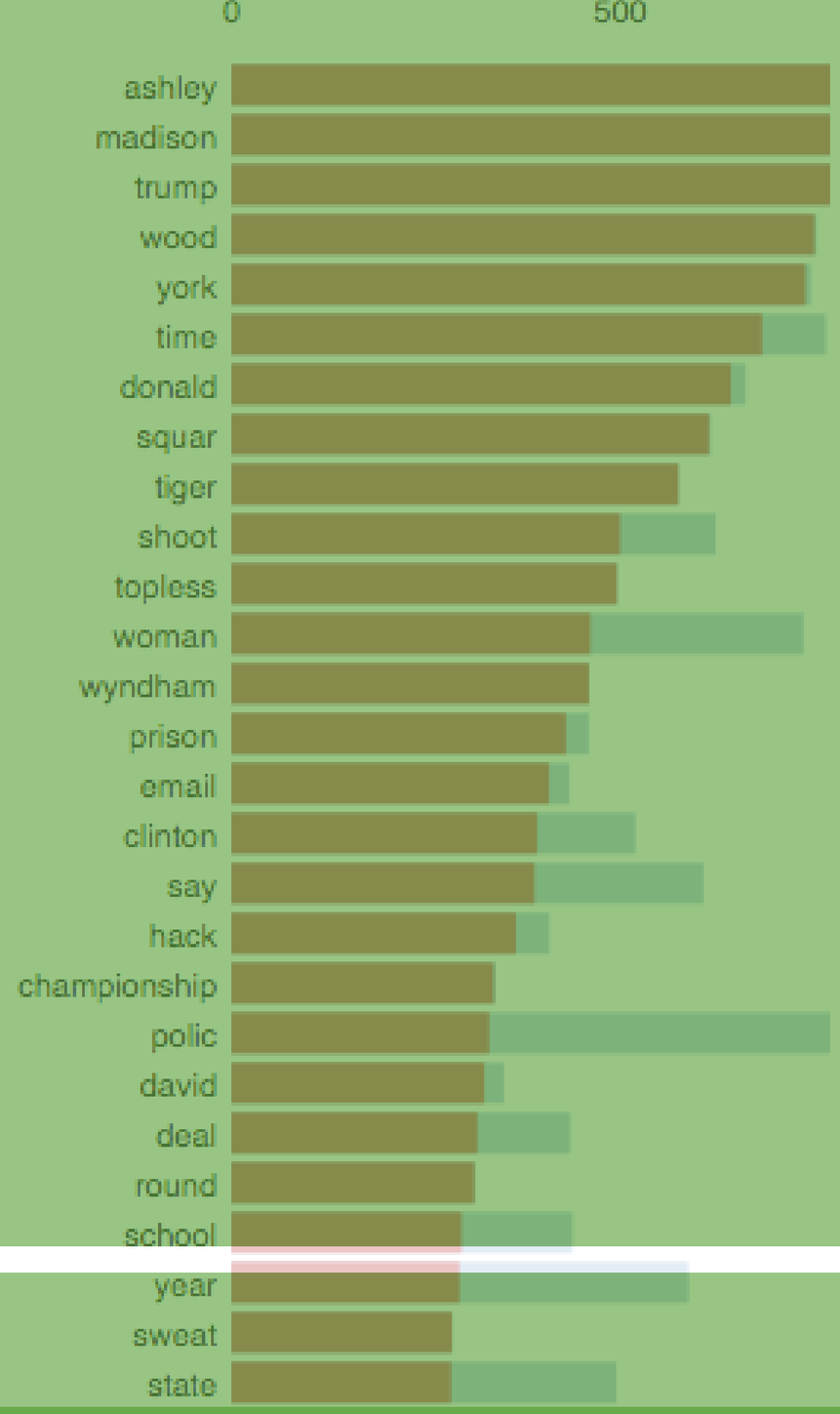

Detecting Seven Basic Emotions: 2014 -- 2019
Normalized
Emotional Composition. Tweets
0.00
0.05
0.10
0.15
0.20
2014
2015
2016
2017
2018
2019
disgust
surprise
neutral
anger
sad
happy
fear


Disgust
4.1%
Fear
21.9%
Surprise
17.3%
Neutral
3.0%
Anger
17.9%
Sad
14.8%
Happy
21.1%

LATENT DIRICHLET ALLOCATION

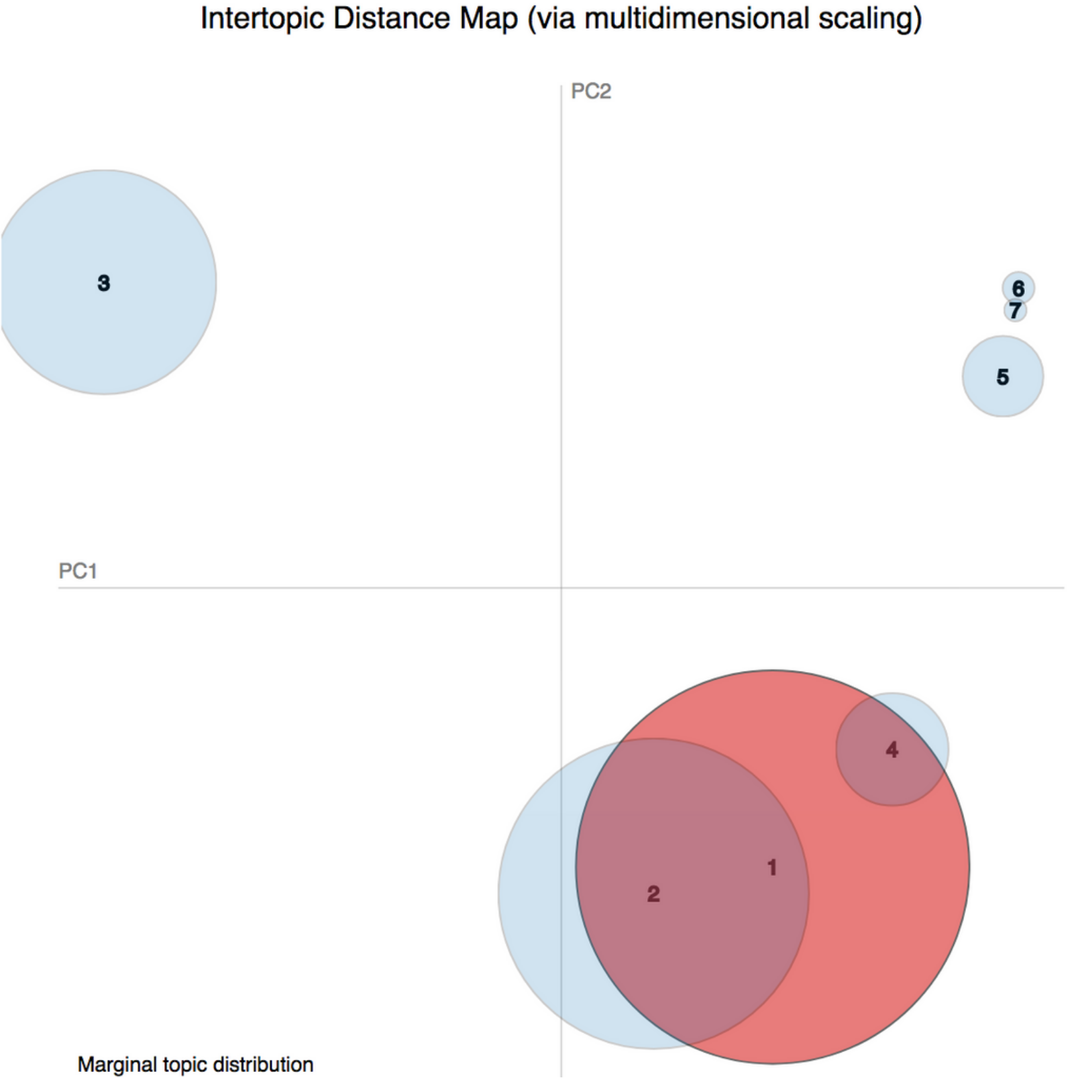
TOPIC MODELING

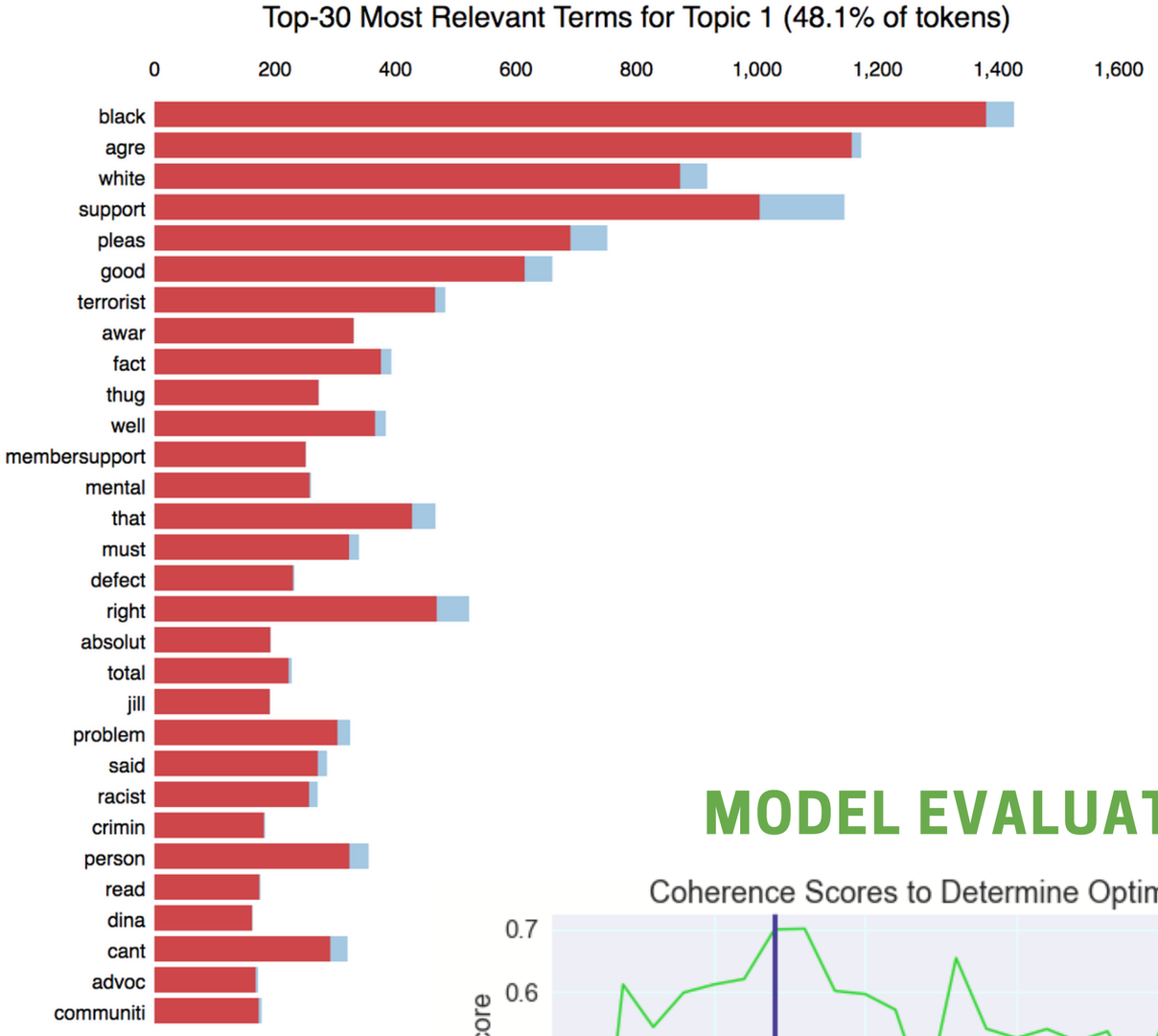




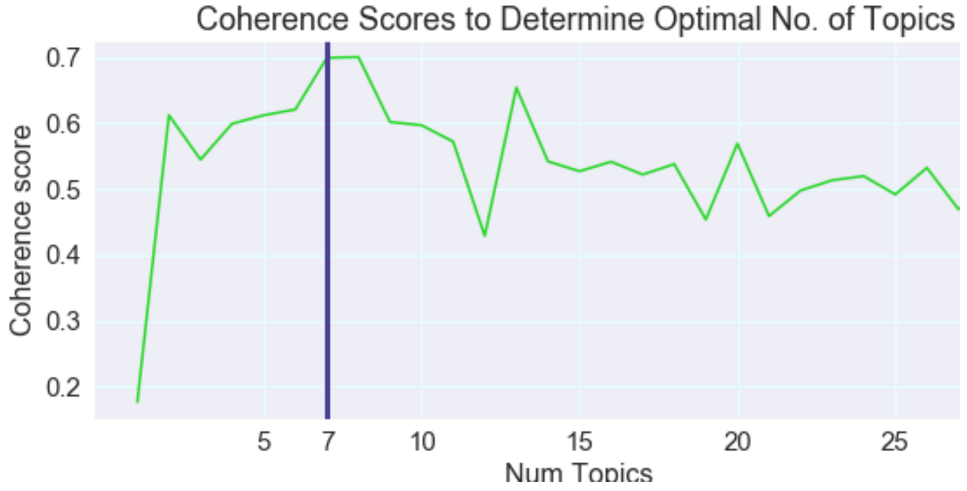
Documents are probability distributions over topics

Topics are probability distributions over words





MODEL EVALUATION



AGGRESSION. RACE. TERROR. DISGUST

48.1 %

SCANDALS. POLITICS. ASHLEY. MADISON. TOPLESS. DONALD

29.9 %

AMERICA. HATE. SHUMER.

15.6 %

ZIMBABWE

3.9 %

EUROPEAN SPORTS. TENNIS.

2 %

SYRIA. TURKEY.AZERI

0.5 %

FUTURE STEPS

- EMOTION RECOGNITION FROM FACIAL EXPRESSION
- NETWORK AND GRAPH ANALYSIS
- CLASSIFIER
- IMPROVE AI EMOTION DETECTOR

Thank You

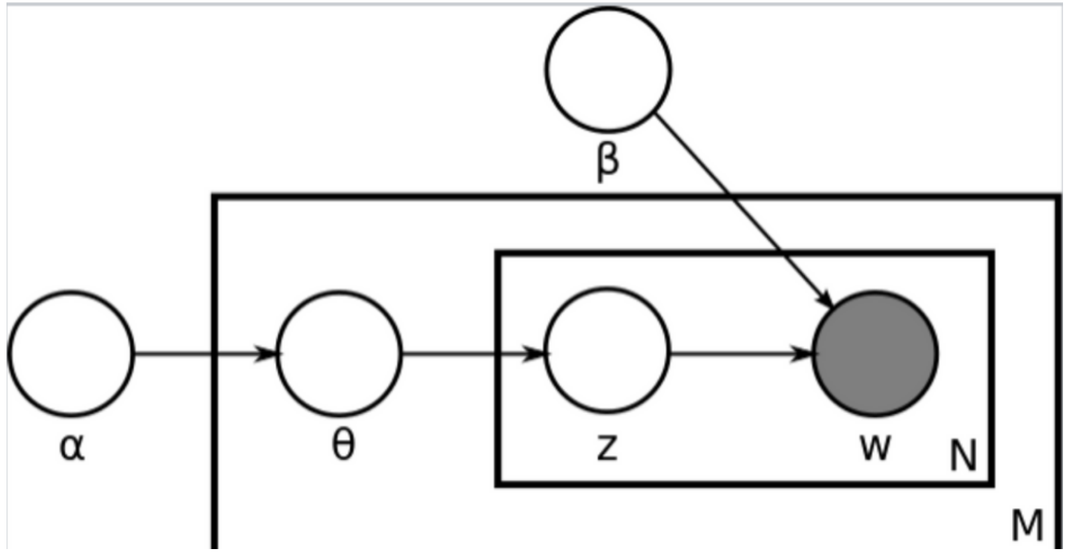
Coherence Score

Mimno et al. [12] proposed to use an asymmetrical confirmation measure between top word pairs (smoothed conditional probability). The summation of *UMass coherence* accounts for the ordering among the top words of a topic.[2]

$$C_{\text{UMass}} = \frac{2}{N \cdot (N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \log \frac{P(w_i, w_j) + \epsilon}{P(w_j)} \tag{4}$$

Word probabilities are estimated based on document frequencies of the original documents used for learning the topics. The calculation for our example would be:

$$\begin{aligned} C_{\text{UMass}} = \frac{1}{6} \Big(& \log(P(sport|game)) + \log(P(ball|game)) \\ & + \log(P(ball|sport)) + \log(P(team|game)) \\ & + \log(P(team|sport)) + \log(P(team|ball)) \Big) \end{aligned} \tag{5}$$



α is the parameter of the Dirichlet prior on the per-document topic distributions,

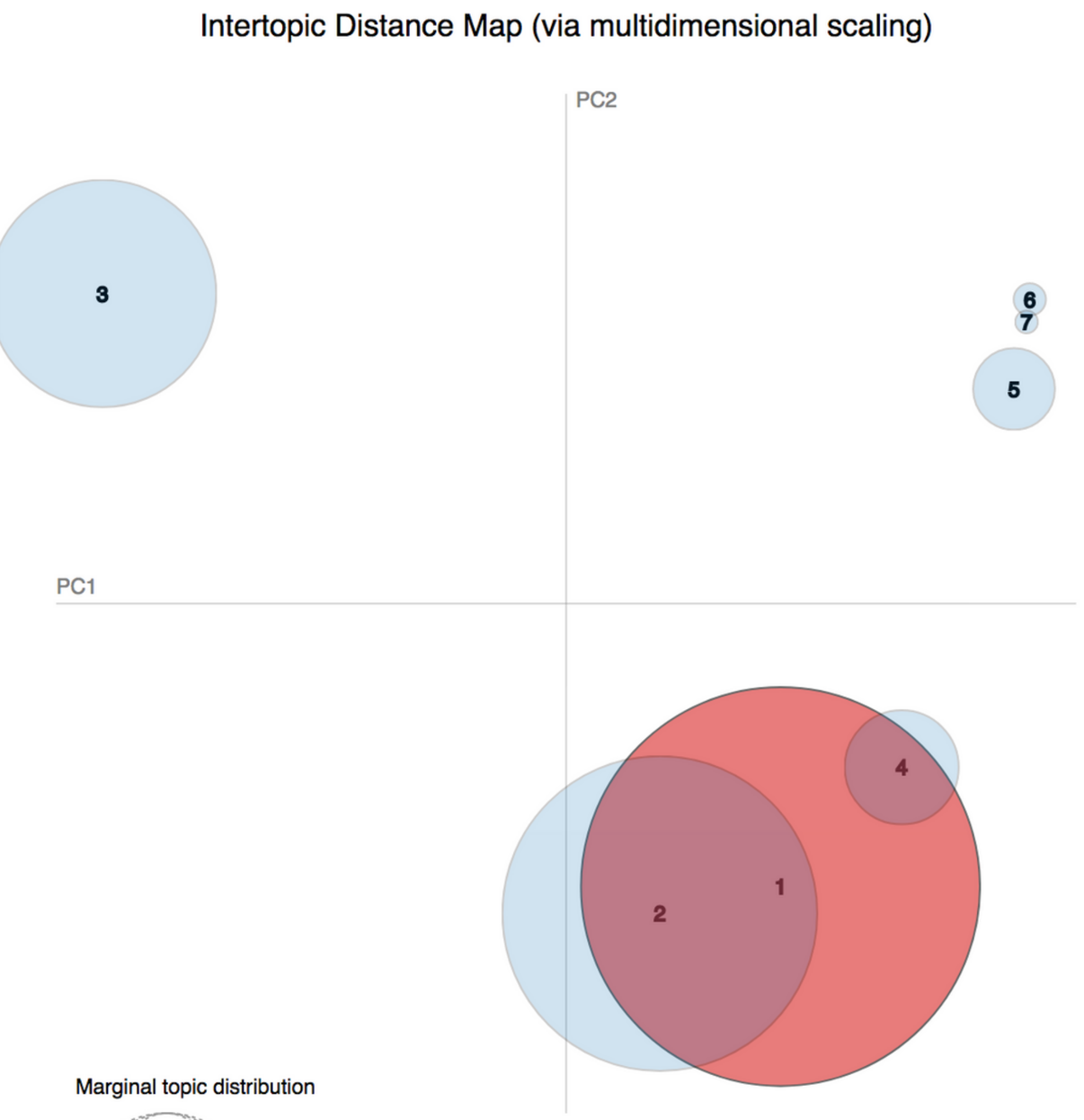
β is the parameter of the Dirichlet prior on the per-topic word distribution,

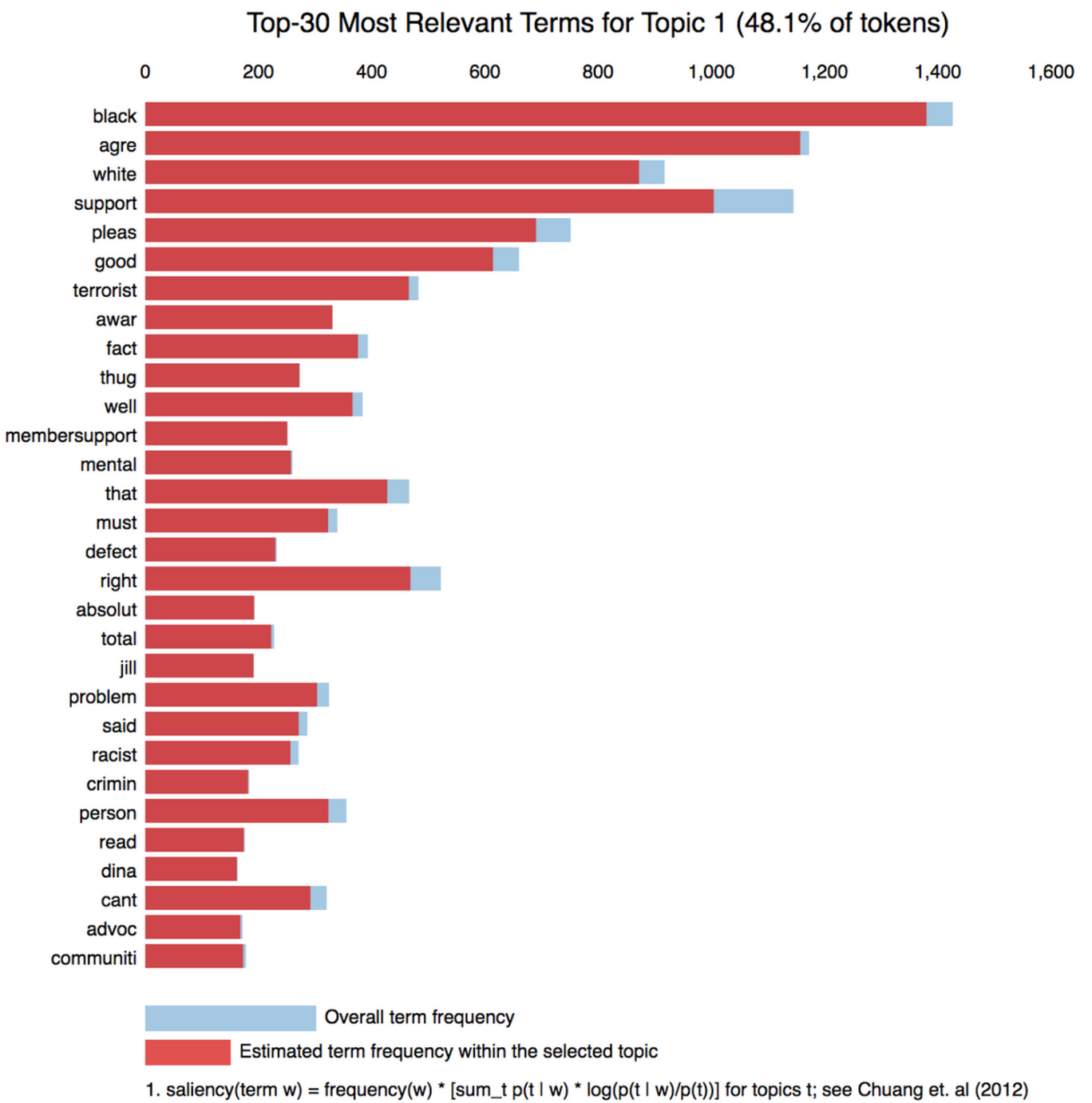
θ_m is the topic distribution for document m,

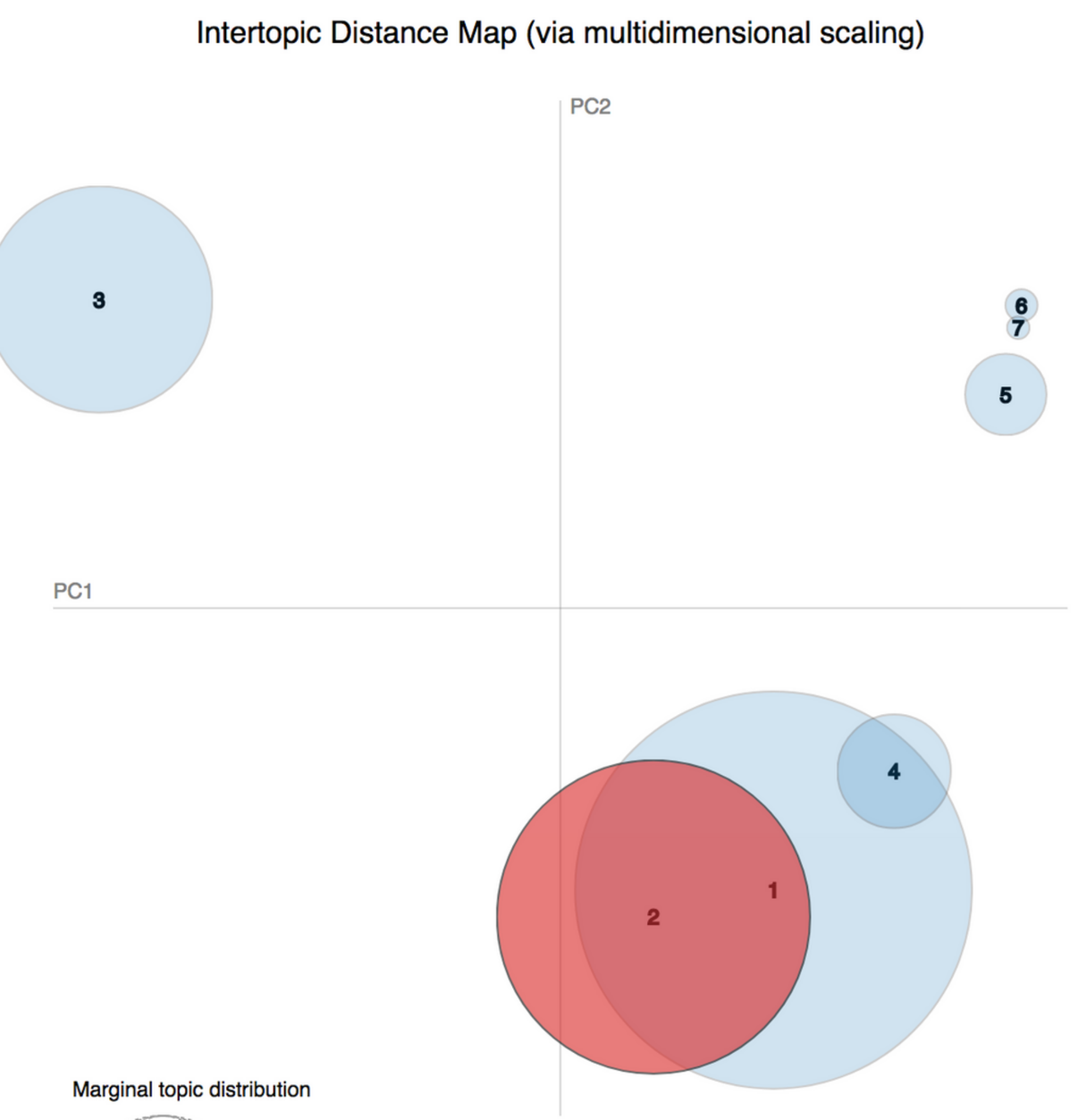
φ_k is the word distribution for topic k,

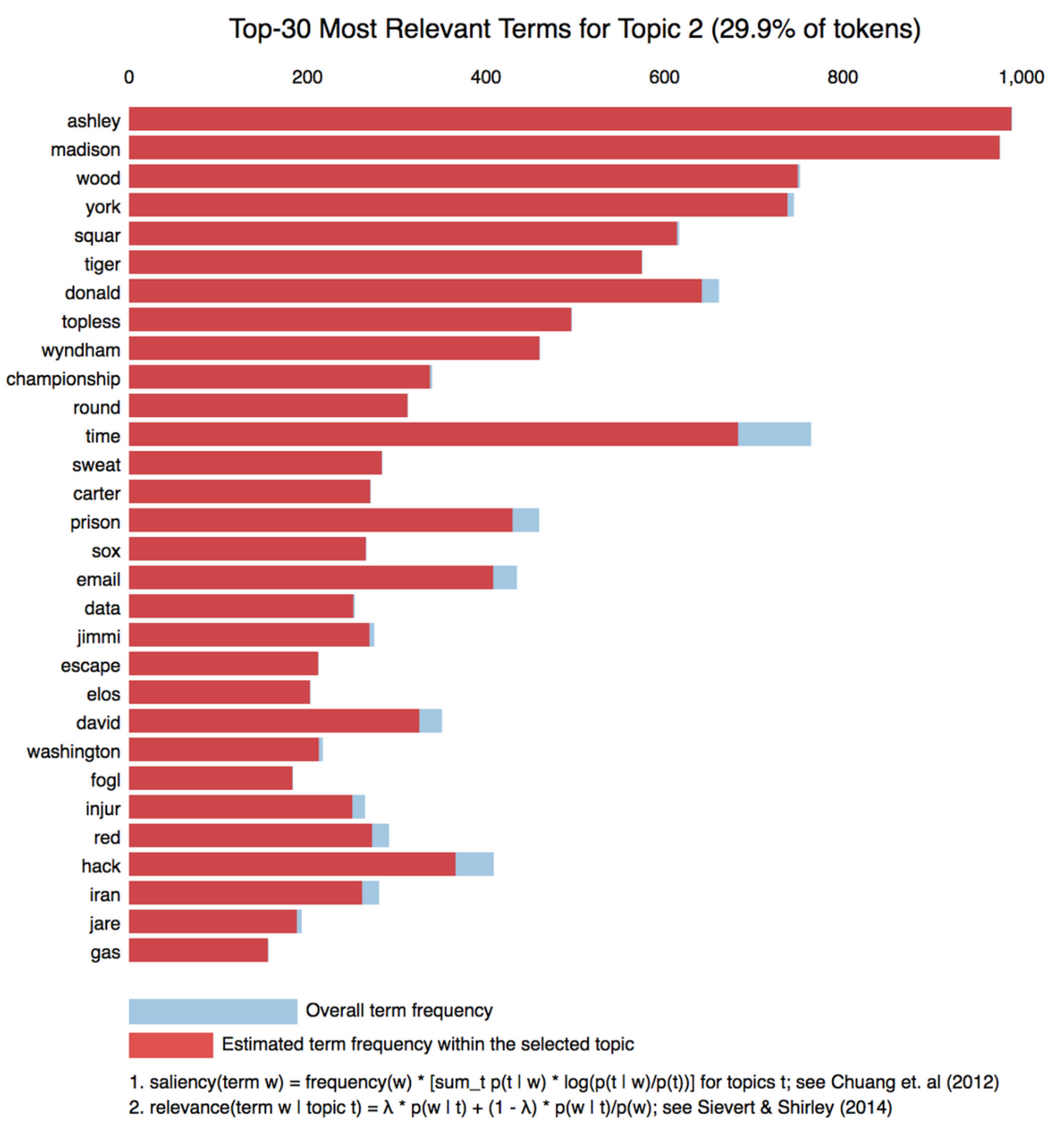
z_{mn} is the topic for the n-th word in document m, and

w_{mn} is the specific word.


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
3
6
7
5
4
1
2
Marginal topic distribution
2%
5%
10%

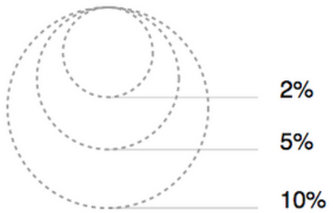

Top-30 Most Relevant Terms for Topic 1 (48.1% of tokens)
0
200
400
600
800
1,000
1,200
1,400
1,600
black
agre
white
support
pleas
good
terrorist
awar
fact
thug
well
membersupport
mental
that
must
defect
right
absolut
total
jill
problem
said
racist
crimin
person
read
dina
cant
advoc
communiti
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
3
6
7
5
4
2
1
Marginal topic distribution
2%
5%
10%

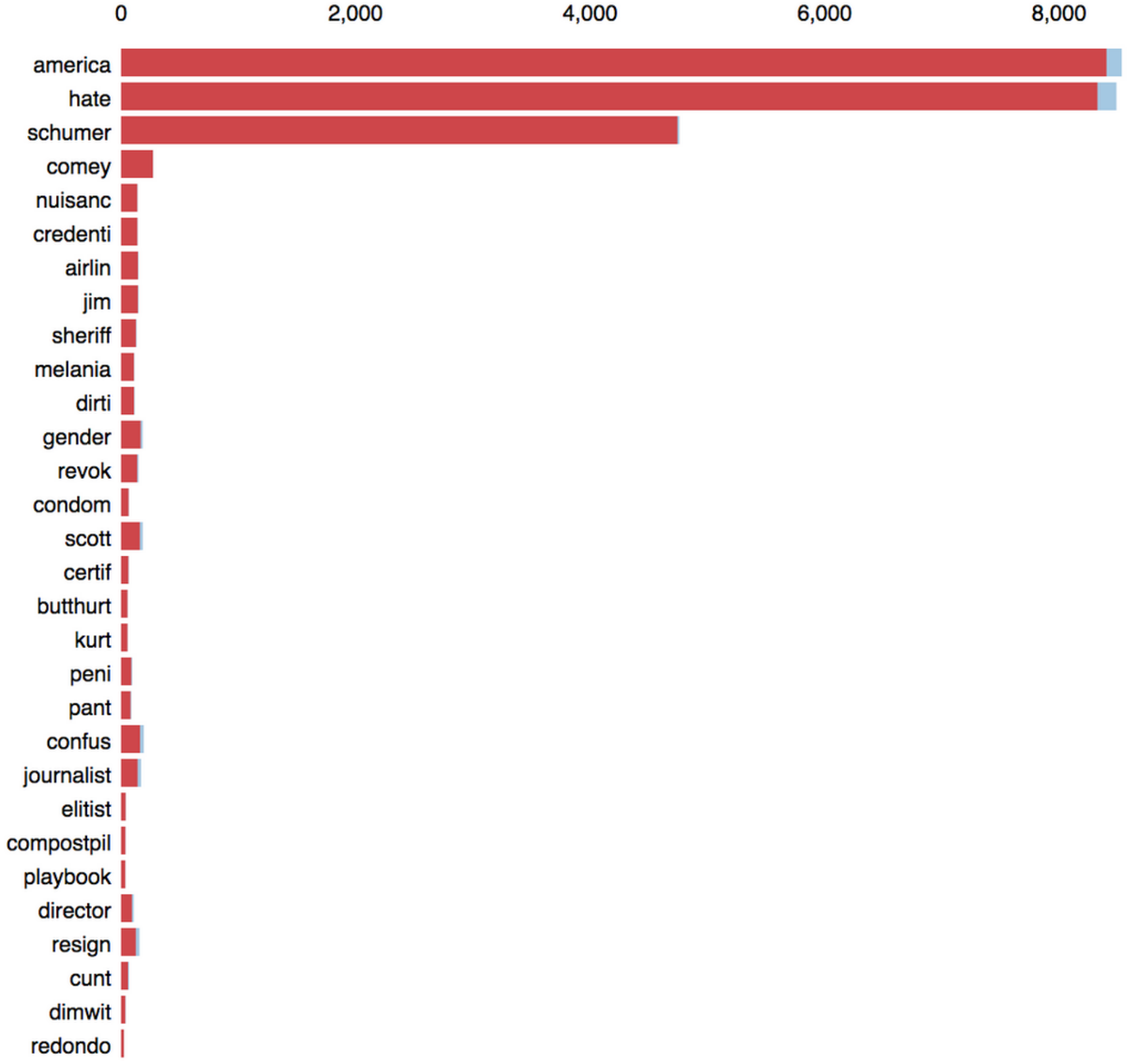

Top-30 Most Relevant Terms for Topic 2 (29.9% of tokens)
0
200
400
600
800
1,000
ashley
madison
wood
york
squar
tiger
donald
topless
wyndham
championship
round
time
sweat
carter
prison
sox
email
data
jimmi
escape
elos
david
washington
fogl
injur
red
hack
iran
jare
gas
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

Marginal topic distribution



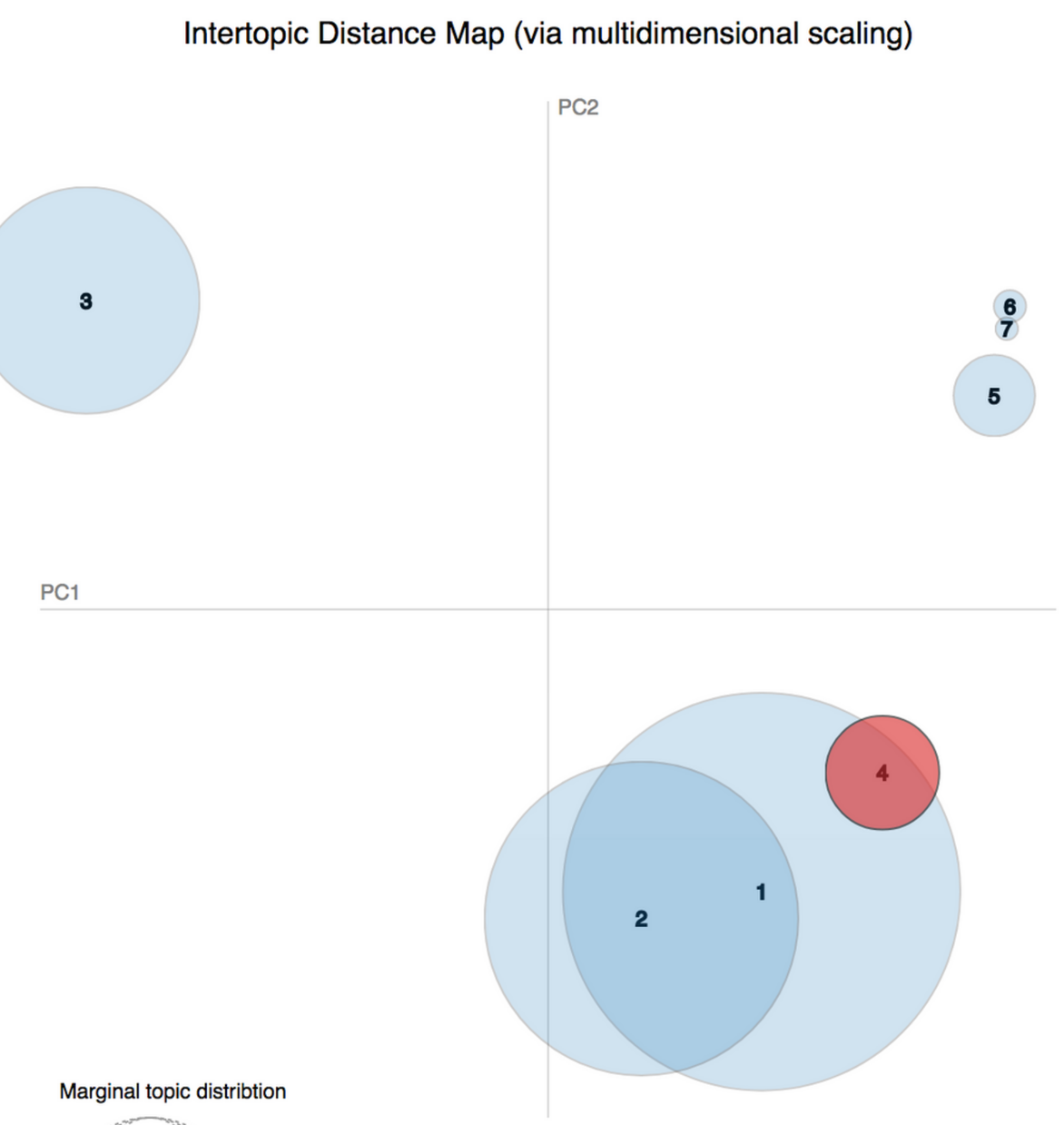
Top-30 Most Relevant Terms for Topic 3 (15.6% of tokens)

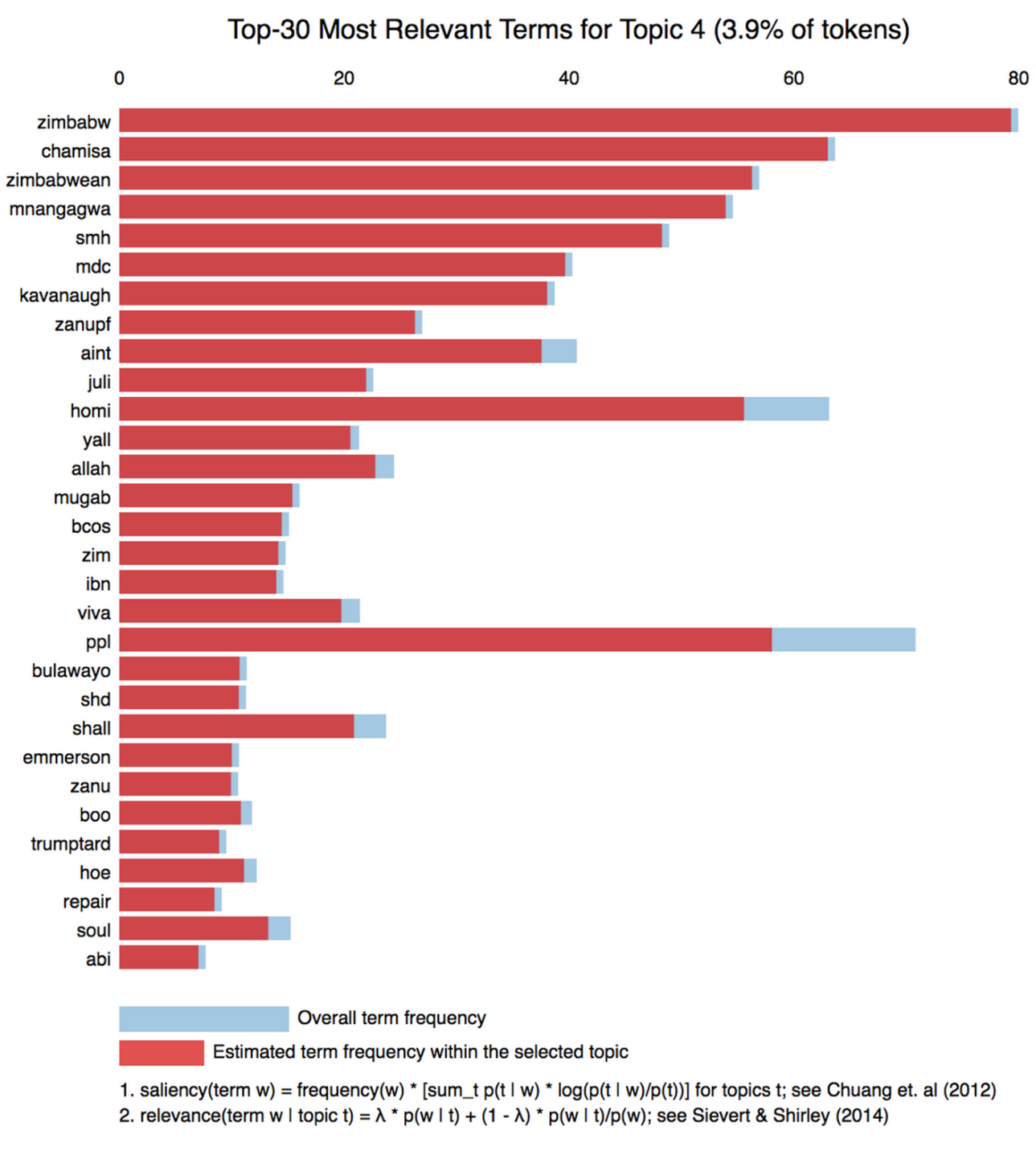


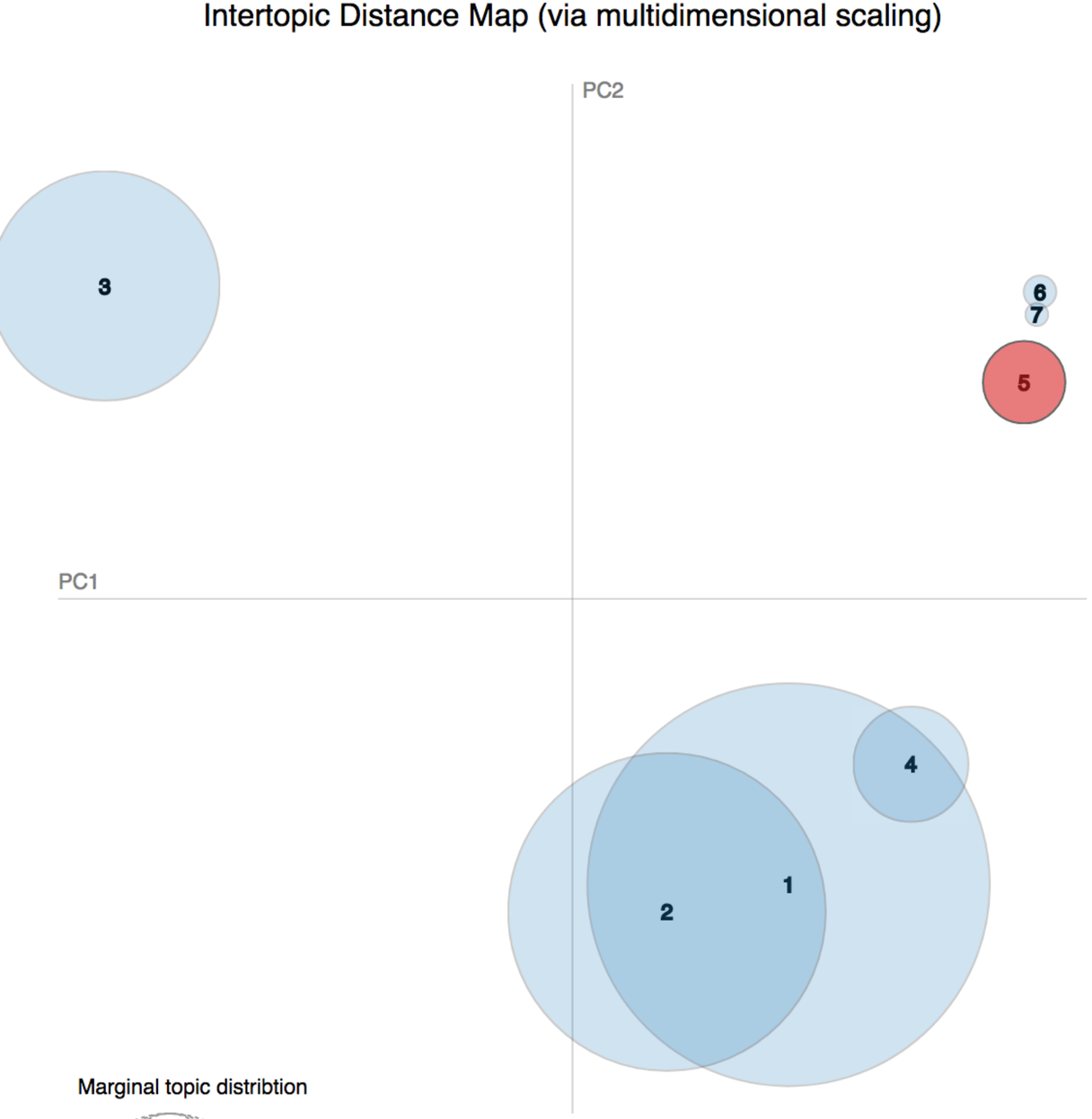
Overall term frequency

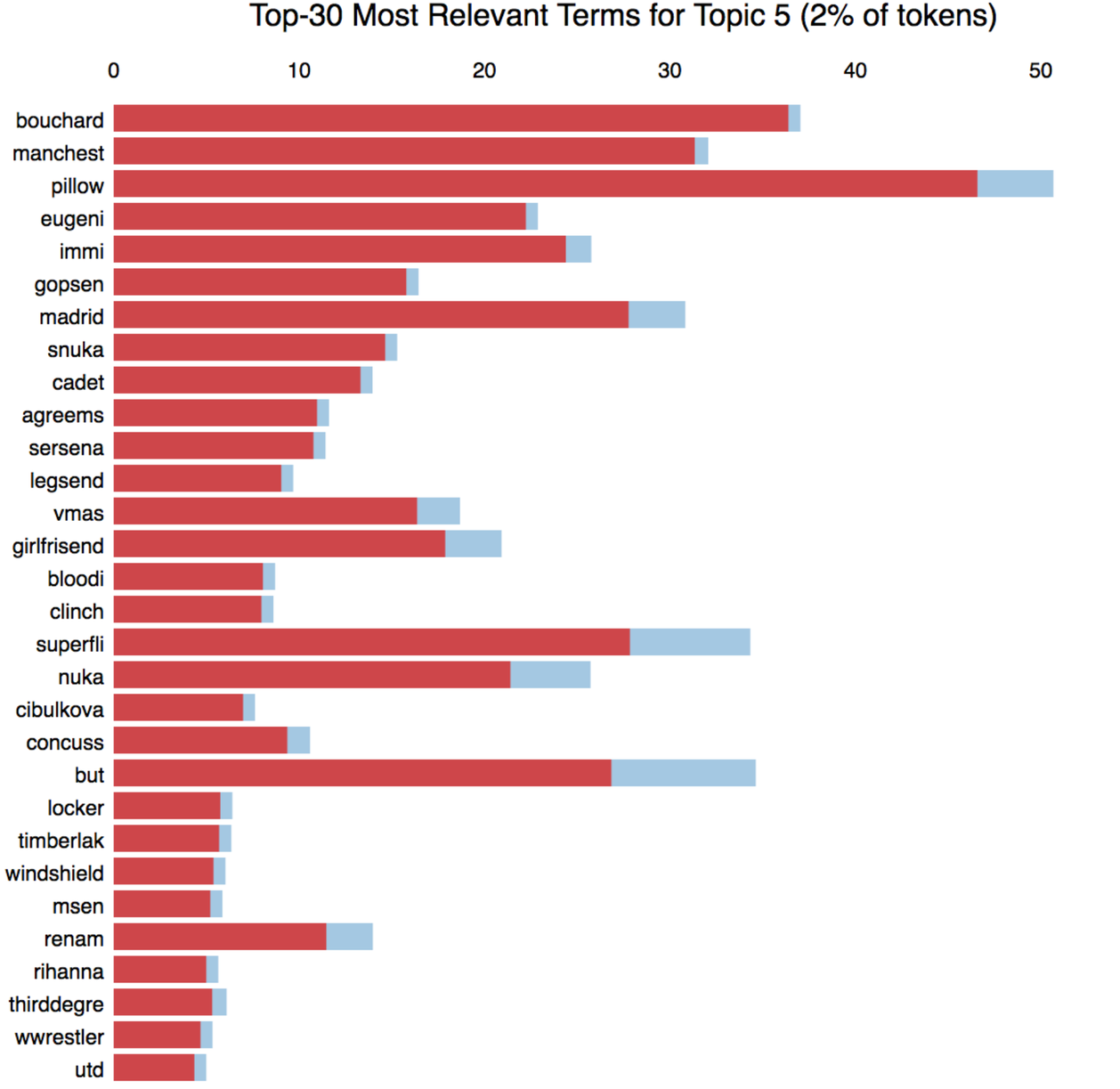
Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
3
6
7
5
4
1
2
Marginal topic distribtion
2%
5%
10%

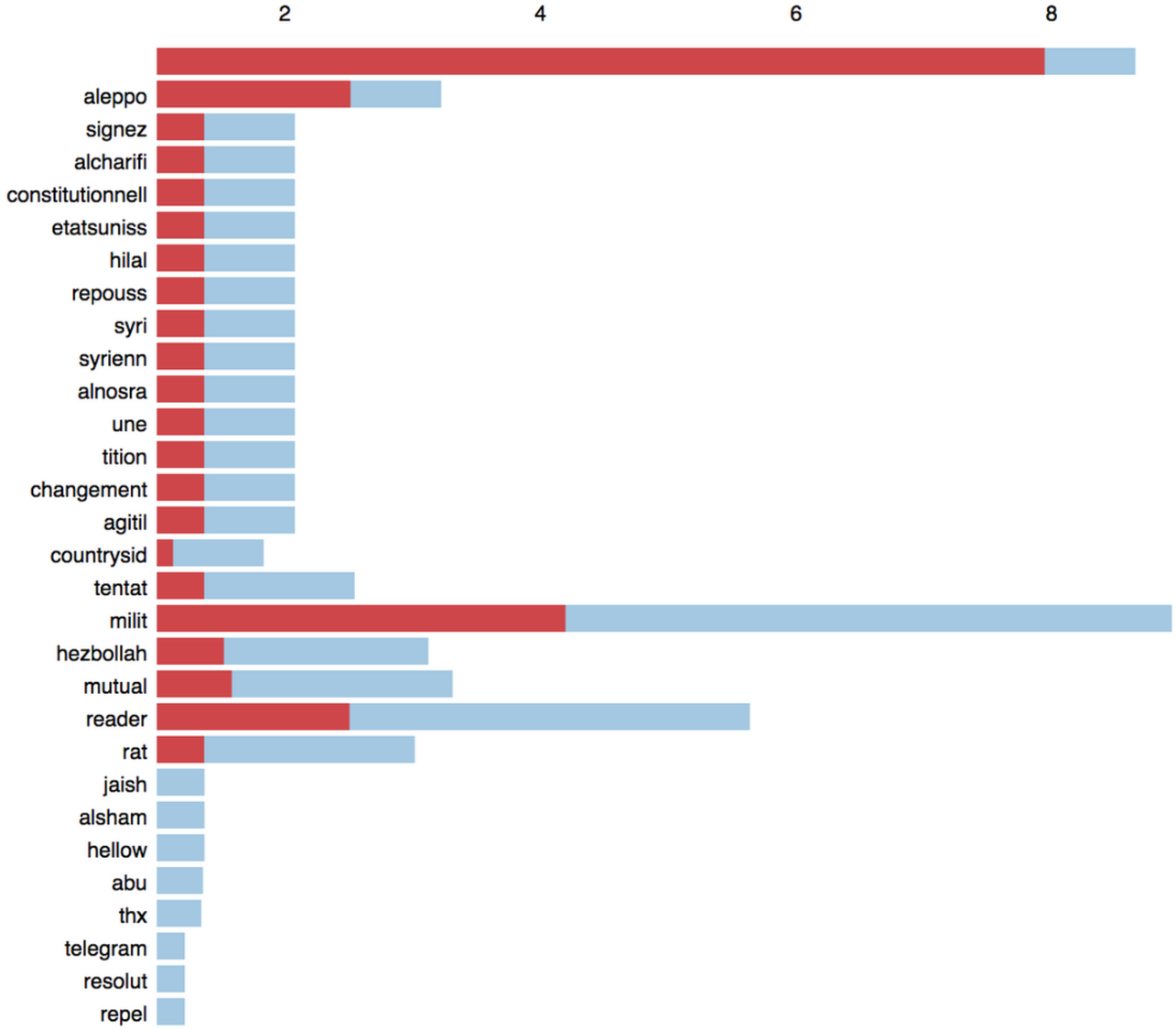

Top-30 Most Relevant Terms for Topic 4 (3.9% of tokens)
0
20
40
60
80
zimbabw
chamisa
zimbabwean
mnangagwa
smh
mdc
kavanaugh
zanupf
aint
juli
homi
yall
allah
mugab
bcos
zim
ibn
viva
ppl
bulawayo
shd
shall
emmerson
zanu
boo
trumptard
hoe
repair
soul
abi
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
3
6
7
5
4
1
2
Marginal topic distribtion
2%
5%
10%

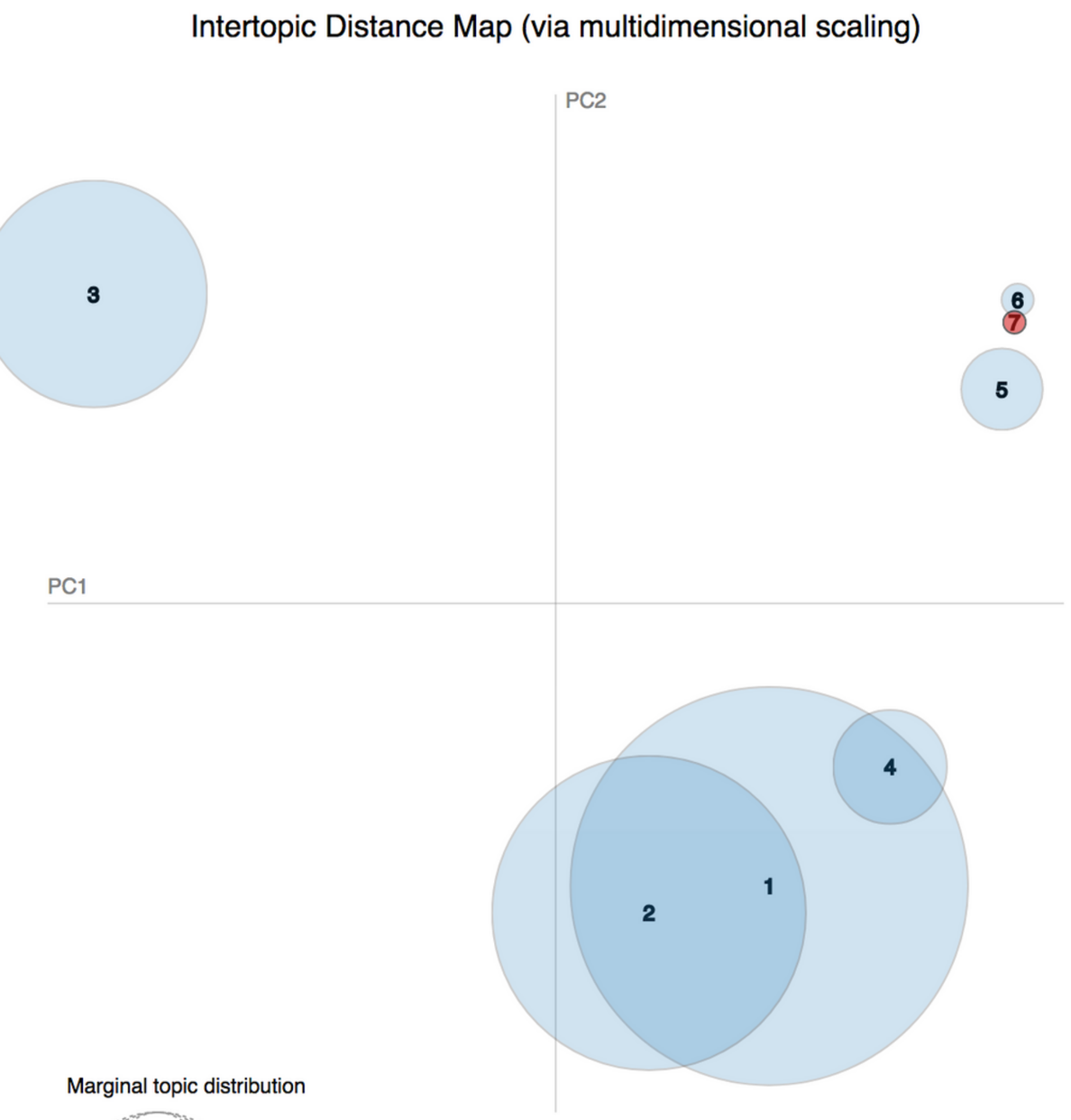

Top-30 Most Relevant Terms for Topic 5 (2% of tokens)
0
10
20
30
40
50
bouchard
manchest
pillow
eugeni
immi
gopsen
madrid
snuka
cadet
agreems
sersena
legsend
vmas
girlfrisend
bloodi
clinch
superfli
nuka
cibulkova
concuss
but
locker
timberlak
windshield
msen
renam
rihanna
thirddegre
wwrestler
utd
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

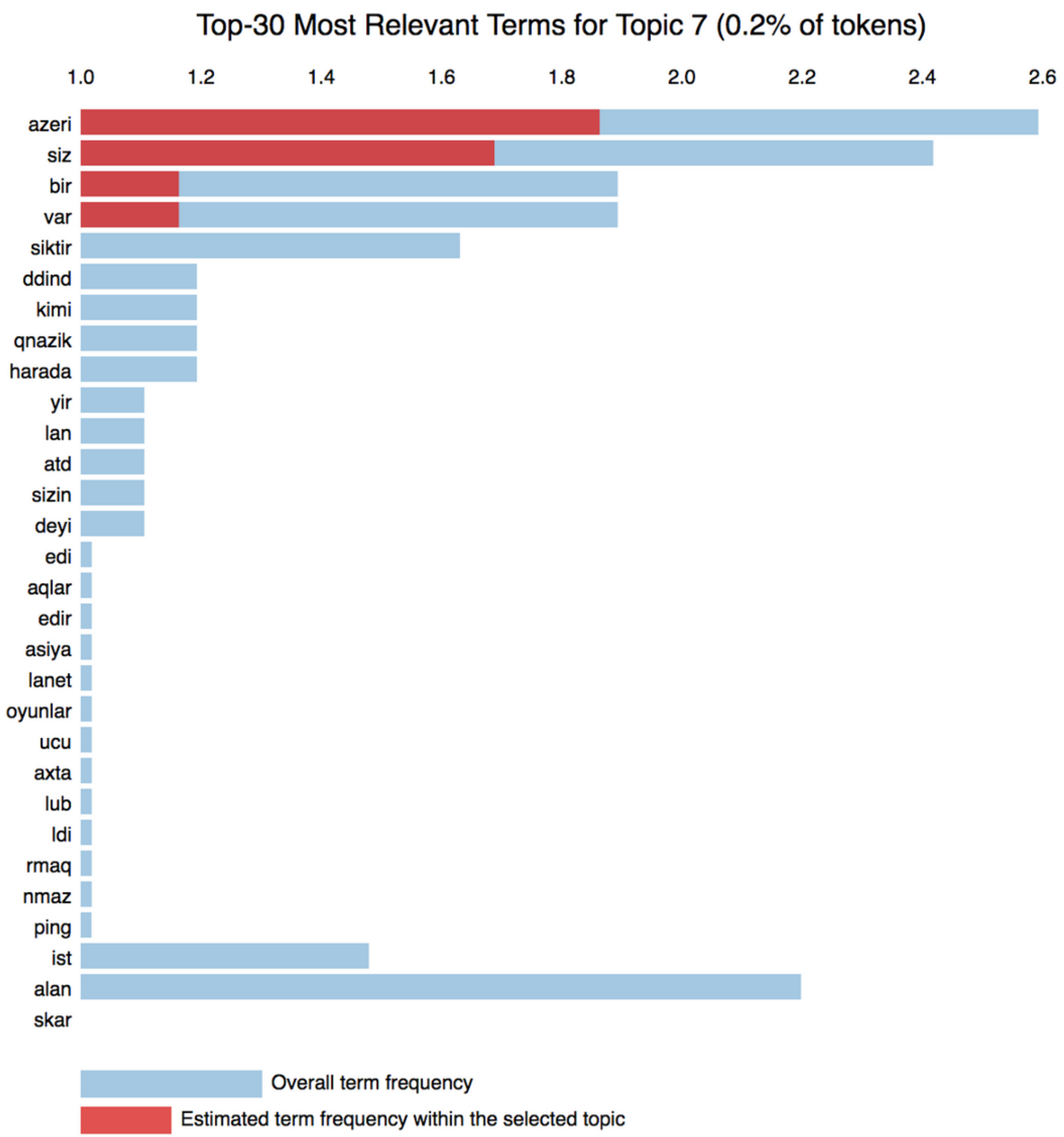
Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 6 (0.3% of tokens)



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)


Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
3
6
7
5
4
1
2
Marginal topic distribution
2%
5%
10%

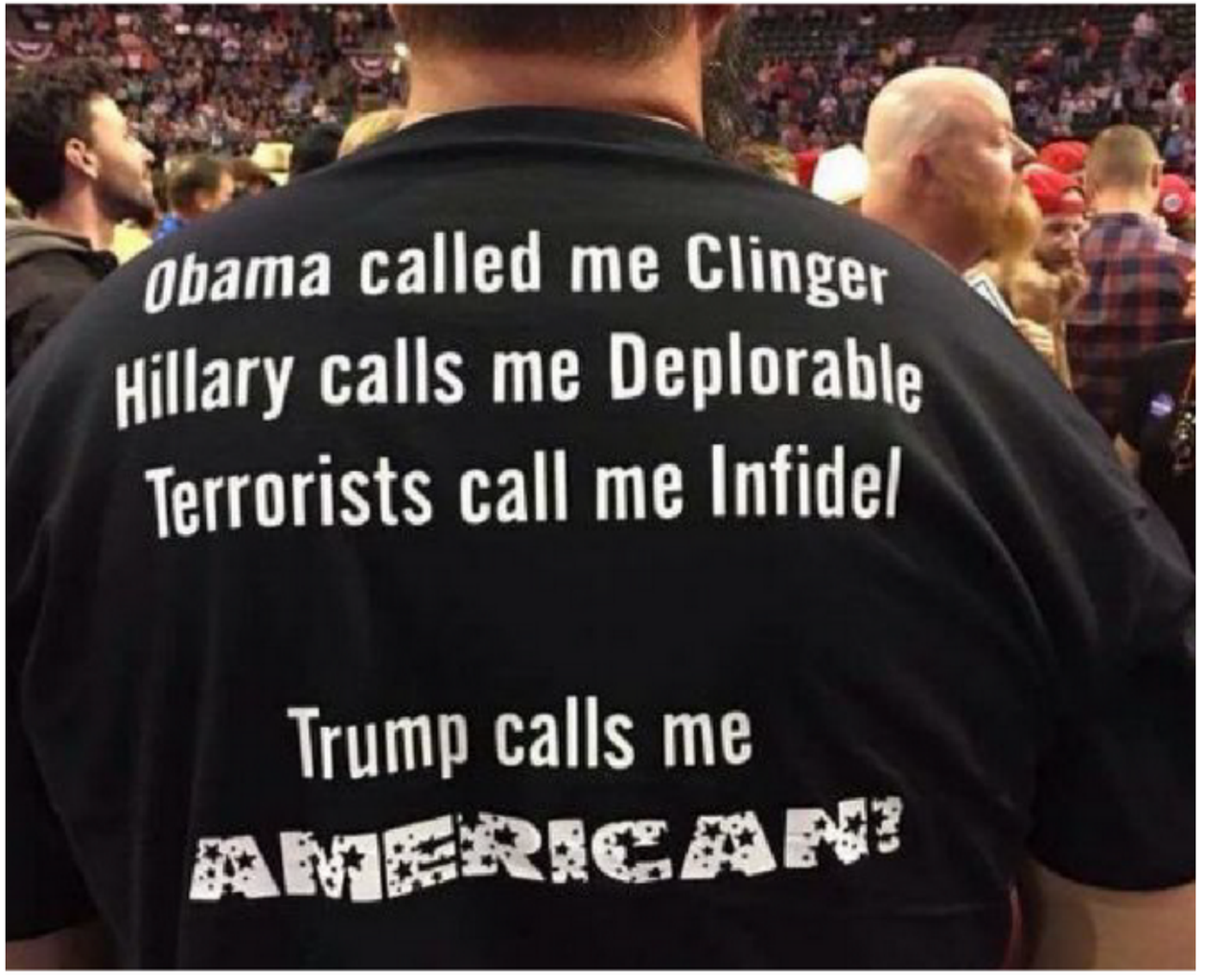

Top-30 Most Relevant Terms for Topic 7 (0.2% of tokens)
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
azeri
siz
bir
var
siktir
ddind
kimi
qnazik
harada
yir
lan
atd
sizin
deyi
edi
aqlar
edir
asiya
lanet
oyunlar
ucu
axta
lub
ldi
rmaq
nmaz
ping
ist
alan
skar
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Tennessee
@TEN_GOP

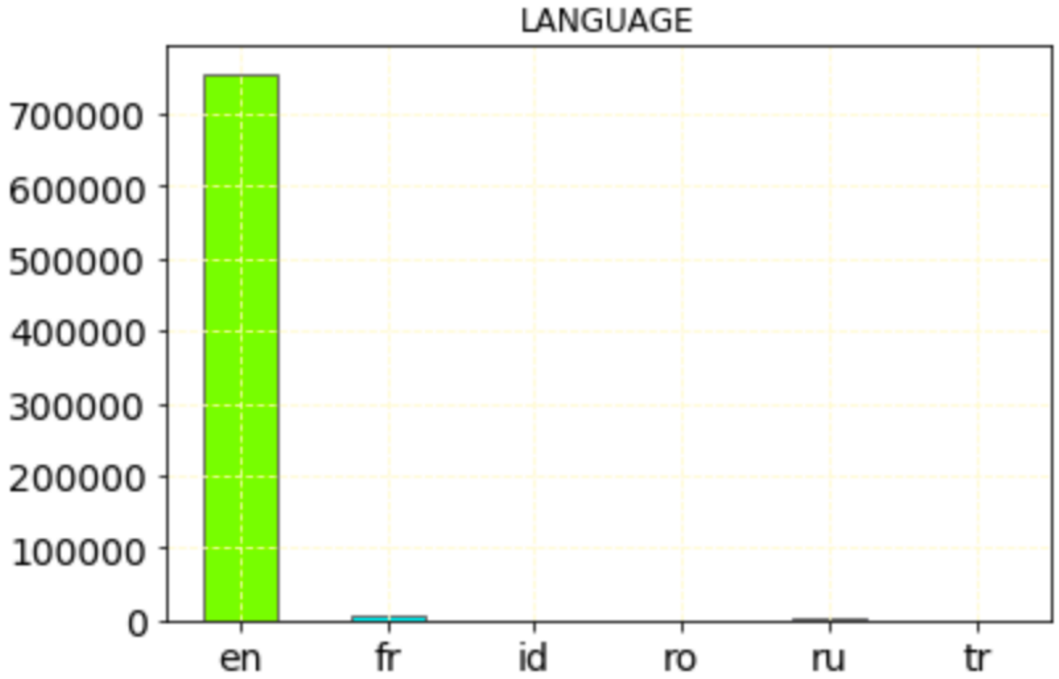
You don't even need these polls..
Donald Trump won! You can read it on Bill Clinton's face..
#debate pic.twitter.com/jCw3loL4EF

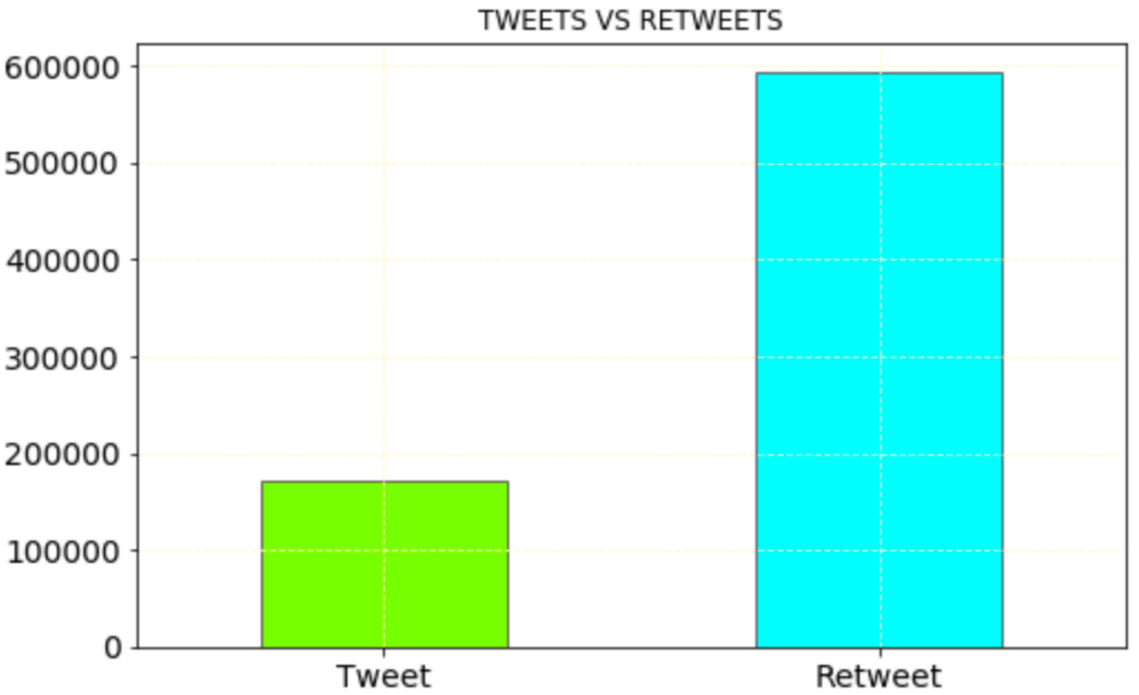
10 Oct 2016

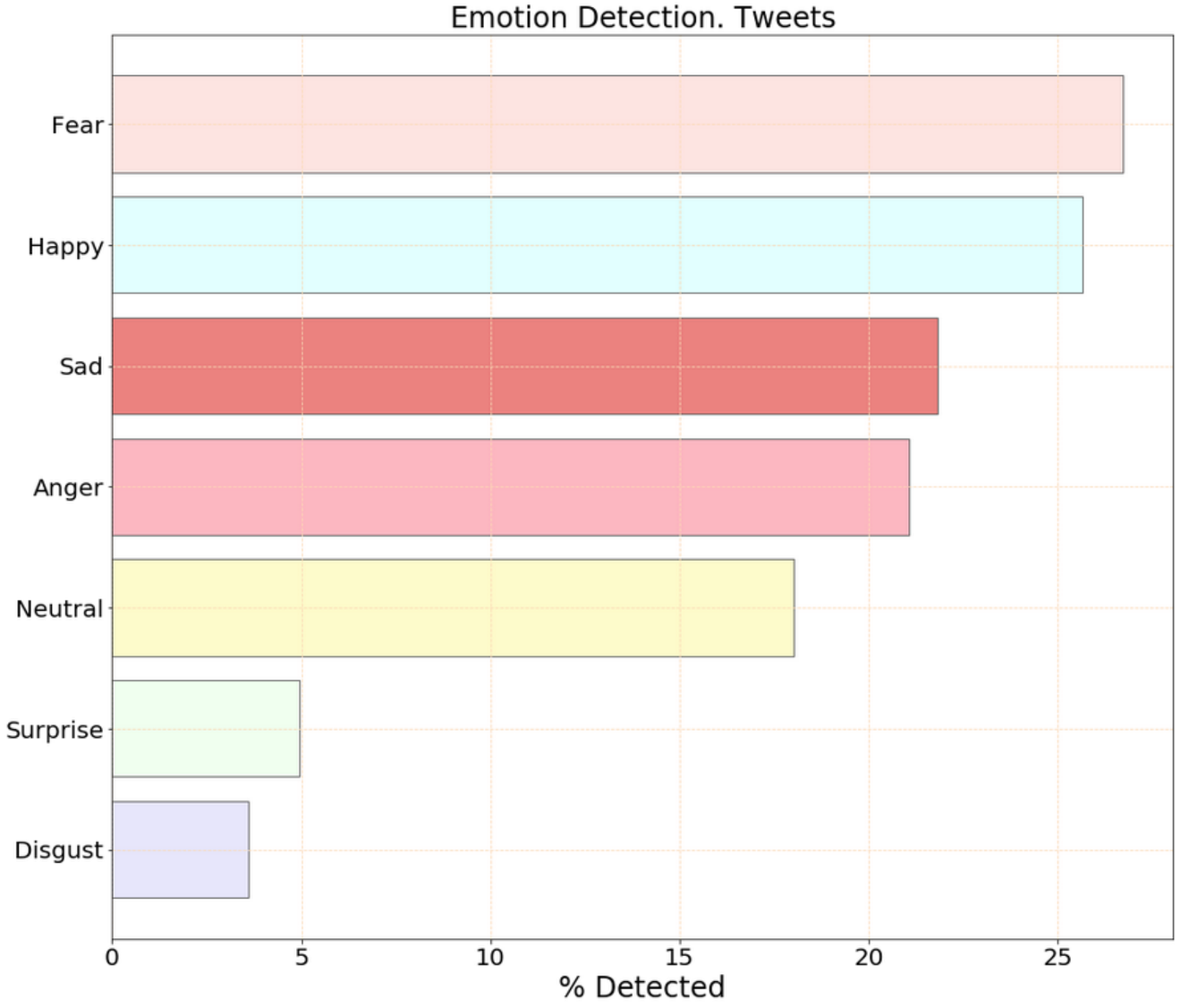
Pamela Moore
@Pamela_Moore13

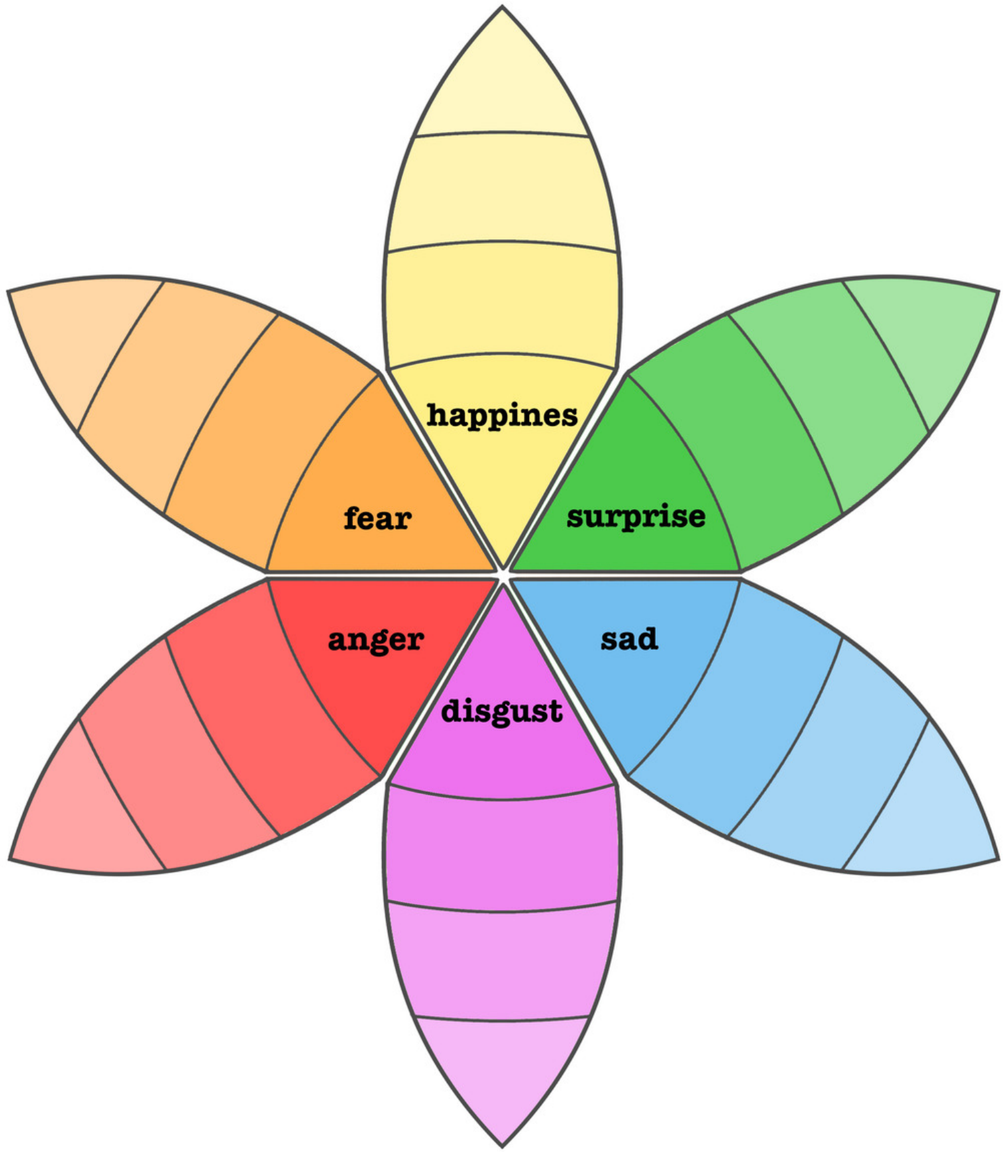



FBI DIRECTOR COMEY
RESIGN
NOW
- The American People


LANGUAGE
700000
600000
500000
400000
300000
200000
100000
0
en
fr
id
ro
ru
tr


TWEETS VS RETWEETS
600000
500000
400000
300000
200000
100000
0
Tweet
Retweet


Emotion Detection. Tweets
Fear
Happy
Sad
Anger
Neutral
Surprise
Disgust
0
5
10
15
20
25
% Detected


happines
fear
surprise
anger
disgust
sad