Characterizing Interconnections and Linguistic Patterns in Twitter

JOHNNATAN MESSIAS

ADVISOR: FABRÍCIO BENEVENUTO

Contents

Introduction

- ✓Social networking sites are powerful
- ✓Facebook: 1.7 bi monthly active users in 2016
- ✓Twitter: 317 mi monthly active users in 2016
- ✓People post everything
- ✓Promote debates
- ✓Demographic information is challenge to obtain
- ✓Why is important to study demographic aspects?

Motivation

- ✓Perspective of Systems
- ✓Sociological point of view
- ✓Linguistic Differences
- ✓Gender and Race Disparities
- ✓Gender and Race Inequalities

- ✓Glass Ceiling
- ✓Not available in Twitter API
- ✓Challenge
- ✓Design Transparent Systems

Goals

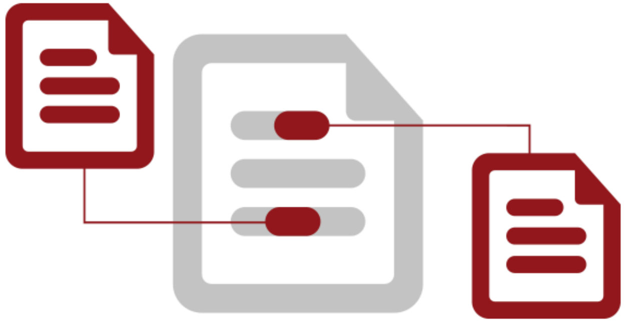
- ✓ Investigate Inequities in Terms of Visibility
- ✓ Investigate Linguistic Aspects and Topics of Interests
- ✓ Characterize Interconnections
- ✓ Design a System that Provides Data Transparency

Contributions

- ✓Clear Insight into How Groups of Users Connect in Twitter
- ✓Linguistic Style of Writing and Topic of Interests of Demographic Groups
- ✓Interconnections and Interactions
- ✓Who Makes Trends? Web-base system
- ✓Published Work

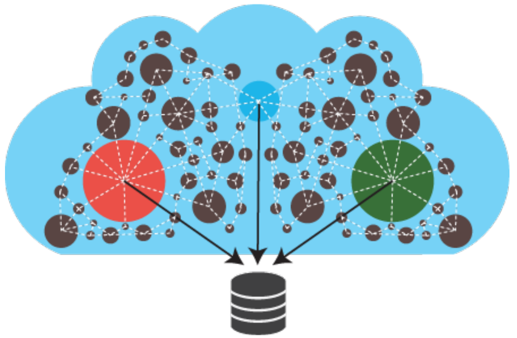
Related Work

- ✓ Demographics in Social Media
- ✓ Inequality in Twitter Visibility
- ✓ Demographics and Linguistic Analysis
- ✓ Algorithmic and Data Transparency
- ✓ Recommendation Diversity
- ✓ Fairness



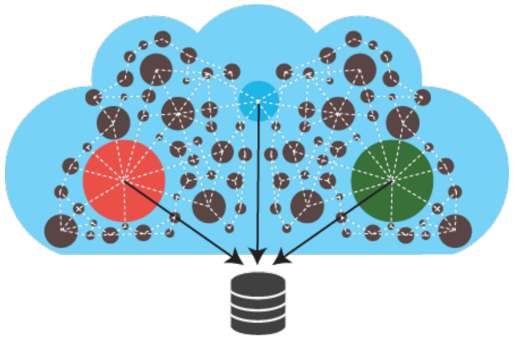
Demographic Information Dataset

- ✓Twitter Dataset
- ✓Crawling Demographic Information
- ✓Baseline Dataset
- ✓Gathering Tweets
- ✓Extraction of Topics
- ✓Linguistic Measures
- ✓Gathering Social Connections and Interactions
- ✓Potential Limitations



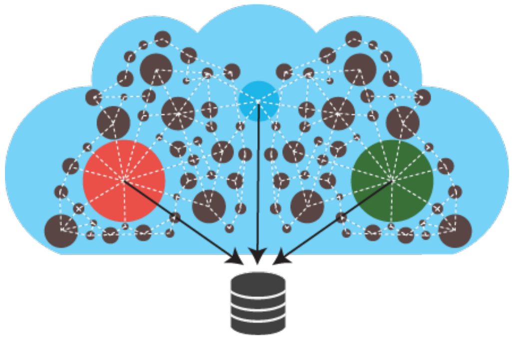
Twitter Dataset

- ✓Twitter Stream API
- ✓1% Random Sample
- ✓July – September 2016
- ✓341,457,982 tweets
- ✓50,270,310 users
- ✓6,286,477 users from U.S. and English tweet
- ✓Time zone filtering



Crawling Demographic Information

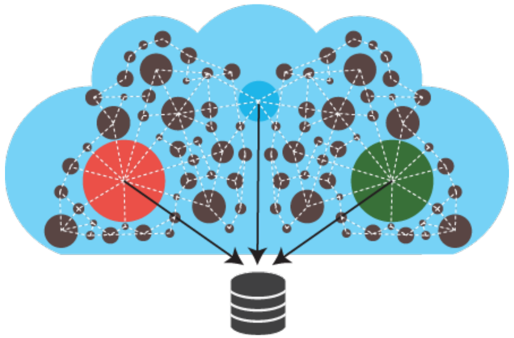
- ✓ Profile Pictures URL
- ✓ Face++ API: Gender, Race, Age, and other attributes
- ✓ 4.6 mi users discarded (73.42%)
  - Users changed their profile picture
  - Pictures. do not have a face
  - Pictures have more than one face
- ✓ 1,670,862 U.S. users with one face



Baseline Dataset

Race	Gender		Total
	Male	Female	
Asian	120,950 (7.24%)	177,205 (10.61%)	298,155 (17.85%)
Black	130,954 (7.84%)	107,827 (6.45%)	238,781 (14.29%)
White	538,625 (32.23%)	595,302 (35.63%)	1,133,927 (67.86%)
Total	790,529 (47.31%)	880,334 (52.69%)	1,670,863 (100%)

- ✓ 1.6 mi users
- ✓ U.S.
- ✓ 1 recognized face

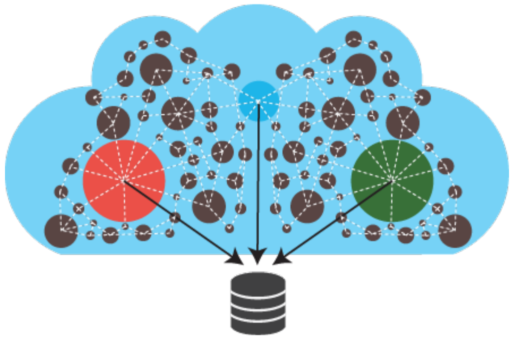


Baseline Dataset

Race (%)	Gender (%)		Total (%)
	Male	Female	
Asian	7.07 (−3.85)	10.05 (−11.28)	17.12 (−10.90)
Black	8.17 (8.53)	6.74 (7.68)	14.91 (11.69)
White	32.88 (8.49)	35.09 (−7.69)	67.97 (1.20)
Total	48.12 (10.91)	51.88 (−10.91)	100.00

- Limitations
- 304,477 random users
- Null model
- 100 random samples

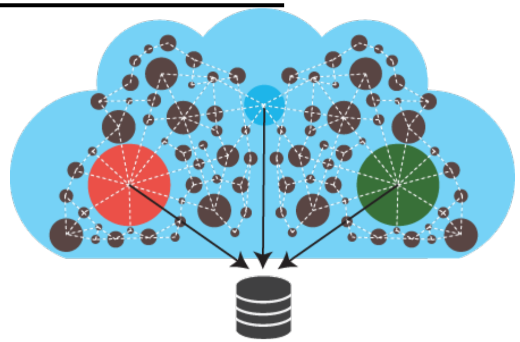
$$Z_{White} = \frac{|U_{White}| - mean(|S_{White}|)}{std(|S_{White}|)}$$



Baseline Dataset

Demographic	Mean	Z-value	S.D.	Min	25-perc	Median	75-perc	Max
Male	$144,035.1 \pm 44.86$	10.91	228.88	143,544	143,883.00	144,054.5	144,156.50	144,680
Female	$160,441.9 \pm 44.86$	-10.91	228.88	159,797	160,320.50	160,422.5	160,594.00	160,933
Asian	$54,311.5 \pm 39.17$	-10.90	199.87	53,907	54,177.25	54,296.5	54,444.00	54,803
Black	$43,514.01 \pm 31.72$	11.69	161.85	43,196	43,380.75	43,503.5	43,633.50	43,887
White	$206,651.49 \pm 46.82$	1.20	238.91	205,921	206,490.25	206,666.5	206,789.25	207,110
Asian Male	$22,043.64 \pm 26.24$	-3.85	133.88	21,674	21,958.75	22,040.5	22,115.50	22,429
Asian Female	$32,267.86 \pm 28.92$	-11.28	147.56	31,900	32,153.50	32,262.0	32,371.75	32,667
Black Male	$23,857.98 \pm 23.81$	8.53	121.48	23,634	23,777.75	23,858.0	23,930.00	24,197
Black Female	$19,656.03 \pm 21.82$	7.68	111.34	19,342	19,585.25	19,660.5	19,737.75	19,944
White Male	$98,133.48 \pm 45.61$	8.49	232.73	97,538	97,995.25	98,130.5	98,297.50	98,623
White Female	$108,518.01 \pm 43.04$	-7.69	219.62	108,025	108,348.25	108,501.5	108,688.00	109,015

✓ 95% confidence level

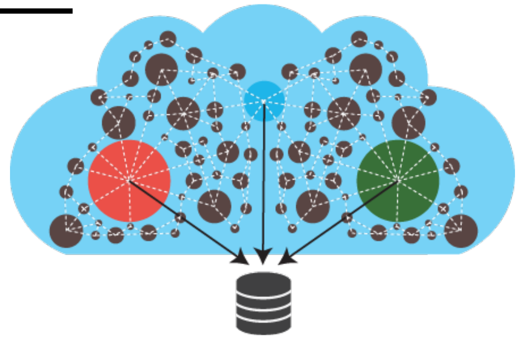


Gathering Tweets

Demographic	Mean	Median	Max
Male	$11,624.76 \pm 109.40$	$3,874$	$1,683,948$
Female	$12,933.40 \pm 105.89$	$4,885$	$1,132,964$
Asian	$14,020.92 \pm 183.73$	$5,544$	$1,108,525$
Black	$18,949.91 \pm 248.46$	$8,245$	$973,225$
White	$10,432.49 \pm 85.28$	$3,637$	$1,683,948$

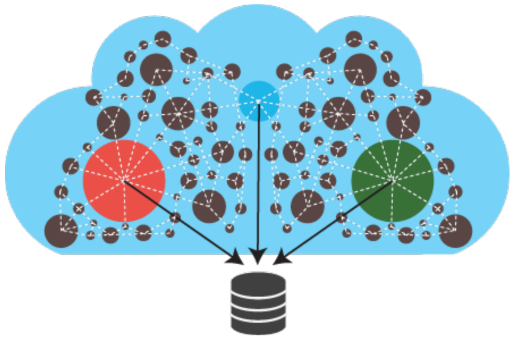
- ✓ Twitter Rest API
- ✓ 3,200 more recently tweets
- ✓ 304,477 users
- ✓ Twitter Limitation

- ✓ 95% confidence level



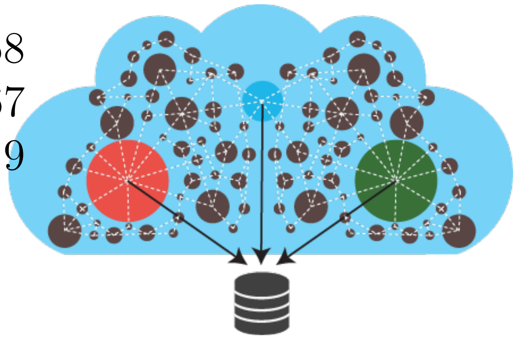
Extraction of Topics

- ✓ Who Likes What Web-base Service
- ✓ List of the friends
- ✓ Manually cleaned sub-topics into:
  - *celebrities == famous*
  - *actors == actor*
  - *business == biz*
  - Removed: best, br, bro, new
- ✓ Top 20: by frequency



Extraction of Topics

Topic	Sub-Topics	Total
World	world, earth, hollywood, usa, canada, texas, international, nyc, country, city, boston, san francisco, france, america, los angeles, brasil, london, india	290, 030
Celebrities	celebrities, famous, stars, celebs, celebrity, star, celeb	245, 125
Entertainment	entertainment	244, 956
Music	music, pop, hip hop, rap, gospel, hiphop	227, 986
TV	tv, television	225, 682
Life	life, lifestyle, health, healthcare, fitness, food, style, smile, drink	157, 032
Fun	fun, funny, humor, lol, laugh	154, 058
Info	info, information	147, 567
Artists	musicians, singers, artist, singer, musician, rappers, bands	141, 519

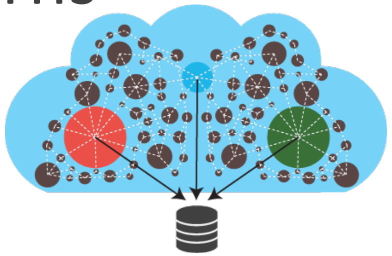


Actors	actors, actresses, actress, actor	140, 647
Media	sports news, tech news, newspapers, music news, breaking news, world news, news media, radio, internet, social media, youtube, sports media, magazines, magazine	135, 849
Writers	writers	126, 051
Bloggers	bloggers, blogs, blog	110, 699
Business	business, biz, businesses	107, 361
Sports	sports, football, basketball, baseball, soccer, futbol, basket, martial arts, sport, mma, golf, cricket, boxing, motorsports, f1, racing	93, 611
Movie	movie, movies, film, films	88, 863
Organizations	organizations, nfl, nba, mlb, nhl, ufc, lfc, lgbt	82, 568
Technology	technology, tech, iphone, digital, geek, software, computer, electronic, android, xbox, mac, gadgets, programming, geeks	72, 137
Politics	politics, government, political, politicians, politician	64, 735
Companies	companies, apple, company, microsoft, google	53, 128

Linguistic Measures

- ✓Linguistic Inquiry and Word Count (LIWC)
- ✓Super text of tweets
- ✓3 categories:
  - Affective
  - Cognitive
  - Linguistic Style
- ✓36 features

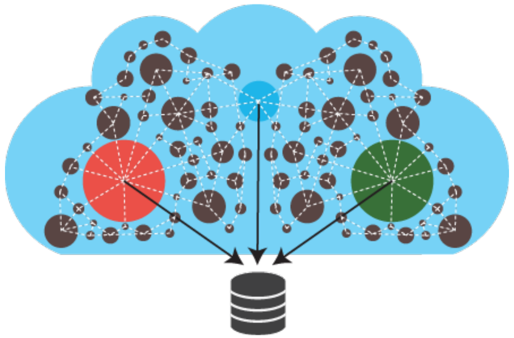
- ✓6 groups:
  - Affective Attributes
  - Cognitive Attributes
  - Lexical Density and Awareness
  - Temporal References
  - Social/Personal Concerns
  - Interpersonal Focus



Gathering Social Connections and Interactions

- ✓ Followers and Friends
- ✓ Unfeasible due to Face++
- ✓ Randomly Select 6,000 users
- ✓ Gather their friends (max of 5,000)
  - Most recent
  - All friends: 98.51%
- ✓ Gather demographic information
  - At least 5% of users
  - Avg. 10.15% and median: 9.40%

- ✓ Interactions based on RT and mentions
- ✓ Crawled all tweets (max of 3,200) for each user
- ✓ Identified users mentioned or retweeted
- ✓ Gather Demographic Information
  - 5% of retweeters and who mentioned



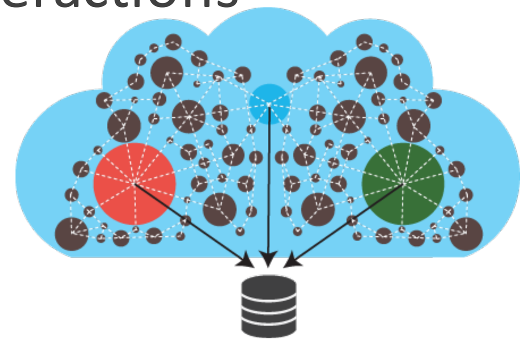
Gathering Social Connections and Interactions

	White	Black	Asian	Total
Male	151,840	52,437	24,299	228,576
Female	137,010	31,011	32,100	200,121
Total	288,850	83,448	56,399	428,697

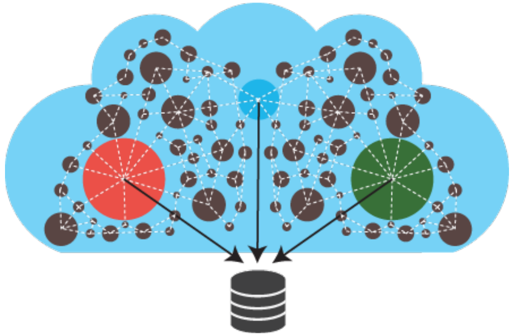
- Number of Friends

	White	Black	Asian	Total
Male	246,879	109,744	51,370	407,993
Female	202,338	60,108	71,137	333,583
Total	449,217	169,852	122,507	741,576

- Number of Interactions

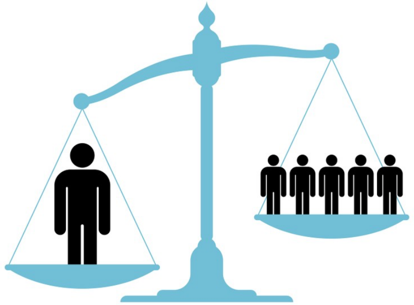


Potential Limitations



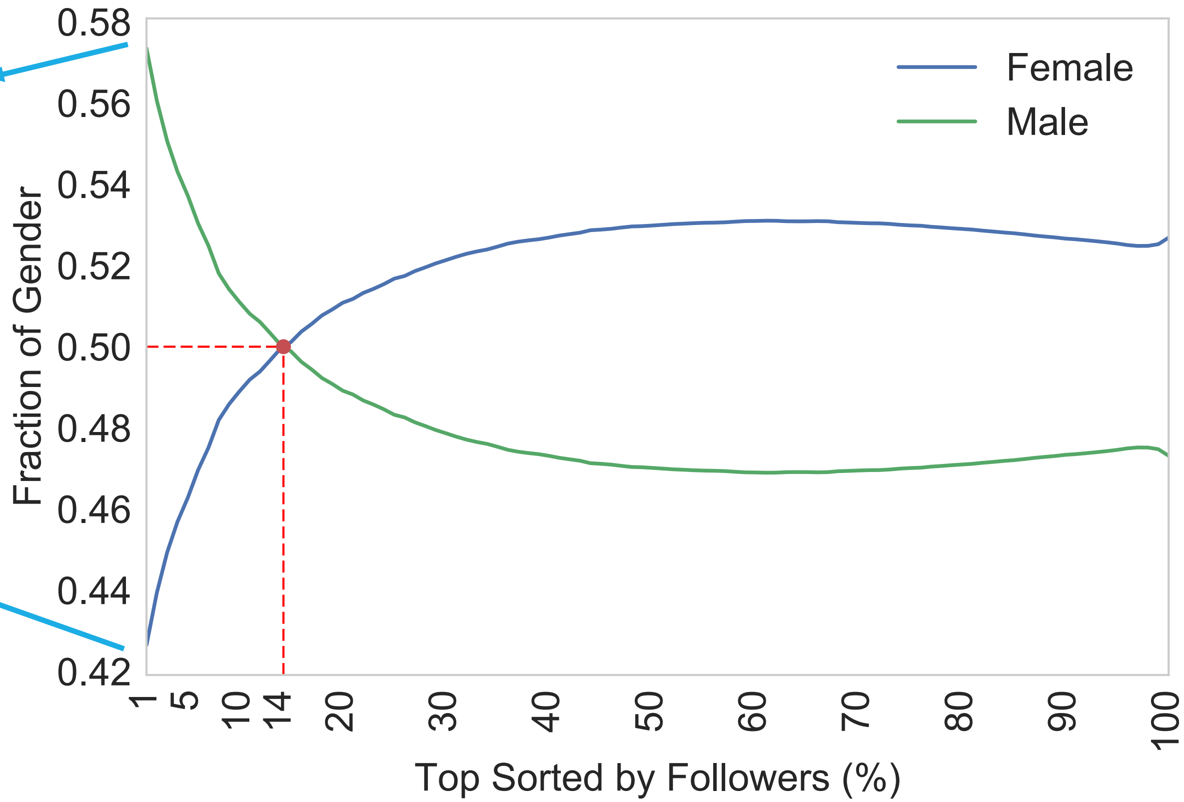
Inequality in Visibility

- ✓Analyze the Association of Demographic Aspects with Visibility
- ✓Discover Possible Inequalities
- ✓Audience Size: Followers and Lists
- ✓Gender Inequality
- ✓Race Inequality
- ✓Taking Together Gender and Race Inequality



Gender Inequality

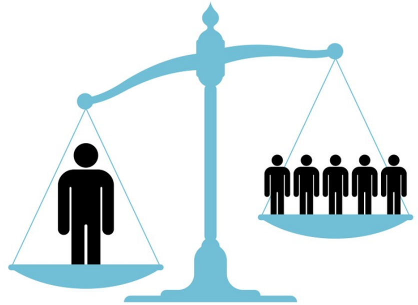
- 57.34%
- 42.65%



✓Males tend to be more followed

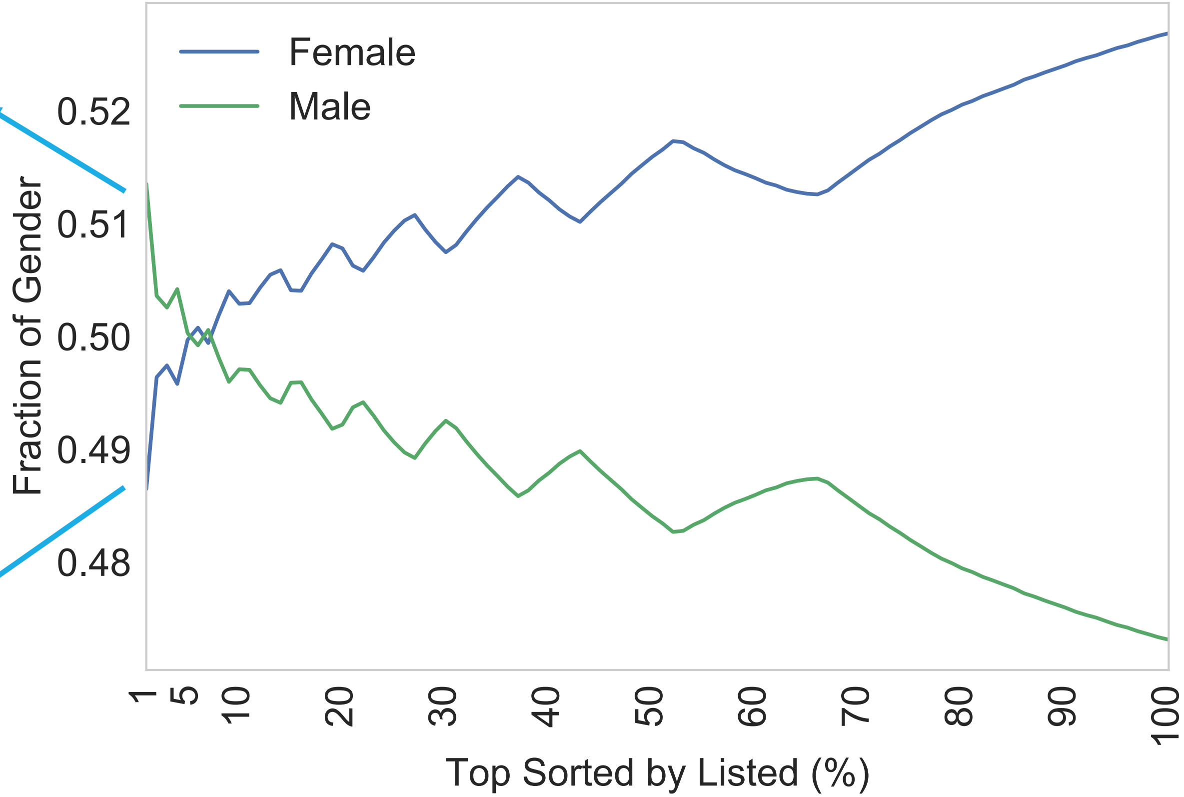
✓Glass Ceiling

✓Gender Disparity

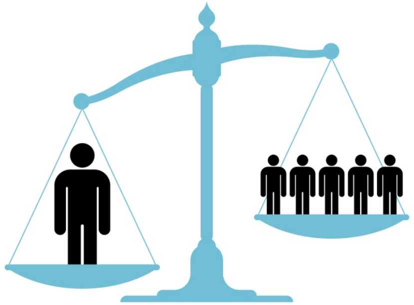


Gender Inequality

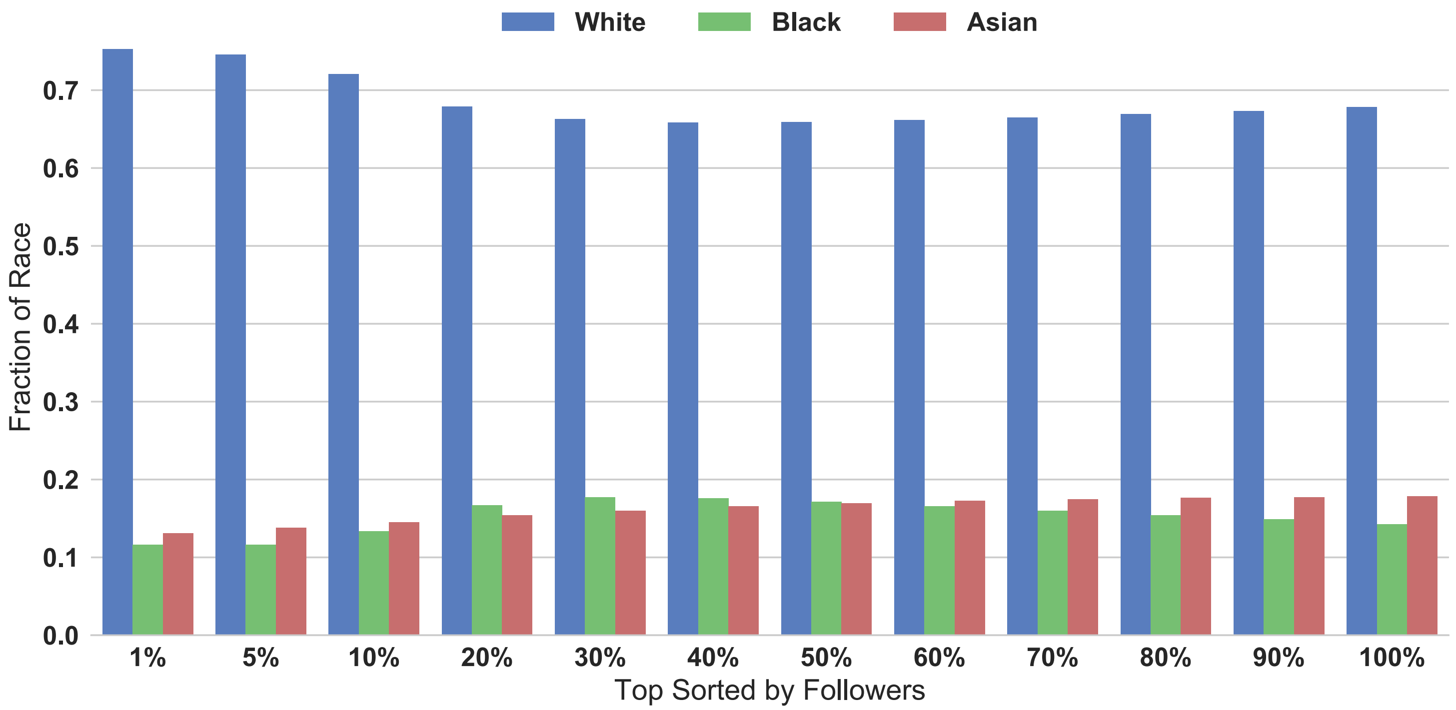
- 51.23%
- 48.64%



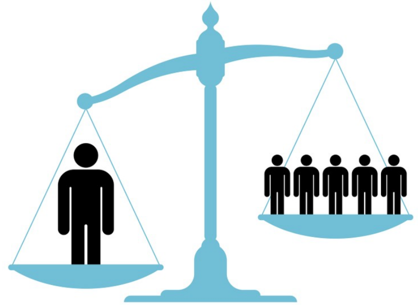
- ✓ Males tend to be more listed
- ✓ Glass Ceiling
- ✓ Gender Disparity



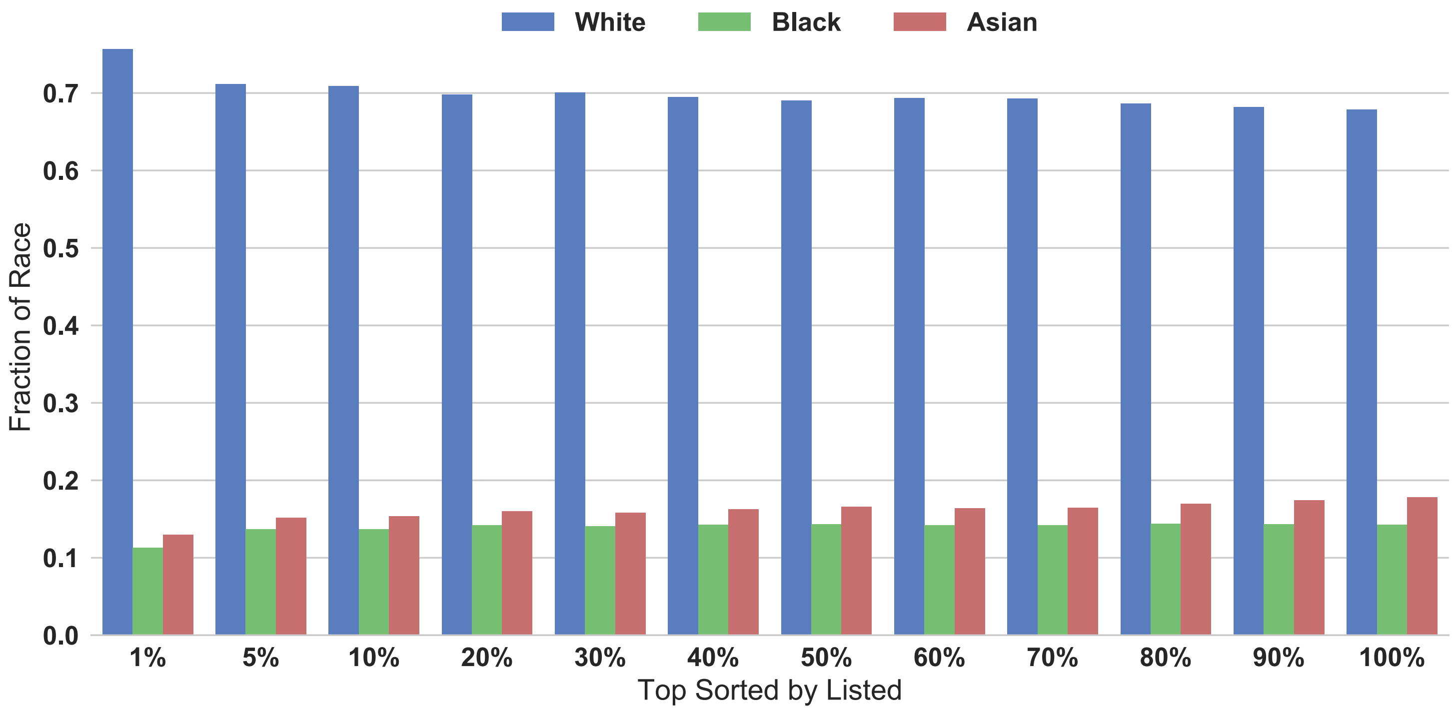
Race Inequality



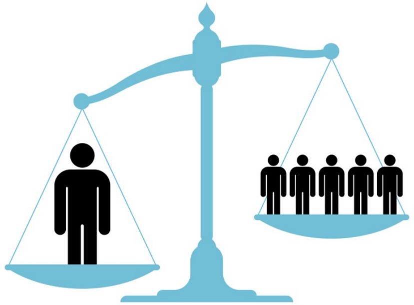
- ✓White tend to be more followed
- ✓Glass Ceiling
- ✓Race Disparity



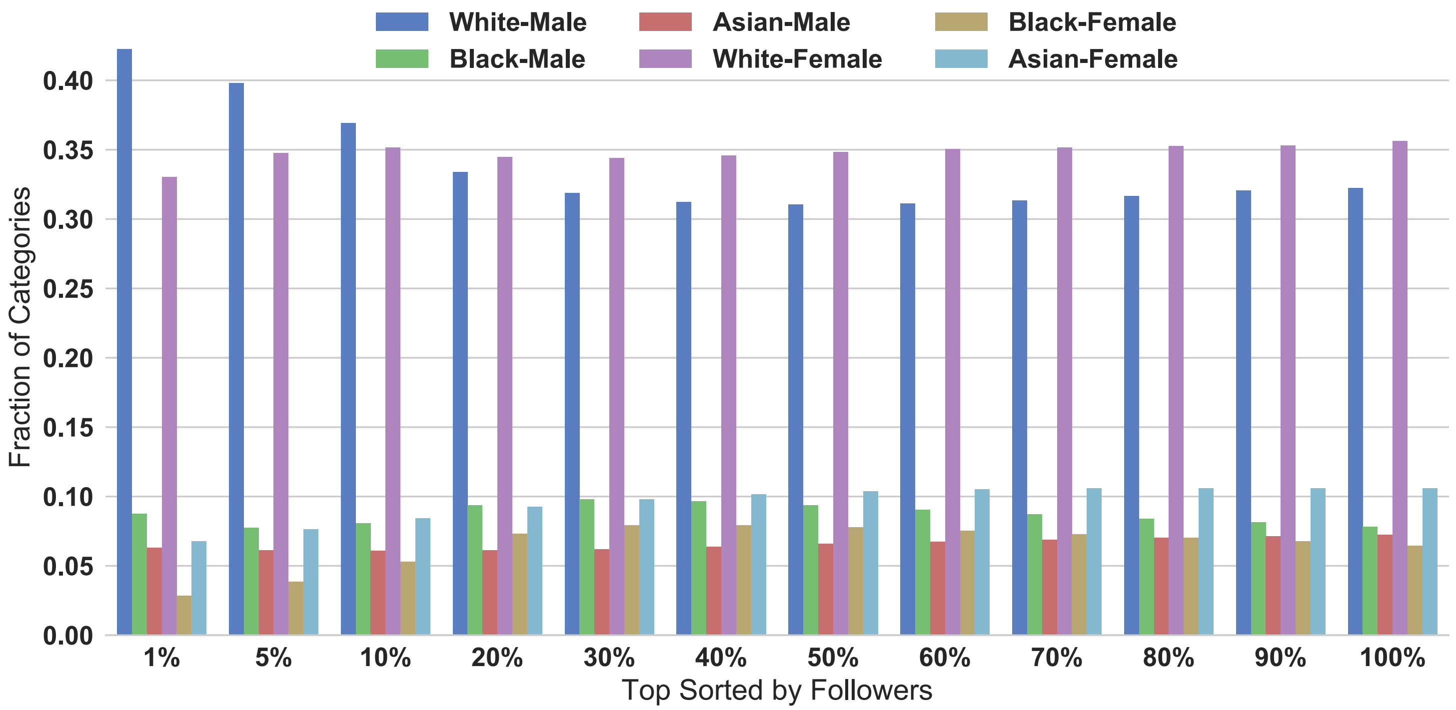
Race Inequality



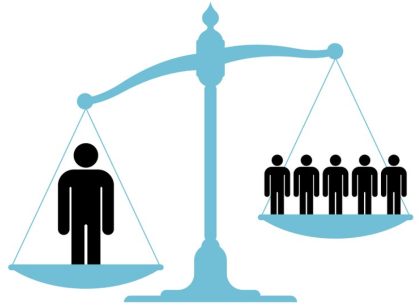
- ✓White tend to be more listed
- ✓Glass Ceiling
- ✓Race Disparity



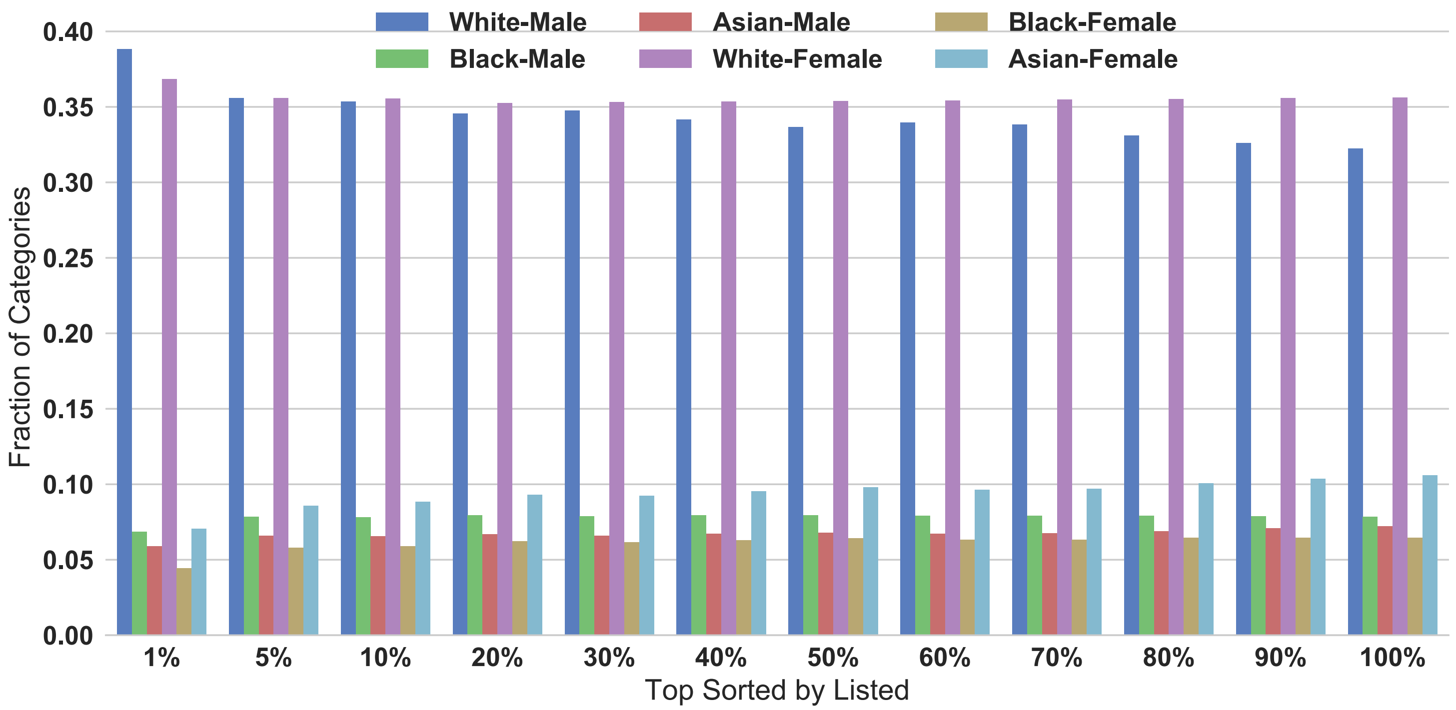
Taking Together Gender and Race Inequality



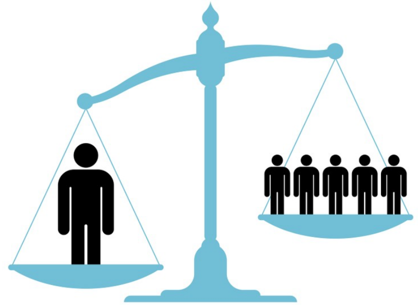
- ✓White male tend to be more followed
- ✓Also Glass Ceiling for males
- ✓Group Disparity



Taking Together Gender and Race Inequality



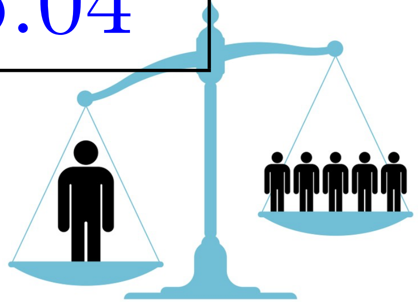
- ✓White male tend to be more listed
- ✓Also Glass Ceiling for males
- ✓Group Disparity



Taking Together Gender and Race Inequality

Race	Followers		Listed	
	Male	Female	Male	Female
Asian	-10.60	-32.70	-16.36	-29.61
Black	+7.17	-57.73	-15.90	-34.20
White	+28.56	-5.84	+18.15	+5.04

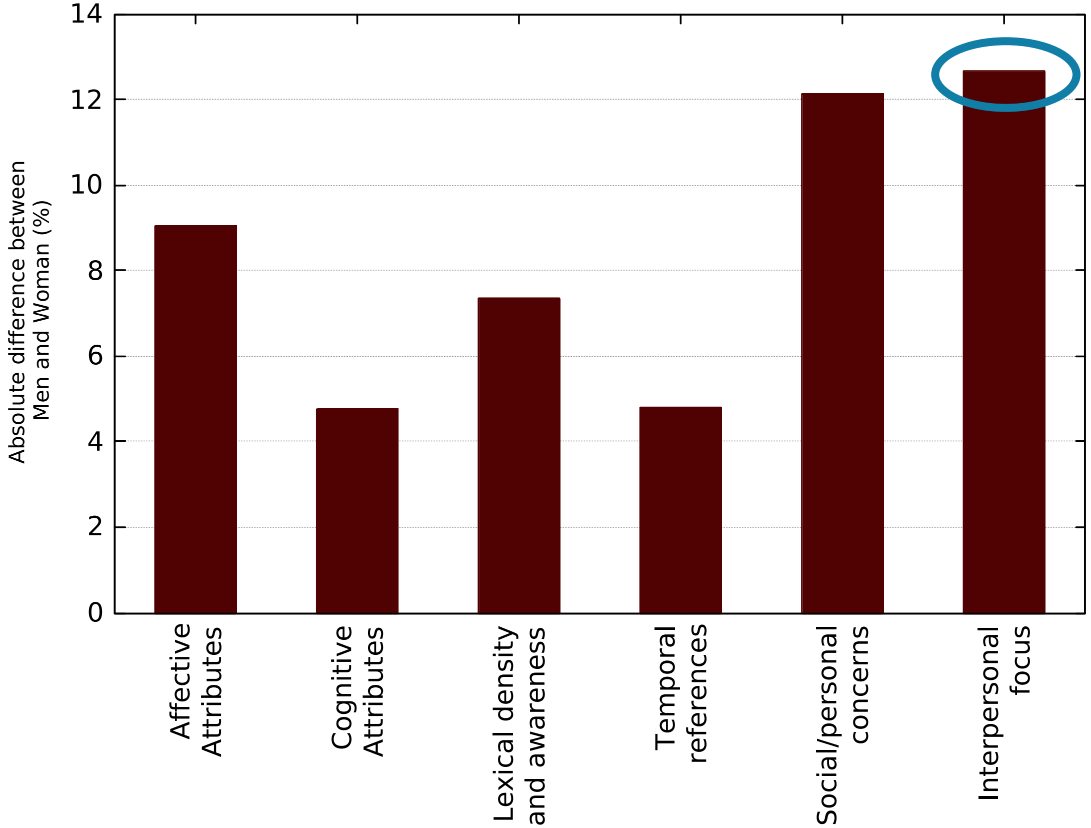
✓Top 1%



Linguistic Patterns

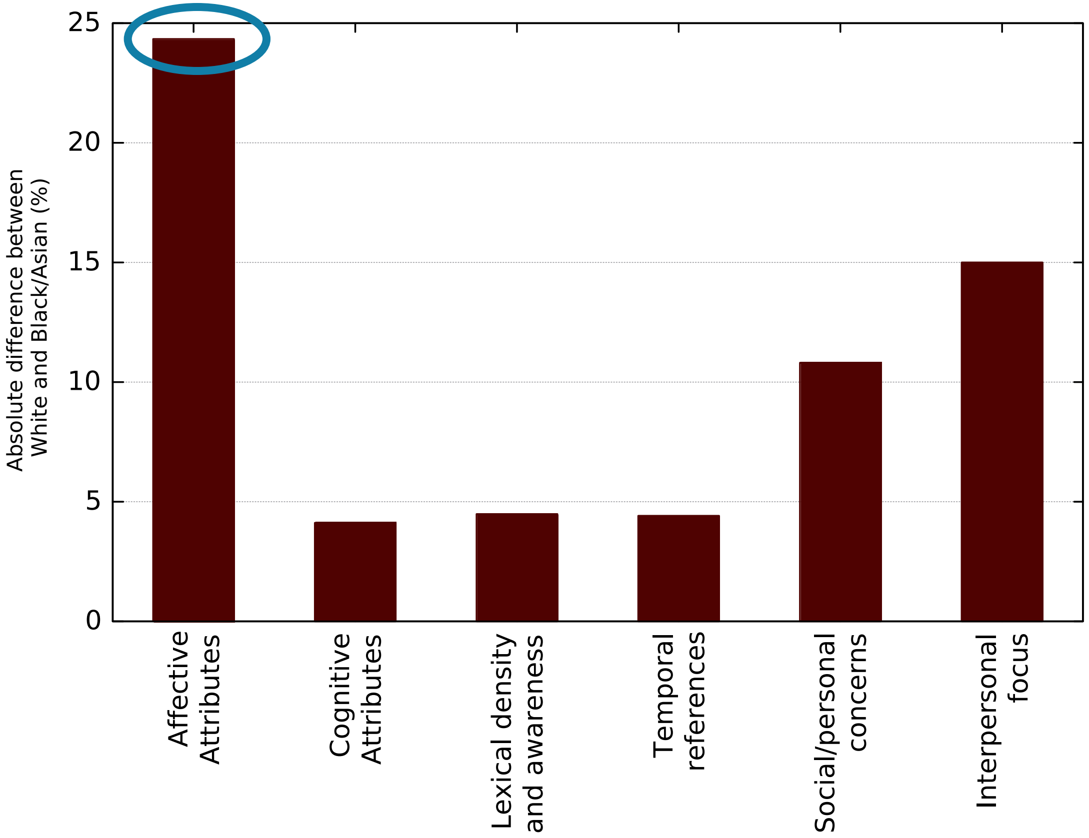
- ✓ Linguistic Differences
  - Mean Absolute Differences
  - Wilcoxon Rank Sum Test
  - Attributes
    - Affective
    - Cognitive
    - Lexical Density and Awareness
    - Temporal References
    - Interpersonal Focus
- ✓ Differences in Topic Interests

Linguistic Differences



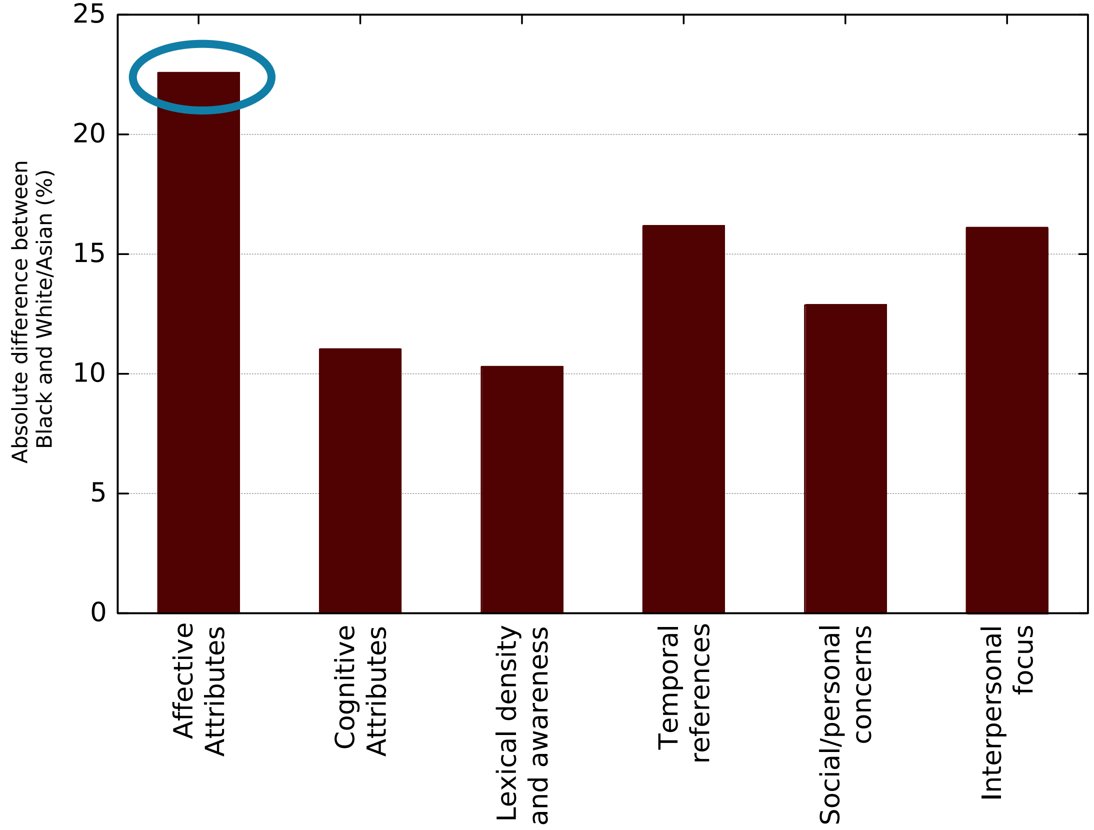
Mean Absolute Differences Between Male and Female Users

Linguistic Differences



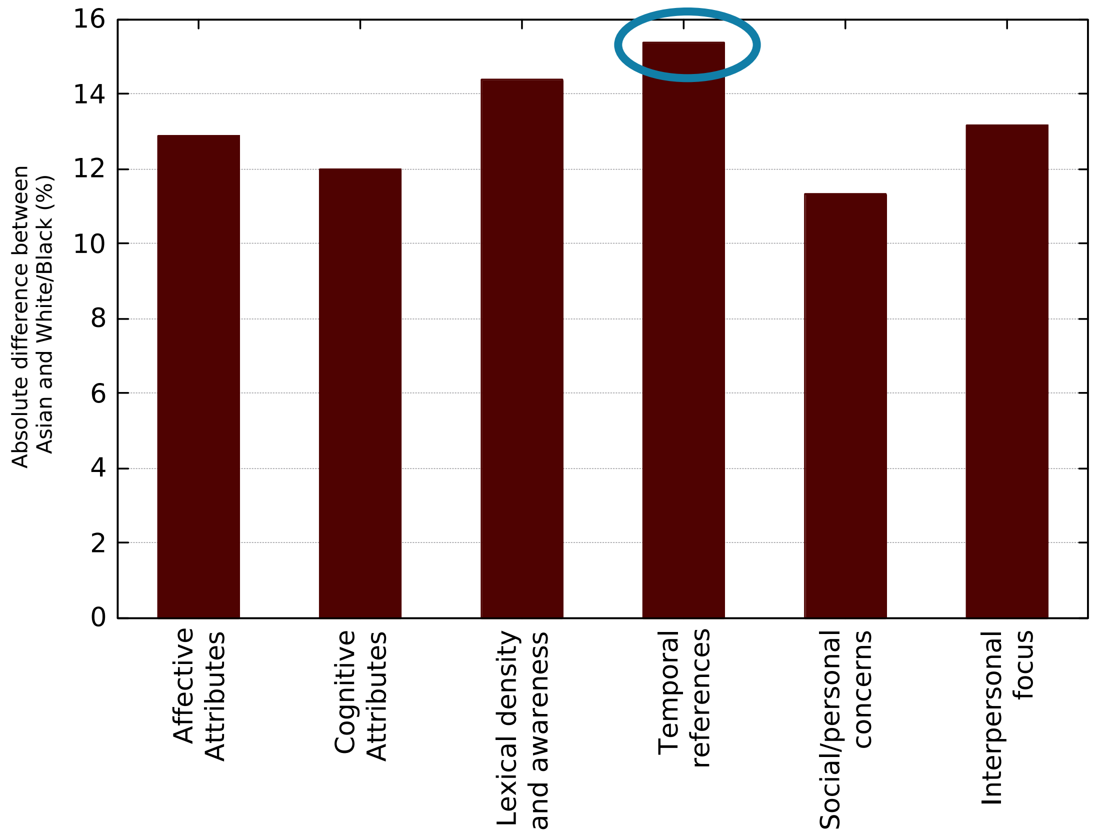
Mean Absolute Differences Between White and Black/Asian Users

Linguistic Differences



Mean Absolute Differences Between Black and White/Asian Users

Linguistic Differences



Mean Absolute Differences Between Asian and White/Black Users

	$\mu(male)$	$\mu(female)$	z
Affective attributes			
anger	0.0055	0.0056	4.733
anxiety	0.0016	0.0019	-74.534
sadness	0.0029	0.0034	-74.394
swear	0.0023	0.0026	-7.411
Cognitive attributes			
Cognition			
causation	0.0101	0.0104	-18.627
certainty	0.0101	0.0111	-60.593
tentativeness	0.0136	0.0141	-14.641
Perception			
see	0.00957	0.0099	-24.538
hear	0.0055	0.0056	-0.033*
feel	0.0035	0.0041	-70.766
percepts	0.0207	0.0218	-41.373
insight	0.0115	0.0125	-46.806
relative	0.1014	0.0999	18.026
Lexical Density and Awareness			
verbs	0.1103	0.1170	-45.808
auxiliary verbs	0.0539	0.0583	-46.441
articles	0.0370	0.0340	77.303
prepositions	0.0843	0.0817	32.596
conjunctions	0.0279	0.0314	-72.098
adverbs	0.0317	0.0355	-66.915
Temporal references			
present tense	0.0802	0.0871	-62.110
future tense	0.0103	0.0106	-15.118
Social/Personal Concerns			
family	0.0026	0.0034	-93.252
friends	0.0028	0.0033	-66.168
social	0.0938	0.1021	-77.896
health	0.0037	0.0044	-76.446
religion	0.0024	0.0025	-26.485
bio	0.0157	0.0203	-102.681
body	0.0045	0.0056	-58.386
achievement	0.0116	0.0105	65.265
home	0.0022	0.0026	-74.049
sexual	0.0011	0.0012	-18.691
death	0.0014	0.0013	29.463
Interpersonal focus			
1st p. singular	0.0245	0.0340	-97.329
1st p. plural	0.0046	0.0045	4.309
2nd p.	0.0160	0.0198	-88.482
3rd p.	0.0030	0.0031	-3.371***

- females tend to use anxiety and sadness terms and phrases.
- males express with anger in their tweets
- females are more likely to write phrases that express cognition and perception.
- females express more confidence and feelings in their writing.
- females make more use of verbs, auxiliary verbs, conjunctions, and adverbs , while males use more articles and prepositions .
- The temporal references attributes are more present in the females.
- Social/Personal Concerns such as family , bio, friends, social, health, are used more by females
- Concern of achievement is expressed more in male
- Females also have a higher tendency to write in the first person singular and in second person
- Males use the first person plural

	$\mu(White)$	$\mu(Black)$	$\mu(Asian)$	$z_{W/B-A}$	$z_{B/W-A}$	$z_{A/W-B}$
Affective attributes						
anger	0.0051	0.0081	0.0056	-67.261	94.610	-5.236
anxiety	0.0017	0.0019	0.0016	-0.696	33.789	-30.517
sadness	0.0031	0.0034	0.0032	-20.814	28.205	-0.625
swear	0.0021	0.0064	0.0027	-90.375	107.344	11.329
Cognitive attributes						
Cognition						
causation	0.0104	0.0105	0.0096	29.931	19.465	-54.832
certainty	0.0105	0.0116	0.0101	-19.404	62.239	-33.955
tentativeness	0.0138	0.0152	0.0130	-8.958	55.174	-40.226
Perception						
see	0.0098	0.0098	0.0095	18.756	6.970	-29.506
hear	0.0055	0.0062	0.0054	-26.349	62.137	-25.331
feel	0.0037	0.0044	0.0039	-44.180	63.963	-5.128
percepts	0.0212	0.0223	0.0210	-14.067	43.711	-23.308
insight	0.0122	0.0128	0.0112	11.133	40.420	-51.201
relative	0.1020	0.1012	0.0936	50.614	15.841	-76.870
Lexical Density and Awareness						
verbs	0.1125	0.1222	0.1082	-16.435	64.214	-39.436
auxiliary verbs	0.0554	0.0612	0.0529	-12.202	58.285	-39.130
articles	0.0366	0.0339	0.0314	96.532	-26.056	-94.363
prepositions	0.0851	0.0817	0.0743	77.024	1.032	-95.556
conjunctions	0.0291	0.0319	0.0286	-11.852	43.571	-25.898
adverbs	0.0329	0.0363	0.0325	-17.239	48.159	-23.542
Temporal references						
present tense	0.0825	0.0912	0.0798	-21.972	69.126	-37.196
future tense	0.0103	0.0119	0.0099	-28.333	79.181	-38.719
Social/Personal Concerns						
family	0.0029	0.0040	0.0032	-74.318	86.721	10.755
friend	0.0031	0.0033	0.0033	-26.248	25.332	8.717
social	0.0956	0.1101	0.0971	-60.389	90.830	-10.166
health	0.0040	0.0044	0.0039	-9.579	45.973	-30.920
religion	0.0024	0.0031	0.0024	-53.672	85.163	-13.154
bio	0.0176	0.0204	0.0179	-32.215	53.914	-10.492
body	0.0048	0.0067	0.0052	-62.906	86.903	-3.428
achievement	0.0114	0.0109	0.0097	69.227	-1.632	-83.506
home	0.0025	0.0024	0.0022	50.362	-4.554	-57.624
sexual	0.0011	0.0019	0.0012	-51.768	71.799	-3.084
death	0.0014	0.0015	0.0013	4.356	31.454	-34.554
Interpersonal focus						
1st p. singular	0.0268	0.0355	0.0296	-51.874	63.492	4.760
1st p. plural	0.0048	0.0042	0.0039	77.425	-28.107	-68.994
2nd p.	0.0169	0.0227	0.0177	-63.930	95.495	-10.148
3rd p.	0.0030	0.0039	0.0028	-36.070	87.717	-37.143

- Black users tend to express more anger and swear than White/Asian.
- Cognitive attributes, almost all features were more present in Black users texts
- Black users have more presence in features like verbs, auxiliary verbs, conjunctions, and adverbs
- Prepositions are more present among White users.
- Black people tend more to use terms related to family, social, religion , and body.
- There is a predominance in the use of first person plural for White
- first person singular, second person and third person are more prominent in the Black group.

Linguistic Differences

	Rank(female)	Rank(male)	Diff(F-M)
i do n't	1	1	0
i ca n't	2	2	0
you do n't	3	3	0
i 'm not	4	4	0
ca n't wait	5	8	3
i 'm so	6	19	13
i love you	7	15	8
do n't know	8	11	3
i want to	9	24	15
more for virgo	10	55	45
more for cancer	11	29	18
i wan na	12	28	16
! i 'm	13	25	12
you ca n't	14	16	2
more for libra	15	39	24
it 's a	16	10	6
and i 'm	17	33	16
more for pisces	18	ne	-
i need to	19	34	15
do n't have	20	27	7

- 6,000 users

- Phrases expressing negation are in the top positions for both males and females. It is also clear to see that
- Females are more into signs than males since phrases with this kind of content present higher differences in the gender ranking.
- It is common the usage of slangs like ”do n’t”, ”ca n’t” and ”wan na” for both genders.

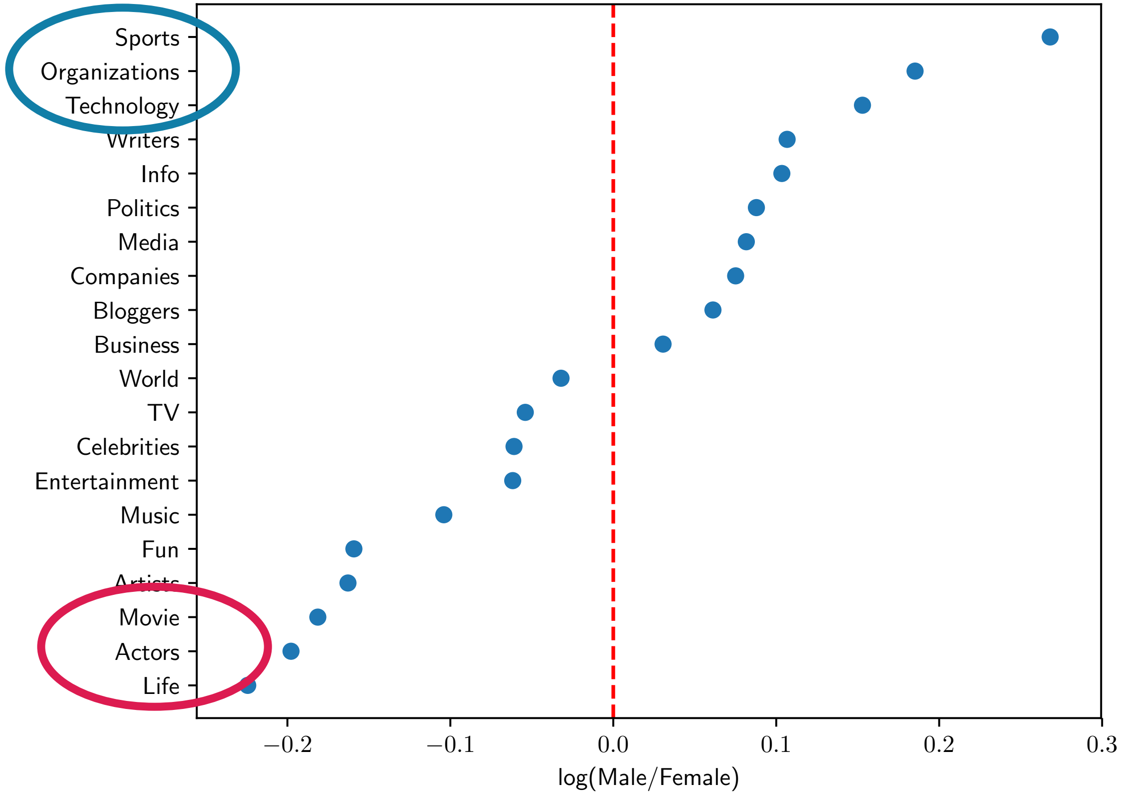
Linguistic Differences

	Rank(White)	Rank(Black)	Rank(Asian)	Diff(W-B)	Diff(W-A)	Diff(B-A)
i do n't	1	1	1	0	0	0
i ca n't	2	2	2	0	0	0
ca n't wait	3	18	7	15	4	11
you do n't	4	4	3	0	1	1
i 'm not	5	8	6	3	1	2
i love you	6	33	4	27	2	29
i 'm so	7	16	6	9	1	10
do n't know	8	19	11	11	3	8
it 's a	9	26	16	17	7	10
one of the	10	48	20	38	10	28
i want to	11	47	10	36	1	37
! i 'm	12	46	29	34	17	17
if you 're	13	28	19	15	6	9
thank you for	14	126	28	112	14	98
it 's not	15	34	32	19	17	2
and i 'm	16	58	21	42	5	37
you ca n't	17	17	17	0	0	0
i 'm at	18	53	26	35	8	27
n't wait to	19	100	51	81	32	49
i liked a	20	7	ne	13	-	-

- 6,000 users

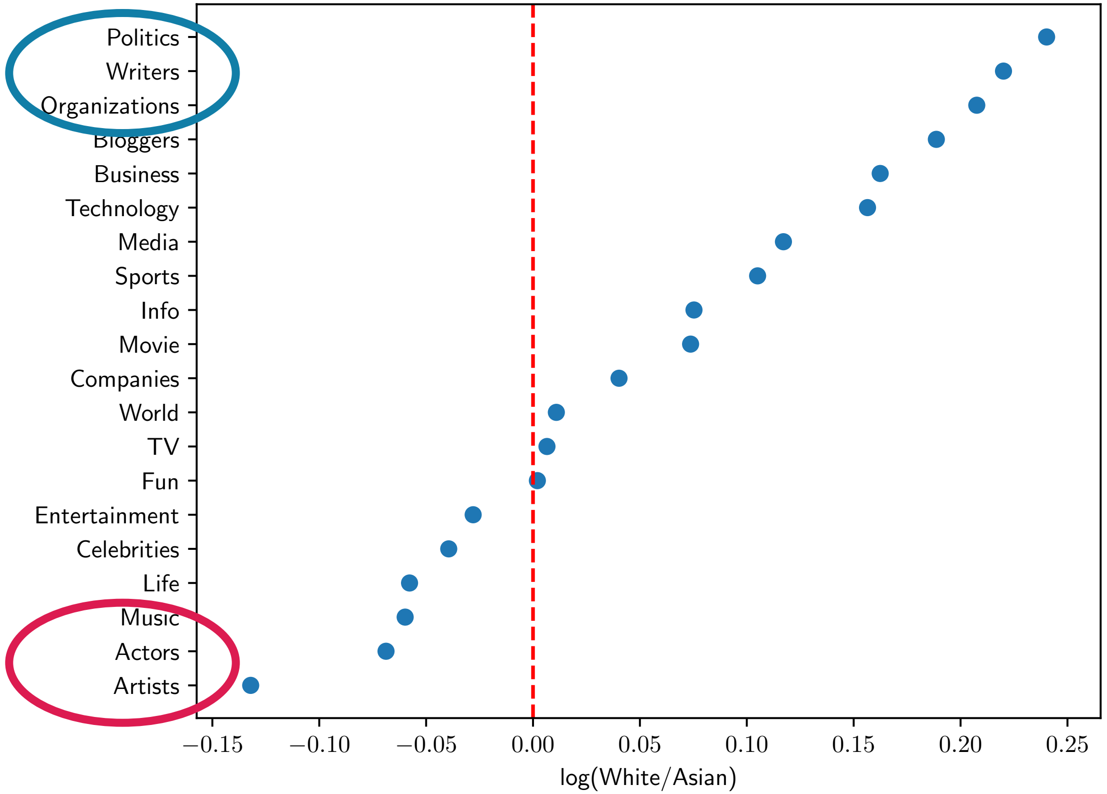
- Phrases containing expressions like "i don't", "i can't" and "i'm not" appear in the top positions for all the racial groups.
- Difference in ranking of the expression "i love you"
  - White and Asian users seem to be more likely to tweet contents with this expression than Black users.
- The expression "i want to" appears more often in the writing of White and Asian users than in the Blacks.

Differences in Topics



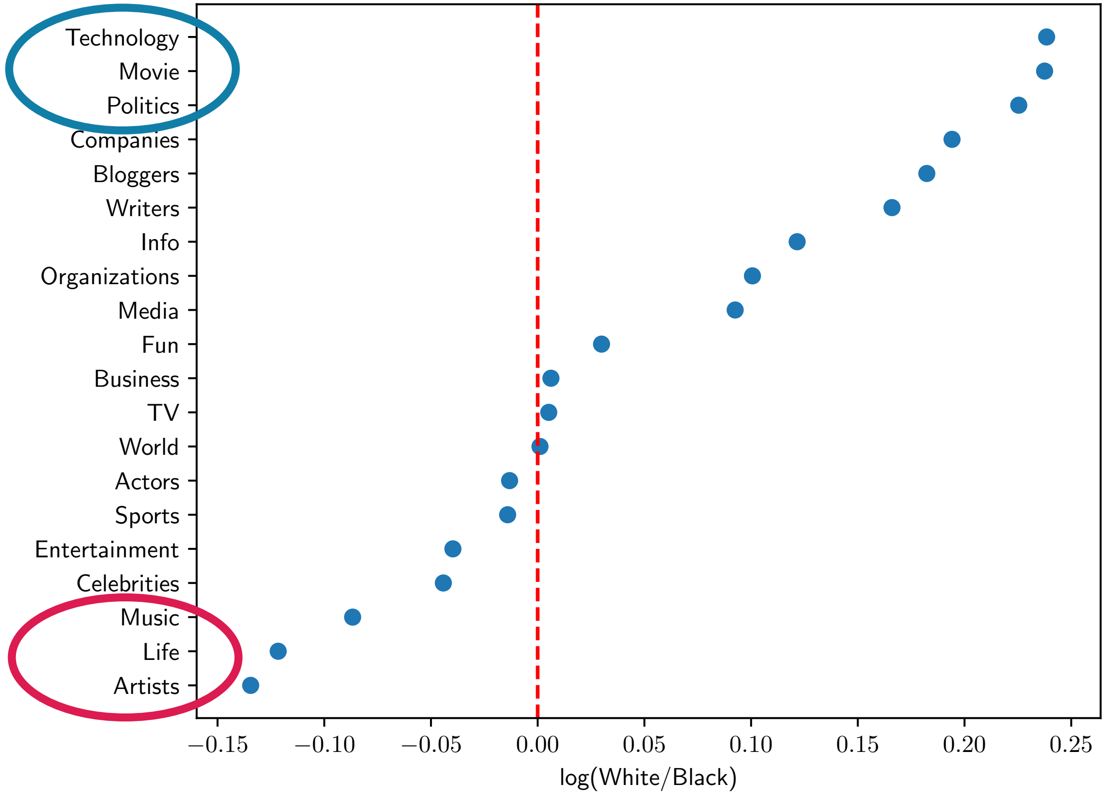
Male vs Female

Differences in Topics



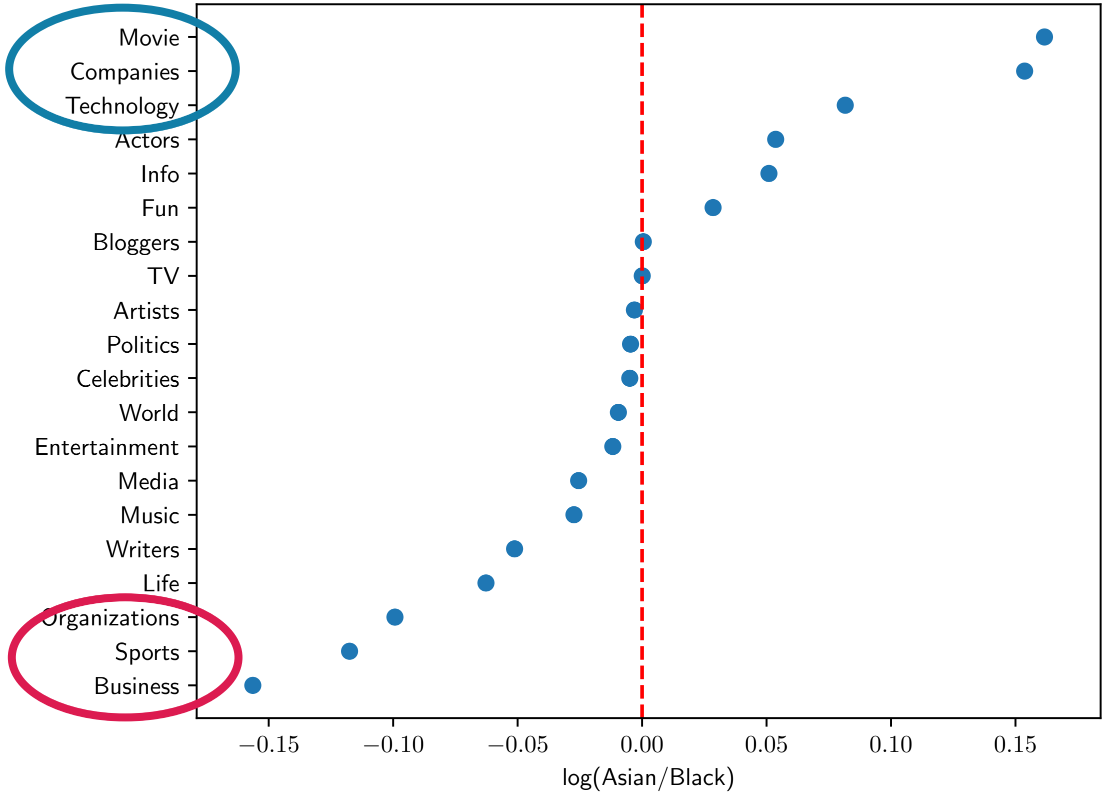
White vs Asian

Differences in Topics



White vs Black

Differences in Topics



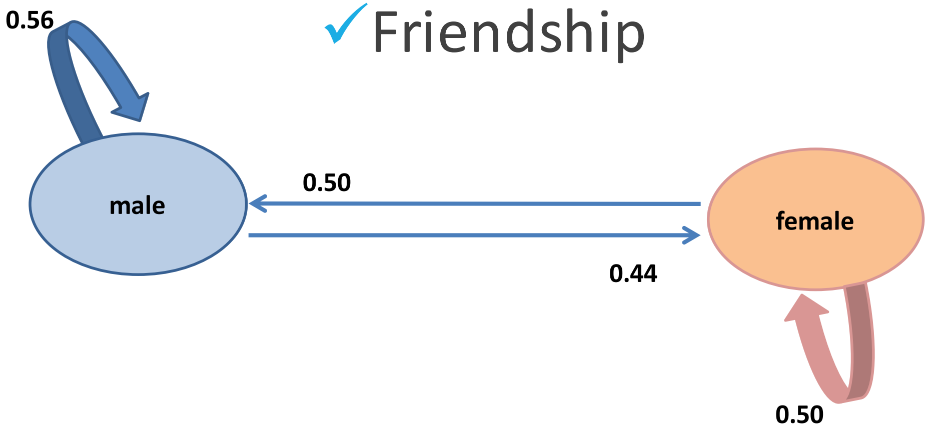
Asian vs Black

Demographic Group Interconnections

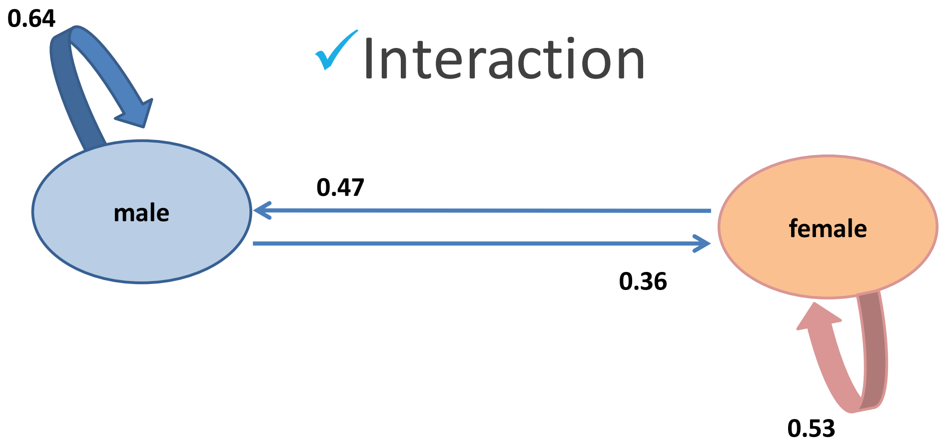
- ✓ Analyze the Interconnections and Interactions of Demographic Groups
- ✓ Gender and its Interconnections
  - Probabilistic Graph
- ✓ Race and its Interconnections
  - Probabilistic Graph
- ✓ Demography of Interconnections
  - Relative Increase or Decrease from What We Would Expect
- ✓ Dataset
  - 448,697 users

Gender and its Interconnections

✓Friendship



✓Interaction

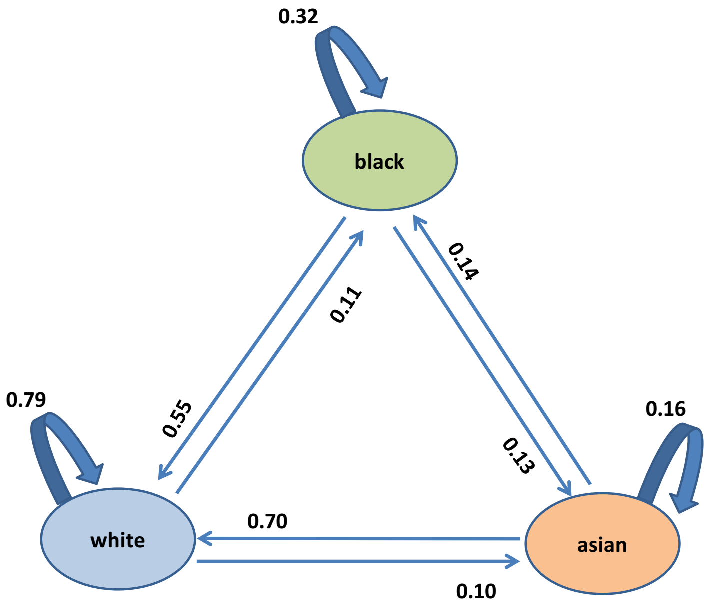


✓What we would expect

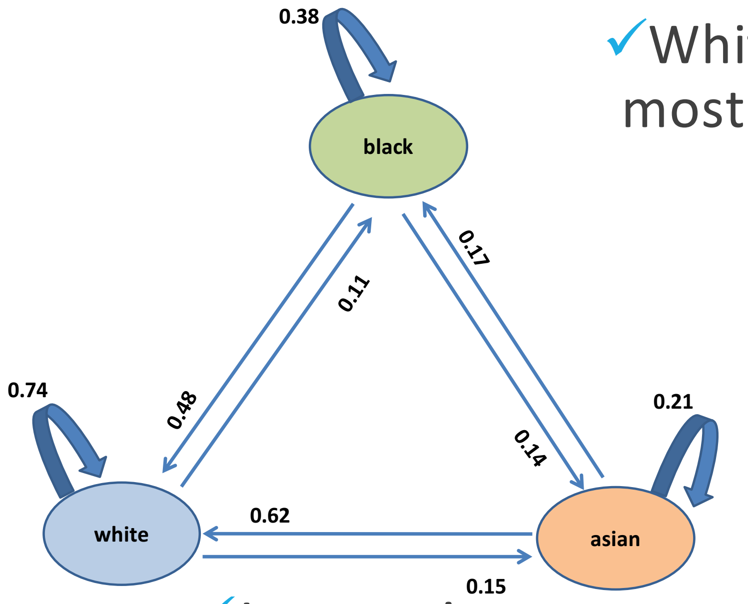
Race (%)	Gender (%)		Total (%)
	Male	Female	
Asian	7.07 (−3.85)	10.05 (−11.28)	17.12 (−10.90)
Black	8.17 (8.53)	6.74 (7.68)	14.91 (11.69)
White	32.88 (8.49)	35.09 (−7.69)	67.97 (1.20)
Total	48.12 (10.91)	51.88 (−10.91)	100.00

✓Male and female users take responsibility

Race and its Interconnections



✓Friendship



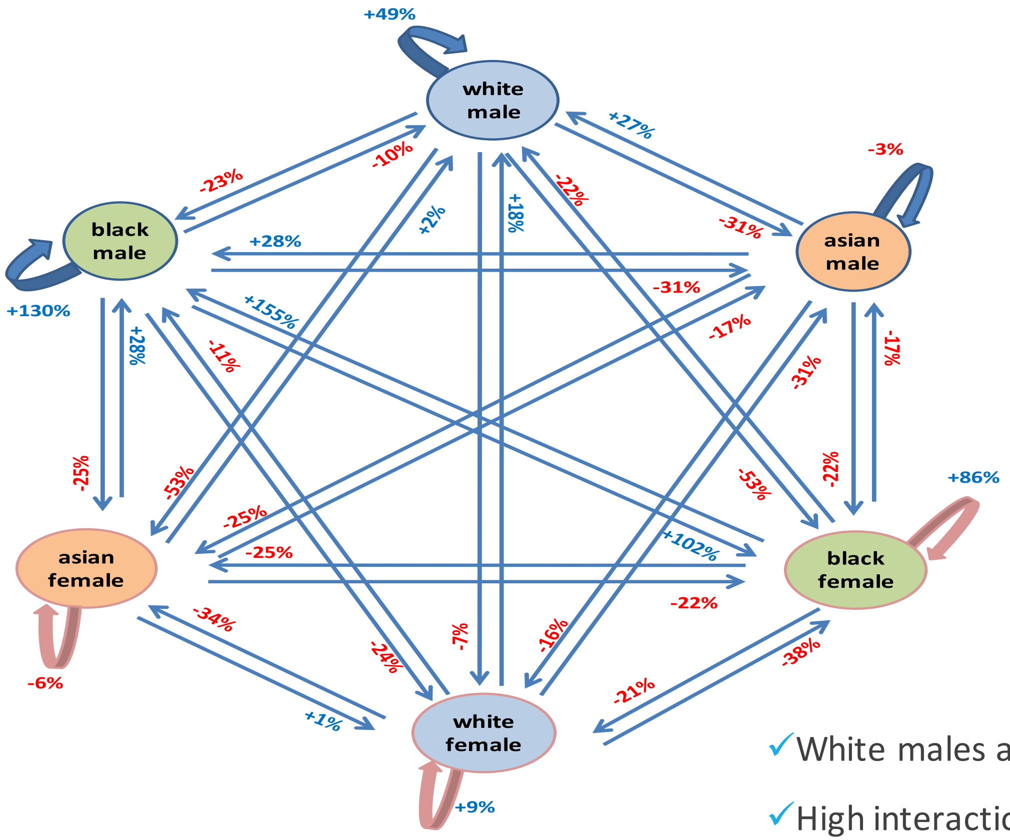
✓Interaction

✓White users tend to be the most followed by users

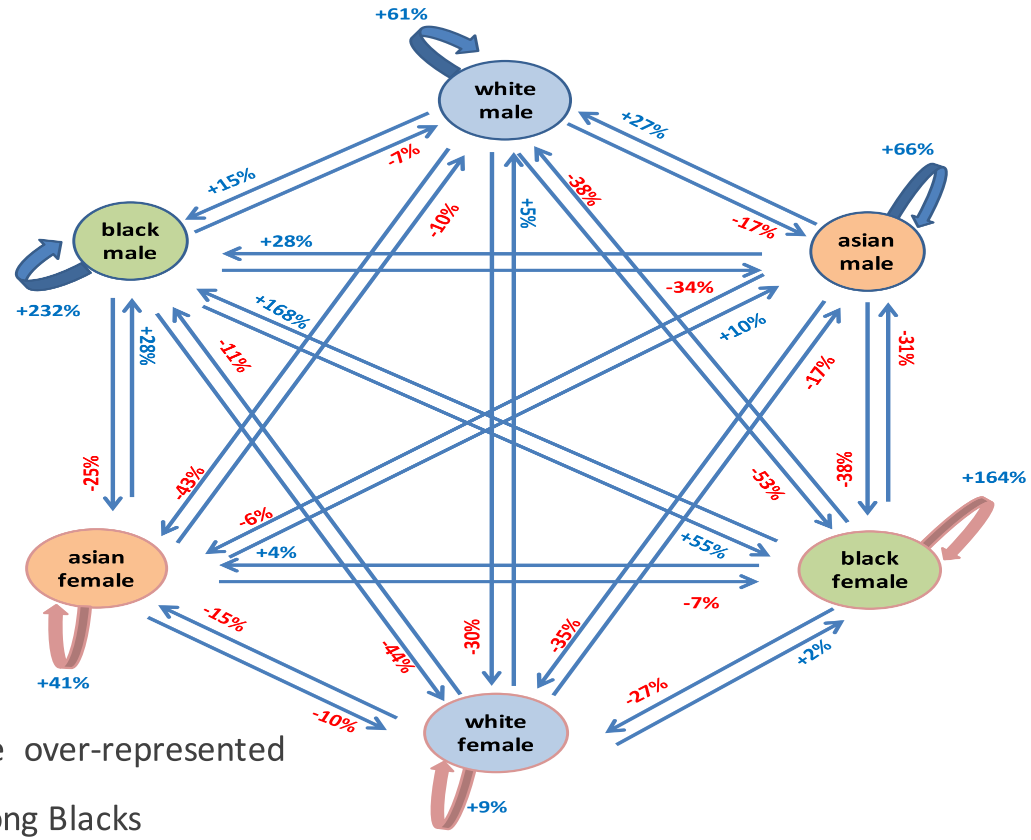
Race (%)	Gender (%)		Total (%)
	Male	Female	
Asian	7.07 (−3.85)	10.05 (−11.28)	17.12 (−10.90)
Black	8.17 (8.53)	6.74 (7.68)	14.91 (11.69)
White	32.88 (8.49)	35.09 (−7.69)	67.97 (1.20)
Total	48.12 (10.91)	51.88 (−10.91)	100.00

✓What we would expect

Demography of Interconnections



✓Friendship



✓Interaction

- ✓White males are more over-represented
- ✓High interactions among Blacks
- ✓Asians are under-represented

Leverage Demographic Aspects to Design Transparent Systems

- ✓ Demographics aspects are valuable to provide transparency
- ✓ White House Suggests More Transparency in Systems
- ✓ Twitter Trending Topics
  - Who Makes Trends? Web-based System
- ✓ Google Suggestion

Who Makes Trends?

- ✓Real-time Web-based System
- ✓Trend Promoters
- ✓Trend Adopters
- ✓Gender, Race, and Age
- ✓US-based Twitter Users
- ✓1% Random Sample
- ✓http://twitter-app.mpi-sws.org/who-makes-trends/

Who Makes Trends? *Discover the Demographics of Twitter Trend Promoters*

Search Trends by Text

Hashtag such as #obamafarewell, #oscars

Search Trends by Date

Select the date

Sample Trends with High Demograhic Bias

High Gender Bias: #wwefastlane #footballmovies #ufcphoenix #janethevirgin #thebachelor

High Racial Bias: #thankyoutrump #obamacare #neweditionbet #dow20k #scotus

High Age Bias: #healthiercelebs #dangerouswomantour #presidentialtvshows #wednesdaywisdom #nationalloveyourpetday

How it Works

Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations 11th International AAAI Conference on Web and Social Media (ICWSM). Montreal, Canada. May 2017.

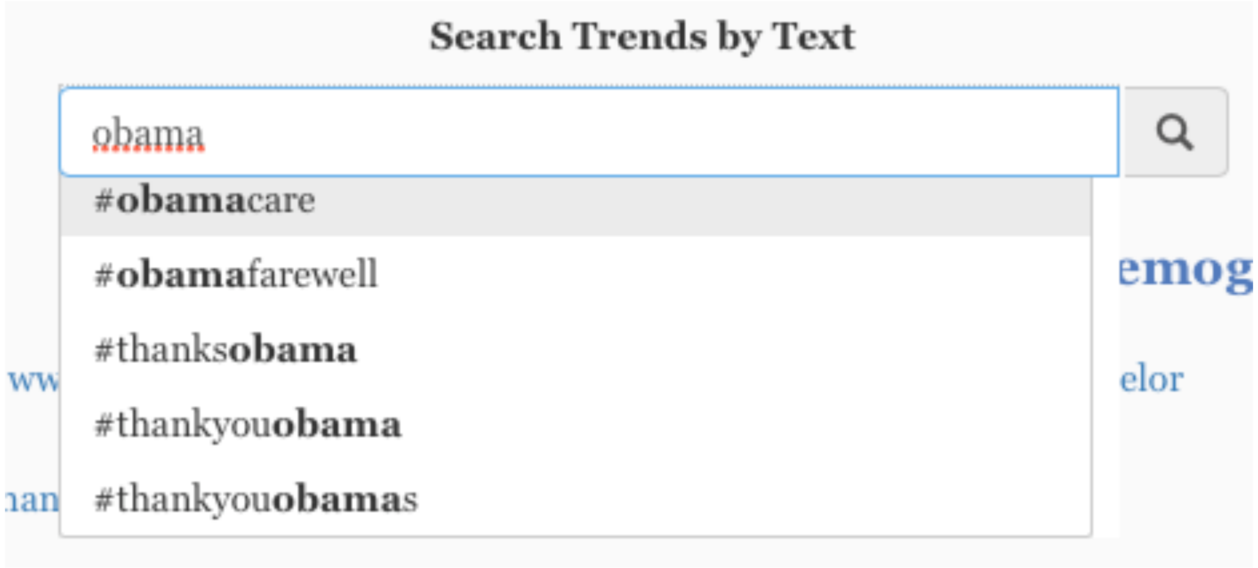
Who Are We?

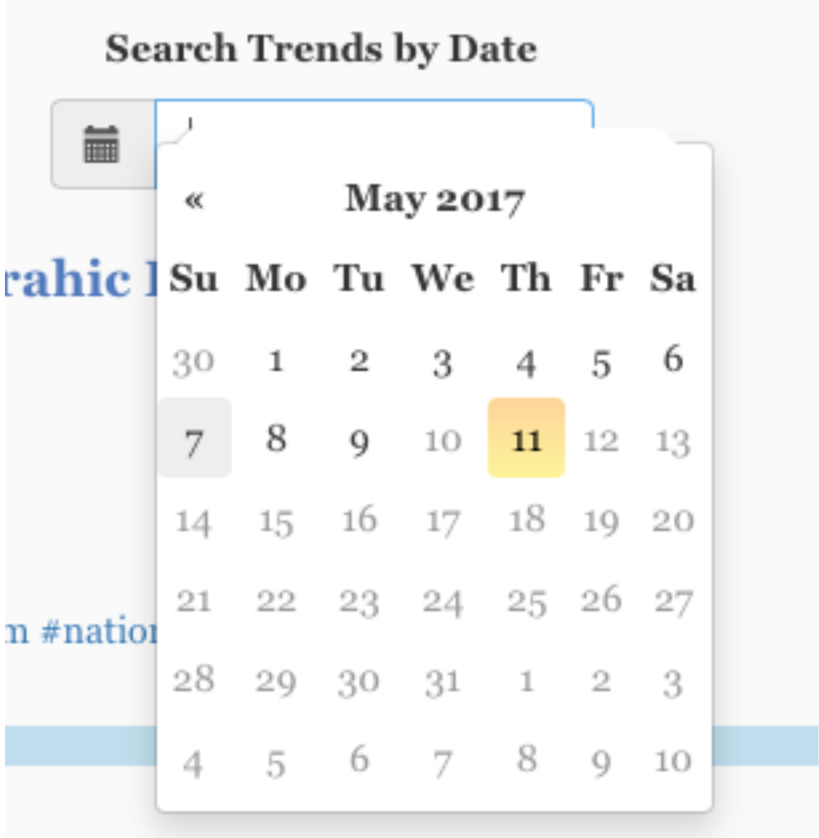
MPI-SWS, Germany
Krishna P. Gummadi

IIT Kharagpur, India
Abhijnan Chakraborty
Saptarshi Ghosh
Niloy Ganguly

UFMG, Brazil
Johnnatan Messias
Fabricio Benevenuto

Who Makes Trends?





Data collection

- ✓1% Random Sample US Tweets in English
  - ✓1% Worldwide < 1% US
- ✓Bounding Box
- ✓Trending Topics of Twitter (every 5-min)
- ✓EST Time Zone
- ✓Twitter Stream API
- ✓Since January 2017
- ✓Demographic Information From Face++

Trending Topic Analysis

Baseline	Gender (%)		Race (%)			Age Group (%)			
	Male	Female	White	Black	Asian	Adolescent	Young	Mid-aged	Old
U.S. Population	49.20	50.80	72.40	12.60	4.80	13.60	26.70	33.20	13.50
Twitter Population	45.97	54.03	73.05	12.25	14.70	26.37	62.58	10.80	0.25

Hashtag	Date	#Promoters With Demographic Inference	#Promoters Without Demographic Inference	#Adopters With Demographic Inference	#Adopters Without Demographic Inference	#Total With Demographic Inference	#Total Without Demographic Inference
#mayday2017	02-05-2017	609	532	162	165	736	660
#metgala	01-05-2017	1563	616	491	257	1988	830
#wwepayback	30-04-2017	862	660	88	70	912	695
#climatemarch	29-04-2017	637	582	155	142	753	694
#fyrefestival	28-04-2017	32	15	1363	925	1384	936
#nfldraft	27-04-2017	4846	3635	2318	1873	6183	4813
#wednesdaywisdom	26-04-2017	317	341	77	57	383	392
#dwts	25-04-2017	360	188	89	51	435	232
#mondaymotivation	24-04-2017	673	717	141	131	803	840
#sundayfunday	23-04-2017	678	613	153	112	812	712
#earthday	22-04-2017	2636	2719	3105	2618	5531	5123
#ripprince	21-04-2017	453	300	171	104	603	393
#happy420	20-04-2017	805	567	132	121	918	661
#bostonmarathon	19-04-2017	789	598	282	225	1005	778
#unicornfrappuccino	18-04-2017	36	13	1712	996	1737	1005
#cleveland	17-04-2017	709	442	406	355	1049	711
#eastersunday	16-04-2017	64	77	1570	1233	1616	1300
#aprilthegiraffe	15-04-2017	810	421	76	56	872	467
#goodfriday	14-04-2017	1674	1422	862	640	2452	1996
#stanleycup	13-04-2017	369	303	600	493	842	709
#bucciovertimechallenge	12-04-2017	171	244	770	992	867	1137
#nationalpetday	11-04-2017	2637	1887	1455	891	4056	2744
#nationalsiblingsday	10-04-2017	3828	1837	2512	1202	6296	3023
#sundayfunday	09-04-2017	775	585	181	130	939	705
#nationalbeerday	08-04-2017	1188	1581	368	333	1529	1876
#syria	07-04-2017	1263	856	654	472	1771	1217
#themasters	06-04-2017	420	452	3015	2513	3225	2771
#13reasonswhy	05-04-2017	280	87	985	363	1225	439
#nationalchampionship	04-04-2017	3376	2561	238	174	3533	2682
#finalfour	03-04-2017	4146	3448	851	618	4739	3887
#openingday	02-04-2017	1732	1342	4129	3437	5461	4506

Demo

Disparate Demographics

Hashtag	Demographics of Promoters							
	Gender (%)		Race (%)			Age Group (%)		
	Male	Female	White	Black	Asian	Adolescent	Young	Mid-aged
#footballmovies	**65.82**	*34.18*	**83.55**	*5.06*	11.39	*10.13*	**70.88**	**18.99**
#ufcphoenix	**77.03**	*22.97*	73.65	10.81	15.54	*16.89*	**71.62**	11.49
#thebachelor	*15.61*	**84.39**	**84.69**	*4.94*	10.37	29.82	64.94	*5.24*
#thankyoutrump	49.55	50.45	**81.98**	8.11	9.91	21.62	*54.96*	**22.52**
#obamacare	**58.11**	*41.89*	**83.78**	*6.76*	*9.46*	*13.51*	*51.26*	**32.43**
#neweditionbet	*40.66*	**59.34**	*28.27*	**58**	13.73	**33.51**	59.93	6.49
#dangerouswomantour	*36.67*	**63.33**	71.67	8.33	**20**	**43.33**	*50*	6.67
#presidentialtvshows	**68.31**	*31.69*	**80.33**	10.93	*8.74*	*8.20*	**72.67**	**18.58**
#nationalloveyourpetday	*28.49*	**71.51**	**80.27**	8.04	11.69	26.94	63.29	9.77

- High Gender Bias
- High Race Bias
- High Age Bias

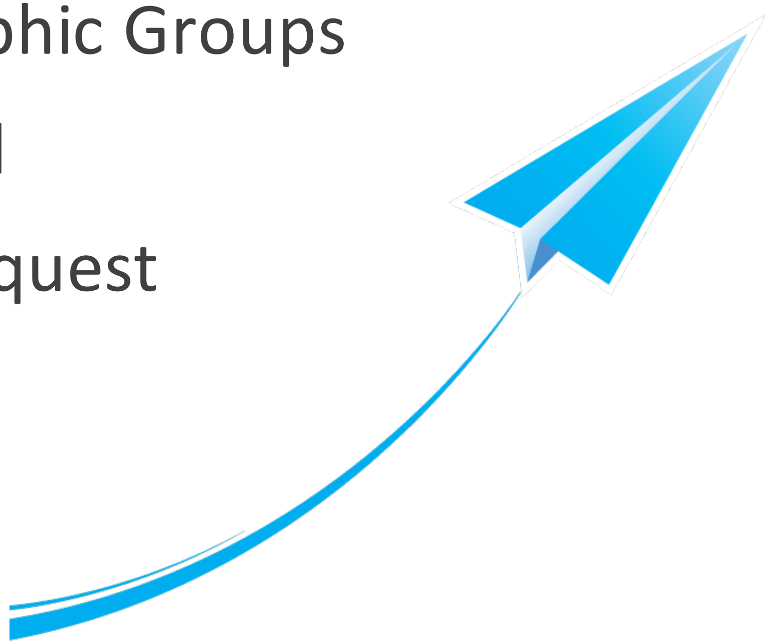
Conclusion

- ✓ Demographic Aspects are Valuable
- ✓ Gender and Race Inequality Exists in Twitter
- ✓ Glass Ceiling also Happens for Male Users
- ✓ Demographic Groups have its Own Preferences
  - Linguistic Style
  - For Topic Interests
- ✓ The Connections Among Demographic Groups Help to Explain Inequality
- ✓ Provide Transparent Systems is Important
  - Who Makes Trends?
- ✓ Potential Limitations

Conclusion

Future Work

- ✓ Explore Age as a Demographic Aspect
- ✓ Linguistic and Social features for Gender and Race Prediction
- ✓ Information Propagation Through Demographic Groups
- ✓ Compile the Results and Submit to a Journal
- ✓ Release our Demographic Dataset under Request



Publications

13 papers

- Conferences
  - 2 x IEEE/ACM ASONAM 2016
  - BraSNAM 2015
  - ACM CSCW 2017
  - 2 x ACM Hypertext 2017
  - AAAI ICWSM 2017
  - SBBD 2015
  - SOUPS 2016
  - Webmedia 2015
  - WI 2017

- Journals
  - IEEE Internet Computing 2017
  - Springer SNAM 2017

The full list of the papers are available at http://johnnatan.me

Publications: Demographics

White, Man, and Highly Followed: Gender and Race Inequalities in Twitter. **Johnnatan Messias**, Pantelis Vikatos, and Fabrício Benevenuto. In Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'17). Leipzig, Germany. August 2017 .

Demographics of News Sharing in the U.S. Twittersphere. Julio C. S. Reis, Haewoon Kwak, Jisun An, **Johnnatan Messias**, and Fabrício Benevenuto.In Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT'17). Prague, Czech Republic. July 2017.

Linguistic Diversities of Demographic Groups in Twitter. Pantelis Vikatos, **Johnnatan Messias**, Manoel Miranda, and Fabrício Benevenuto. In Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT'17). Prague, Czech Republic. July 2017.

Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. Abhijnan Chakraborty, **Johnnatan Messias**, Fabrício Benevenuto, Saptarshi Ghosh, Niloy Ganguly, and Krishna P. Gummadi. In Proceedings of the Int'l AAAI Conference on Web and Social (ICWSM'17). Montreal, Canada. May 2017.

Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media. Juhi Kulshrestha, Motahhare Eslami, **Johnnatan Messias**, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, and Karrie Karahalios. In Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'17). Portland, Oregon, USA, February 2017.

Publications: Other Topics

An Evaluation of Sentiment Analysis for Mobile Devices. **Johnnatan Messias**, João P. Diniz, Elias Soares, Miller Ferreira, Matheus Araújo, Lucas Bastos, Manoel Miranda, and Fabrício Benevenuto. In Springer Nature Social Network Analysis and Mining. Volume 7, Issue 1, 2017.

Longitudinal Privacy Management in Social Media: The Need for Better Controls. Mainack Mondal, **Johnnatan Messias**, Saptarshi Ghosh, Krishna P. Gummadi, and Aniket Kate. IEEE Internet Computing (Special Issue on Usable Privacy & Security). Volume 21, Issue 3, May-June, 2017.

From Migration Corridors to Clusters: The Value of Google+ Data for Migration Studies. **Johnnatan Messias**, Fabrício Benevenuto, Ingmar Weber, and Emilio Zagheni. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16). San Francisco, USA. August 2016.

Towards Sentiment Analysis for Mobile Devices. **Johnnatan Messias**, João P. Diniz, Elias Soares, Miller Ferreira, Matheus Araújo, Lucas Bastos, Manoel Miranda, and Fabrício Benevenuto. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16). San Francisco, USA. August 2016.

Publications: Other Topics

Forgetting in Social Media: Understanding and Controlling Longitudinal Exposure of Socially Shared Data. Mainack Mondal, **Johnnatan Messias**, Saptarshi Ghosh, Krishna P. Gummadi, and Aniket Kate. In Proceedings of the 12th Symposium on Usable Privacy and Security (SOUPS'16), Denver, CO, USA, June 2016.

Algoritmos de Aprendizado de Máquina para Predição de Resultados das Lutas de MMA. Leandro A. A. Silva, **Johnnatan Messias**, Mirella M. Moro, Pedro O. S Vaz de Melo, and Fabrício Benevenuto. In Proceedings of the 30th Brazilian Symposium on Databases (SBBD'15). Petrópolis, Brazil. October, 2015.

Brazil Around the World: Characterizing and Detecting Brazilian Emigrants Using Google+. **Johnnatan Messias**, Gabriel Magno, Fabrício Benevenuto, Adriano Veloso, and Virgílio Almeida. In Proceedings of 21st Brazilian Symposium on Multimedia and the Web (WebMedia'15). Manaus, Brazil. October, 2015.

Bazinga! Caracterizando e Detectando Sarcasmo e Ironia no Twitter. Pollyanna Gonçalves, Daniel Dalip, Julio C. S. Reis, **Johnnatan Messias**, Filipe Ribeiro, Philipe Melo, Leandro A. A. Silva, Marcos Gonçalves, and Fabrício Benevenuto. In Proceedings of the Proceedings of the Brazilian Workshop on Social Network Analysis and Mining (BraSNAM). Recife, Brazil. July, 2015.

THANK YOU!

HTTP://JOHNNATAN.ME