

Why Swear?

Analyzing and Inferring the Intentions of Vulgar Expressions

Eric Holgate[†], Isabel Cachola[†], Daniel Preotiuc-Pietro^{*}, Junyi Jessy Li[†]

[†]University of Texas at Austin

^{*}Bloomberg LP

Motivation

The image displays three tweets from different users, each with a green egg icon profile picture. The tweets are arranged in a 2x2 grid, with the bottom-right cell empty. Each tweet includes a settings gear icon, a 'Follow' button, a text body, interaction icons (reply, retweet, star, share, more), and a timestamp.

- Positive_User (@User1):** "EMNLP 2018 is the shit!"
- Negative_User (@User2):** "The weather today is utter shit."
- Abusive_User (@User3):** "Don't @ me you piece of shit."

Motivation

- 1 Vulgarity is employed purposefully
- 2 Vulgarity is used for various pragmatic goals
- 3 Vulgarity is prevalent in daily communication

Pragmatic Roles

Aggression (15.2%):

- *The word is used in order to harm the person or group the tweet is about*



Pragmatic Roles

Emotion (24.8%)

- The word is used to express emotions (positive or negative) related to the user's internal states, exclamations, feelings or attitude towards an object.
 - If the vulgar token is removed, the emotion is too.



Pragmatic Roles

Emphasis (29.8%)

- The word is used to emphasize a statement or feeling



Pragmatic Roles

Auxiliary Use (17.0%)

- The use of this word is simply a manner of speaking.
 - Descriptions of external emotions (i.e., those of someone else) fall into this category



Pragmatic Roles

Signaling Group Identity (4.7%)

- The word is used to mark membership in a social group
 - This includes reappropriative usage of slurs



Pragmatic Roles

Non-Vulgar (8.2%)

- The use of this word is not vulgar
 - i.e., named entities



Pragmatic Roles

1. Aggression
2. Emotion
3. Emphasis
4. Auxiliary Use
5. Signaling Group Identity
6. Non-Vulgar

Research Questions

Do demographic factors impact why users employ vulgarity?

1

Can we predict why users employ vulgarity?

2

Is modeling vulgar intent useful for NLP tasks?

3

Data

- We introduce a data set of 8,524 instances of vulgar words annotated with one of six roles
 - Across 7,800 tweets
 - Sourced from 4,132 users with demographic info (Preotiuc-Pietro et al., 2017)
 - Gender, age, education, income level, faith, political ideology
 - Vulgarity defined with a list from www.noswearing.com
 - Regular expressions include spelling variation and self-censorship e.g., *damnnnnn* or *a\$\$*

Data

- Annotated for vulgar intention
 - MTurk
 - IAA - Krippendorff's Alpha of 0.506
 - 7 annotations/instance
 - QC - Excluded annotators with <0.2 agreement with the majority of others
 - Majority vote aggregation, ties were split by one of the co-authors
- Available at: <https://github.com/ericholgate/VulgarFunctionsTwitter>
- Sentiment annotation for 6,800 tweets from the same corpus is also available (Cachola et al., 2018): <https://github.com/ericholgate/vulgartwitter>

Research Questions

Do demographic factors impact why users employ vulgarity?

1

Can we predict why users employ vulgarity?

2

Is modeling vulgar intent useful for NLP tasks?

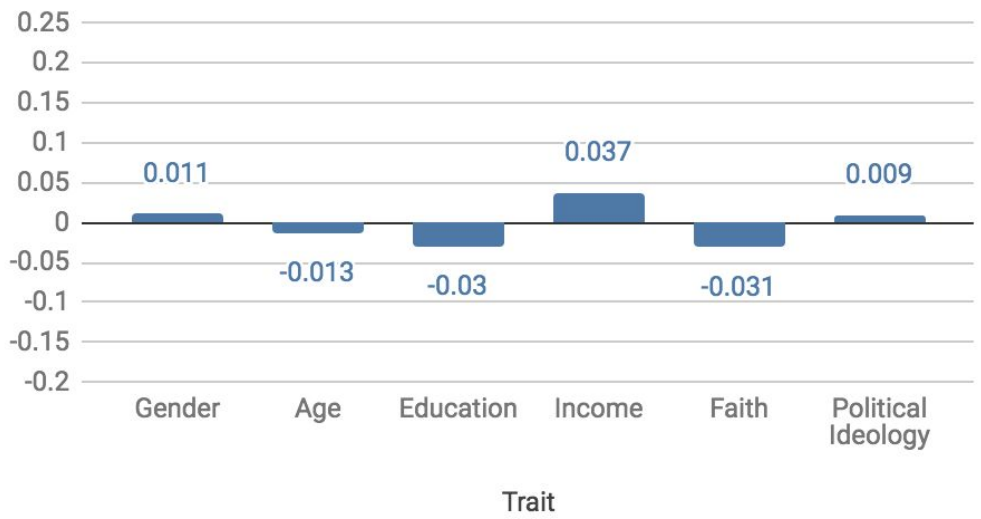
3

Demographic Analysis

Pearson correlation

- Dependent variable
 - fraction of vulgar function use
- Controlled for age & gender
- Bonferroni corrected
 - account for multiple comparisons

Pearson Correlation - Aggression

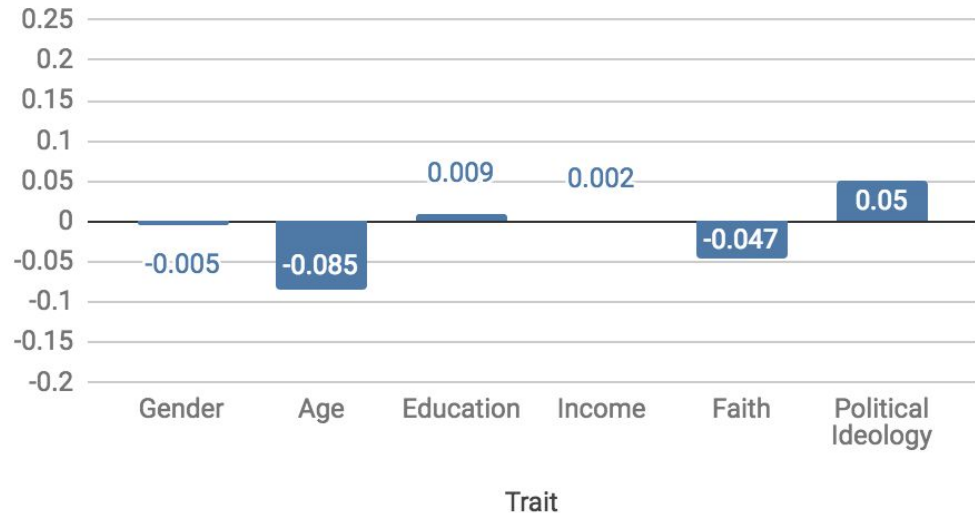


Demographic Analysis

Pearson correlation

- Dependent variable
 - fraction of vulgar function use
- Controlled for age & gender
- Bonferroni corrected
 - account for multiple comparisons

Pearson Correlation - Emotion

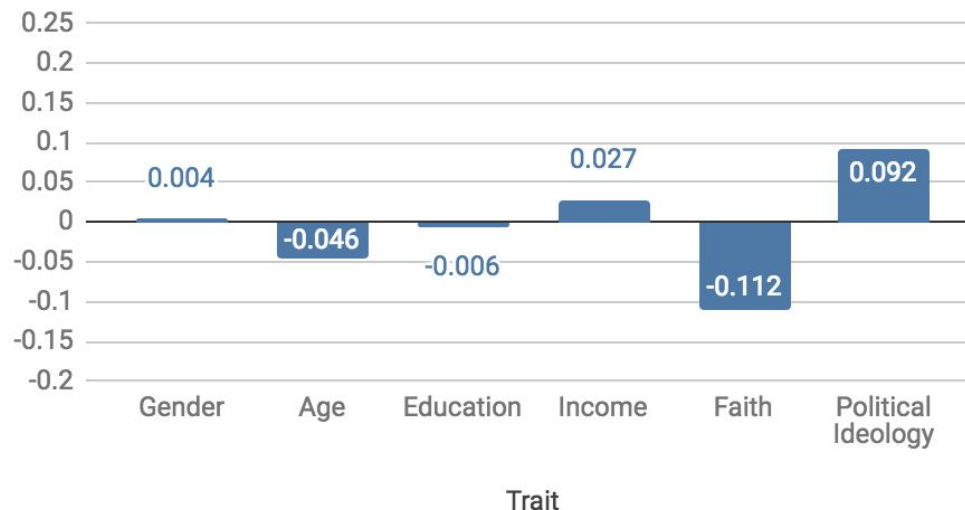


Demographic Analysis

Pearson correlation

- Dependent variable
 - fraction of vulgar function use
- Controlled for age & gender
- Bonferroni corrected
 - account for multiple comparisons

Pearson Correlation - Emphasis

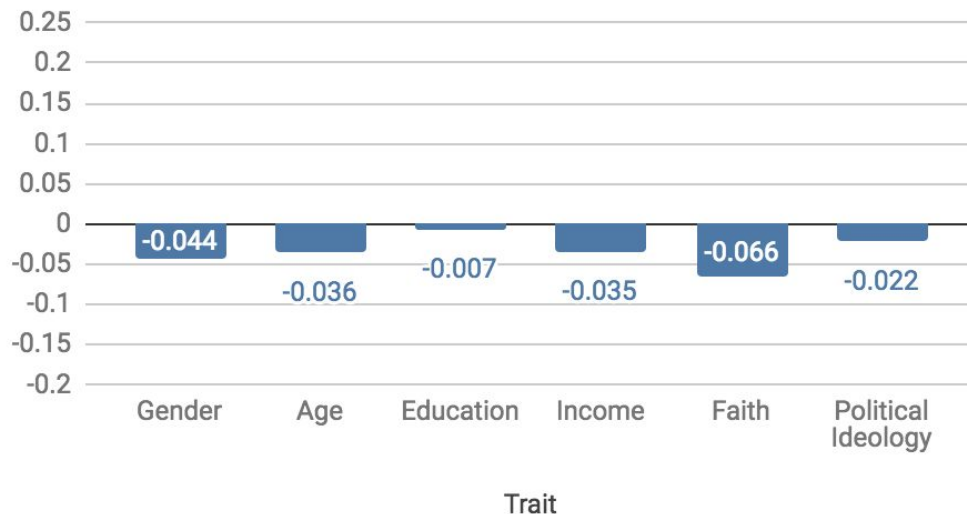


Demographic Analysis

Pearson correlation

- Dependent variable
 - fraction of vulgar function use
- Controlled for age & gender
- Bonferroni corrected
 - account for multiple comparisons

Pearson Correlation - Auxiliary

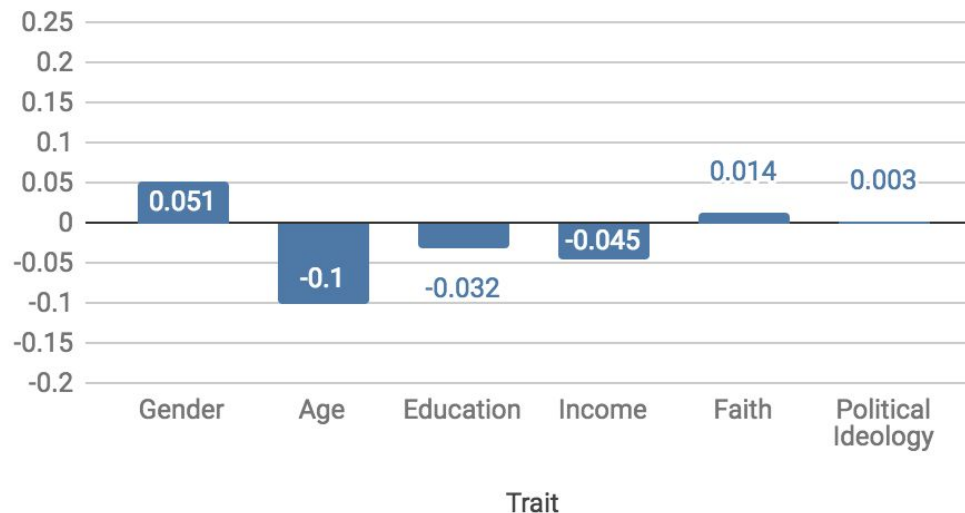


Demographic Analysis

Pearson correlation

- Dependent variable
 - fraction of vulgar function use
- Controlled for age & gender
- Bonferroni corrected
 - account for multiple comparisons

Pearson Correlation - Group Identity

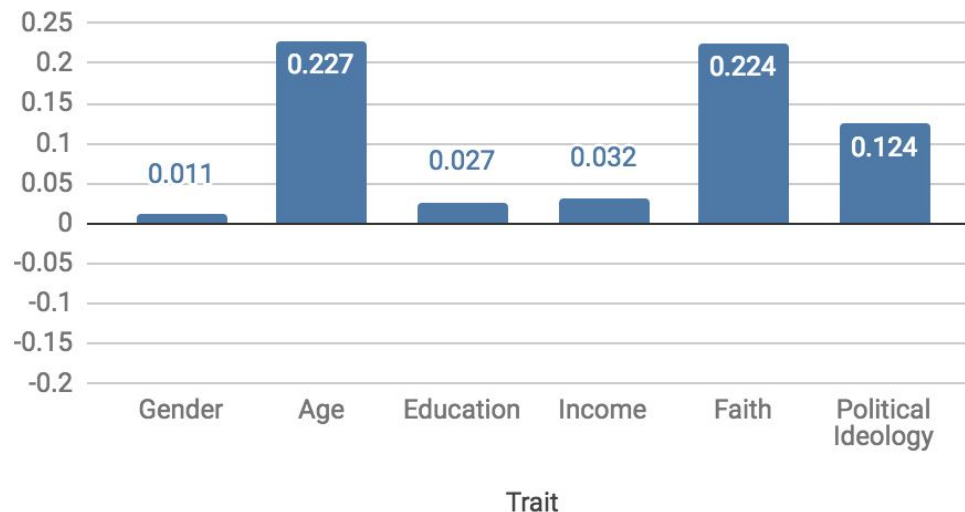


Demographic Analysis

Pearson correlation

- Dependent variable
 - fraction of vulgar function use
- Controlled for age & gender
- Bonferroni corrected
 - account for multiple comparisons

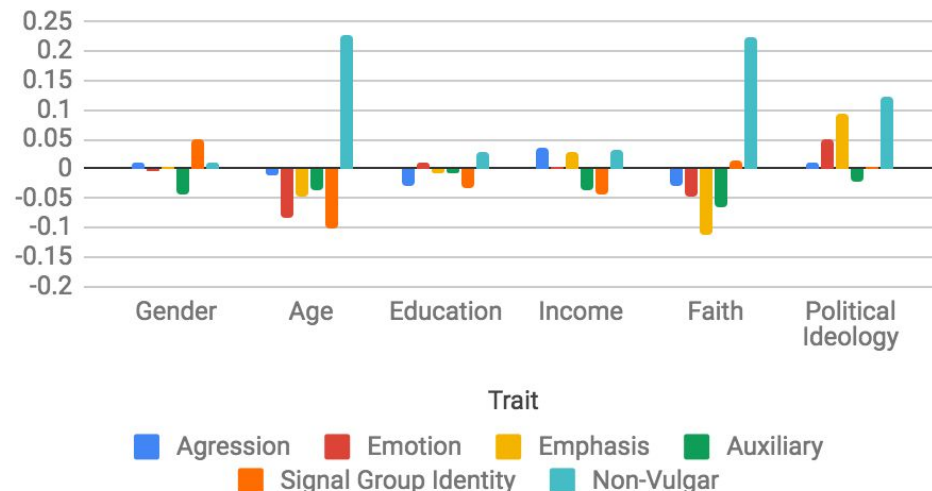
Pearson Correlation - Non-Vulgar



Demographic Analysis

- Younger users
 - [+] signal group identity
 - [+] emotion
 - [-] non-vulgar
- Politically liberal users:
 - [+] emphasis
- Religious users
 - [-] emphasis
- Gender, Education, & Income
 - No effects

Pearson Correlation



Research Questions

Do demographic factors impact why users employ vulgarity?

1

Yes!

Research Questions

Do demographic factors impact why users employ vulgarity?

1

Can we predict why users employ vulgarity?

2

Is modeling vulgar intent useful for NLP tasks?

3

Prediction: Features

- Vulgar token features
 - Intention distribution from training data
- Global tweet features
 - Tweet content (average GloVe embeddings)
- Local word context
 - Sentiment
 - Part-of-Speech (trigrams around the target token)
 - Brown Clusters (previous and next token)

Predicting Vulgar Function

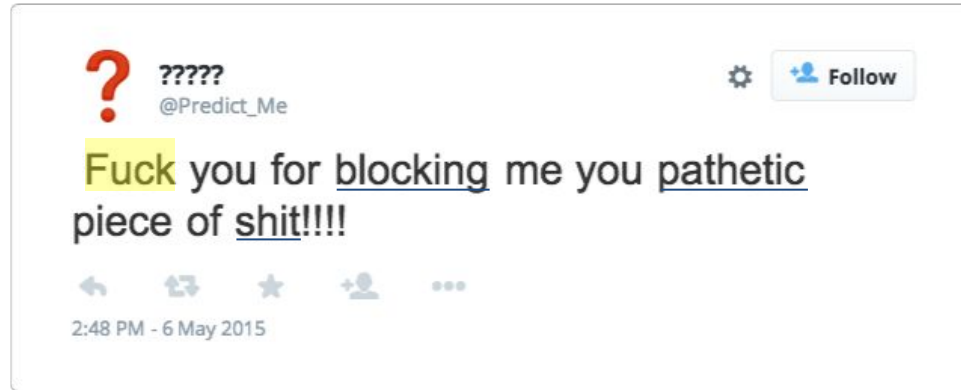


- Some words are used with predominately one function
- The most frequent, however, are distributed amongst all the functions

Predicting Vulgar Function

- Features
 - Vulgar token features
 - Intention distribution from training data
 - Global tweet features
 - Tweet content (average GloVe embeddings)
 - Local word context
 - Sentiment
 - Part-of-Speech (trigrams around the target token)
 - Brown Clusters (previous and next token)

Predicting Vulgar Function



- Vulgar instances centered around product reviews tend to be emotive or emphatic
- Conversational tweets are more frequently auxiliary

Features

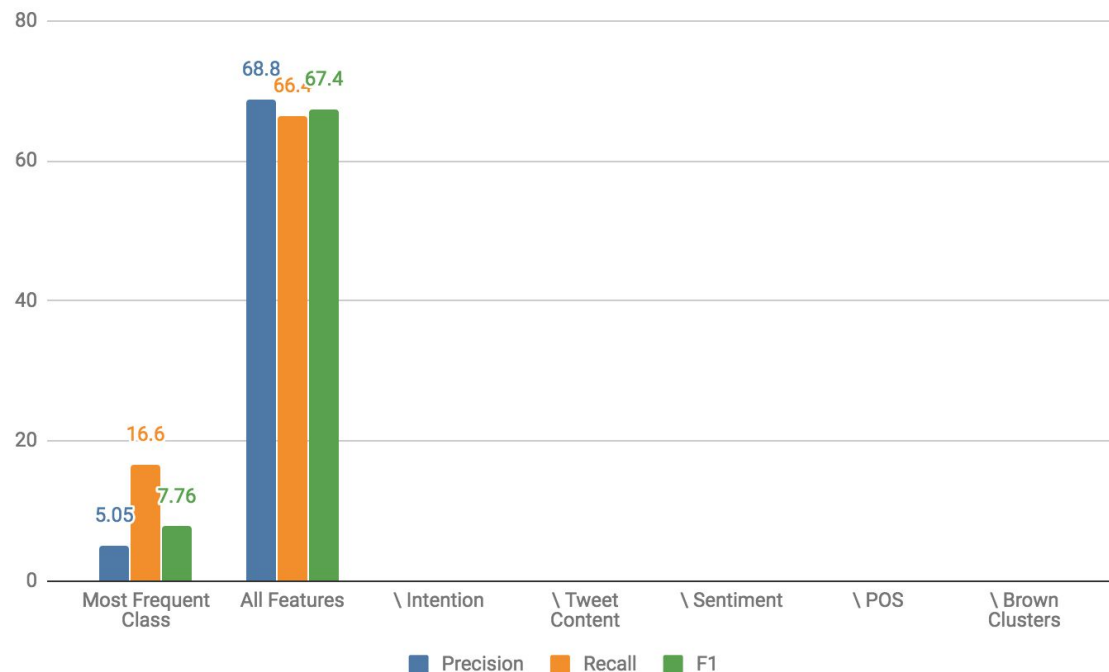
- Features
 - Vulgar token features
 - Intention distribution from training data
 - Global tweet features
 - Tweet content (average GloVe embeddings)
 - Local word context
 - Sentiment
 - Part-of-Speech (trigrams around the target token)
 - Brown Clusters (previous and next token)

Intuition

- Many vulgar words can even be substituted for one another, even when they don't have any concrete meaning:
 - Who the *hell/fuck*
 - I don't give a *damn/shit/fuck*
- Pinker (2007) calls these *strange synonyms*

Predicting Vulgar Functions

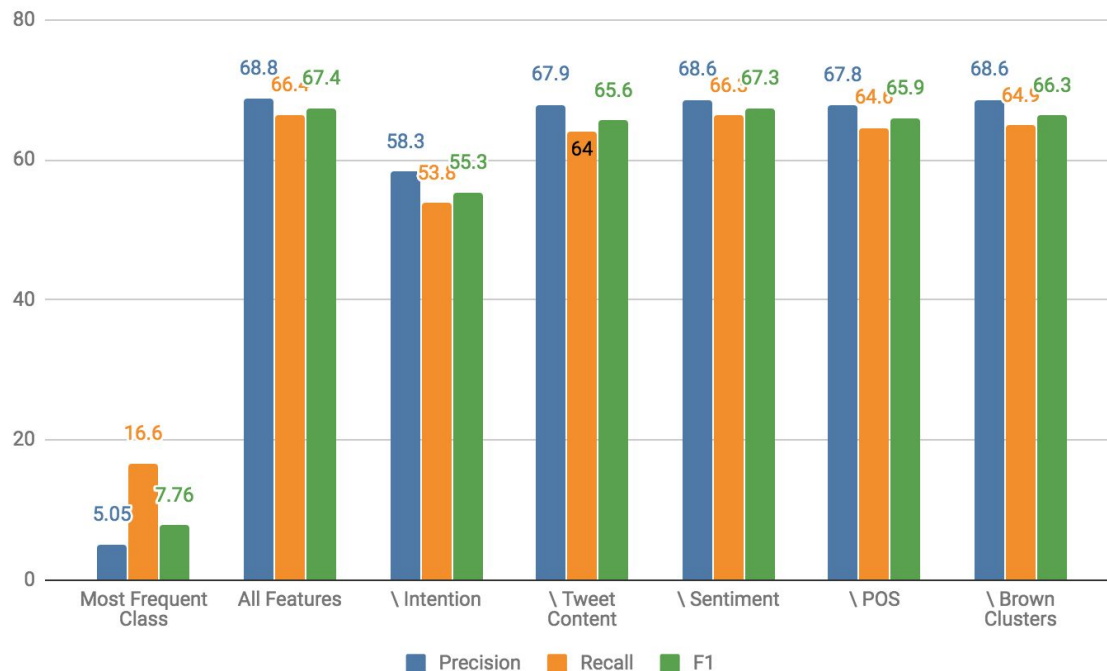
- Logistic regression classification
 - six one vs. all binary classifiers
- Data Split:
 - Train: 6,883
 - Test: 1,087
 - Val: 554
- A BiLSTM-based approach did not yield improvement



Predicting Vulgar Functions

Ablation Study

- Intention distribution contributes unique information
- Other features are complementary



Research Questions

Can we predict
why users
employ
vulgarity?

2

Yes!

Research Questions

Do demographic factors impact why users employ vulgarity?

1

Can we predict why users employ vulgarity?

2

Is modeling vulgar intent useful for NLP tasks?

3

Vulgar Intention and Hate Speech

- Hate Speech detection
 - Downstream task for vulgar function prediction
- Dataset introduced by Davidson et al. (2017)
 - 24,802 tweets labeled with one of the three classes:
 - hate speech
 - offensive
 - neither
 - all tweets contain vulgar words
 - <https://github.com/t-davidson/hate-speech-and-offensive-language>

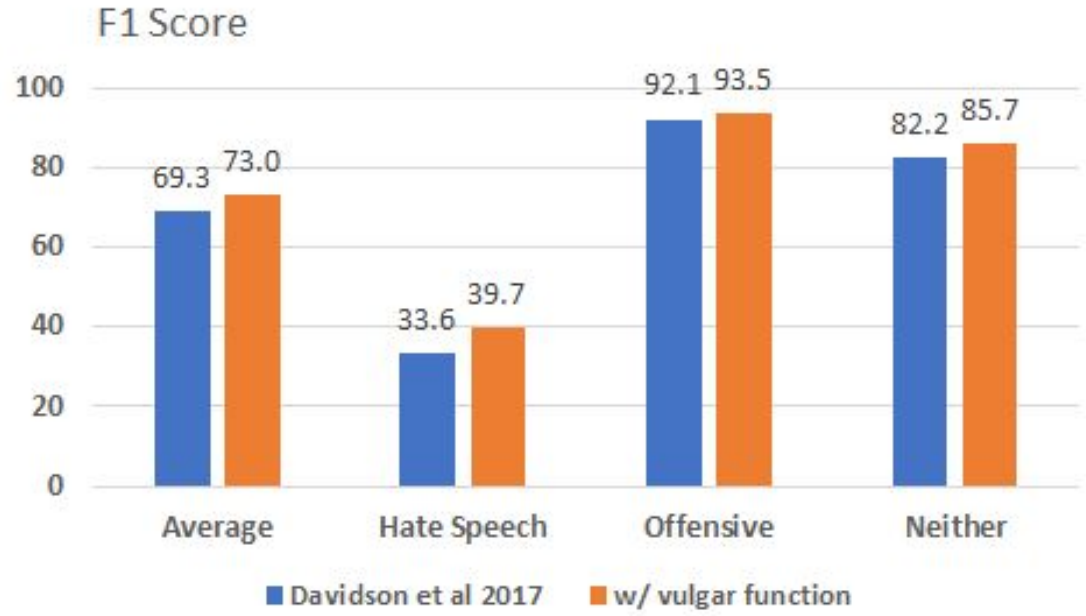
Vulgar Intention and Hate Speech

- Logistic regression model from Davidson et al (2017) using:
 - TF-IDF weighted token features
 - POS unigram to trigrams
 - reading level metrics
 - sentiment information
 - Twitter features (hashtags, mentions, etc.)
 - generic tweet features (character, word and syllable counts)
- We add an explicit vulgarity feature group:
 - The predicted distribution over vulgar functions (6 features)
 - averaged if more than one vulgar token/tweet

Vulgar Intention and Hate Speech

Results

- The addition of vulgarity features yields improvement in all three classes
- These features are most influential for detecting hate speech, the class with the lowest accuracy



Research Questions

Is modeling vulgar
intent useful for
NLP tasks?

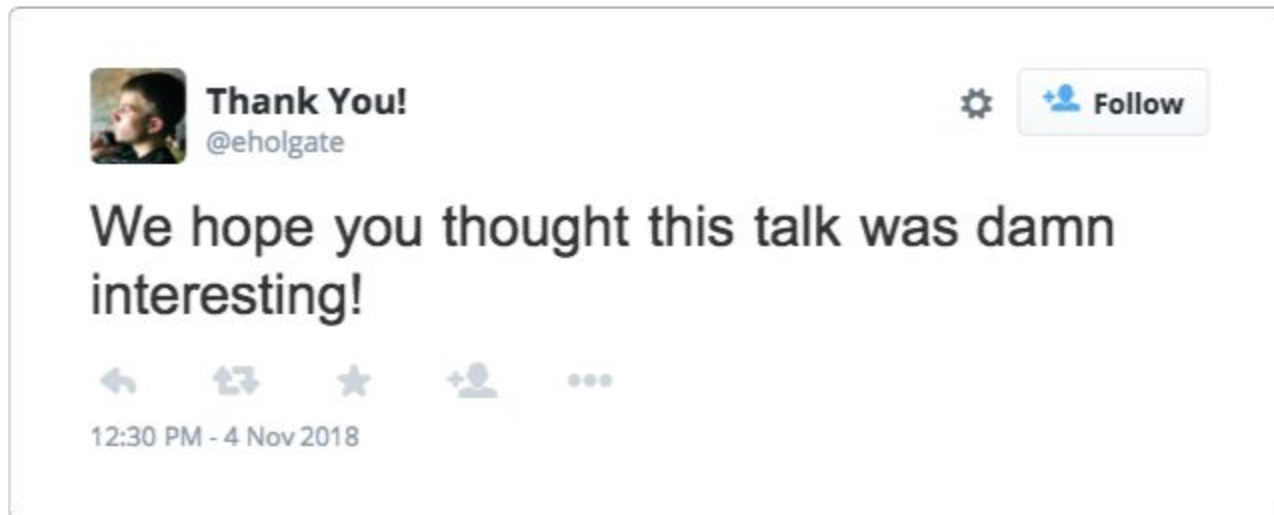
3

Yes!

Take Aways

- Vulgarity is used with several pragmatic functions
 - We can predict these from context
 - Vulgar intent is useful for downstream tasks like hate speech detection
- New data set focused on vulgarity functions
- Sociodemographic features impact vulgar role usage

Thank You!



holgate@utexas.edu

Annotation Confusion Matrix

	Agr	Emo	Emp	Aux	Sig	Non
Agr	0.63	0.11	0.09	0.07	0.10	0.01
Emo	0.07	0.59	0.20	0.13	0.01	0.01
Emp	0.04	0.18	0.68	0.07	0.01	0.02
Aux	0.07	0.16	0.15	0.56	0.03	0.03
Sig	0.17	0.06	0.07	0.11	0.57	0.02
Non	0.02	0.04	0.04	0.10	0.02	0.77

Most Frequent Terms by Function

Aggression		Express Emotion		Emphasis		Auxiliary		Signal Group Identity		Non-Vulgar	
Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq
cunt	86.9%	pissed	84.4%	fucking	84.7%	asses	73.9%	bitches	88.9%	mick	100%
asshole	86.3%	bullshit	64.2%	fuckin	84.0%	shitting	69.2%	nigga	85.7%	cracker	97.5%
asshit	83.0%	fucked	61.3%	goddamn	70.0%	arse	69.2%	slut	26.0%	dyke	92.8%
faggot	81.8%	shitty	52.6%	damn	62.3%	cock	62.9%	whore	25.0%	coon	92.8%
fag	73.3%	shit	42.5%	hell	51.4%	pussy	52.3%	hoe	23.8%	ho	88.3%

- Some words are very consistently used with a specific function
 - For these words, this feature will be very predictive
- For words like ass, which are very diverse, this feature will be less informative.