Sentiment and Emotion Analysis for Education

Cristina Garbacea
University of Michigan

Student emotional wellbeing

- Essential for both academic and social development
- American College Health Association: 32% of students depressed
  UC Berkeley
    2015: 47% of graduate students suffer from depression
    2005: 10% had considered suicide
- Student communications: academic engagement, satisfaction, goals, well-being, and career expectations

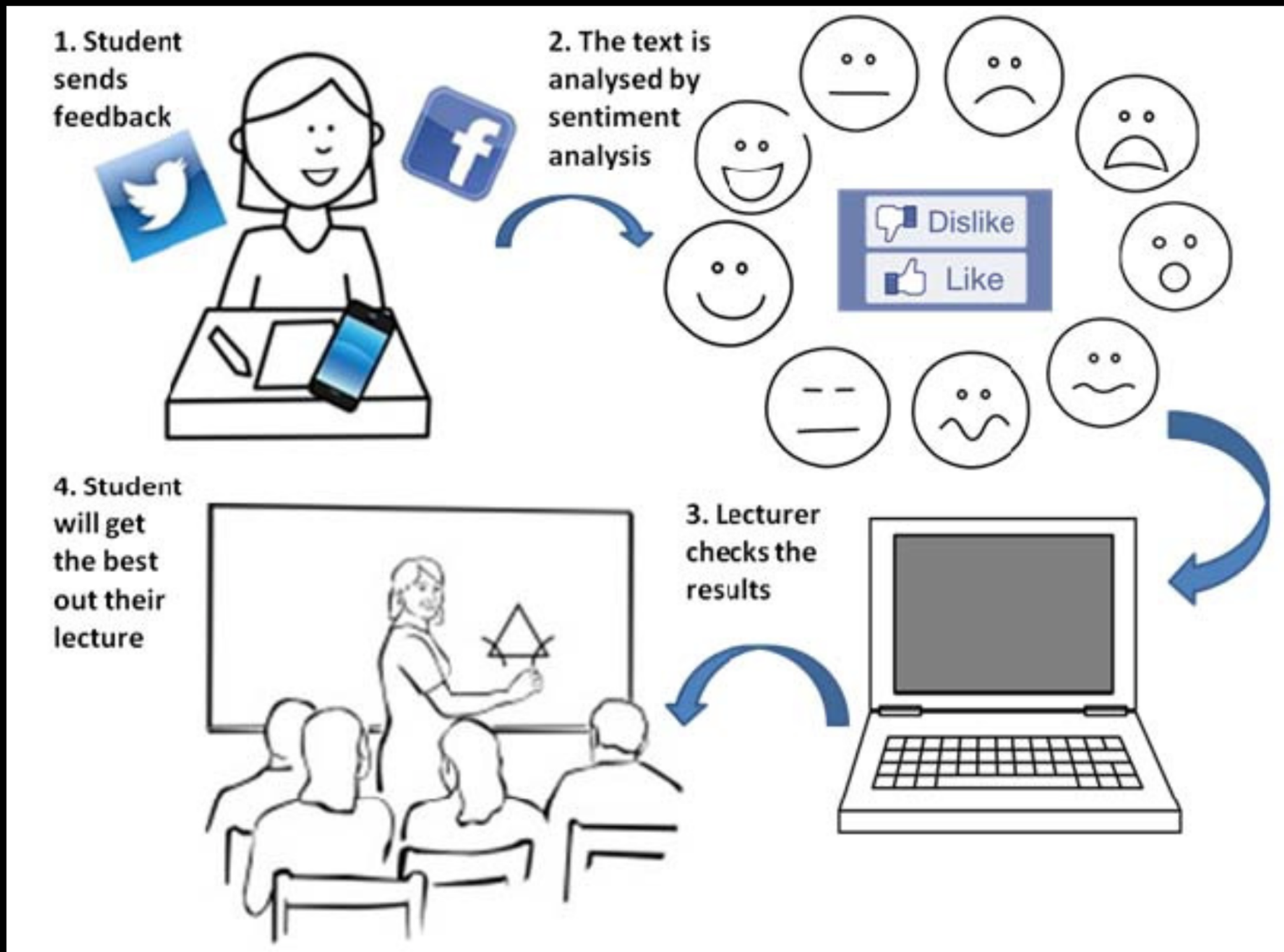


Figure 1: SA-E System Architecture

SA-E: Sentiment Analysis for Education (Nabeela Altrabsheh, Mohamed Medhat Gaber, Mihaela Cocea, IDT2013)

Student Academic Discourse

Prior research:

- primarily relies on survey responses
    few studies analyze online emotions and opinions

- little understanding of academic discourse at large scale:
    small sample sizes (100-200 students)
    small scale contexts (single classroom)

Using Social Media to Measure Student Wellbeing:
A Large-Scale Study of Emotional Response in Academic Discourse

(Svitlana Volkova, Kyungsik Han, and Courtney Corley, SocInfo 2016)

- **Emotions** = states in which internal sensations are experienced
  - can be triggered by the external environment
  - short in duration
  - *{ joy, sad, fear, disgust, anger, surprise } (Ekman)*

- **Sentiments** = likes and dislikes
  - involve a person-object relationship
  - display themselves over long periods of time
  - *{ positive, negative, neutral }*

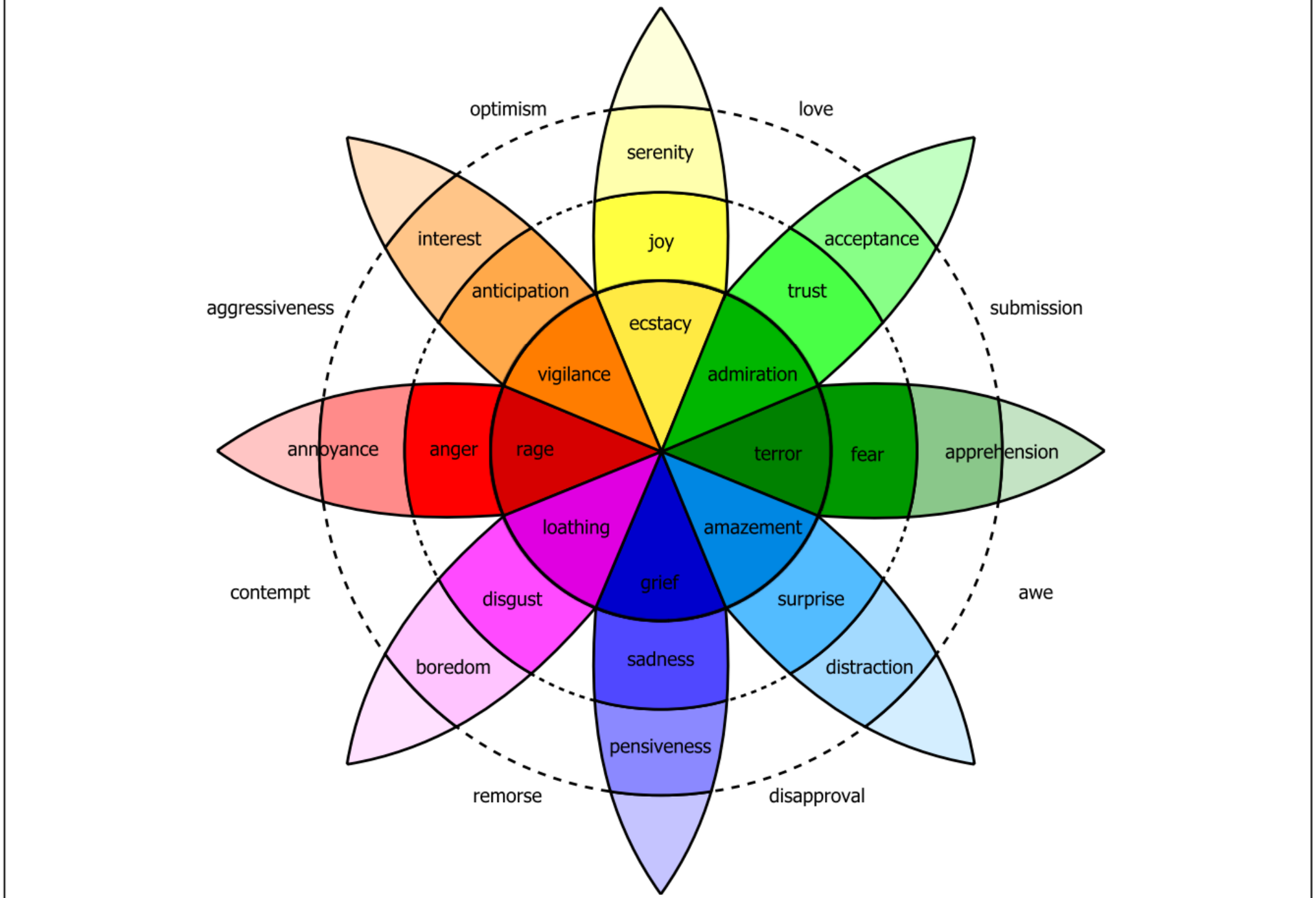


Figure 21.11 Plutchik wheel of emotion.

(Speech and Language Processing, Dan Jurafsky and James H. Martin, 2016)

- RQ: How students use social media broadly, and Twitter specifically, for academic purposes?

- Goals:
  - classify discourse types as social/personal/general
  - analyzing emotions and sentiments (affects)
  - measure variations in emotions and sentiment
  - correlate affects in social media with public survey data

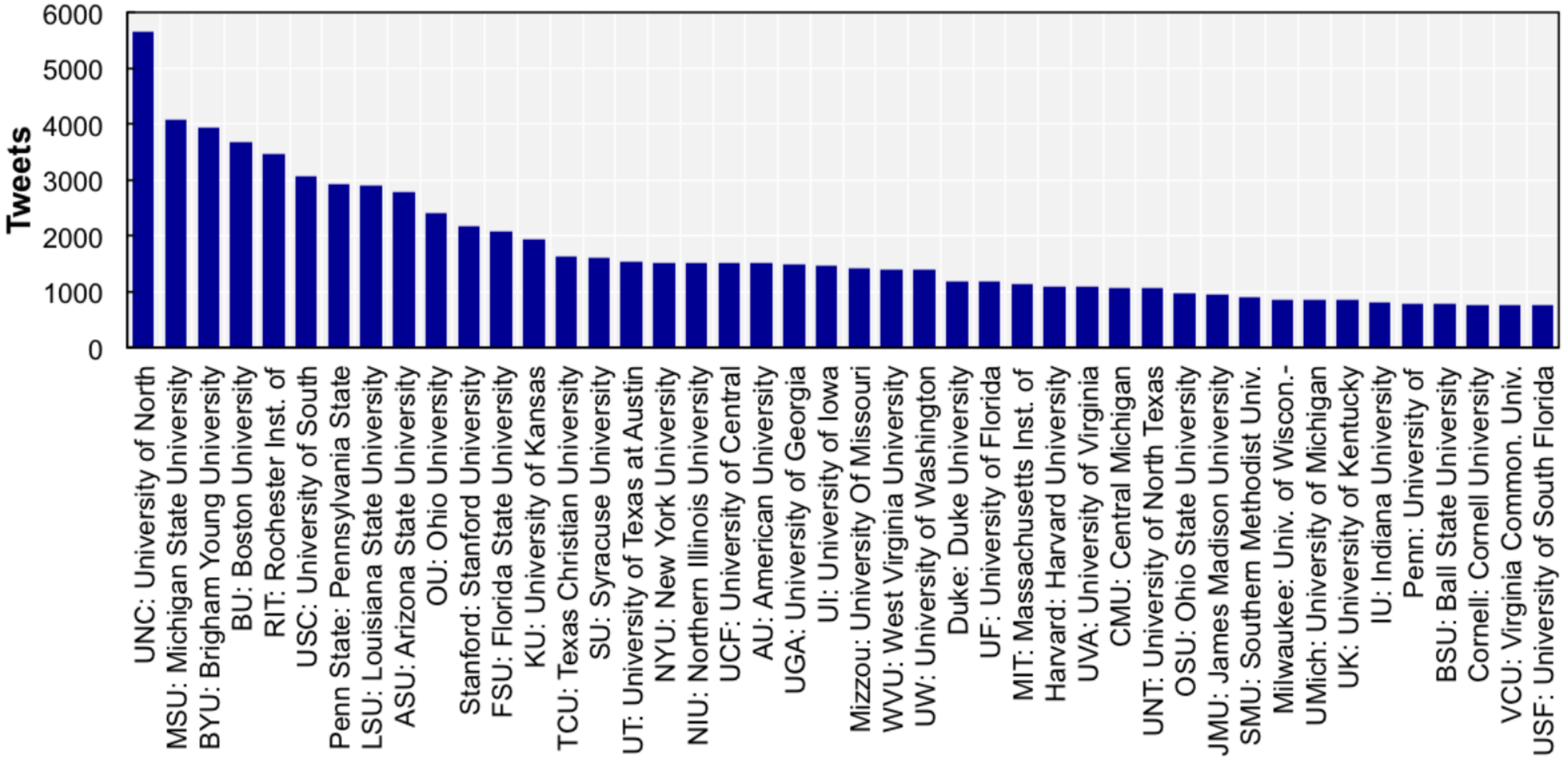


Fig. 1: The list of 44 U.S. universities and the number of tweets per university.

- Dataset:
  - 124,801 education related tweets from 7,208 users
  - bioinformatics, computer science, social science, psychology, economics, political science

- **Discourse-type classification models**
  (general/ personal/ social)

  - Features: word n-grams (unigrams, bigrams, trigrams), tf.idf, text embeddings (Word2Vec, GLoVe, NPMI)
  - Classification Model: logistic regression, 10 fold c.v.
  - F1 = 0.64

- **Sentiment and Emotion classification models**

  - publicly available emotion and sentiment models
  - Features: lexical (word ngrams), syntactic, stylistic (elongations, positive and negative emoticons, hashtags, punctuation, and negation)
  - F1= 0.6 for sentiment, F1 = 0.78 for emotion

Results - Discourse Type (I)

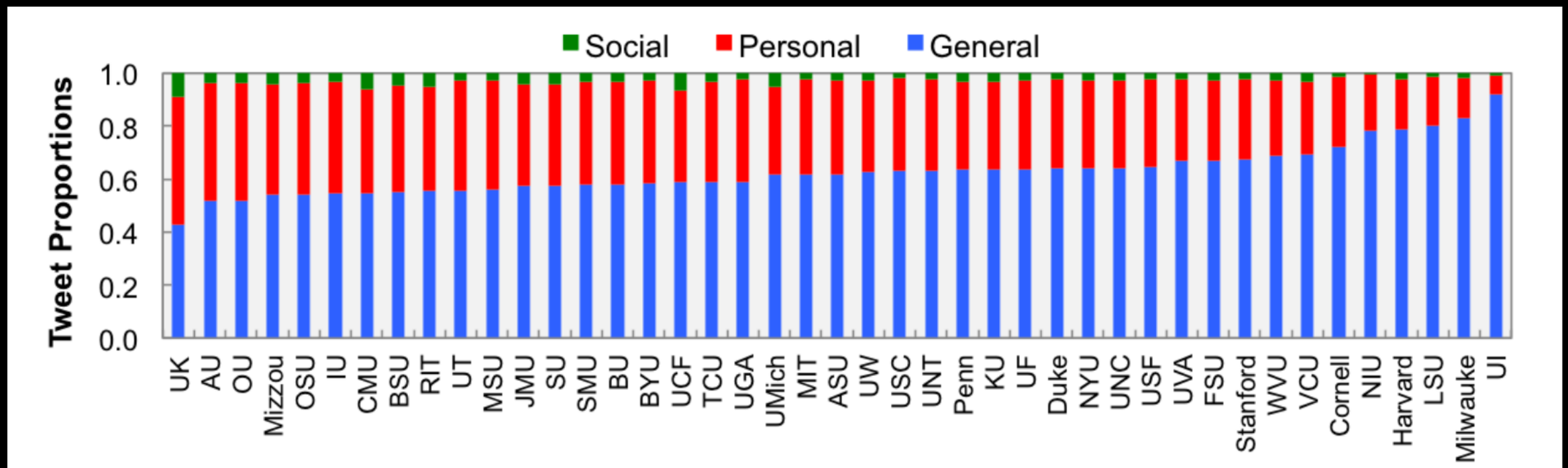


Fig. 2: Proportion of tweets per university classified as general, personal, and social. Proportions are sorted by general type in an ascending order.

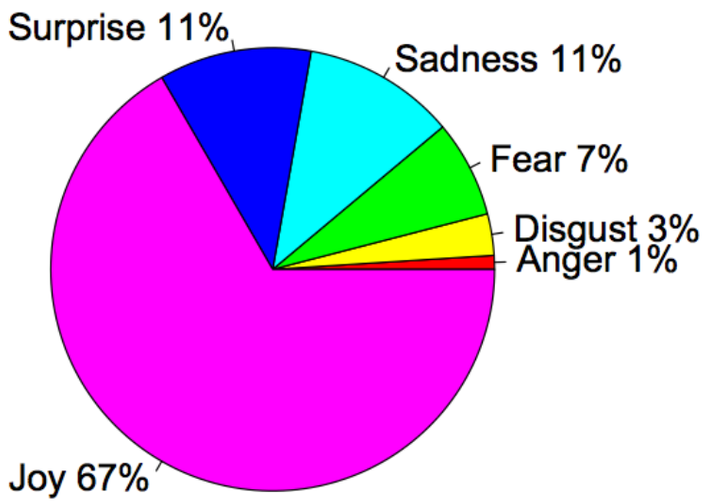
- general (62%), personal (34%), and only 4% of all tweets were labeled as social.

Results - Discourse Type (I)

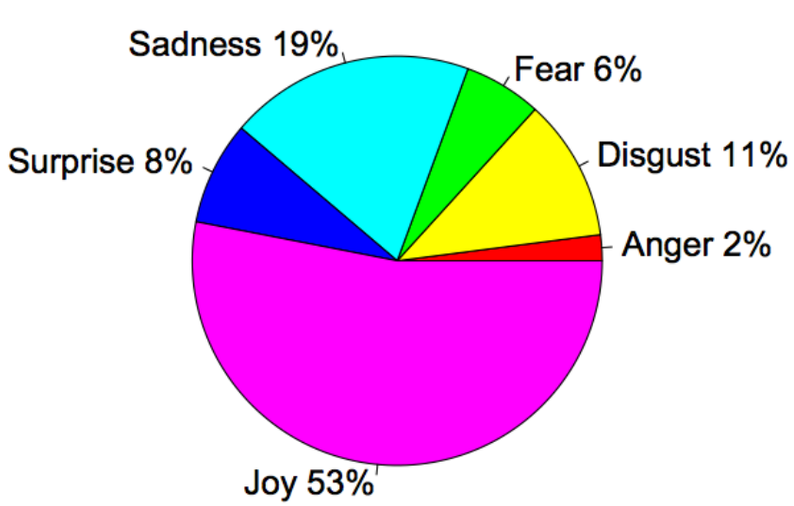
Table 1: The example tweets annotated with three types of academic discourse – social, personal, and general. The total number of annotated tweets is 1,569.

Main category	**Sub-category (Tweet Example)**
General (850)	(1) General academic information [46] *New 2 year postgraduate positions [name] in social studies of algorithms and data! 1st deadline Sept 26th.* (2) Other research studies, reports, or activities (804) *Insightful comparison of the two statistical cultures (data & algorithmic) by [name]. [URL]*
Personal (495)	(3) Personal academic achievements (32) *I'm excited to intern at @Microsoft this summer!* (4) Personal academic activity updates (210) *Submission complete... First conference submission as a first author. Woo!* (5) Personal academic thoughts or questions (253) *Is it weird that while I'm so tired w/ work, completing tasks somehow makes me feel good? #school*
Social (224)	(6) Academic interest dialogues (224) *@[name1] @[name2] Seems there is a theoretical limit on the number of citations you can fit into 10 pages...*

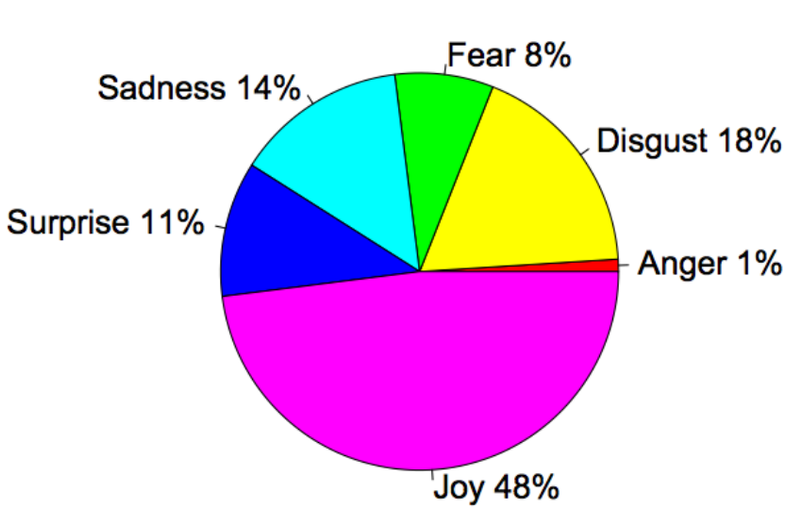
Results - Emotion (II)



(a) General: 53,190 tweets



(b) Personal: 28,461 tweets



(c) Social: 2,498 tweets

- joy most prevalent emotion across all discourse types, then sadness
- disgust expressed more often in social discourse
- surprise expressed the least in personal discourse
- anger and fear expressed equally across all types

Results - Sentiment (III)

Table 3: Discourse-based analysis results: sentiment proportions extracted from 79,329 tweets over 44 universities.

Discourse Type	General	Personal	Social
Neutral	29%	**38%**	**44%**
Negative	**51%**	36%	24%
Positive	20%	25%	32%
Subjective	**71%**	62%	56%

- Subjective opinions (positive and negative) were expressed more in general discourse compared to social and personal tweets.

- Positive sentiment is expressed the least

Results - Universities (IV)

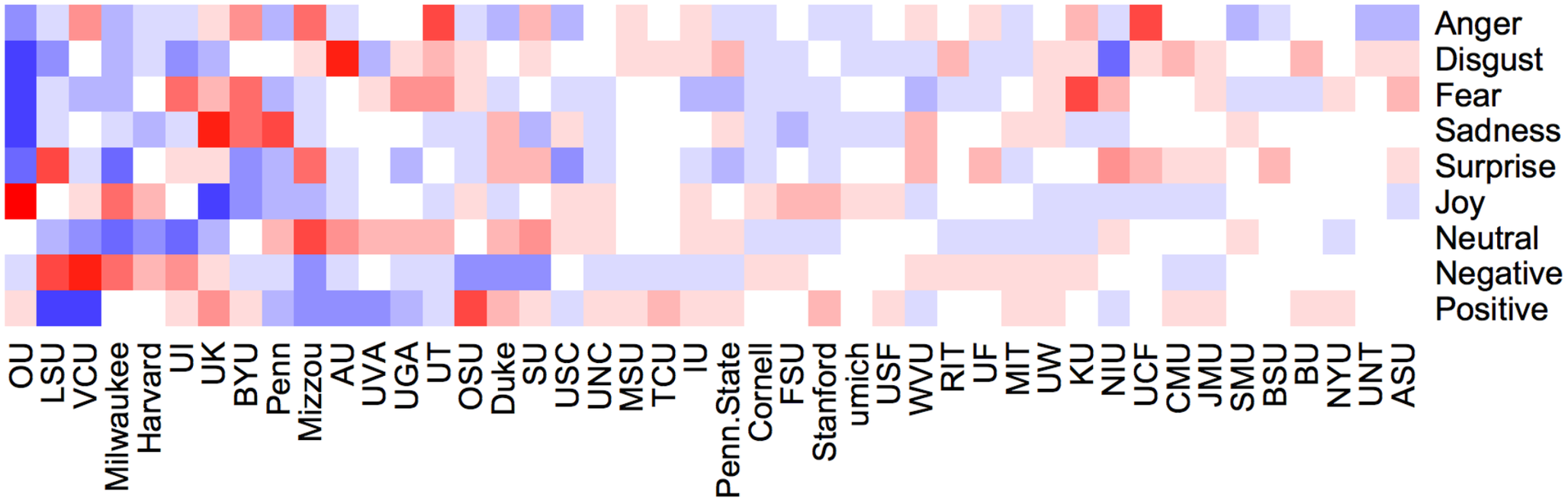


Fig. 4: University-based analysis results: sentiment and emotion proportions expressed across universities extracted from all discourse types. Red↑ represents high, blue↓ represents low sentiment and emotion proportion values.

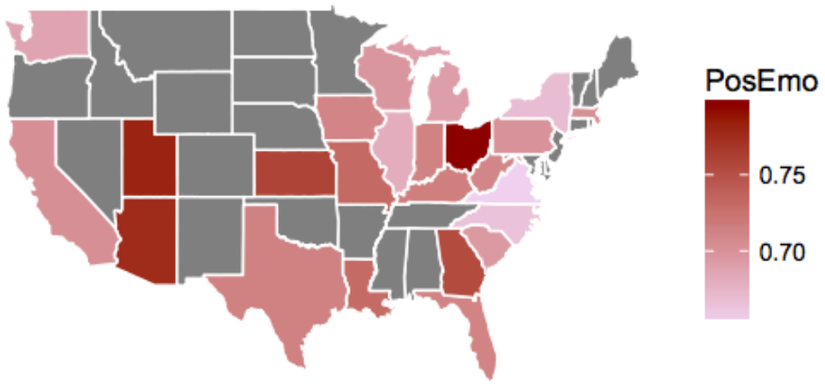
- differences in the degree of expressing emotions and sentiments across universities

Table 4: Top 10 colleges that are the most subjective vs. neutral, produce most positive vs. negative opinions, and produce most positive vs. negative emotions. Colleges in bold show the overlap with the top happiest schools[11], and schools in italic show the overlap with the top stressed schools[12] obtained using survey data, demonstrating that affects depicted in students tweets could be used to understand their happiness (or stress) in their school life.

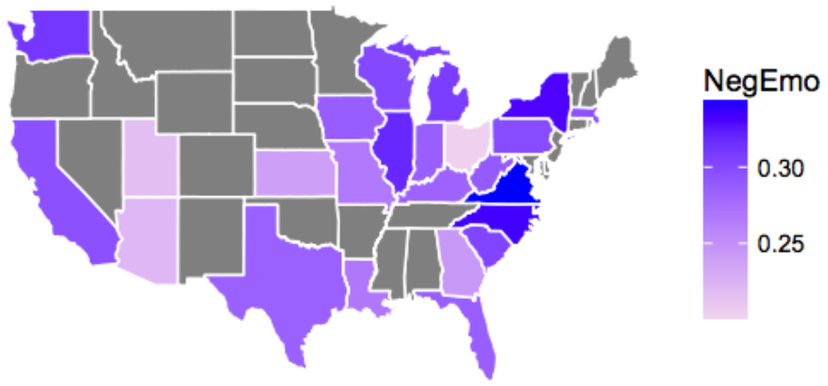
Opinionatedness		**Sentiments**		**Emotions**	
NEUTRAL	SUBJECTIVE	POSITIVE	NEGATIVE	POSITIVE	NEGATIVE
Mizzou	UI	UST	VCU	OU	BYU
USC	Milwaukee	**OSU**	LSU	Milwaukee	UK
AU	VCU	UK	Milwaukee	**FSU**	*Penn*
SU	Harvard	Duke	UI	**Harvard**	*UNC*
UGA	UK	**Stanford**	Harvard	USC	*AU*
UVA	Carolina	TCU	RIT	LSU	UT
Penn	LSU	**BYU**	*UNC*	Cornell	KU
UT	UW	USF	*UF*	**Stanford**	UGA
Duke	Stanford	SU	WVU	UNC	UW
Penn State	UF	JMU	Cornell	UMich	*JMU*

Results - Geolocation (V)

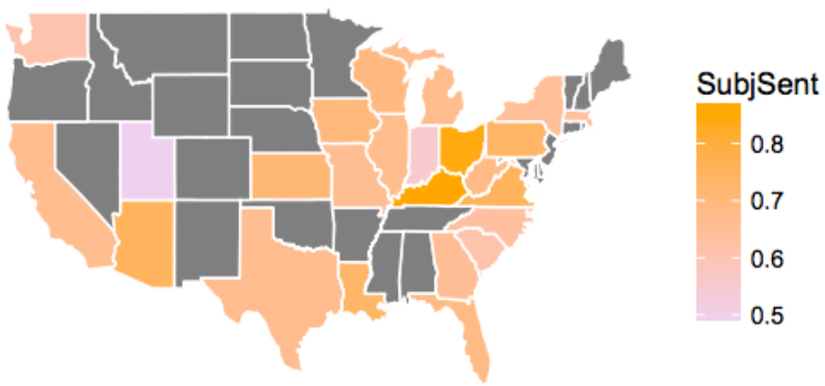
(a) Positive emotion proportions



(b) Negative emotion proportions



(c) Subjective opinion proportions



(d) Neutral opinion proportions

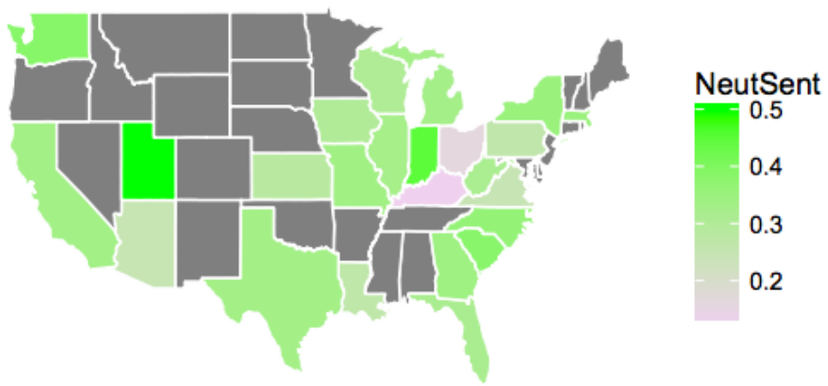


Fig. 6: Geo-located positive and negative emotion proportions, neutral and subjective sentiment proportions aggregated across universities and U.S. states (grey color shows states with no emotion and sentiment data available).

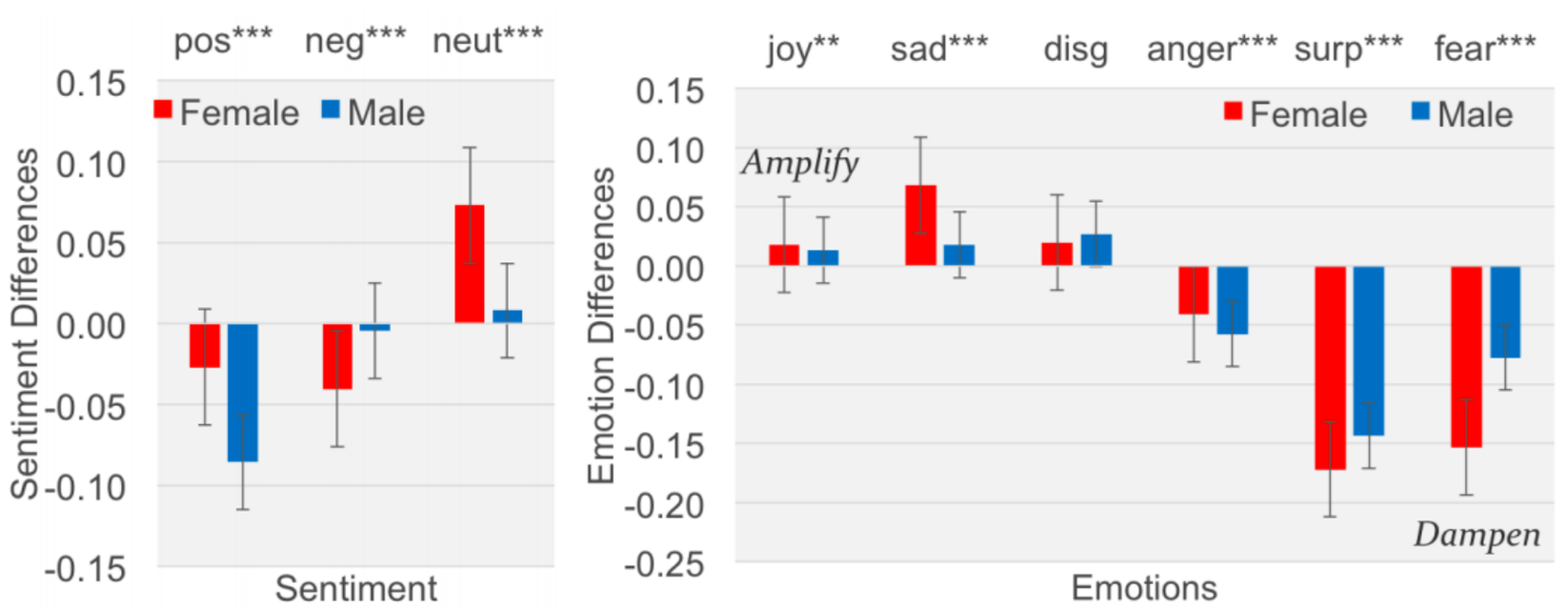
Conclusion

- emotion and opinion signals expressed on Twitter correlate with university tuition, and student satisfaction

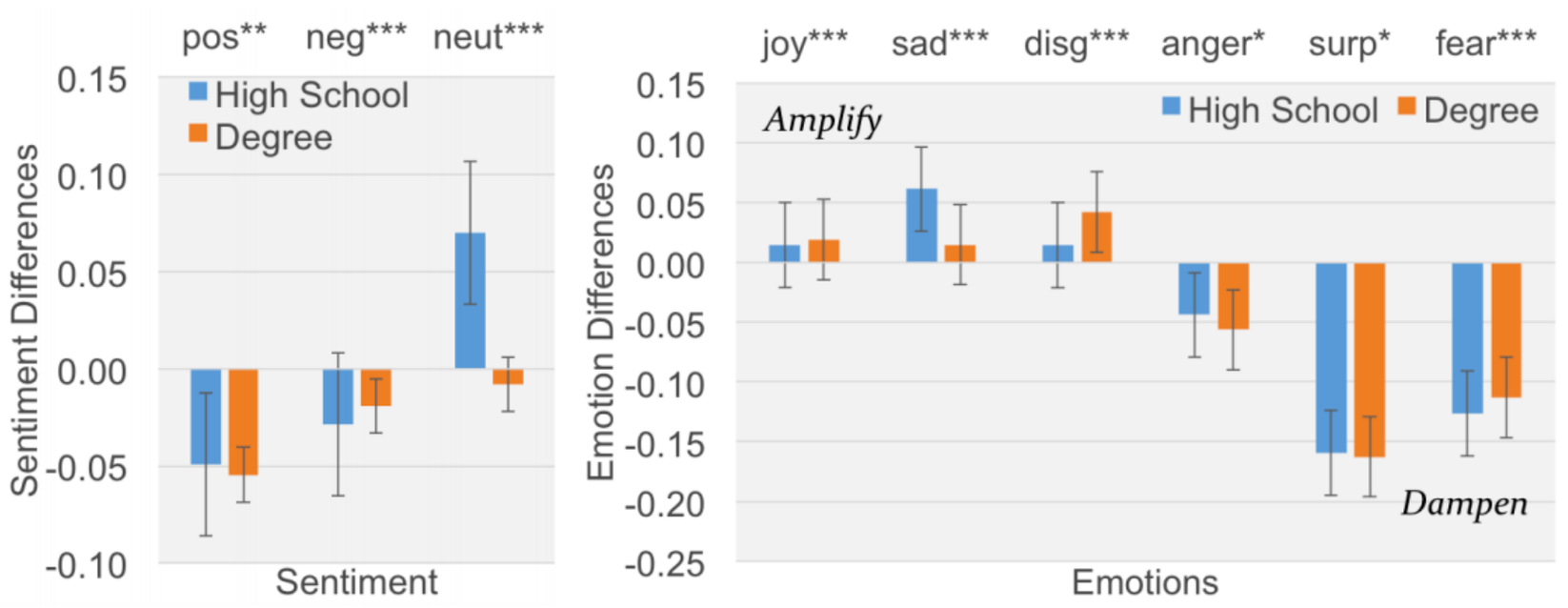
- emotions and sentiments expressed in academic discourse across universities correlate with public survey data on the most happy and stressed colleges in the U.S.

- universities differ in terms of students expressing anger, disgust, fear, and surprise, in addition to joy and sadness

Future work:
- other variables that can potentially influence students' mood (e.g. weather)



(a) Male vs. female



(c) College degree vs. high school education