



Who is bragging more online? A large scale analysis of bragging in social media



Mali Jin[◇], Daniel Preoțiu-Pietro[♠], A. Seza Doğruöz[♣] and Nikolaos Aletras[◇]
◇University of Sheffield, ♠Bloomberg, ♣Ghent University

{m.jin, n.aletras}@sheffield.ac.uk, dpreotiucpie@bloomberg.net, as.dogruoz@ugent.be



Bragging

- Bragging is a speech act that involves **disclosing positive content** about oneself or their close network

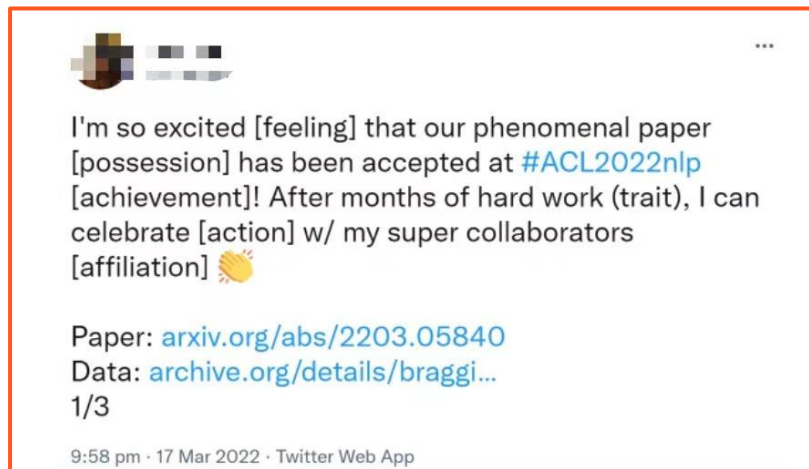


Figure 1: An example of bragging tweet.

Bragging

- Bragging is a speech act that involves **disclosing positive content** about oneself or their close network
- Bragging behavior was quantitatively studied in specific contexts through **manual analysis** of **small data sets** of hundreds of posts

Bragging

- Bragging is a speech act that involves **disclosing positive content** about oneself or their close network
- Bragging behavior was quantitatively studied in specific contexts through **manual analysis** of **small data sets** of hundreds of posts
- Our work employs **computational sociolinguistics methods** to conduct the first **large scale study** of bragging behavior on Twitter (U.S.) by focusing on its **overall prevalence, temporal dynamics** and impact of **demographic factors**.

Measuring Bragging Prevalence

- **Data for Model Training**

- We use a bragging data set consisting of **6,696 English tweets** and each tweet is manually annotated as **bragging (781)** and **non-bragging (5,915)** (Jin et al., 2022)

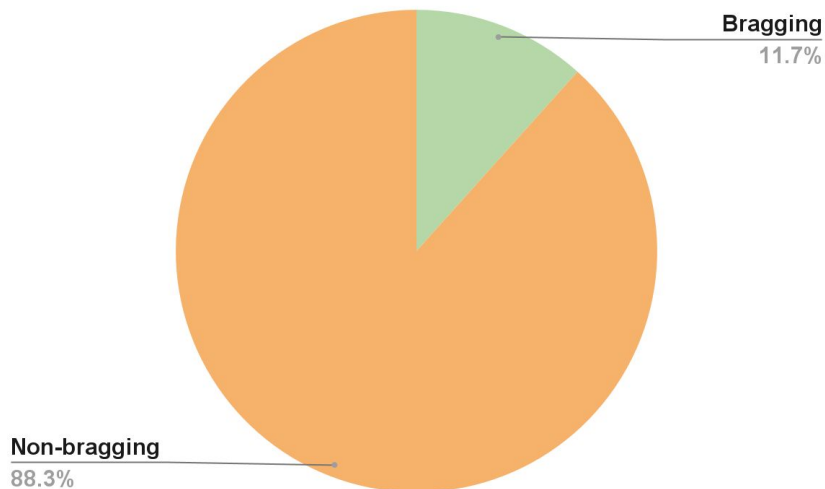


Figure 2: Statistics of the bragging data set.



Measuring Bragging Prevalence

- Data for Model Training
- Predictive Model
 - We re-implement the best performing predictive model, **BERTweet-LIWC**, in the original paper ([Jin et al., 2022](#)) on identifying whether a tweet contains bragging or not (**72.51** macro F1)

Measuring Bragging Prevalence

- Data for Model Training
- Predictive Model
- Bragging Prevalence Metrics
 - **Overall Bragging Prevalence** - We compute the distribution of **average bragging percentage** across users for each time window (**each month**)

$$A(a_1, \dots, a_n) = \frac{P(p_1, \dots, p_n)}{Q(q_1, \dots, q_n)} \quad n: \text{number of months}$$

Measuring Bragging Prevalence

- Data for Model Training
- Predictive Model
- Bragging Prevalence Metrics
 - **Overall Bragging Prevalence** - We compute the distribution of **average bragging percentage** across users for each time window (**each month**) (n: number of months)

$$A(a_1, \dots, a_n) = \frac{P(p_1, \dots, p_n)}{Q(q_1, \dots, q_n)} \quad n: \text{number of months}$$

- **User Bragging Prevalence**
 - We obtain a **time-normalized** bragging distribution for **each user** in each month

$$D_u(d_{u_1}, \dots, d_{u_n}) = \frac{B_u(b_{u_1}, \dots, b_{u_n})}{T_u(t_{u_1}, \dots, t_{u_n}) * A}$$

Measuring Bragging Prevalence

- Data for Model Training
- Predictive Model
- Bragging Prevalence Metrics
 - **Overall Bragging Prevalence** - We compute the distribution of **average bragging percentage** across users for each time window (**each month**) (n: number of months)

$$A(a_1, \dots, a_n) = \frac{P(p_1, \dots, p_n)}{Q(q_1, \dots, q_n)} \quad n: \text{number of months}$$

- **User Bragging Prevalence**
 - We obtain a **time-normalized** bragging distribution for **each user** in each month

$$D_u(d_{u_1}, \dots, d_{u_n}) = \frac{B_u(b_{u_1}, \dots, b_{u_n})}{T_u(t_{u_1}, \dots, t_{u_n}) * A}$$

- We **average** the normalized bragging distribution for each user

$$l = \frac{\sum_{i=1}^n d_{u_i}}{n}$$

Analysis Data

- **A Set of Twitter Users**

- We combine three data sets where users associated with self-reported **socio-demographic characteristics** (Preotiuc-Pietro et al., 2016; Guntuku et al., 2017; Jaidka et al., 2020), which contains **6,369** users in total

Analysis Data

- **A Set of Twitter Users**

- We combine three data sets where users associated with self-reported **socio-demographic characteristics** (Preotiuc-Pietro et al., 2016; Guntuku et al., 2017; Jaidka et al., 2020), which contains **6,369** users in total
- We collect **all historical tweets** from these users resulting in more than **9.7 million tweets**

Analysis Data

- **A Set of Twitter Users**
- **Data Filtering**
 - Non-English tweets, duplicate tweets, replies and retweets

Analysis Data

- A Set of Twitter Users
- Data Filtering
 - Non-English tweets, duplicate tweets, replies and retweets
 - Tweets that are **automatically generated from third parties**



Figure 3: an example of tweet generated by third parties automatically.

Analysis Data

- **A Set of Twitter Users**
- **Data Filtering**
 - Non-English tweets, duplicate tweets, replies and retweets
 - Tweets that are **automatically generated from third parties**
 - Users that have posted **fewer than 20 tweets**

Analysis Data

- **A Set of Twitter Users**
- **Data Filtering**
 - Non-English tweets, duplicate tweets, replies and retweets
 - Tweets that are **automatically generated from third parties**
 - Users that have posted **fewer than 20 tweets**
 - Finally, it contains **2,685 users** and **1,031,276 tweets** in total

Analysis Data

- **A Set of Twitter Users**
- **Data Filtering**
- **Computing Bragging Ratio**
 - We use the predictive model to identify bragging or not of all tweets (**78.55 macro F1** on a batch of 100 tweets in the analysis data)
 - Then we compute the normalized bragging ratio for each user using bragging prevalence metrics

Analysis Data

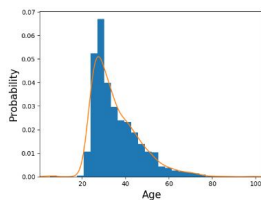
- **A Set of Twitter Users**
- **Data Filtering**
- **Computing Bragging Ratio**
- **User Demographics Traits**
 - Gender
 - Age
 - Education
 - Income

Analysis Data

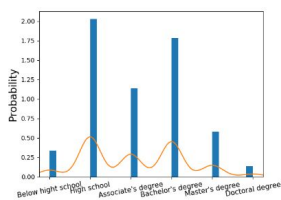
- **A Set of Twitter Users**
- **Data Filtering**
- **Computing Bragging Ratio**
- **User Demographics Traits**
- **User Popularity Traits**
 - No. Followers
 - No. Friends/No. Followers
 - No. Listings

Analysis Data

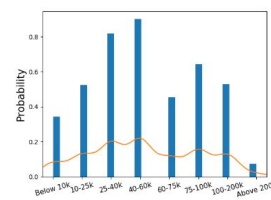
- A Set of Twitter Users
- Data Filtering
- Computing Bragging Ratio
- User Demographics Traits
 - Gender (males-33.81%, females-66.19%), Age, Education, Income
- User Popularity Traits
 - No. Followers, No. Friends/No. Followers, No. Listings



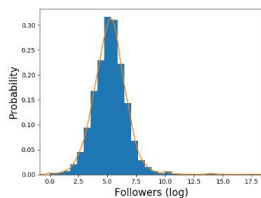
(a) Age.



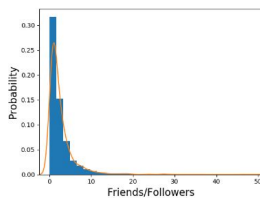
(b) Education.



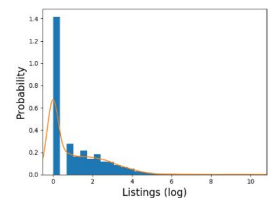
(c) Income.



(d) Log-scaled no. followers.



(e) Friends/followers.



(f) Log-scaled no. listings.

Figure 4: Statistics of user demographics and popularity traits

Bragging Prevalence over Time

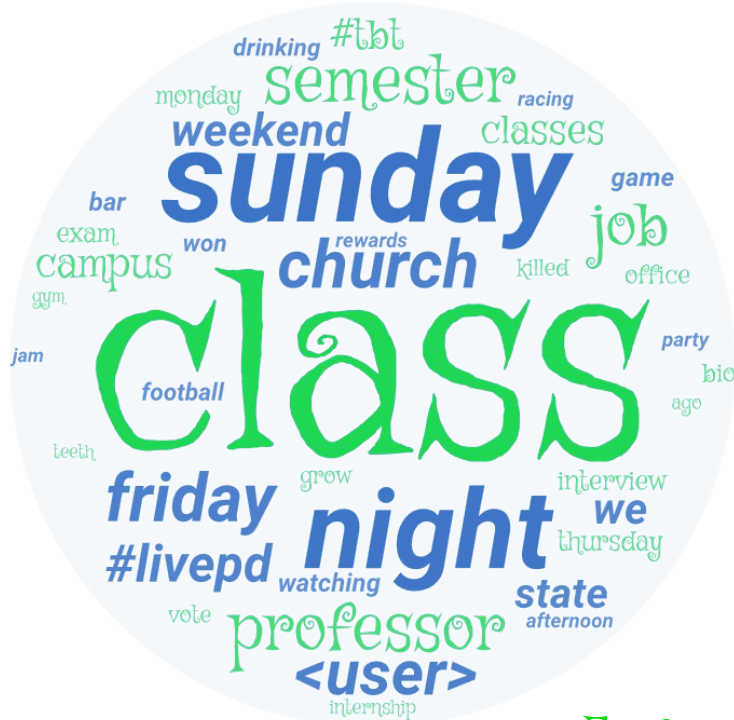
The **mean** bragging ratio is **0.1114** and the **median** bragging ratio is **0.0968** (0.0020 and 0.0013 after normalization) for all users.



Figure 5: Bragging percentage by year and month.

When to Brag and How

Day of the Week



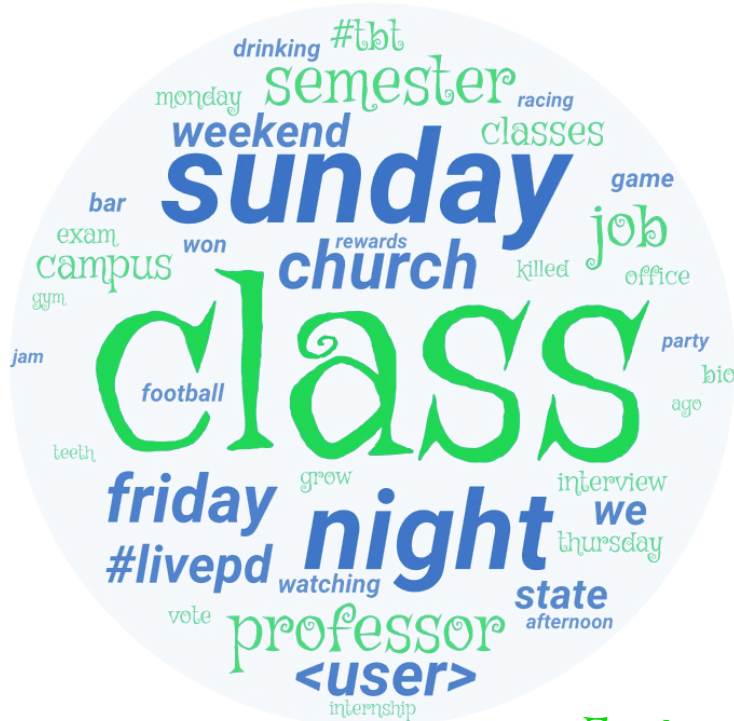
- Both weekdays and weekends involve words related to **time** (*monday, sunday*)

Figure 6: Unigram feature correlations with bragging between weekday and weekend.

Female
Male

When to Brag and How

Day of the Week



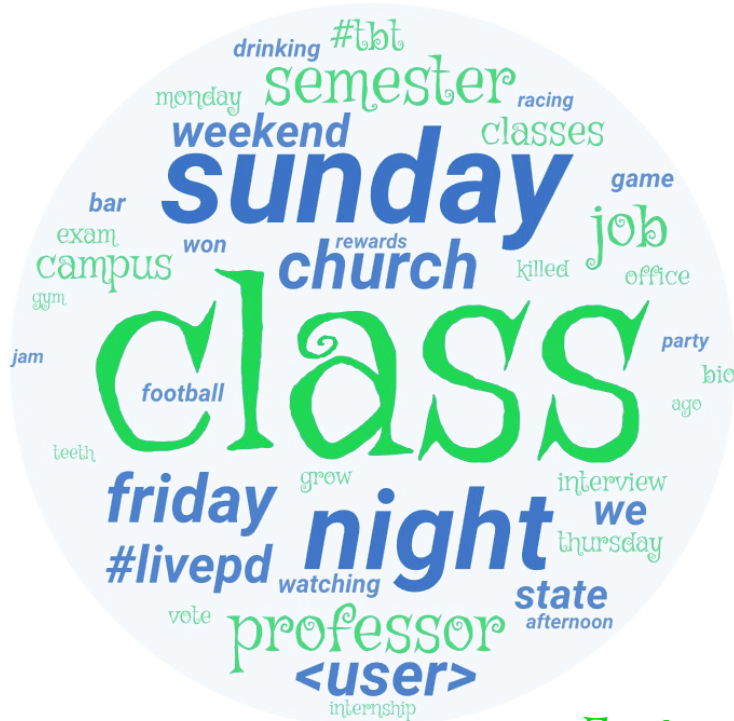
- Both weekdays and weekends involve words related to **time** (*monday, sunday*)
- User mostly brag about their **school life** or **work** on weekdays (*class, office*)

Figure 6: Unigram feature correlations with bragging between weekday and weekend.

Female
Male

When to Brag and How

Day of the Week



- Both **weekdays** and **weekends** involve words related to **time** (*monday, sunday*)
- User mostly brag about their **school life** or **work** on **weekdays** (*class, office*)
- Another popular bragging topic on weekdays is about going to **gym**
- Bragging on weekends usually focus on certain **entertainment, recreation** and **worship activities** (*church, bar, football*)

Figure 6: Unigram feature correlations with bragging between weekday and weekend.

Female
Male

Demographic and Popularity Factors

Trait	r	p_{unc}	p_{corr}
User Demographics			
Gender ($\text{♀}-1, \text{♂}-0$)	0.10	<.001	<.001
Age	-0.16	<.001	<.001
Education	0.14	<.001	<.001
Income	0.07	<.003	<.002
User Popularity			
No. Followers	0.12	<.001	<.001
No. Friends/Followers	-0.10	<.001	<.001
No. Listings	0.09	<.001	<.001

Table 1: Pearson correlations between user-level traits and their bragging metric. p_{unc} and p_{corr} refer to uncorrected and corrected (Bonferroni correction) p-values.

Main Findings

- Bragging **decreases over time** within the same population of users

Main Findings

- Bragging **decreases over time** within the same population of users
- **Female, younger, more educated** and **popular** users and users with a **higher income** in the U.S. are more likely to brag

Main Findings

- Bragging **decreases over time** within the same population of users
- **Female, younger, more educated** and **popular** users and users with a **higher income** in the U.S. are more likely to brag
- **Male** users and users with **higher income** brag more about leisure activities; **Female** users and users with **lower education** focus more on themselves when bragging. Bragging by **older** users and users who have **higher education** are more likely to involve others. Emojis are more frequently used by **female** and **younger** users while bragging

Takeaways

- We present the the **first large-scale** quantitative study of bragging on social media, focusing on understanding the **prevalence, temporal dynamics** and **user factors** impacting bragging prevalence

Takeaways

- We present the the **first large-scale** quantitative study of bragging on social media, focusing on understanding the **prevalence, temporal dynamics** and **user factors** impacting bragging prevalence
- The analysis involves more than 1 million English Twitter posts from a **group of 2,685 users** in the U.S. over ten years

Takeaways

- We present the the **first large-scale** quantitative study of bragging on social media, focusing on understanding the **prevalence, temporal dynamics** and **user factors** impacting bragging prevalence
- The analysis involves more than 1 million English Twitter posts from a **group of 2,685 users** in the U.S. over ten years
- We conduct an extensive **linguistic analysis** to unveil specific **bragging themes** associated with different traits

Thank you!

- Please check out our paper for more details!
- Contact Information
 - Mali Jin: m.jin@sheffield.ac.uk

