

# An Analysis of Twitter Discourse During the Tokyo 2020 and Beijing 2022 Olympics

Jordan Rodrigues & Naman Bhargava  
Georgia Institute of Technology



### Story

Diverse Twitter discourse is not always beneficial (hate speech against athletes, nations and identity)

Are certain topics inherently toxic?  
 Do regions vary in sentiment and toxicity?  
 Is there variance across Olympic years?

Why now?

### Methods

#### Modeling Pipeline

395K Tweets → 
 Unigram and bigram decomposition → 
 Outlier removal → 
 Topic Modeling with LDA\* (6 topics\*\*)

\*Tuned to minimize UMass Coherence Score  
\*\*Principal words as topic labels

#### Statistical Pipelines (Non-parametric $\alpha = 0.001$ )

Event:  $\chi^2$  Goodness of Fit test  
 Popularity: Kruskal Wallis Test (median difference)  
 Nation: Standard error  
 Topic: Kruskal Wallis Tests

### Event Analysis

• 2022 has stat sig **greater** toxicity than 2020  
 • 2022 has stat sig more **polar** sentiment than 2020

### Nation Analysis

### Data

Data Sources (unchanged from midpoint)

	Dataset #1	Dataset #2	Dataset #3
Source	Kaggle	Kaggle	Twitter API
Event	Tokyo	Tokyo	Beijing
Search Query	#Tokyo2020	#Olympics	#Beijing2022 #Olympics
Dates	7/24/21-7/27/21	7/24/21-8/01/21	2/17/22-2/21/22
Rows	160K	291K	170K
Common Attributes	User - Username, location, description, creation date, followers, followees/friends Tweet - Retweets, favorites, text contents		

621K Tweets → 
 English Filter → 
 Lemmatize & stop word removal → 
 Sentiment Analysis → 
 Toxicity Analysis → 
 395K Tweets

	Followers	Friends	Toxicity	Sentiment	Char
Mean	97K	1.4K	0.05	0.16	114
STD	1.2M	7K	0.18	0.43	62
Min	0	0	0	-1	8
Max	78M	773K	1	1	955

### Key Findings

- Sentiment polarization increases in 2022 compared to 2020
- People are generally positive with low toxicity, and toxicity is unattractive to broader followings
- Countries have different sentiment and toxicity with some correlated with event performance
- Topics can have significantly different toxicity associations
- Results and modeling pipeline can inform future moderation to control toxicity in global events

	Low Toxicity	High Toxicity
- Sent.	Sports can be so cruel at times #TableTennis	Annoying c*nt stop whistling in the crowd
+ Sent.	Congratulations Mirabai proud of you	That was so f*cking awesome!!!

### Popularity Analysis

- Toxicity **decreases** with user popularity (stat sig)
- Positivity **increases** with popularity (stat sig)
- Extremely popular users use more **neutral** sentiment
- Avg. users have **positive** sentiment and **low** toxicity

### Topic Analysis

#### Topics Summary

U\_Mass Coherence Score: -3.3

**Kamila Valieva** - figure\_skating, win\_gold, kamila\_valieva,  
**China** - Taiwan, Propaganda, Sovereignty, rebuked\_china  
**FrozenFin** - frozen\_penis, finnish, skier\_suffers  
**Ceremony** - ceremony, closing, yuzu, dwen, eileen\_gu