

Micromobility in Smart Cities: A Closer Look at Shared Dockless E-Scooters via Big Social Data

Yunhe Feng¹, Dong Zhong², Peng Sun³,
Weijian Zheng⁴, Qinglei Cao², Xi Luo⁵,
and Zheng Lu²

¹University of Washington

²University of Tennessee

³Chinese University of Hong Kong, Shenzhen

⁴Purdue University

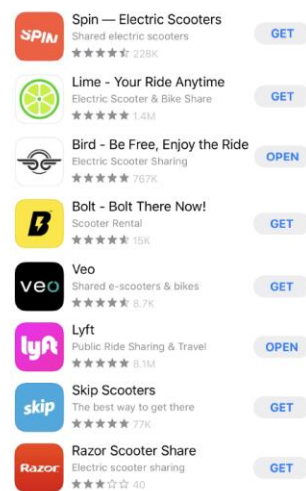
⁵SLAC National Accelerator Lab

Background

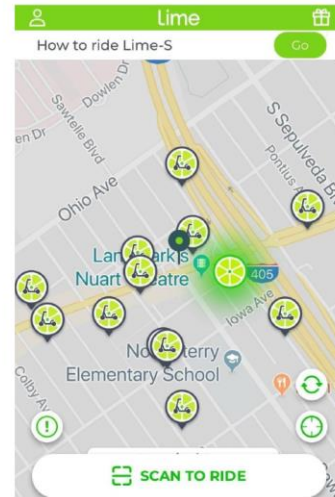
- **Pros:** affordability, easy accessibility via an app, and zero emissions.
- **Cons:** traffic rules, public safety, and parking regulations.
- **11%** of Paris residents reported using e-scooters either frequently or from time to time [1].
- According to NACTO[2], people took **38.5 million trips** on shared e-scooters in 2018.



(a) E-scooter



(b) E-scooter apps



(c) App UI


Motivation

- A missing research perspective from the riders' comments shared via smartphones, especially via social media apps.
- Social media is a good data source to investigate the e-scooter usages.
 - ❑ **Diversity:** multimodal social data (text, image, timestamps, GPS, emoji, etc.) enables detailed profiles of e-scooter sharing services in various aspects.
 - ❑ **Scalability:** performing a large-scale study of micromobility with social media data is flexible and effortless.
 - ❑ **Transparency:** public social media platforms (e.g., Twitter) addressing potential concerns of research non-reproducibility.

Methodology

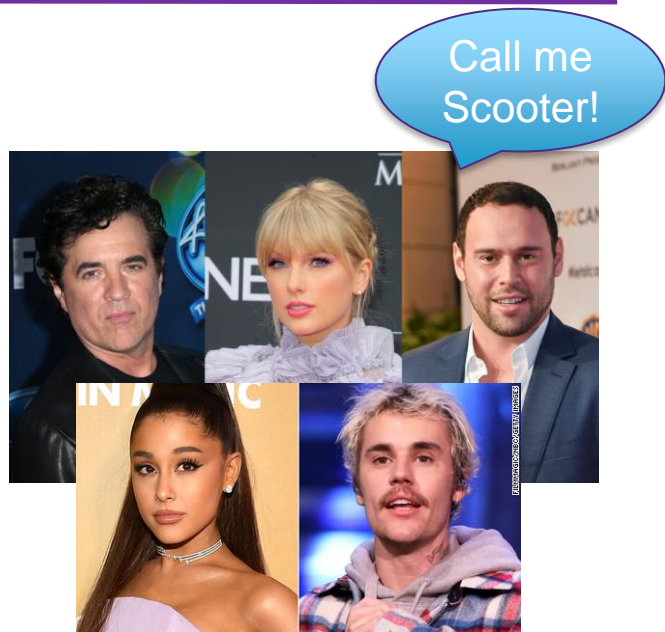
- Use Twitter to collect data
 - ☐ Free APIs to collect tweets of interest
 - ☐ The fourth most active social networking platform
 - ☐ Users expect the data they post on Twitter to be publicly available
- Big data analytics
 - ☐ Spatio-temporal Visualization
 - ☐ Topic discovery - latent Dirichlet allocation (LDA)
 - ☐ Optical character recognition (OCR)
 - ☐ Logo detection and recognition
 - ☐ Sentiment analysis

Data Collection & Bot Detection

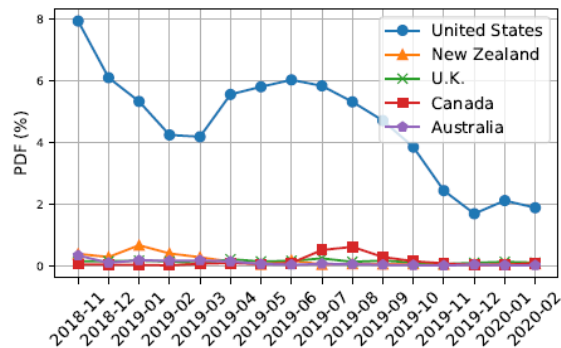
- Twitter's Streaming APIs with keywords of *scooter* and emoji 
 - ❑ 5.8 million tweets generated by 2.7 million users
 - ❑ 178,048 different image URLs (33,851 expired)
 - ❑ October 6, 2018 to March 14, 2020
- Remove bots [3]
 - ❑ Those who posted more than 525 scooter-tagged tweets, i.e., more than one such tweets per day, during the data collection period
 - ❑ Those who posted over 100 scooter-tagged tweets in total and the top three frequent posting intervals covered at least their 90% tweets.

Data Cleanup

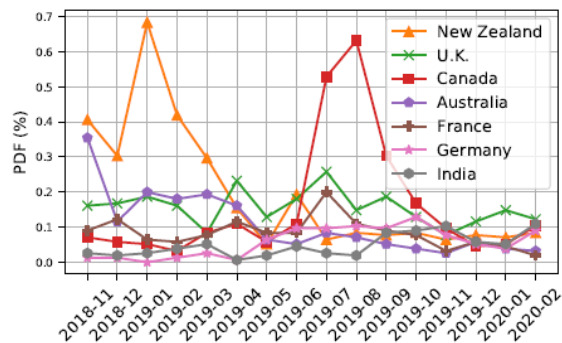
- Word co-occurrence analysis
 - ❑ Taylor, Swift, Justin, Bieber, Scott, Samuel, Braun, Ariana, Grande, and Borchetta
 - ❑ 1,541,815 related tweets were deleted
- Further reduce false positives
 - ❑ Tweets containing the word of **Share**
 - ❑ Tweets containing shared e-scooter brands including **Bird, Lime, Spin, Bolt, gruv, Lyft, Sherpa, VeoRide, Taxify, Jump, RazorUSA, Scoot Networks**, and **Skip**.



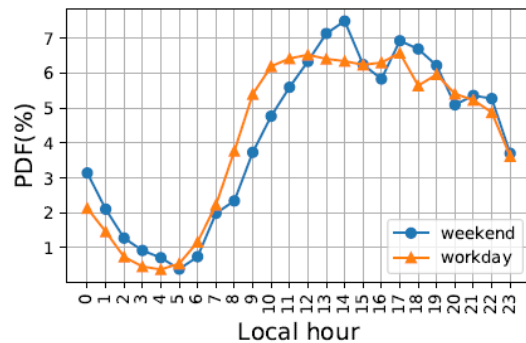
Temporal Distribution



(a) Monthly dist. by country



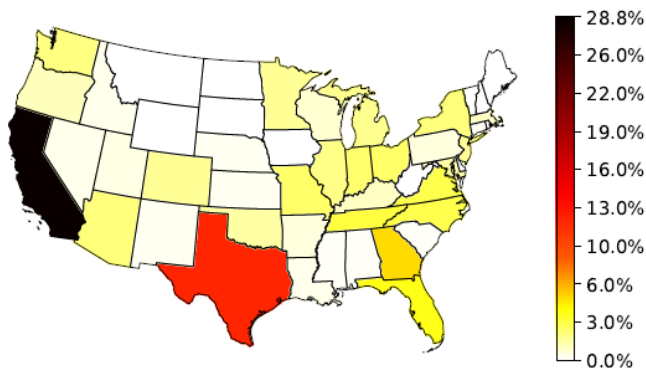
(b) Monthly dist. w/o the U.S.



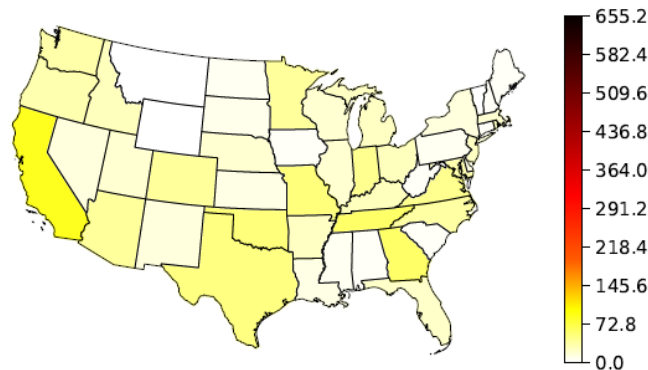
(c) Hourly dist. in the U.S.

- The United States accounts for more than 82% of all collected tweets.
- Usage peaks in summer and drops in winter.
- New Zealand & Australia ↓, Germany & India ↑, and United Kingdom & France →.
- The most active time during weekdays (weekends) is between 10:00 am to 5:00 pm (12:00 pm and 7:00 pm).

Geospatial Distribution



(a) Percentage by state



(b) Normalized by population

- California (28.8%) and Texas (11.7%) account for more than 40% of all tweets.
- Six (CA, TX, GA, FL, NC, OH) out of the top ten states with the highest percentages were among the ten most populous states.
- After normalizing by state population, we obtained a relatively smooth distribution.

Topic Discovery

- Data preprocessing
 - ❑ Each tweet content is a single document.
 - ❑ Corrected misspelled words, removed stop words, tokenized words, and lemmatized words.
 - ❑ Performed the term frequency inverse document frequency (TF-IDF).
- Latent Dirichlet Allocation (LDA)
 - ❑ # of topics = 12 achieving the highest coherence score
 - ❑ We clustered the 12 topics into four categories
 - Deployment
 - Stakeholder
 - Operation
 - Emotion

Topic Discovery

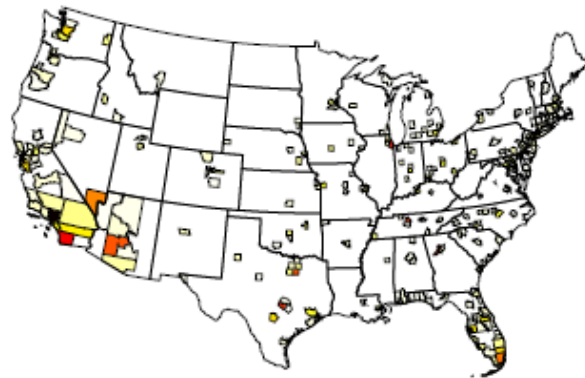
TABLE I: The extracted topics using the LDA topic model

	Stakeholder			Emotion		Deployment			Operation			
Rank	Rider	Gig Worker	Company	Positive	Negative	Transport.	City	Regulation	Parking	Transaction	Injury	Product
1	kid	gig	electr	ride	ride	bike	santa	citi	sidewalk	app	injuri	balanc
2	kick	worker	startup	around	one	share	monica	program	park	ride	fire	self
3	adjust	built	compani	fun	got	transport	st	compani	bike	charg	rider	electr
4	wheel	contractor	san	day	saw	car	loui	pilot	peopl	use	accid	bo
5	height	gen	market	love	shit	citi	montreal	electr	road	free	recal	skateboard
6	amp	economi	via	downtown	time	transit	paul	new	lane	unlock	caus	smart
7	child	lemon	new	san	fuck	trip	toronto	council	use	code	man	board
8	light	kmh	launch	time	guy	use	joe	bike	block	tri	injur	inch
9	adult	stabl	ford	rode	hit	mobil	canal	share	ride	one	via	fold
10	boy	libbi	tech	one	someone	electr	german	come	pedestrian	minut	report	bluetooth

Deployment in Cities



(a) Exact (*lat, lon*) coordinates



(b) Geospatial distr. by county

- 3359 exact GPS coordinates located in 579 cities.
- Most extensive scooter deployments occurred in large cities at East Coast and West Coast, and other metropolises.
- Three cities (Los Angeles, San Francisco, San Diego) in California contributed more than 15.6% of all GPS-tagged tweet.

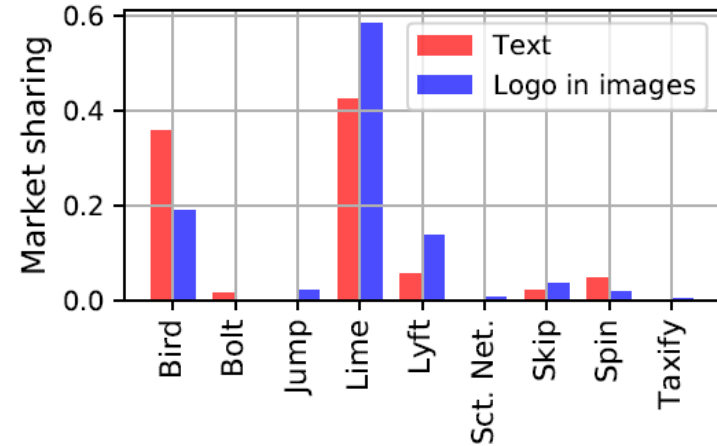
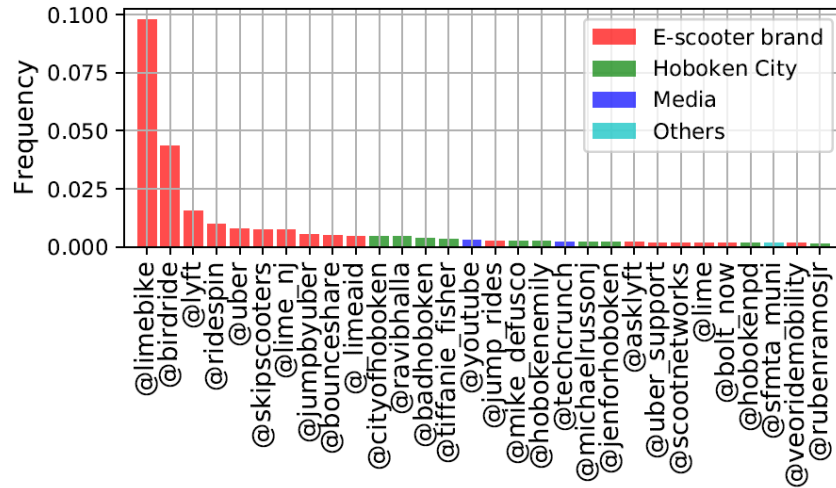
Policies and Regulations

- We surveyed policies on shared e-scooters in ten cities generating the most GPS data.
- Common rules
 - ☐ Self-protection requirements, e.g., wearing helmets
 - ☐ Riding behaviors, e.g., no phone usage while riding
 - ☐ Traffic restrictions, e.g., reasonable max speeds from 15 mph to 30 mph
 - ☐ Parking rules, e.g., not blocking sidewalks

Rider Profiles

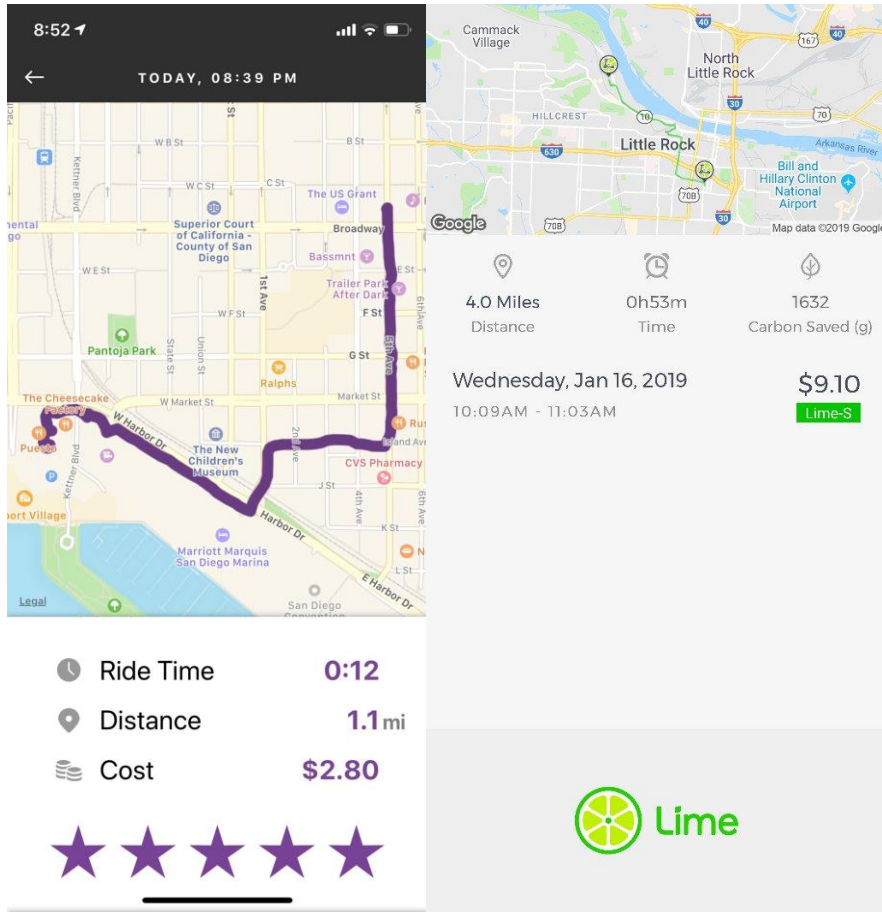
- A great gender gap in shared e-scooters with 34.86% identified as female and 65.14% as male, being consistent with a study [4] by Portland State University, which reported 34% riders identified as a woman, 64% as a man, and 2% as transgender or non-binary.
- Only 4.17% riders were recognized as kids.
- 83.51% users did not wear a helmet when riding, which might be one of the most common risky behaviors.

Scooter Companies



- All the top 10 most frequent mentions are e-scooter brands.
- The top 3 of market shares based on tweet text and recognized logos, i.e., Lime, Bird, and Lyft, agree with the top @mentioned accounts

Transaction Analysis



- Optical Character Recognition (OCR) to extract the trip information.
- Regular expressions to extract the payment and trip duration.
- Median payment and median duration were \$3.8 and 15.0 minutes.
- Average payment and duration were \$8.9 and 44.3 minutes with standard deviations of 14.6 and 95.7.

Self-reported Injury

TABLE II: Self-reported injury categories

Head (22.88%)					Trunk & Hands (27.45%)					Leg & Feet (49.67%)				
Chin	Eye	Mouth	Nose	Others	Arm	Elbow	Finger	Hand	Others	Ankle	Heel	Knee	Thigh	Others
5.23%	7.19%	5.88%	5.88%	11.76%	8.50%	5.88%	2.61%	11.76%	0.65%	11.11%	0.65%	24.84%	3.92%	10.46%

- **Knee protective gears** are required because of their highest injury frequency in all body parts.
- **Fingerless gloves** can be a good choice for riders to avoid hand bruises, the second most common injuries, and enable touching smartphone screens at the same time.
- **A helmet with chin protection** is a must because over half of the reported chin wounds were very serious.

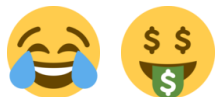
Parking Behaviors

- 37.39% e-scooters were docked at right places properly, and the rest **62.61% were in wrong places.**
- Among those e-scooters parked improperly, 34.78% were parked in the middle of sidewalks; 4.78% were placed indoors; 5.65% were vandalized; and 17.39% were parked in other wrong areas.
- **Blocking sidewalks** was the most common improper e-scooter parking behavior.

Emotion Analysis

- **Emoji:** 62.6% positive, 16.2% neutral, and 21.2% negative

☐ Positive



☐ Negative



- **Emoticon:** the positive emotions override the negative ones

TABLE III: Emoticons by categories (numbers represent frequency)

Positive	:) 490	:3 467	:) 137	:-) 87	8) 79	:p 35	;-) 33	:b 16	*) 9	:* 8	=) 7	:^ 6	x-p 6	:-)) 4	:] 3
Negative	:/ 325	:(321	:\ 245	:- (22	:o 18	:/ 14	:\$ 10	:c 9	8-0 6	:[3	:{ 2	=\ 2	:-c 0	:-i 0	:j 0

Conclusion

- Leveraged massive volumes of heterogeneous Twitter data, including text, @mentions, GPS data, general photos, screenshots of e-scooter apps, emojis, and emoticons, to study e-scooter ridesharing services on a large scale.
- After performing a comprehensive data preprocessing to remove noise and reduce false positives, we summarized 12 popular topics using the LDA topic model.
- For each of the extracted topics, we reported the profound insights and patterns, such as the popularity in different cities, the gender gap of riders, e-scooter market shares, transaction information, injury types, parking behaviors, and emotions from the public.

Reference

1. ODOXA, New modes of travel are disrupting transport in Ile-de-France, 2019. [Online]. Available: <https://shorturl.at/nBLUO>
2. N. A. of City Transportation, 84 Million Trips Taken on Shared Bikes and Scooters Across the U.S. in 2018, 2020. [Online]. Available: <https://nacto.org/shared-micromobility-2018/>
3. N. Ljubešić and D. Fišer, “A global analysis of emoji usage,” in Proceedings of the 10th Web as Corpus Workshop, 2016, pp. 82–89.
4. J. Dill, The e-scooter gender gap, 2020. [Online]. Available: <https://jenniferdill.net/2019/02/01/the-e-scooter-gender-gap/>

Thank you!
Q&A