Chasing Total Solar Eclipses on Twitter: Big Social Data Analytics in Once-in-a-lifetime Events

Yunhe Feng[*], Zheng Lu[*], Zhonghua Zheng[†], Peng Sun[*‡], Wenjun Zhou[*], Ran Huang[§], and Qing Cao[*]
[*]University of Tennessee, [†]University of Illinois at Urbana-Champaign, [‡]Zhejiang University, [§]Indiana University Bloomington
Email: [*]{yunhefeng, zlu12, wzhou4, cao}@utk.edu, [†]zzheng25@illinois.edu, [‡]sunpengzju@zju.edu.cn, [§]huangran@iu.edu

***Abstract*—With the popularity of social networking services, big social data analytics emerged in various applications, such as discovering trending topics, monitoring public sentiment, and identifying human mobility patterns. In this paper, we take the opportunity of *The 2017 Great American Eclipse*, a once-in-a-lifetime event, to look into its potential social, emotional, and human movement impacts at the national level. Specifically, we collected more than five million English eclipse-mentioning tweets in a real-time manner using Twitter Streaming APIs. Then we profiled spatio-temporal distributions of the data, extracted both hashtagged and latent topics, analyzed emotions using polarized words, emojis and emoticons, and revealed both interstate and intrastate eclipse-chasing travel patterns. Our study provides a comprehensive example of understanding big social data and its associated influence from diverse perspectives.**

***Index Terms*—social media, human behavior, online event tracking, big social data, social sentiment analysis**

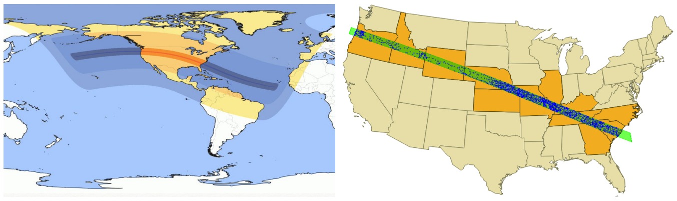
I. Introduction and Background

The rapid development of social networking services in recent years has boosted the growth of big social data analytics, a key emerging technique in understanding how people communicate and interact in social contexts. Diverse interdisciplinary fields, such as health care [1], [2], business intelligence [3], team sports [4], political science [5], and disaster management [6], are benefiting from social network based big data analysis and applications. However, the big data generated by social media users during once-in-a-lifetime events, which is a unique and important type of big social data, is still underexplored.

Few studies on similar topics have been conducted so far due to the following challenges. First, as the name implies, the once-in-a-lifetime event is extremely rare since it only occurs once or very limited times during the whole human life. Its appearing rarity means a low chance of being observed. Second, only a small portion of such events involve large-scale populations and regions at the same time. However, the population and the region are two key factors in generating the big representative data. Finally, it is indispensable that people are willing to publicly share such experiences online, making it possible to collect massive amounts of data for further study.

A natural phenomenon of *The Great American Eclipse* happened on Aug 21, 2017 (see Figure 1) offered a perfect chance to study how the extremely rare event influenced the people's life in large scale. This coast-to-coast total solar eclipse crossed the continental United States (U.S.) for the first time since 1918, and the next one with a similar path will occur in 2024. This rare and spectacular event attracted wide attention from the entire continent. According to The Washington Post [7], nine in ten adults in the U.S. watched this total eclipse. In addition, the eclipse related topics went viral on social networks before, during, and after this event.

DOI: 10.1109/GLOBECOM38437.2019.9013304 [Preprint Version]



(a) Across the world [8]

(b) Across the U.S. [9]

Fig. 1. Path of the eclipse shadow across the world and United States. The narrow tracks are called the path of totality. Small blue dots inside the narrow track in (b) refer to U.S. cities. The highlighted states in (b) represent areas where total solar eclipse was visible.

To chase this total solar eclipse online, we monitored and tracked English tweets mentioning the word "eclipse" via the Twitter Streaming APIs in a real-time manner. Our collected data satisfied the 5 Vs (volume, variety, velocity, veracity, and value) of big data completely [10]. *Volume*: more than 5.97 million tweets posted by 5.4 million Twitter users worldwide were collected in total within three and a half days. *Variety*: diverse types of Twitter data including structured data (e.g., timestamps, emojis, emoticons) and unstructured data (e.g., plain text, #hashtags, and geolocations) were contained in our dataset. *Velocity*: Twitter users generated more than 2500 tweets per minute on average when the eclipse occurred. *Veracity*: we preprocessed the data by detecting and removing tweets created by bots to ensure data quality. *Value*: the solar eclipse data enabled the first investigation (to our best knowledge) of how once-in-a-lifetime events influence human behaviors in large scale.

In this paper, we first visualized the spatio-temporal distributions of collected data to offer an overall understanding. Then we extracted popular topics using both social media exclusive tools (i.e., #hashtags and @mentions) and general topic models. We also conducted a comprehensive social sentiment analysis via polarized words, facial emojis, and emoticons to measure the general public's emotions and feelings. Finally, both interstate and intrastate human movement patterns in the U.S. were explored. We summarized key findings as follows:

- People all over the world (from 168 countries) chased this eclipse on Twitter, although it mainly occurred in the U.S.
- The event spawned various online topics covering business, science, health, politics, education, and entertainment fields.

- While happiness dominated the public's emotions, some people felt frustrated due to missing the solar eclipse.
- Besides those who lived in large cities, people staying in the path of totality geotagged tweets very intensively.
- To enhance the outdoor watching experience, people tended to travel towards the nearby path of totality regardless of interstate or intrastate trips.

II. Dataset

In this section, we describe the data collection and give an overview of the spatio-temporal distribution of the data.

A. Data Collection & Preprocessing

We used Twitter's Streaming APIs, which enable developers to filter and collect real-time tweets, to crawl all the English tweets containing the word "eclipse" from the entire duration of the event (from Aug. 20, 2017 to Aug. 24, 2017). The collected tweets were formatted in JavaScript Object Notation (JSON) files with named attributes and associated values [11].

As the bots on Twitter emerge to be popular, detecting and removing tweets generated by bots are necessary. Inspired by the bot detection approach proposed by Ljubesic et al. [12], we conceived the two types of Twitter users as bots: (1) those who post more than 100 eclipse-tagged tweets per day; (2) those who post more than 25 eclipse-tagged tweets per day and the top three frequent posting intervals cover at least their 90% tweets. Thus, we removed 8314 tweets created by 20 bots. Finally, we put together 5.97 million tweets including original tweets, replies, retweets, and quoted tweets, and 1.17 million unique tweets after deleting retweets.

B. Time Distribution

We started to collect the data nearly one day prior to the eclipse occurrence and stopped it until the eclipse trend faded away three days later. The number of tweets per hour mentioning the "eclipse" is demonstrated in Figure 2, where the U.S. Central Time time stamps of key events during the solar eclipse are highlighted using different colors. Around 12 hours before the maximum eclipse, i.e., the last night before the eclipse, the eclipse related tweets began to go viral. The most extensive discussions (around 0.15 million tweets posted per hour) occurred during the total eclipse crossed the U.S., as shown by the highlighted bars in Figure 2. The solar eclipse lasted less than six hours, but the event-related topics kept being popular for additional at least six hours. The least active time slots were always between 2:00 am and 3:00 am in the U.S. Central Time as few people post tweets at the very early morning. The eclipse related tweets dwindled 48 hours later after the first partial eclipse can be seen on the earth.

C. Geospatial Distribution

Although this solar eclipse's path of totality was mainly across the U.S., it attracted worldwide attention. Figure 3(a) shows that the tweets mentioning the English word "eclipse" come from 168 countries, even including non-English speaking countries like China and Japan. It is not surprising that a majority (89.5%) of collected tweets come from the United States. Canada (3.4%) ranks as the second and was followed by Britain (2.1%). Any other country except the above three ones accounts for less than 0.4%.

Therefore, we focused on the U.S. to study the spatial distribution of eclipse-tagged tweets at the national level. Figure 3(b) illustrates the states with a large population, such as California, Texas, and New York, contribute most tweets. It is impressive that, after normalizing the number of tweets by the resident population per state [13], those states along the solar eclipse's path of totality became dominating, which indicates that on average people staying close to the path are more likely to share this event on Twitter.

III. Topic Discovery

In this section, we explore the popular eclipse relevant topics using #hashtags, @mentions, and topic models respectively.

A. Hashtags

The #hashtags are mainly used for indexing keywords or topics. According to Twitter, hashtagged words that become very popular are often trending topics. We observed and extracted more than 89,500 unique hashtags in our collected tweets. Unsurprisingly, most hashtags were describing *The 2017 Great American Eclipse* event directly. To be specific, 64 hashtags including individual or combined words of *#Eclipse*, *#Solar*, *#Total*, or *#2017*, accounted for over 43% of the frequencies of all hashtags. The rest popular hashtags spawned by the eclipse event are shown in the Figure 4(a). It is interesting to find political terms, such as *#Trump*, *#MAGA (Make America Great Again)*, and *#TrumpResign*, are very popular. In addition, location-related words like *#USA*, *#Oregon*, and *#Nashville* are frequently mentioned by Twitter users. The topic of science can also be inferred by *#Science*, *#Tech*, and *#Periscope*. The *#EXO*, a boy band which published a song titled *Eclipse*, is also mentioned by more than 2000 times.

B. Mentions

Another useful method to infer the topic is by identifying the mentioned accounts. The Twitter accounts @*YouTube*, @*RealDonaldTrump*, and @*NASA* account for the most, corresponding to the discussions in entertainments, politics, and science areas. Especially, @*YouTube* is repeatedly mentioned by Twitter users who either watched or uploaded the videos on YouTube concerning this event. The Twitter account owned by President Trump, @*RealDonaldTrump* is frequently mentioned, along with the news from the Washington Post reported that "*Trump celebrates solar eclipse by looking up without special viewing glasses*". Not surprisingly, @*NASA* is mentioned many times since this event is a rare astronomical phenomenon. On the other hand, the Twitter accounts owned by the news medias such as @*CNN* and @*FoxNews* are mentioned very often, addressing this social hot spot. We conclude that the top mentioned accounts are in agreement with the top Hashtags.

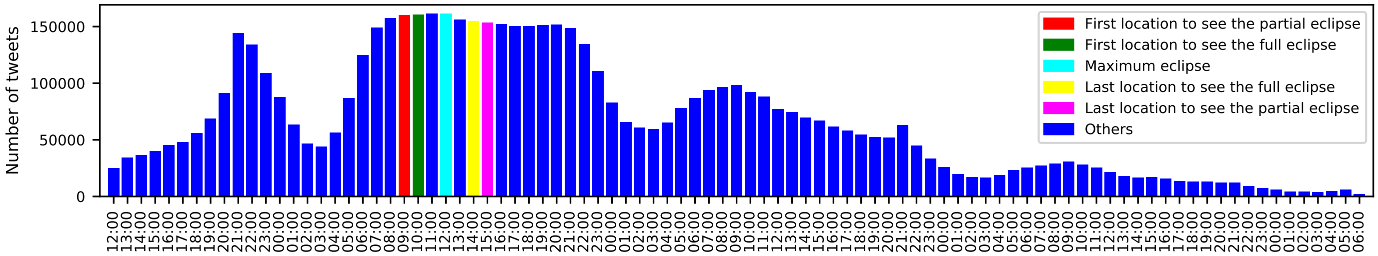
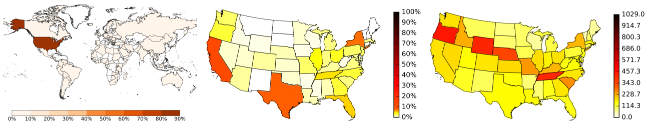
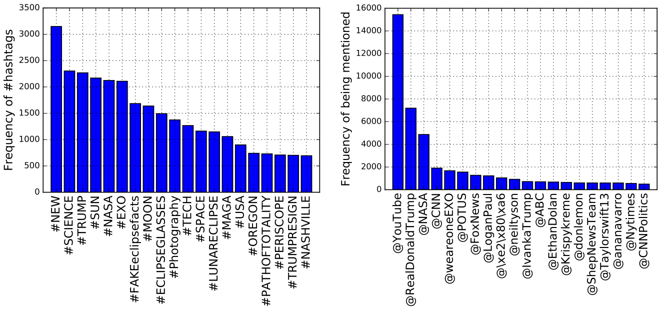


Fig. 2. Distribution by hour (U.S. Central Time). We collected the data from 12:00 pm, Aug 20 to 06:00 am, Aug 24.



(a) World distribution (b) Percentage by state (c) Normalized by POP

Fig. 3. Worldwide and national spatial distribution of tweets. (a) The U.S. contributes more than 89% tweets. (b) States with a large population contribute more. (c) States along the path of totality tweet more per million people.



(a) Hashtags (b) Mentioned Twitter accounts

Fig. 4. Hashtags and mentioned Twitter accounts

C. Topic Models

To establish a deeper understanding of eclipse-related topics on Twitter, we adopted latent Dirichlet allocation (LDA) [14], a widely used topic model, to learn hidden topics distributed among our collected data. Specifically, we treated each tweet content as a single document, and the whole collection of unique tweets as the corpus of documents. For each document, we corrected misspelled words, removed stop words and most frequently occurring words such as "eclipse", tokenized words, and lemmatized words during the text preprocessing. Then we combined all documents and perform the term frequency-inverse document frequency (TF-IDF) to create a TF-IDF based corpus. Finally, we applied the LDA model on the created corpus to extract latent topics.

One of the challenging tasks for LDA is to choose a suitable number of topics. We measured the topic coherence using C_v metric, which was reported as the best coherence measure by combining normalized pointwise mutual information (NPMI) and the cosine similarity [15], to decide the optimal number of topics in our case. We evaluated the topic numbers ranging from one to twenty with 500 passes, and found the topic number of eleven achieved the highest C_v score.

The eleven extracted hidden topics are reported in Table I, where we used brief words to summarize the inferred topics and added comments for further explanation. Since this total solar eclipse was a time-sensitive event in terms of both lasting time and occurring frequency, the topic of "time" was formatted naturally. The weather condition was another factor that might impact eclipse viewing, so it became a topic trend on Twitter. When watching solar eclipses, special devices such as glasses, pinhole projectors, and proper lens were required to prevent blindness. Instead of watching the eclipse outdoors, some people preferred to chase the eclipse via NASA streaming online videos. Along with videos, we detected the theme of music, which corresponded with the popularity of eclipse-related songs, e.g., Bonnie Tyler's *Total Eclipse of the Heart*, during the event [16]. It is interesting to discover the "education" topic consisting of words like "school," "safe," and "class," representing the extensive discussion of various school absence policies on the eclipse day.

In addition, the eclipse aroused the public interest in space science to learn how and why solar eclipses happened. This event also inspired people to discuss the *Fire Nation*, an American animated television series, which contained a major battle fought during a solar eclipse. Another two topics come from the business and politics domains respectively, because product deals (e.g., shirts and glasses) and the news that President Trump stared into solar eclipse without safety glasses went viral online. The last topic is about the public emotions which are further discussed in the next session.

IV. Sentiment Analysis

We analyze the emotions of Twitter users from three angles, i.e., polarized words, facial emojis, and emoticons.

A. Polarized Words

Regarding polarized words, we used several lexicons. First, we followed the Twitter-specific lexicons used in [17]. We found that the most common positive terms included "great" (n=25,036), "like" (n=11,078), "excellent" (n=515) and "rock on" (n=20), and the most common negative terms included "suck" (n=2479), "fail" (n=1460) and "eww" (n=53).

Then, to include more polarized words, we took advantage of the TextBlob [18]. The polarity score of TextBlob is a number in the range $[-1.0, 1.0]$, where -1.0 is the most negative, 1.0 is the most positive, and 0.0 is neutral. We used polarity thresholds 0.5 and -0.5 to identify terms with

TABLE I
THE 11 EXTRACTED TOPICS USING THE LDA TOPIC MODEL

Topic	Words	Comment
Time	year; next; last; wait; hour; time; night; minut; one; day; ago; tomorrow	A time-sensitive (once-in-a-lifetime) event
Weather	see; cloud; cool; cloudi; pretti; beauti; experi; rain; partial	Weather/watch conditions
Devices	glass; use; camera; box; viewer; pair; phone; watch; pinhol; filter; weld; len	Devices used for watching eclipses
Health hazards	look; eye; glass; stare; blind; directli; sun; protect; wear; damag; burn; hurt; sunglass	Eye protection
Video & music	live; watch; via; stream; video; nasa; bonni; tyler; playlist; song	Bonnie Tyler: a famous singer
Education	school; safe; class; parti; student; readi; fun; learn; babi; teacher; hope	Students watch eclipses
Popular science	moon; power; earth; shadow; sun; energi; flat; light; super; tree; crescent; shine	Scientific explanation for eclipses
TV series	fire; nation; star; inspir; throw; ring; attack; edit; watch; sunset; corona; footag; lord	Fire Nation: American animated series
Business	deal; medium; social; shirt; product; blast; glass; hous; account	Deals and sales
Political news	trump; news; fake; presid; donald; report; racist; cnn; articl; trend; blame; speech; hair	News about the President of the U.S.
Emotion	like; shit; fuck; realli; lol; damn; excit; talk; hype; feel	Positive and negative emotions

clear polarity. Therefore, a term is considered positive if its TextBlob polarity score is above 0.5, and negative if its TextBlob polarity score is below -0.5. Figure 5 visualizes the polarized words using word clouds. When expressing both positive and negative emotions, Twitter users prefer certain words such as “great,” “good,” “f**king,” and “bad” above all other terms. It is interesting to note that initialisms like “LOL” (laughing out loud) and “LMAO” (laughing my ass off) are also frequently used to express emotions. Figure 5(c) shows that positive emotions dominate the whole polarized words.

B. Facial Emojis

The facial emojis are officially categorized as positive, neutral, and negative sentimental groups by the Unicode Consortium. We summarize the top 15 most frequent facial emojis in each group (65% positive, 21% neutral, and 14% negative) in Table II. The most widely used emoji in this study is the face with tears of joy 😂, the most used emoji globally, which is consistent with many other research findings [19]. The second and third popular emojis are also positive. It is worthwhile to note that both of them wearing either sunglasses 😎 or “heart” 😍. Considering this event required people to wear special glasses or lenses to protect eyes, we suspect that people used these emojis heavily not only because of the popularity of these emojis, but also the suitability to be in accordance with the context of the solar eclipse. When missing the eclipse due to the cloudy weather conditions or personal reasons, Twitter users tended to post negative emojis to express their depressions and even furies.

C. Emoticon

Although the emojis are gradually taking over the emoticons on social networks, the emoticons are still ubiquitous because of their simplicity and platform independence. We summarized the usage of all positive and negative emoticons in the list of sideways Latin-only emoticons [20], as shown in Table III. The most popular positive emoticons in this study are “:3” ($n = 4916$), “:)” ($n = 4057$), and “;)” ($n = 949$), while the most popular negative ones are expressed by“:/” ($n = 2514$) and “:-(” ($n = 2294$). Similar to polarized words, Twitter users preferred a limited set of both positive and negative emoticons. For example, the total count of “:3” and “:)” is greater than the remaining top six positive emoticons, and the total count of “:/” and “:(” overwhelm other negative emoticons. Correspondingly, the main pattern of emoticons (71% positive and 29% negative) is in line with the facial emojis, where most people were excited about this event, while some were bothered by the terrible traffic or missing this event.

V. Human Mobility

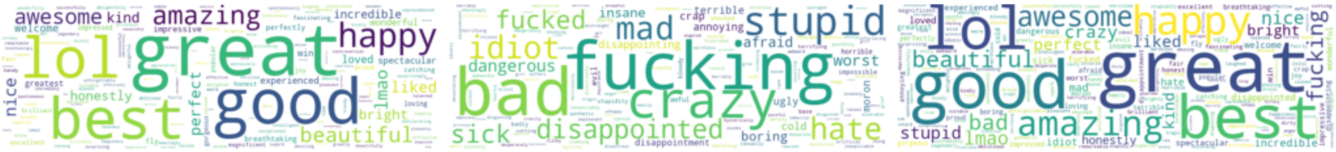
During *The 2017 Great American Eclipse*, it was reported a million people traveled to Oregon to chase this natural event in the path of totality. In this section, we analyze human travel behaviors based on two types of Twitter geographical information, namely, profile locations and tweet locations.

A. Profile Location

The profile location refers to the residential address where Twitter users specified in their public account profiles. Since Twitter users are allowed to use free-form text to describe profile locations, it is challenging to extract precise geographical information from such messy, flexible and diverse profile locations using rule-based approaches. Instead, we resorted to Nominatim [21], a search engine for OpenStreetMap data, to infer the most likely locations by submitting queries via Nominatim online APIs. Places at different levels of detail (e.g., city, county, and street) were fetched through Nominatim depending on the address precision in the user profile. For example, the exact latitude/longitude point coordinate of a coffee house titled “Precision Pours” at Louisville, Colorado can be determined by querying the detailed location of “1030 E South Boulder Rd”. Note that some account locations, such as “17th & Fontain”, “Anywhere, Earth”, and “From old houses, to old towns”, are too casual to be geo-referenced.

B. Tweet Location

When posting tweets, Twitter users can geo-tag their tweets optionally. The tweet location can be either a box polygon of coordinates defining general areas like cities and neighborhoods, or an exact GPS latitude and longitude coordinate. Both of the two types of locations enable us to identify the real-time location of users when tweeting. Figure 6 shows the distribution of tweets geo-tagged with exact latitude and longitude coordinates. In our dataset, most of such tweets are posted either along the eclipse’s path of totality, or from densely populated largest cities such as New York City, Boston, Washington, D.C., Seattle, San Francisco, and Los



(a) Positive word cloud (b) Negative word cloud (c) Positive and negative word cloud

Fig. 5. Polarized word cloud. Positive words with a polarity larger than 0.5 and negative words with a polarity less than -0.5 in TextBlob.

TABLE II
FACIAL EMOJI USAGE IN ECLIPSE-RELATED TWEETS

Emoji Type	Emoji Count														
Positive	😂 4644	😎 3858	😍 1241	😊 697	😉 566	😁 480	🤣 373	😃 317	😅 288	😆 212	🤗 211	🙂 192	😀 189	😄 189	😘 165
Neutral	🤔 1118	🙄 765	😒 381	🙃 351	😜 314	😏 238	😐 216	😑 195	😕 148	😔 141	😮 139	😲 121	😌 108	😫 102	😴 97
Negative	😭 713	😱 426	😳 376	😩 373	😤 189	😬 169	😢 154	😡 106	😞 95	😵 84	☹️ 58	😟 53	😨 47	😰 39	🙁 28

TABLE III
EMOTICONS USAGE IN ECLIPSE-RELATED TWEETS

Type	Emoticon Count														
Positive	:3 4916	:) 4057	;) 949	:-) 701	:p 362	;-) 227	8) 161	=) 67	d: 59	:^) 46	:b 41	:-)) 33	:* 33	=p 30	:-p 25
Negative	:/ 2514	:(2294	:o 160	:-(144	:c 126	:'(77	:-/ 45	:\ 30	:[22	=/ 18	:$ 15	8-0 12	:-o 9	:{ 5	=\ 3

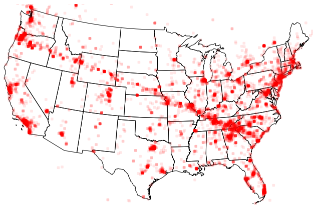


Fig. 6. Distribution of tweets with exact latitude and longitude coordinates.

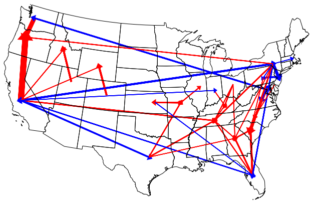


Fig. 7. The 50 most popular interstate trips during the event.

Angeles. The dense distributions in less populated states such as Idaho, Wyoming, and Missouri inspired us to study how people travel between states. On the other hand, the overcrowded distributions in Tennessee, Oregon, South Carolina, and Georgia where the path of totality crossed motivated us to look deeper into people's movements inside states.

TABLE IV
THE 20 MOST FREQUENT INTERSTATE MOVEMENTS

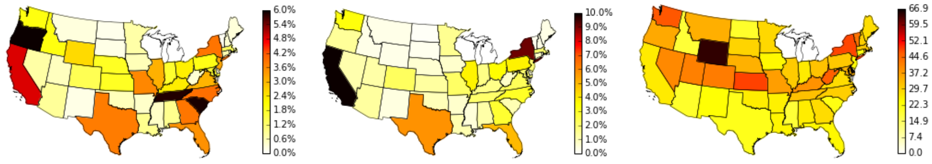
From	To	Count	From	To	Count
California	Oregon	252	Pennsylvania	New Jersey	71
Washington	Oregon	176	D.C.	Washington	66
North Carolina	S. Carolina	131	California	Texas	66
New York	New Jersey	122	Florida	Georgia	62
D.C.	Virginia	101	Missouri	Kansas	60
D.C.	Maryland	84	New York	S. Carolina	57
Florida	S. Carolina	80	New York	Florida	55
New York	California	79	New York	Tennessee	54
Colorado	Wyoming	78	Georgia	S. Carolina	52
Utah	Idaho	77	Texas	Tennessee	52

C. Interstate Movement

We treated Twitter users' profile locations as their places of residence. If Twitter users posted eclipse-related tweets outside the state of residence, we assumed they traveled between states to chase the eclipse. Table IV illustrates the 20 most popular interstate trips for chasing eclipse. The destinations of red trips are states where the total solar eclipse occurred. On the contrary, the blue ones represent cross-state travel towards states where the only partial eclipse was visible. It is interesting to note that the top three blue trajectories, i.e., New York → New Jersey, D.C. → Virginia, and D.C. → Maryland, are more likely to be generated by commuters because these pairs of states are so close to each other that people may even live and work in two different states.

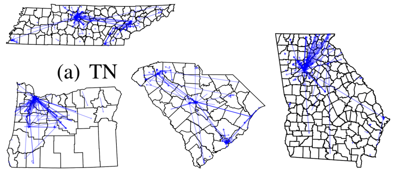
We then visualized the 50 most frequent interstate trips, as shown in Figure 7. The red arrows represent the trajectories to states through which the eclipse's path of totality crossed. The blue arrows are the trajectories to states where the total eclipse did not occur. The width of the arrows indicts the frequencies of travels. The red arrows dominate the blue ones greatly, suggesting people did prefer states where the totality can be observed. However, the red coast-to-coast movements are less than the blue ones, which infers that most people probably chase the eclipse in nearby states instead of far-away states.

Finally, we summarized all 10,334 interstate trips of Twitter users during the event with respect to destination states, origin states, and interstate travel likelihoods. Figure 8(a) shows that all of the three most attractive destinations, i.e., Oregon, South Carolina, and Tennessee, are located in the path of totality, where people can experience the totality. The eight most popular origin states except Washington, D.C. have the largest populations, as demonstrated in Figure 8(b). However, after normalizing the number of interstate trips by state populations, we find the two most active origin states are Delaware and



(a) Destination states (b) Origin states (c) # trips per million people

Fig. 8. Interstate movement



(a) TN (b) OR (c) SC (d) GA

Fig. 9. Intrastate movement

Wyoming which have the seventh least and the first least population among all states (see the number of people taking interstate trips per million people in Figure 8(c)). In addition, it seems that people living in the northern parts of the U.S. made more frequent interstate trips on average during this event.

D. Intrastate Movement

To further investigate the chasing behaviors at a detailed level, we select four states, i.e., Tennessee (TN), Oregon (OR), South Carolina (SC), and Georgia (GA) along the eclipse totality path to study human intrastate movements. As shown in Figure 9, people also sought to find a suitable place to watch the eclipse inside states for convenience. In all of the four states, most of the starting locations are from major cities. To be more specific, Twitter users living in Nashville, Knoxville, Memphis and Chattanooga in Tennessee, Portland and Eugene in Oregon, Atlanta in Georgia, and Columbia, Charleston, and Greenville in South Carolina, were more likely to take a trip to chase the eclipse. Directions of intrastate movements are generally towards the path of totality. For example, Twitter users in Atlanta trended to move to the northeast board of Georgia where the total eclipse occurred. It should be noted that moving ranges of all three major cities in South Carolina are much smaller than cities in other states, which could be explained by the fact that all of the three cities were located in the path of totality.

VI. Conclusion

In this paper, we leveraged massive volumes of heterogeneous Twitter data to study an important but underexplored type of big social data, the once-in-a-life event, in large scale for the first time. Nearly six million English eclipse-mentioning tweets generated by 5.4 million Twitter users were collected during the once-in-a-life *The 2017 Great American Eclipse*. We demonstrated the capability of big social data analytics to investigate this particular event's potential social, emotional, and travel impacts on human activities. Specifically, we utilized #hashtags, @mentioned accounts, and topic models to extract and summarize popular topics spawned by the total solar eclipse. By means of sentiment analysis, we found that while most people were excited about this natural phenomenon, there were also complaints because people missed this rare event or encountered with traffic jam. Furthermore, according to the analysis of human mobility, we investigated that people chased the eclipse through both interstate and intrastate trips to gain better watching experience. Our study provides a comprehensive example of understanding big social data and its associated influence from diverse perspectives.

References

[1] T. Huang, L. Lan, X. Fang, P. An, J. Min, and F. Wang, "Promises and challenges of big data computing in health sciences," *Big Data Research*, vol. 2, no. 1, pp. 2–11, 2015.

[2] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2, no. 1, p. 3, 2014.

[3] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact." *MIS quarterly*, vol. 36, no. 4, 2012.

[4] D. Lusher, G. Robins, and P. Kremer, "The application of social network analysis to team sports," *Measurement in physical education and exercise science*, vol. 14, no. 4, pp. 211–224, 2010.

[5] E. Colleoni, A. Rozza, and A. Arvidsson, "Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data," *Journal of communication*, vol. 64, no. 2, pp. 317–332, 2014.

[6] J. R. Ragini, P. R. Anand, and V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis," *International Journal of Information Management*, vol. 42, pp. 13–24, 2018.

[7] B. Guarino, *Wildlife fled the sky and bees went silent during the 2017 total solar eclipse*, 2018. [Online]. Available: https://wapo.st/2Ok4G2D

[8] TimeAndDate, *August 21, 2017 — Great American Eclipse (Total Solar Eclipse)*, 2017. [Online]. Available: https://www.timeanddate.com/eclipse/solar/2017-august-21

[9] R. E. A. Jr, *US Cities in Path of Total Solar Eclipse*, 2017. [Online]. Available: http://robslink.com/SAS/democd94/eclipse_2017.htm

[10] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2013, pp. 48–55.

[11] Twitter, *Introduction to Tweet JSON*, 2018. [Online]. Available: https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/intro-to-tweet-json

[12] N. Ljubešić and D. Fišer, "A global analysis of emoji usage," in *Proceedings of the 10th Web as Corpus Workshop*, 2016, pp. 82–89.

[13] W. P. Review, *United States Population 2018*, 2017. [Online]. Available: http://worldpopulationreview.com/countries/united-states-population/

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[15] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proceedings of the eighth ACM international conference on Web search and data mining*. ACM, 2015, pp. 399–408.

[16] D. Cassel, *Eclipse 2017 Mania: All the Data*. [Online]. Available: https://thenewstack.io/eclipse-mania-data

[17] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi, "Bad news travel fast: A content-based analysis of interestingness on twitter," in *Proceedings of the 3rd International Web Science Conference*. ACM, 2011, p. 8.

[18] S. Loria, P. Keen, M. Honnibal, R. Yankovsky, D. Karesh, E. Dempsey *et al.*, "Textblob: simplified text processing," *Secondary TextBlob: Simplified Text Processing*, 2014.

[19] Dictionary.com, *What does the face with tears of joy emoji mean*, 2017. [Online]. Available: http://www.dictionary.com/e/emoji/face-with-tears-of-joy-emoji/

[20] Wikipedia, *Sideways Latin-only emoticons*, 2018. [Online]. Available: https://en.wikipedia.org/wiki/List_of_emoticons

[21] OpenStreetMap, *Nominatim: a search engine for OpenStreetMap data*, 2018. [Online]. Available: https://wiki.openstreetmap.org/wiki/Nominatim