Preparing for 2024 Total Solar Eclipse: Social Media Data Mining for Understanding Spectator Experience in Southern Illinois

Ruopu Li*, Joseph Kalinzi

Geography and Environmental Resources, School of Earth Systems and Sustainability, Southern Illinois University, Carbondale, IL 62901

Phone: (618) 453-6038; Fax: (618) 453-6465; Email: ruopu.li@siu.edu

* Correspondence directed to Ruopu Li. Mailing Address: 1000 Faner Drive, Carbondale, IL Tel: 618-453-6038, Email: Ruopu.Li@siu.edu

Abstract

This study analyzes visitor dynamics during the 2017 Total Solar Eclipse in Southern Illinois using data from X (formerly known as Twitter). Focusing on spatial and temporal patterns, we identified key visitor clusters and sentiments. The majority of visitors originated from Chicago, IL, Nashville, TN, and St. Louis, MO. Findings revealed concentrated activities in specific locations, with generally positive experiences shared on social media. Insights gained will aid in planning for the upcoming 2024 eclipse, enhancing visitor experiences and economic benefits for the region. This research underscores the value of social media data in understanding and managing large-scale events in rural areas.

Keywords: Total Solar Eclipse; Social Media; Mobility; Twitter; X; Southern Illinois

1. Questions

Southern Illinois holds a unique distinction as the only place in the world that intersects the paths of two Total Solar Eclipse (TSE) events, one that occurred in 2017 and the next in April 2024. The 2017 eclipse attracted an unprecedented 50,000 ~ 100,000 spectators to the region (Robaugh and Staff, 2017). This surge of visitors significantly boosted tourism revenue and economic prospects for rural cities and towns in the area. While the event was economically beneficial, the influx of visitors placed substantial strain on rural infrastructure and services, leading to challenges like traffic congestion, overloaded cellular networks, and stretched public health and safety resources.

To better prepare for the upcoming 2024 TSE, understanding the patterns of these visitors is crucial. Detailed insights into their arrival and departure times, origins, accommodation choices, dining and shopping habits, and overall experiences are invaluable. Such data will assist the region in enhancing its tourism infrastructure and maximizing economic advantages for the next eclipse. In recent years, social media applications have flourished, serving as social sensors that detect human dynamics, a phenomenon extensively studied (Hamstead et al., 2018; Li et al., 2019; Galesic et al., 2021). However, the bulk of these studies predominantly focus on metropolitan areas, where social media user density is high. In contrast, human dynamics in rural areas have been less explored through the lens of social media due to the data scarcity.

We hypothesize that the 2017 TSE presented an exceptional opportunity to study data-rich human dynamics during a period of significant tourist influx. The purpose of this study is to examine the spatial and temporal patterns of the spectator crowds and their experiences during the 2017 TSE. The results are expected to inform the strategies for improved visitor experiences in the upcoming 2024 TSE event.

2. Methods

The study area focuses on Jackson and Williamson Counties in Southern Illinois with a large state university and an approximate population of 120,000. We selected X, previously known as Twitter, as our primary source for social media data, a resource extensively utilized in tracking human dynamics in previous studies (Hamstead et al. 2018; Liu et al. 2019). We used Tweepy Python library to retrieve geo-tagged tweets within the geographic boundary of the study area via the Twitter API v2 endpoints, provided through Twitter Academic Research Access. In our study, we extracted and analyzed specific data points from each tweet: geolocations, time stamps, and discussion contents. TextBlob, a Python Natural Language Processing (NLP) library, was used to parse the tweet contents and eliminate the stopwords and compute the frequency of conversational words to determine popular topics. The sentiment polarity was obtained using the Textblob sentiment.polarity module which is based on a Naïve Bayes classification algorithm to evaluate each tweet's sentiment scores. As X account users include local residents, visitors, and those undetermined, only those people with X profile locations outside the area of interest were labeled as visitors. To pinpoint areas with tweet hotspots, we mapped the geotagged tweets and applied the Kernel Density tool using ArcGIS Pro. The results of our data mining were empirically verified by contacting local entities, including businesses like vineyards and restaurants, governmental bodies, and non-profit organizations.

3. Findings

We gathered 503 geotagged tweets from the study area. Compared with the estimated total visits, we acknowledge that this sample may not represent the entire spectators. Figure 1 shows the tweet frequency over a one-week period. Notably, the daily volume of location-tagged tweets remained relatively low until it surged on the day preceding the TSE, reaching its peak the day of the eclipse,

before diminishing rapidly. This pattern suggests that most tourists visited the region within a narrow three-day window centered around the event. For most days, evenings witnessed the highest volumes of tweets, suggesting active evening gatherings among those spectators.

Out of the 503 tweets, the vast majority displayed positive or neutral sentiments. About 9% of these tweets expressed highly positive sentiments. An example of such tweets is: "This place is pretty awesome. Taken a couple of minutes before things went dark @ Blue Sky Vineyards". ~ 90% of the tweets were categorized as neutral or slightly positive. Notably, only two tweets carried negative sentiments: "Wow. This theater smells bad.", and "Battled awful traffic and even worse heat and humidity to see the #totaleclipse but it was so?"

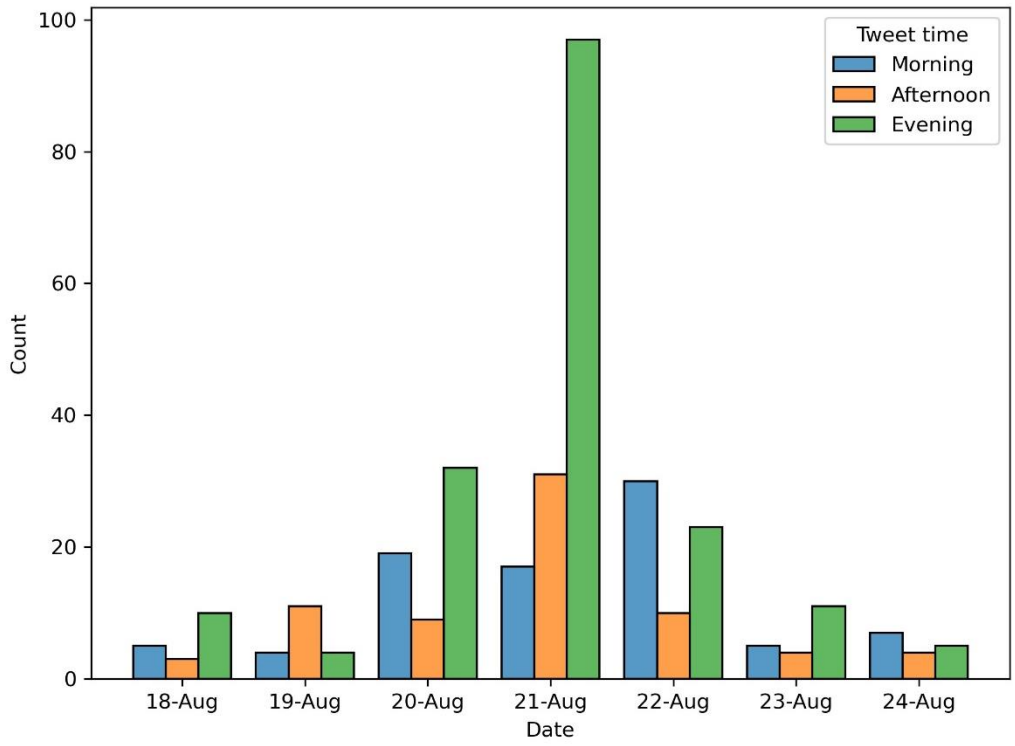

Figure 1 A histogram showing the frequency of tweets by morning (12am-noon), afternoon (noon-6pm), and evening (6pm-12am) within a one-week period (August 21 is the TSE day)

Our spatial analysis revealed the locations where spectators were most concentrated during the 2017 TSE. We identified seven primary spatial clusters: Blue Sky Vineyard, Walker's Bluff Vineyard, downtown Carbondale, the SIU Campus, Rainmaker Art Studio (Makanda), City of Herrin, and Village of Goreville. These locations were often associated with TSE events, such as Moonstock Music Festival held at Walker's Bluff vineyard, the spectator event hosted at Saluki Stadium at SIU, and a similar event at Blue Sky Vineyard. Additionally, our findings indicated that tourists also frequented restaurants near Marion's strip mall and the University Mall.

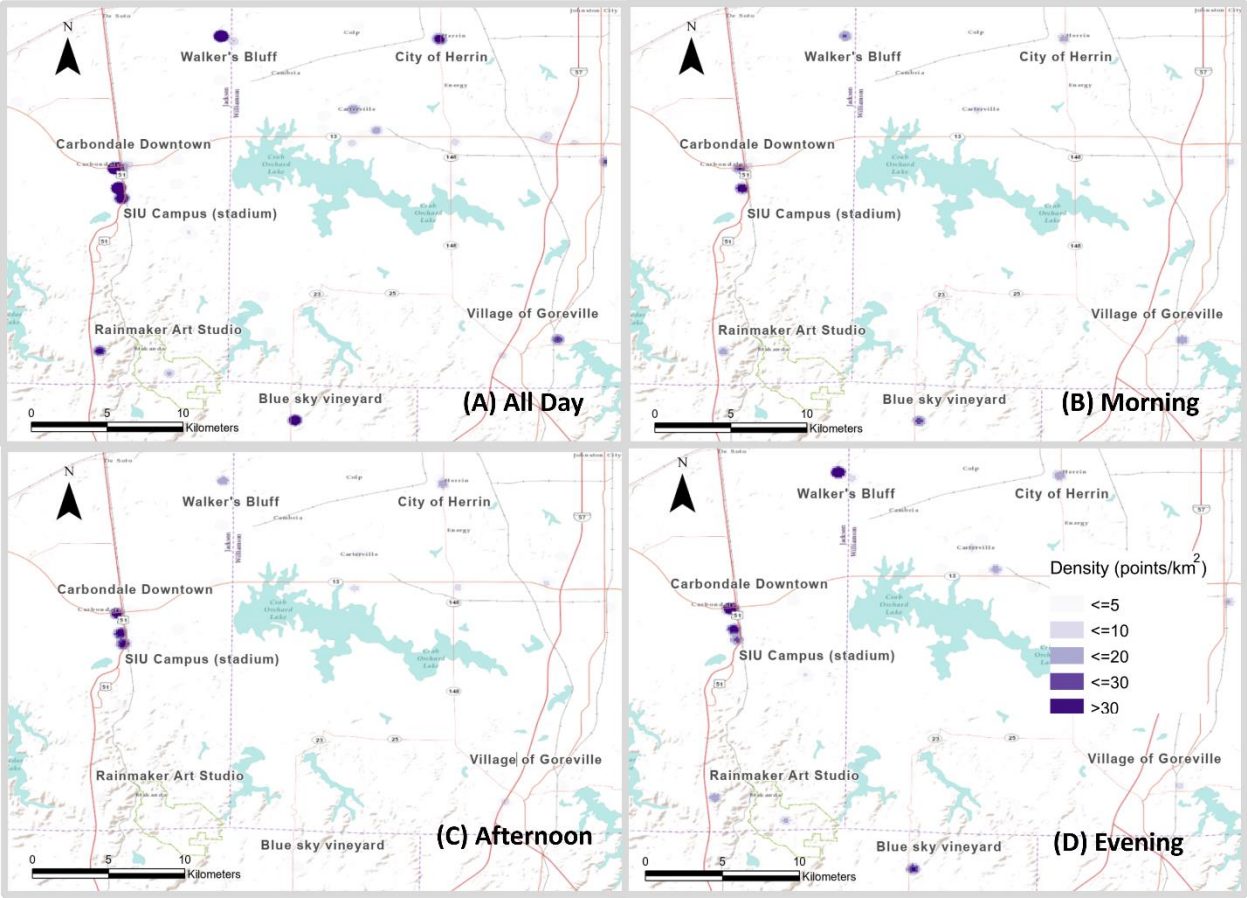

Figure 2 The hotspots of geotagged tweets on August 21

An analysis of those spectators' X profiles indicates that the majority traveled from the Chicago area (Figure 3). Nashville, TN and St. Louis, MO ranked second. These results suggest that advertising campaigns may focus on key cities to attract tourists.

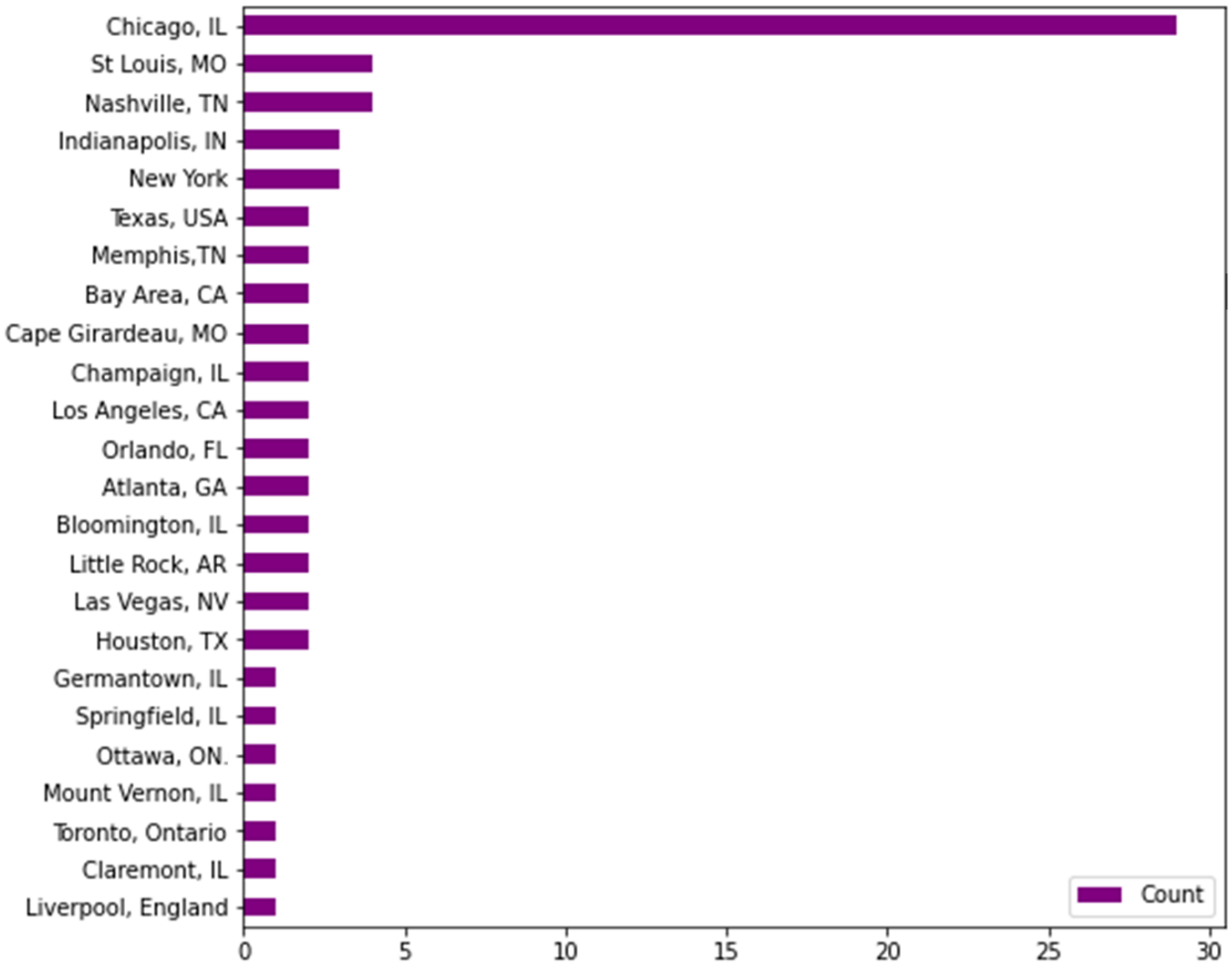

Figure 3 The location origins of spectators who visited southern Illinois during the 2017 TSE

Based on the major spatial clusters identified (Figure 2), we reached out to ~ 20 agencies and businesses at these clusters. Their accounts provide insights into visitor experiences during the 2017 TSE and for the next event. For example, Saluki Stadium witnessed around 30,000 visitors, selling 13,284 tickets and 4,571 parking passes. Rainmaker Art Studio in Makanda saw visitor

numbers around 10,000 ~ 15,000, and is planning commemorative stickers for the next TSE, though it faces planning challenges like parking and porta-potty rental. Walker's Bluff, hosting Moonstock 2017 with about 12,000 visitors, is planning for the next eclipse, hoping a new road will ease previous traffic issues. However, negative media coverage about overcrowding adversely affected their previous event.

Our research demonstrates that social media offers a novel lens for understanding spectator visitation patterns and experiences during the 2017 Total Solar Eclipse event. By analyzing spatial clusters and engaging directly with local agencies and businesses, we have gleaned valuable insights for future event planning, particularly for the upcoming 2024 TSE.

Acknowledgment

This study was funded by the SIU Foundation Board Research Award. We appreciate the access to historical tweets provided by Twitter's Academic Research Access. We also acknowledge the valuable contributions of Summer Stahl, a dearly missed undergraduate student.

References

Galesic, M., Bruine de Bruin, W., Dalege, J., et al. (2021). Human social sensing is an untapped resource for computational social science. Nature, 595(7867), 214–222. https://doi.org/10.1038/s41586-021-03649-2

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. Computers, Environment and Urban Systems, 72, 38-50. https://doi.org/10.1016/j.compenvurbsys.2018.01.007

Li, R., Crowe, D., Leifer, D., Zou, L., & Schoof, J. (2019). Beyond big data: Social media challenges and opportunities for understanding social perception of energy. Energy Research & Social Science, 56, 1012-1017. https://doi.org/10.1016/j.erss.2019.101217

Liu, X., Huang, Q., & Gao, S. (2019). Exploring the uncertainty of activity zone detection using digital footprints with multi-scaled DBSCAN. International Journal of Geographic Information Science, 33(6), 1196-1223. https://doi.org/10.1080/13658816.2019.1594824

Robaugh, D., & Staff, P. (2017). Solar eclipse in Carbondale, Illinois: When you can see total eclipse along path of totality. Patch. Retrieved from https://patch.com/illinois/across-il/solar-eclipse-2017-carbondale-illinois