ORIGINAL PAPER

Open Access

Social big data mining for the sustainable mobility and transport transition: findings from a large-scale cross-platform analysis

Michael Stiebe[1,2*]

Abstract

The paper reports findings from a study that examining how cross-platform social media analysis can help to map the digital discourse on sustainable mobility and sustainable transport, and enhance the understanding of sociotechnical low-carbon transport transitions. Using the hashtag search queries #sustainabletransport and #sustainablemobility, 33,121 Tweets (2013–2021) and 8,089 Instagram images including captions (2017/2018–2021) were scraped using the Python modules Twint and Instaloader. Quantitative text and sentiment analyses were applied to the Tweets and image captions. Additionally, an automated machine learning-based image analysis of the Instagram images was conducted using object detection via OpenCV. Synthesized results formed the base for a cross-platform analysis inspired by Rogers' method comprising hot topics/key themes, user mentions, sentiment polarity, and co-hashtags. Notably, electromobility emerged as a prominent theme, particularly on Instagram, while #sustainabletransport was closely associated with active travel, notably bicycling, and #sustainablemobility showcased a dominance of electromobility discourse. The study demonstrates the investigative potentials of cross-platform social media analysis studies to enhance the understanding of sociotechnical low-carbon transport transitions. Drawing on key results, the paper suggests an adapted version of the Geelsean Multi-Level Perspective on Sociotechnical Transitions.

Keywords Cross-platform analysis, Social media analysis, Instagram, Twitter, Social big data, Sustainable mobility, Sustainable transport, Sociotechnical transitions

1 Introduction

The formal concept of sustainable mobility was introduced about three decades ago in the 1992 EC Green Paper on the Impact of Transport and Environment drawing on the perspectives of the 1987 Brundtland Report (European [20, 26]). According to Haas et al. [25], the sustainable mobility transition is strongly influenced by four megatrends, namely climate change, digitization, urbanization, and extractivism. Furthermore, recent research identifies three Grand Narratives for sustainable mobility, 1) low mobility societies, 2) collective transport 2.0, and 3) electromobility [26]. The urgency as well as the increasing societal and political acknowledgment of the progressing climate crisis, the related need for innovative sustainable transport technologies, policies, and strategies have catalyzed the emergence of a fast-growing academic field of sustainable transport research. This research field has evolved from an earlier STEM-dominated into a highly dynamic and interdisciplinary realm. As a consequence of increasing acknowledgment of the complexity of sustainability transitions, the number of studies with sociotechnical focus to understand the

*Correspondence:
Michael Stiebe
michael.stiebe@hslu.ch
[1] Institute of Tourism and Mobility (ITM), Competence Center for Mobility (CCM), Lucerne University of Applied Sciences and Arts (HSLU), Rösslimatte 48, 6002 Lucerne, Switzerland
[2] Graduate School of Business and Economics, Faculty of Business and Economics, University of Basel, Peter Merian-Weg 6, 4002 Basel, Switzerland

dynamics of sociotechnical sustainable mobility transitions has grown over recent years.

Since the platformization of the internet has given rise to global social media platforms, such as Facebook, Instagram, and Twitter, there is a constantly growing pool of user-generated data concerning a very broad array of economic, social, technological and political issues, for instance, sustainable mobility and sustainable transport [15, 31, 41]. These data pools are often called social big data [16, 32, 34, 38]. Social big data analysis has become a popular means of scholarly inquiry, particularly in the digital humanities (DH) [13]. It is inherently interdisciplinary and incorporates areas such as data mining, machine learning, statistics, graph mining, and natural language processing [7].

Since most social media data mining studies draw on only one social media platform, there are growing concerns in the academic community regarding these mono-social media platform studies since no single social media platform is representative of the general population [9]. Thus, Rogers [41] advocates for the enhancement of methodological frameworks and a transition towards cross-platform social media analysis approaches [41].

This research employs a cross-platform social media analysis methodology using an inquiry strategy based on hashtags which are widespread technical feature of various social media platforms, especially Twitter and Instagram, providing a great means of conducting a comparative analysis of the digital discourse on sociotechnical phenomena across platforms.

The project explores potential synergies between the fields of DH and sociotechnical sustainable transport research, and how cross-field collaboration can enhance sociotechnical sustainable transport research and the general understanding of sociotechnical low-carbon transport transitions. Thus, this research asks the overarching question:

How can social media cross-platform analysis help enhance the understanding of sociotechnical low-carbon transport transitions?

Two sub-questions guide the research:

1) What are the key themes and differences in the social media discourse about #sustainablemobility and #sustainabletransport on Twitter and Instagram?
2) How can social media analysis help to further the understanding of sociotechnical phenomena and processes in low-carbon transport transitions?

2 Methodology

This chapter outlines the theoretical perspective, briefly summarizes the research workflow, and provides a detailed account of the methods employed in this study.

2.1 Theoretical perspective

This work is embedded within the pragmatic research paradigm which assumes that reality and truth are under constant renegotiation and subject to behavior, social norms, and beliefs [30]. Taylor and Bogdan [52] argue that people's words and actions are a product of how they personally define their world. Likewise, Furlong [22] argues that reality is in essence a result of our own making. Thus, the positivist notion that social science could uncover the truth about the real world is being rejected and a strong emphasis is placed on the workability in research by adopting a worldview that permits a research design and methodologies sufficing the purpose and goal of this study. The latter concerns understanding sustainable transport transitions as a sociotechnical phenomenon.

This paper draws on the Geelsean multi-level perspective (MLP) framework on sociotechnical transitions which allows researchers to employ a holistic perspective to the dynamics of sociotechnical systems which the MLP structures into three interconnected levels, i.e., 1) Sociotechnical Landscape (Exogenous Context), 2) the Sociotechnical Regime, and 3) Niche Innovations [23, 24, 49] (see Fig. 1).

2.2 Cross-Platform analysis

This paper employs cross-platform social media analysis on Twitter and Instagram. Both platforms share core features of social media, including internet-based applications, user-generated content, and networking, contributing to the phenomenon of big data [10, 55]. While Twitter has been extensively used for mobility and transport related research [5, 8, 15, 29, 31, 39, 51], studies utilizing Instagram data analysis are comparatively limited [19, 40, 44, 47]. Despite the existence of various social media platforms with common features such as geotagging, @mentions, and hashtags, Twitter often takes precedence in social media platform analysis studies [53]. However, this mono-platform focus in research poses challenges to deriving generalizable assumptions and valid results due to platform-specific user cultures and demographics [9, 42, 53]. Therefore, cross-platform analysis is essential to obtain complementary samples and enhance the representativeness of social big data
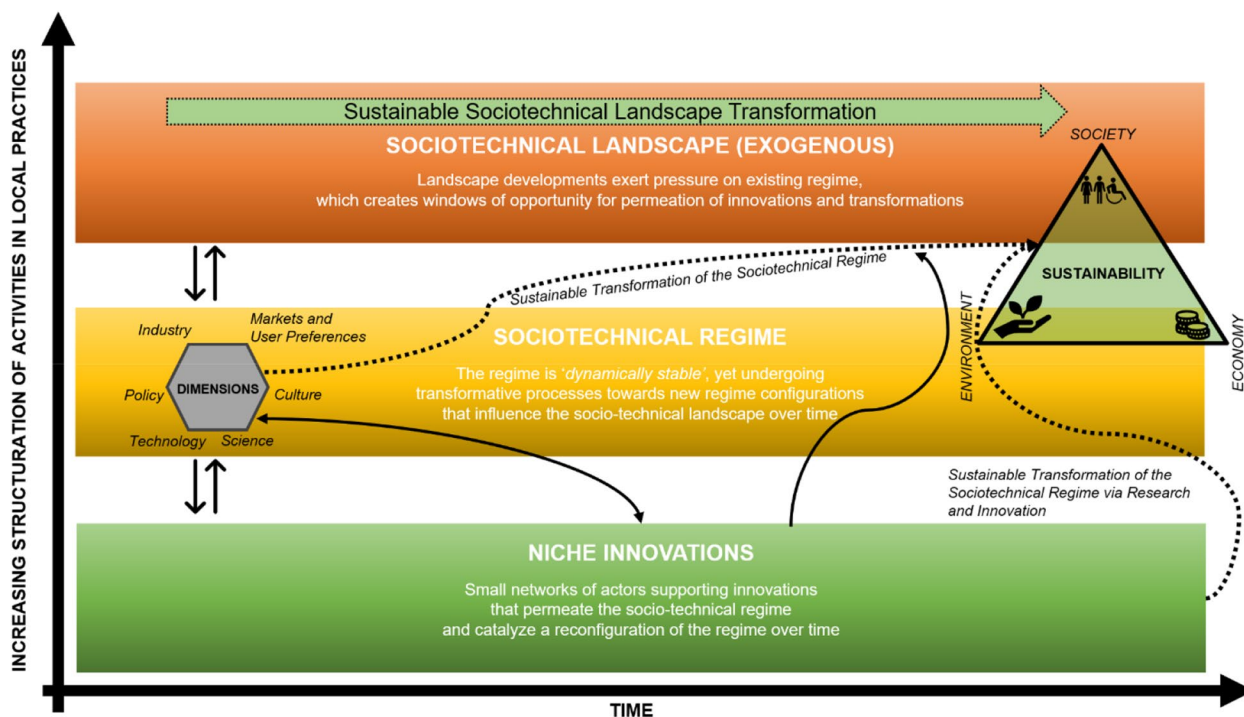
Fig. 1 Multi-Level Perspective (MLP) on Low Carbon Transitions, adapted from Geels ([23], p. 28)

Table 1 Elements of cross-platform analysis, Rrogers ([41], p. 11)

No	Element	TWITTER	INSTAGRAM	FACEBOOK
1	Query Design	Hashtags, Keywords, Locations, Users	Hashtags, Locations	Groups, Pages
2	Data Capture	In Advance (For Overtime Data); On Demand (For Very Recent Data)	On Demand (For Overtime Location Data and Recent Hashtag Data)	On Demand (for Overtime and Recent Data)
3	Platform User Accounts (With Primary Actions)	User (Follow)	User (Follow)	User (Friend, Follow), Group (Join), Page (Like)
4	Content (Media Contents And Digital Objects)	Tweet (Text, Photo, Video, Hashtag, @mention, URL, Geotag)	Photo, Video (Text, Hashtag, Geotag, @mention)	Post (Text, Video, Photo, URL)
5	Activities (Resonance Measures)	Like (Fav), Retweet	Like, Comment	Like, Comment, Share

research. Rogers [41] advocates for cross-platform analysis to capitalize on practical similarities and technical feature overlaps between social media platforms, outlining five core elements (see Table 1) and six steps (see Table 2) including 1) choosing a contemporary issue, 2) designing a query strategy, 3) developing an analytical strategy, 4) considering the configuration of use, 5) cross-platform analysis, and 6) discussing the findings.

2.3 Research workflow
Figure 2 provides an overview of the research process via a flowchart. Further details on the research workflow are provided in the subsequent sections.

2.4 Data collection
The initial steps in Rogers' [41] methodology for cross-platform analysis entail selecting a contemporary issue and designing an appropriate query strategy. Given the focus on social media discourse concerning sustainable mobility and sustainable transport, the hashtags #sustainabletransport and #sustainablemobility were chosen. Holden et al. [27] noted the interchangeable use of "sustainable transport" and "sustainable mobility" in academic literature, with a preference for "sustainable mobility" in Europe and "sustainable transport" in North America. Despite #sustainabletransport potentially sufficing for analysis, both hashtags were selected to collect

Table 2 Steps in cross-platform analysis, adapted from Rogers ([41], pp. 12–13)

No	Step	Description
1	Choosing a Contemporary Issue	e.g., Revolution, Disaster, Election, Social Cause etc
2	Designing a Query Strategy	Identifying the right hashtags or other queries to be mined
3	Developing an Analytical Strategy	Consider spatiotemporal and contextual aspects
4	Considering the Configuration of Use	Consider whether one deals with pages, individuals and whether comments or interactions are allowed
5	Cross-Platform Analysis	Undertake the platform analysis, according to the query design strategy as well as the analytical strategy discussed above, across two or more platforms. For each platform consider engagement measures, such as the sum of likes, shares, comments (Facebook), likes and retweets (Twitter) and co-hashtags (Instagram). Which (media) content resonates on which platforms? Consider which content is shared across the platforms (co-linked, inter-liked and cross-hashtagged), and which is distinctive, thereby enabling both networked platform content analysis as well as medium-specific (or platform-specific) effects
6	Discussing Findings	Discussion of findings with respect to medium research, social research, or a combination of the two. Does a particular platform tend to host as well as order content in ways distinctive from other platforms? Are the accounts of the events distinctively different per platform or utterly familiar no matter the platform?

larger data samples and examine potential discrepancies in digital discourses, challenging the assumed synonymity observed by Holden et al. [27].

2.4.1 Twitter and Instagram data scraping

Twitter is a social network and microblogging platform that allows users to publish brief public messages, known as *tweets*, which can include text and media. This platform has gained significant traction among data scientists and researchers for its utility in capturing large datasets of public opinion and discourse on a wide array of topics [54]. Additionally, tweets can be enriched with geotags and hashtags, further expanding their utility for detailed, location-specific, and topic analyses.

There are two common approaches to obtaining Twitter data, i.e., either directly through the official Twitter REST API, or through web scraping-based applications. This study employed the web-scraping method drawing on Python scripts and the library Twint[1] which, according to its official GitHub description, is an advanced Twitter scraping and Open Source Intelligence (OSINT) tool that does not use Twitter's API, allowing to scrape Twitter data and evade most API limitations. Apart from their text content, the tweets were collected including their associated meta-data including date, time, geotag, applied hashtags. A total of 33,121 Tweets from 2013–01-01 to 2021–04-01 were mined, 16,608 for #sustainabletransport and 16,513 for #sustainablemobility (see Table 3).

Instagram is a social networking platform that distinguishes itself from Twitter by focusing primarily on user-generated multimedia content, including photos and videos, rather than text-based posts. Users frequently enhance their content with hashtags and captions, encouraging interaction through comments and discussions among users [28]. Despite Instagram's growing restrictions on the types and amounts of content it permits to be scraped or mined, resourceful open-source developers provide solutions, such as the Python module Instaloader[2] which is a tool to download pictures (or videos) along with their captions and other metadata from Instagram.

Leveraging the capabilities of the Instaloader Python module for this research, a dataset of 8,089 public Instagram posts was collected, with half of the posts (4,054) tagged #sustainabletransport and the other half (4,035) tagged #sustainablemobility. This dataset, encompassing captions, hashtags, and metadata, spans a period from 2017 to 2021 (see Table 3). Due to practical constraints, such as heightened internet data traffic and computer memory limitations, video content was not scraped. Additionally, the research necessitated the creation of a new Instagram account, as Instaloader's functionality requires user authentication. This approach to data collection is not without its challenges since Instagram's algorithms are designed to detect and potentially block or restrict accounts engaging in scraping activities deemed inappropriate by the platform. This is a common challenge in social media research since social media platform operators are making it increasingly difficult for academics to obtain comprehensive access to their data [6]. During the scraping process, the Instagram account used for user authentication in Instaloader was blocked

---

[1] https://github.com/twintproject/twint (project no longer maintained).
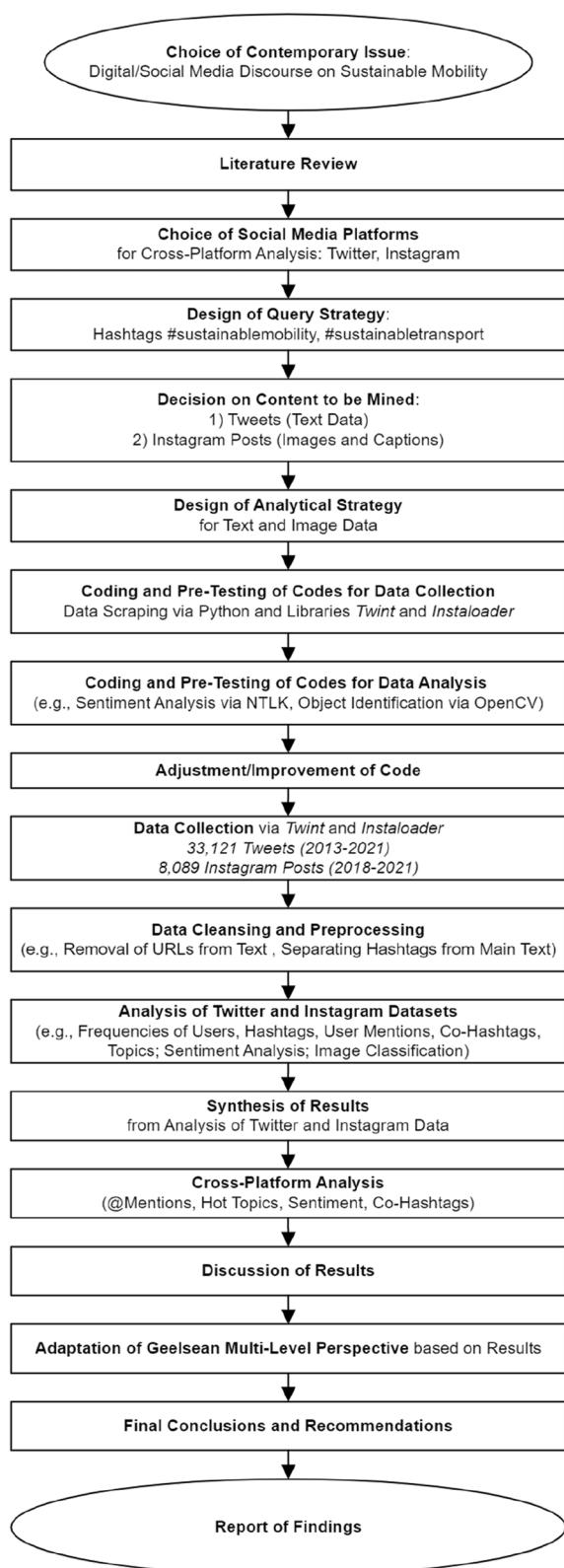
[2] https://instaloader.github.io/

Fig. 2 Research workflow

Table 3 Quantitative overview of scraped Tweets and Instagram posts

| Year | #sustainabletransport | | #sustainablemobility | |
	Twitter	Instagram	Twitter	Instagram
2013	527	0	342	0
2014	1055	0	660	0
2015	1242	0	1010	0
2016	2880	0	992	0
2017	1639	1	1263	0
2018	2489	5	2689	1
2019	2942	4	3683	1
2020	3068	2831	4550	1727
2021 (*only until 04/2021)	766	1213	1324	2306
Total	**16,608**	**4054**	**16,513**	**4035**

several times, which was partially mitigated through changing the IP using proxy connections Table 4.

2.5 Data analysis
This section details the analytical methods used to address the research questions Table 5.

2.5.1 Data cleansing and pre-processing
The initial step in the analytical journey involved the cleaning and preprocessing of the collected social media data, a process critical to the integrity and success of social media data analysis [54, 55].

Leveraging Python, a rigorous cleaning process for the textual content harvested from Twitter and Instagram was employed. This entailed the removal of links and URLs, a common source of noise in textual data, ensuring a focus on meaningful content. Hashtags, while removed from the main text to purify the dataset, were preserved in a separate column within the CSV files. This dual approach permitted to maintain the contextual relevance of hashtags without cluttering the primary textual analysis. Similarly, Instagram captions and comments underwent a rigorous cleaning process, with URLs excised and hashtags meticulously separated. Stop-words, i.e., linguistically ubiquitous yet analytically trivial words, were also removed to distill the essence of the discourse.

Beyond these foundational steps, advanced preprocessing techniques were integrated. Textual content was normalized to a uniform case, facilitating consistent analysis, followed by tokenization to dissect the text into analyzable components. This step is crucial for identifying and evaluating the sentiment-bearing elements of the text. Recognizing the complexity of human communication as well as the inherent challenge of detecting sarcasm

Table 4 CSV table excerpt of image classification results for #sustainabletransport Instagram dataset

Year	Month	Day	Time	Filename	Shown Objects
2020	4	15	15–15-53	2020–04-15_15-15-53_UTC.jpg	person
2020	4	15	15–15-53	2020–04-15_15-15-53_UTC.jpg	bicycle
2020	4	15	15–15-53	2020–04-15_15-15-53_UTC.jpg	motorcycle
2020	4	15	16–51-46	2020–04-15_16-51-46_UTC.jpg	potted plant
2020	4	15	17–00-48	2020–04-15_17-00-48_UTC.jpg	truck
2020	4	15	17–00-48	2020–04-15_17-00-48_UTC.jpg	truck
2020	4	15	19–31-59	2020–04-15_19-31-59_UTC.jpg	bicycle
2020	4	16	05–19-00	2020–04-16_05-19-00_UTC.jpg	car
…	…	…	…	…	…

Table 5 Sentiment polarity averages 2013–2021

| | Twitter | | Instagram | |
Year	#sustainabletransport	#sustainablemobility	#sustainabletransport	#sustainablemobility
2013	0.6167	0.6316	N/A	N/A
2014	0.7400	0.6394	N/A	N/A
2015	0.7172	0.7228	N/A	N/A
2016	0.6760	0.7641	N/A	N/A
2017	0.7016	0.6997	1.0000	N/A
2018	0.6924	0.7125	0.8000	1.0000
2019	0.7262	0.7475	1.0000	1.0000
2020	0.7268	0.7767	0.7879	0.8077
2021	0.6997	0.7508	0.7995	0.7701
Total Avg	0.6996	0.7161	0.8775	0.8945

and irony, that can act as potential sentiment polarity reversers in textual content [11], this was born in mind during the preprocessing steps. The automatic detection of rhetorical devices is a very interesting yet one of the most challenging NLP tasks when using microblogging platform posts, and thus requires advanced text preprocessing and analytical strategies [1, 11, 21]. However, given the scope and available resources for this research as well as the findings from Dimovska et al. [18] that text preprocessing has very little impact on results in automated sentiment detection, the preprocessing strategy was not specifically adjusted. Nevertheless, this simplified preprocessing strategy may have also incurred a certain degree of inaccuracy in the sentiment classification process, which is kept in mind during the discussion of findings.

In parallel with the textual data preparation, the mined Instagram images were systematically reviewed for integrity. This process involved scanning for and eliminating broken JPEG files to ensure a seamless batch-processing experience for subsequent automatic image classification tasks. Through this meticulous examination, a solitary broken image file was identified and removed, thereby safeguarding the quality of the image dataset.

2.5.2 Quantitative text analysis

To systematically uncover the prevailing themes and focal topics within posts and discussions on sustainable transport and mobility, a comprehensive quantitative text analysis was conducted on the data collected from both Twitter and Instagram. This analysis aimed to catalog and compare the frequency of specific keywords, shedding light on the subjects that dominate conversations on each platform. By pinpointing the most frequently mentioned keywords along with manual identification of thematic clusters, the analysis illuminates the focal interests and concerns of the online discourse surrounding sustainable transport.

Moreover, the quantitative approach extended beyond mere keyword frequency, offering deeper insights into user engagement across these platforms. By evaluating the most active users and most frequently mentioned (via @ function) within the context of sustainable transport and mobility, key influencers and contributors to

the discourse were identified. This aspect of the analysis not only reveals who is driving the conversation but also provides a measure of the engagement level surrounding various topics.

The temporal dimension of the analysis further reveals how discussions and priorities have evolved over a specified period. By tracking changes in keyword frequency and user activity over time, the study uncovers

$$\text{classify}(word_1, word_2, \ldots, word_n) = \underset{cat}{\arg\max}\, P(cat) \times \prod_{i=1}^{n} P(word_i | cat) \tag{1}$$

trends in the public digital discourse, offering a dynamic view of the shifting landscape of sustainable transport discussions.

2.5.3 Sentiment analysis

The pre-processed textual content from the Twitter and the Instagram data provided a foundation for the employment of sentiment analysis by "(…) which the level of subjective content in information is quantified" [55], p. 168). According to Batrinca and Treleaven [6], sentiment analysis is about mining attitudes, emotions, feelings, and subjective impressions rather than facts, and aims to determine the attitude expressed with respect to the topic or the overall contextual polarity of a text. In this research, an analysis of sentiment polarity (sometimes also called "sentiment orientation") was conducted, i.e., deciding whether an opinion in a text is *positive* or *negative.*

Using the Python package Natural Language Toolkit (NLTK) and a supervised machine-learning-based approach via the Naïve Bayes classification algorithm, the sentiment polarity of the Tweets as well as Instagram text data was automatically classified. The classification model was trained using NLTK's Twitter corpus named "twitter_samples" that contained a sample of 20,000 Tweets retrieved from the Twitter Streaming API, together with another 10,000 which are divided into negative and positive tweets according to their sentiment [37].

The Naïve Bayes classifier is of general purpose, simple implementation, and advantageous because it requires relatively little training data to estimate the necessary parameters for classification [6, 17]. It is based on conditional probability, and despite its simplicity and the assumption of independence between words, performs well across many domains [17]. This technique calculates the probability of categories given a document by utilizing the joint probabilities of words and categories, based on the principle of word independence. The foundation of this method is Bayes' theorem, which allows for the combination of prior knowledge and observed data.

Specifically, it assumes that the attributes of a data point are independent within a class, enabling the estimation of a class's probability for a given data point through the product of the individual probabilities of its attributes. The classifier calculates the probability for a text to belong to each of the defined sentiment categories. The category with the highest probability for the given text wins, which can be denoted as in Eq. 1:

Naïve Bayes Classifier, adapted from Batrinca et al. [6].

The Naïve Bayes sentiment classification algorithm has been successfully applied in several sentiment analysis studies on Twitter and Instagram [36, 43, 48, 50].

Assigning a value of 1 to positive and 0 to negative posts, it was possible to calculate the sentiment polarity averages for each social media dataset permitting a comparison between the discourse and topics associated with sustainable mobility and sustainable transport on each network within the given timeframes. In the context of sentiment analysis, identifying messages with closely competing sentiment probabilities can be critical for nuanced understanding. A rule was formulated to identify "questionable" messages – those that neither strongly exhibit positive nor negative sentiment. This rule is defined by the criteria where both the probability of a message being positive (P_{pos}) and negative (P_{neg}) fall within an intermediate range. Given a dataset of messages M, each message $m_i \in M$ is analyzed for sentiment, yielding two probabilities $P_{pos}(m_i)$ and $P_{neg}(m_i)$, representing the probabilities of the message being positive and negative, respectively. A message is classified as "questionable" if both probabilities fall within a specified intermediate range (specifically between 0.4 and 0.6), formally defined as follows:

Let Q be the subset of M where each $m_i \in M$ satisfies:

$$0.4 \leq P_{pos}(m_i) \leq 0.6 \text{ and } 0.4 \leq P_{neg}(m_i) \leq 0.6 \tag{2}$$

Classification Rule for "Questionable" Messages.

The use of lexicons as an alternative method for sentiment analysis was considered during this research. Lexicon-based approaches rely on a predefined list of words each associated with a sentiment score, which can be used to evaluate the sentiment of a text without the need for training data. However, Naïve Bayes sentiment classification stands out in the analysis of social media posts, mainly due to its capacity to grasp the context in which words are used, an area where lexicon-based methods fall short. This ability is crucial on social media, where the sentiment of words can vary greatly with context [3].

Fig. 3 Image classification via OpenCV applied to a photo from #sustainabletransport Instagram dataset

Moreover, Naïve Bayes adapts effectively to the dynamic nature of social media language, learning from evolving expressions and slang, unlike lexicon-based methods that require constant updates to their sentiment dictionaries [46]. This classifier also excels in processing speed, essential for analyzing large datasets in real-time. Additionally, it can deal with ambiguous sentiments more adeptly through probabilistic models and integrate with various data sources for enhanced accuracy, offering a more comprehensive approach than lexicon-based analysis [4, 35]. These aspects made the Naïve Bayes classification the preferred method for sentiment analysis in this social media analysis context.

2.5.4 Image analysis
In the exploration of social media for academic research, image analysis emerges as a pivotal technique to uncover trends and discussions related to specific topics, search queries, or hashtags. Using Python and libraries, e.g., TensorFlow, Keras, OpenCV, ImageAI, for image analysis and object detection has become an established method in data science [14].

In this study, OpenCV was used for object detection. The scope of image analysis was specifically directed towards the Instagram dataset, considering that the extraction from Twitter was restricted to textual content, omitting audio-visual elements.

Initially, images underwent a preprocessing phase to standardize dimensions and normalize pixel values across the dataset. This crucial step enhances the analytical quality of the images and prepares them for further

processing. Image segmentation, facilitated by a pretrained TensorFlow model, played a key role in isolating distinct objects within the images, enabling detailed examination.

Rather than engaging in manual feature engineering or annotation, the study harnessed the capabilities of the pre-trained TensorFlow model "frozen_inference_graph". This model, adept at recognizing 90 different object classes, including various vehicles relevant to mobility and transport such as bicycles, trains, buses, and cars, provided an extensive set of features for object detection. This strategic choice streamlined the analysis by leveraging existing, comprehensive features for object detection, thereby simplifying the process.

The use of OpenCV allowed for the processing of images through the model to detect and classify objects. Images were transformed into a compatible format, set as inputs to the model, and the output was analyzed to identify and classify objects within the images. Each detected object was assigned a label from the model's predefined set of classes (see Fig. 3).

Following object detection, the identified objects underwent meticulous cataloging. This process involved capturing and recording multiple objects that may coexist within a single image, ensuring comprehensive data collection. Subsequently, the extracted data underwent systematic organization and was stored in a CSV file, facilitating quantitative analysis. This analysis aimed to examine the prevalence of various modes of transport and vehicles within the Instagram dataset.

This methodological rigor underscores the application of advanced machine learning techniques in
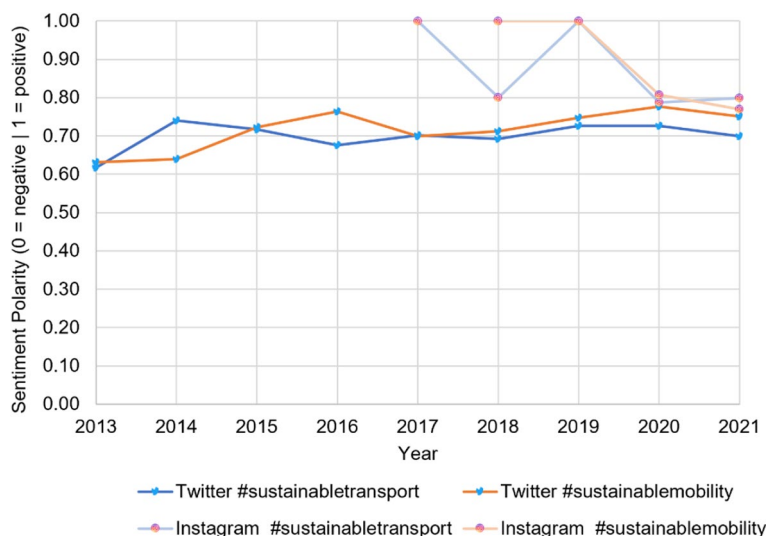
Fig. 4 Sentiment polarity development 2013 – 2021

dissecting social media content, thereby offering profound insights into the discourse surrounding sustainable transport and mobility. Through this analysis, the study delves deeper into public engagement and perceptions regarding sustainable practices as manifested in social media platforms. For instance, it examines the frequency with which low-carbon modes of transport are associated with sustainable mobility or sustainable transport in social media posts, thereby enriching our understanding of societal attitudes and behaviors towards sustainability.

3 Results and discussion

This section presents and critically discusses pertinent analysis results.

3.1 Sentiment polarity

Evident fluctuations in the average sentiment polarity from 2013 to 2021 were identified in posts tagged with either #sustainabletransport or #sustainablemobility. The average annual sentiment for posts tagged with either one of the two hashtags were always above the neutral 0.5-threshold, which suggests that both hashtags were prevalently used in a positive context. Noteworthily, both hashtags have become more positive from 2013 to 2021. The hashtag #sustainabletransport increased from 0.6167 in 2013 to 0.6997 in 2021 (+8 percent points), and #sustainablemobility increased from 0.6316 in 2013 to 0.7508 in 2021 (+11 percent points). The positive peaks for #sustainabletransport and #sustainablemobiltiy were in 2014 and 2020, respectively. Whereas both curves look relatively similar and close to each other in general,

there are two visible gaps that occurred in 1) 2014 when the average sentiment of #sustainabletransport was approximately 10 percent points higher than #sustainablemobility, and 2) in 2016 when the average sentiment of #sustainablemobility was approximately 9 percent points higher than #sustainabletransport. The abovementioned phenomena are visible in Fig. 4. Based on the rule for the classification of questionable messages defined in Sect. 2.5.3., about 13.6% and 13.8% of the tweets tagged with #sustainablemobility and #sustainabletransport showed competing polarity probabilities, respectively. For the Instagram captions tagged with #sustainablemobility and #sustainabletransport the shares of questionable captions were lower with approximately 8.1% and 8.0%, respectively.

3.2 Key themes, co-hashtags, users mentions

In addressing the question, "What are the key themes in social media discourse around #sustainablemobility and #sustainabletransport on Twitter and Instagram?", eleven major thematic clusters were identified from tweets and Instagram captions through quantitative text analysis and subsequent manual clustering. These clusters include: 1) mobility/transport, 2) active travel (bicycling, walking), 3) public transport (bus, train), 4) automotive/cars, 5) electromobility, 6) urban and smart mobility, 7) future mobility/innovation, 8) sustainable development, 9) conferences, 10) covid-19, and 11) global issues. Significant overlap across both platforms was noted for themes of active travel, electromobility, urban and smart mobility, and public transport.

Further analyses of co-hashtags and dataset overlaps confirmed a stronger linkage between the electromobility

Table 6 Cross-platform analysis results for #sustainabletransport

#sustainabletransport

Rank	Hot Topics		Co-Hashtags		Mentions	
	Twitter	Instagram	Twitter	Instagram	Twitter	Instagram
1	cycling	bike	cycling	sustainability	recycle4gm	teslamotors
2	new	sustainable	sustainability	electricvehicle	UNDESA	marty12tesla
3	sustainable	electric	transport	comment4comment	UN	teslaclubsocal
4	electric	link	climatechange	ev	Transport_EU	myteslavibez
5	bike	bio	electricvehicles	ebike	embarqnetwork	myevcom
6	mobility	travel	ev	electricvehicles	sustrans	unpluggedperformance
7	sustainability	car	publictransport	cycling	recycleforbury	carlacargoengineering
8	great	day	activetravel	electriccar	guardian	theteslasquad
9	cities	people	sdgs	electricride	elonmusk	electriccarbeauty
10	future	make	mobility	sustainableliving	OfficialTfGM	spacex

Table 7 Cross-platform analysis results for #sustainablemobility

#sustainablemobility

Rank	Hot Topics		Co-Hashtags		Mentions	
	Twitter	Instagram	Twitter	Instagram	Twitter	Instagram
1	transport	electric	Mobility	Zeroemissions	Groupe_Renault	niu
2	new	mobility	Ev	Mobility	movinonconnect	europeanmobilityweek
3	sustainable	new	Electricvehicles	Futurride	riversimple	mercedesbenz
4	cities	link	smartmobility	ev	Michelin	generalmotorsdesign
5	electric	sustainable	transport	electric	movinonconf	ionity
6	future	bio	urbanmobility	electricmobility	UITPnews	mercedesbenzusa
7	city	know	innovation	socialdistancing	CIVITAS_EU	doctoran
8	innovation	city	emobility	SustainableLuxury	Y4PT	daimler_ag
9	cycling	vehicles	publictransport	EV	Transport_EU	chipganassiracing
10	electricvehicl es	hybrid	cycling	niumobility	EU_Commission	fia

theme with #sustainablemobility, and the active travel theme with #sustainabletransport. this distinction underscores different conceptual focuses within the sustainable mobility discourse.

An examination of more than 8,000 Instagram image posts revealed bicycles were present more than twice as much in #sustainabletransport posts compared to #sustainablemobility. The share of images depicting cars was noticeably higher under #sustainablemobility (approximately 30%) compared to #sustainabletransport (approximately 23%). Findings indicate vehicles, particularly those associated with private motorized transport and active travel, were depicted more frequently than public transport vehicles in images related to both hashtags – appearing in more than half of the images under each

hashtag. These findings challenge initial assumptions about the prominence of public transport in sustainable mobility discussions.

Analysis of top-mentioned users revealed a notable presence of Elon Musk/Tesla/SpaceX mentions on Instagram, particularly within the #sustainabletransport conversation, highlighting a prominent electromobility focus. Conversely, on Twitter, mentions were more varied, including public figures, international organizations (EU, UN), and companies within the mobility sector, indicating a broad engagement with sustainable mobility and transport themes across sectors.

Tables 6 and 7 showing cross-platform analysis results based on the quantitative text analysis, and Table 8 showing the most frequently classified objects in the Instagram

Table 8 Top 15 objects detected in Instagram image posts

Rank	#sustainabletransport		#sustainablemobility	
	Object	% of Images	Object	% of Images
1	person	52.40%	person	46.88%
2	bicycle	27.45%	car	30.03%
3	car	22.98%	motorcycle	16.03%
4	motorcycle	8.20%	bicycle	12.03%
5	truck	5.79%	truck	7.02%
6	bus	4.25%	bus	2.75%
7	boat	2.67%	tv	2.31%
8	train	2.50%	laptop	2.20%
9	book	1.94%	airplane	2.14%
10	laptop	1.55%	book	1.90%
11	airplane	1.51%	clock	1.83%
12	chair	1.51%	chair	1.80%
13	kite	1.28%	kite	1.66%
14	umbrella	1.22%	train	1.42%
15	bench	1.15%	boat	1.19%

images provide comprehensive insights, highlighting the dynamic interplay between personal, technological, and policy dimensions of sustainable mobility and transport in social media discourse.

3.3 Discourse differences Twitter vs. Instagram

The comparative analysis of Twitter and Instagram regarding #sustainabletransport and #sustainablemobility revealed a notable 54% overlap in the top 30 hot topics across both platforms. Interestingly, #sustainabletransport exhibited a slightly higher incidence of exact topic matches across platforms compared to #sustainablemobility. Moreover, a co-hashtag analysis enhanced this finding, demonstrating an even greater overlap of 55%.

Investigation into the topic clustering for each platform, based on the combined hashtags, highlighted Active Travel and Electromobility as the predominant themes within Twitter and Instagram, respectively. This distinction points out a platform-specific nature of discourse surrounding sustainable mobility.

An analysis contrasting sentiment across the two platforms revealed that, despite different analysis periods, Instagram content associated with both hashtags was generally more positive compared to Twitter. Across both platforms and throughout the analysis periods, the sentiment remained positively skewed, maintaining above a 0.5 neutral polarity threshold.

The examination of Twitter and Instagram content revealed not only overlaps in hot topics and co-hashtags but also significant differences in thematic dominance – Active Travel on Twitter and Electromobility on Instagram. The distinction in thematic dominance—Twitter's focus on Active Travel and Instagram's emphasis on Electromobility—might be influenced by the platforms' inherent characteristics. Drawing on Lee et al. [33], who found Twitter to be more oriented towards everyday occurrences, it could be hypothesized that topics of daily mobility, such as Active Travel and Public Transport, naturally gravitate towards Twitter. This contrasts with Instagram, where the visual and aspirational nature of content may favor discussions around Electromobility. However, these speculations remain tentative in the absence of comprehensive sociodemographic data to further elucidate these patterns [45].

Differences in sentiment trends between Twitter and Instagram also emerged, complicated by varying analysis periods. The consistently more positive sentiment in Instagram captions, compared to tweets, may partly result from the sentiment classification algorithm's training predominantly on Twitter data, suggesting platform-specific nuances in content sentiment.

This analysis underscores the complexity of social media discourse on sustainable mobility, highlighting both shared interests and platform-specific discussions.

3.4 Enhanced understanding of sustainable mobility transitions

This section delves into the second research subquestion: "How can social media analysis help to further the understanding of sociotechnical phenomena and processes in low-carbon transport transitions?".

Building on the Geelsean MLP on Sociotechnical Transitions, a cornerstone in the study of sociotechnical systems and sustainable transport [12, 24], this research acknowledges the model's broad applicability. However, it posits an evolution of the MLP to more explicitly encompass the influences of the scientific community and the realm of digital social media discourse. In response to this identified gap, a refined version of the MLP is proposed, wherein the original Niche Innovations level is reimagined as a level of Sociotechnical Sustainable Transport Research. This redefined layer is segmented into four critical phases integral to fostering sociotechnical shifts towards sustainable mobility: 1) conducting research to decode the existing sociotechnical transport regime and its landscape; 2) investigating the prerequisites for sustainable sociotechnical regimes; 3) supporting and steering the transition towards sustainable mobility through targeted research; and 4) enhancing the efficacy and impact of sustainable mobility transitions through continued innovation and study.

Acknowledging the vital role and expanding influence of social media in shaping public discourse and potentially guiding policy and innovation, this study
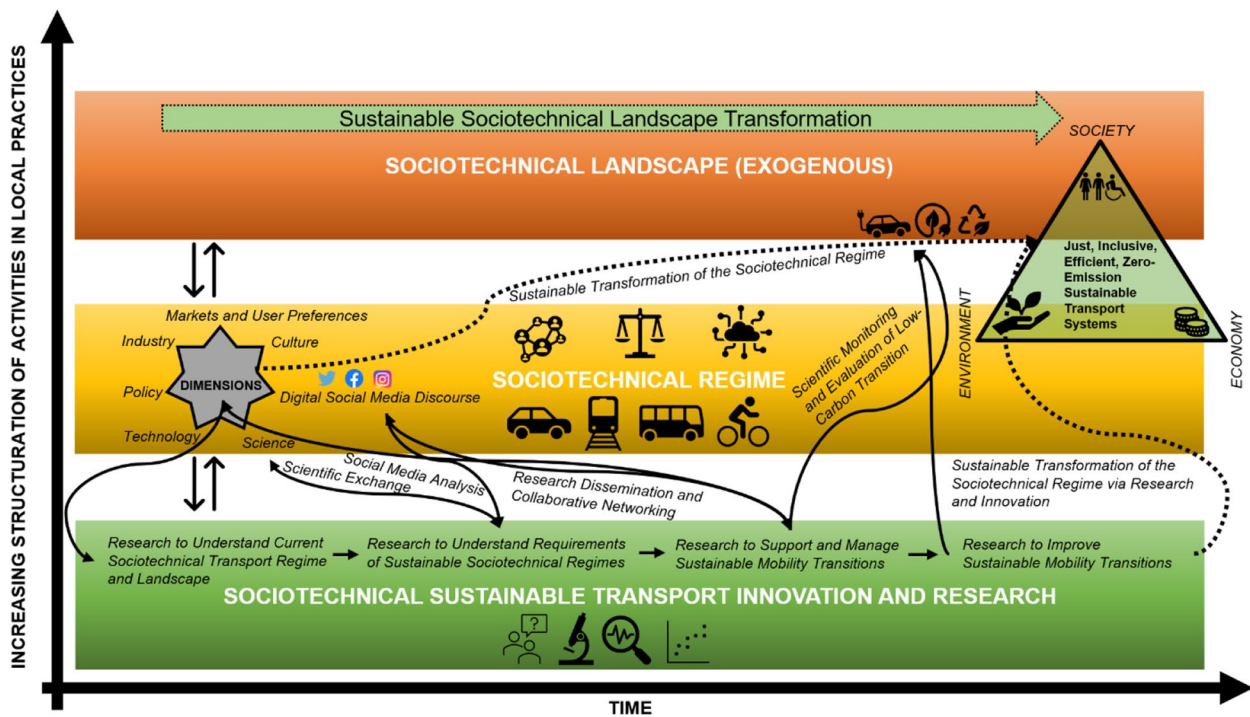
Fig. 5 Sociotechnical sustainable transport research-focused Adaption of MLP, based on Geels and Kemp ([24], p. 474)

introduces Digital Social Media Discourse as a vital seventh dimension to the MLP's original six dimensions at the sociotechnical regime level. This addition highlights the changing landscape of information sharing and community engagement, emphasizing how social media platforms have become crucial battlegrounds for ideas, innovations, and ideologies related to sustainable mobility.

The revised model, illustrated in Fig. 5, represents a step forward in the direction towards understanding and guiding sociotechnical sustainability transitions. By integrating the dynamic and influential sphere of social media, this enhanced MLP model offers a more nuanced and comprehensive framework for analyzing and facilitating the journey towards sustainable transport and mobility.

4 Conclusions

This research investigated how cross-platform social media analysis can help enhance the understanding of sociotechnical low-carbon transport transitions. The study has drawn on an exploratory cross-platform social media analysis approach based on Instagram and Twitter posts under the hashtags #sustainabletransport and #sustainablemobility. In total, 33,121 Tweets and 8,089 Instagram image posts including captions were scraped and analyzed. Some of the core findings of this study

combining the results of the hot topic as well as co-hashtag analyses comprise insights into the main themes and thematic clusters within the sphere of the digital discourse on Twitter and Instagram regarding the concepts sustainable transport as well as sustainable mobility. It has become apparent that only the third of Holden et al.'s [26] Grand Narratives for sustainable mobility, i.e., Electromobility, has been significantly present in the digital discourse on both platforms, especially on Instagram. While the strongest link to #sustainablemobility was the electromobility theme, #sustainabletransport was related the closest to the theme of Active Travel, especially bicycling. Despite not being included in the Grand Narratives, the latter theme, namely Active Travel, has been the most prominent one across the two platforms based on aggregated results from the co-hashtag as well as hot topic analyses. An intriguing finding from the cross-platform analysis of frequently mentioned users is the overwhelming dominance of the Elon Musk/Tesla cluster across both platforms. What's particularly noteworthy is that these mentions are consistently linked with #sustainabletransport, a trend observed on both Twitter and Instagram. Whereas public transport and low-mobility societies are among the main topics of contemporary sustainable transport and mobility research [26, 56], both themes were neither significantly reflected in the digital discourse regarding #sustainabletransport nor

#sustainablemobility. To the author's surprise, alternative fuels/synthetic fuels or hydrogen mobility were not reflected to any considerable extent either.

Based on the analyses, gaps in the Geelsean MLP have been identified, leading to its adaptation. The model now integrates sociotechnical sustainable transport research and digital research methods to better understand and manage sociotechnical sustainable mobility transitions. This enhanced model introduces a seventh dimension, digital social media discourse, at the meso-level, i.e., the sociotechnical regime.

Investigating the digital social media discourse drawing on the social media analysis method, ideally cross-platform analysis, adds to the holistic understanding of sociotechnical low-carbon transport transitions. This signifies the potentials and benefits of future sociotechnical sustainable transport research – DH collaborations, since social media analysis has become a core method in the DH.

The cross-platform analysis of #sustainabletransport and #sustainablemobility identified significant disparities between these concepts on Twitter and Instagram. While academia often treats them interchangeably, this study reveals nuanced differences in public perception and connotations. Contrary to previous assumptions, sustainable transport and sustainable mobility are used in distinct contexts, challenging regional preferences suggested by Holden et al. [27]. The strength of their association with phenomena like electromobility varies substantially, highlighting the need for careful consideration in sociotechnical sustainable transport and mobility research.

4.1 Theoretical implications

The application of a methodological framework that utilizes cross-platform social media analysis for exploring public discourse on sustainable mobility transitions signifies a substantial enhancement to the existing research landscape in sociotechnical sustainable mobility research. This approach has illuminated the pivotal role of digital public spheres in shaping sociotechnical transitions, contributing empirical evidence from social media data to the discourse. Specifically, the augmentation of the Geelsean MLP on Sociotechnical Transitions to incorporate digital social media discourse as a distinct dimension at the sociotechnical regime level signifies a crucial theoretical advancement. This inclusion reflects the growing influence of digital platforms in facilitating societal engagement with issues of sustainable transport [26].

A noteworthy discovery of this investigation was the distinct engagement with the terminologies "sustainable transport" and "sustainable mobility" across social media platforms, challenging the prevailing academic norm of using these terms interchangeably. The analysis insights into dominant themes such as Electromobility and Active Travel, especially the prominence of narratives around Elon Musk and Tesla, offer a nuanced perspective on public interest and discourse not previously covered in scholarly works. Additionally, the exploration of hashtag usage and thematic clusters provides original contributions by delineating the specific dynamics of digital discourse related to sustainable transport, thereby refining the theoretical frameworks guiding sociotechnical transition research.

4.2 Practical implications

The insights from this analysis provide actionable strategies for transport companies, policymakers, and other stakeholders aiming to enhance sustainability practices. Understanding the distinctions between "sustainable transport" and "sustainable mobility" through social media discourse enables tailored communications and policies that align with public perceptions and expectations. This nuanced understanding contests the geographical assumptions posited by Holden et al. [27], emphasizing their distinct contextual associations within sustainable transport phenomena. Identifying gaps and potential enhancements in the MLP model based on this study's findings offers a roadmap for integrating digital social media discourse into sociotechnical transition research methodologies. This highlights the potential for interdisciplinary collaborations between sociotechnical sustainable transport and DH research, emphasizing the critical role of social media analysis. Ultimately, social media cross-platform analysis emerges as a vital tool for advancing understanding and management of sociotechnical low-carbon transport transitions, bridging theoretical insights with practical applications for informed policy-making and strategic planning.

4.3 Future research

This study identifies gaps in sentiment analysis, advocating for Multi-Lingual Sentiment Analysis (MSA) techniques to promote language inclusivity as for instance recommended by Agüero-Torales et al. [2]. Future studies should incorporate the analysis of user comments and interactions, thereby gaining deeper insights into discourse on sustainable transport and mobility. Exploring non-verbal interactions like likes and shares, incorporating sociodemographic and gender analyses, and expanding research to non-Western social media platforms (e.g., Weibo) could enrich understanding. Integrating spatiotemporal dynamics and Geographic Information Systems (GIS) visualizations could provide insights for policy and community engagement. The dynamic nature of social media, exemplified by Elon Musk's acquisition of

Twitter (now X) in 2023, presents challenges and opportunities, highlighting the need for flexibility in research adaptation.

4.4 Limitations

While this study is one of the first to examine the social media discourse on sociotechnical transitions towards low-carbon transport via a cross-platform approach, it faces significant limitations. Primarily leveraging English language data for sentiment analysis may overlook global perspectives on sustainable transport, particularly from non-English speaking communities. The focus on specific hashtags introduces selection bias, capturing only a fraction of the broader conversation. Additionally, analyzing Twitter and Instagram may not fully represent wider public opinion due to platform demographics. Reliance on machine learning for sentiment classification, despite its capabilities, may struggle with nuances like sarcasm, potentially leading to inaccuracies. Neglecting non-verbal interactions such as likes and shares limits understanding of digital discourse dynamics. Lack of analysis on comments and user interactions hinders insights into sustainable transport discourse. Geographical and demographic distribution of social media discourse was not systematically explored, missing regional and sociodemographic influences on discussions.

Authors' contributions
Research Design, Data Collection, Data Analysis, Data Visualization, Writing of Research Manuscript.

Availability of data and materials
The author is committed to the FAIR data stewardship and sharing principles. All datasets used in this study are made freely accessible to other research researchers.

Declarations

Competing interests
There are no competing interests regarding this research.

References
1. Agrawal, A., Jha, A. K., Jaiswal, A., & Kumar, V. (2020, August). Irony detection using transformers. In 2020 International Conference on Computing and Data Science (CDS) (pp. 165-168). IEEE. https://doi.org/10.1109/CDS49703.2020.00040.
2. Agüero-Torales, M. M., Salas, J. I. A., & López-Herrera, A. G. (2021). Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing, 107*, 107373.
3. Akaichi, J., Dhouioui, Z., & Pérez, M. J. L.-H. (2013). *Text mining facebook status updates for sentiment classification*. 2013 17th International conference on system theory, control and computing (ICSTCC). https://doi.org/10.1109/ICSTCC.2013.6689032.
4. Al-Sheikh, E. S., & Hasanat, M. H. A. (2020). *Social media mining for assessing brand popularity*. In *Global Branding: Breakthroughs in Research and Practice* (pp. 803–824). IGI Global. https://doi.org/10.4018/IJDWM.2018010103.
5. Arafat, M. (2020). *A Review of Models for Hydrating Large-scale Twitter Data of COVID-19-related Tweets for Transportation Research*. https://doi.org/10.31124/advance.12192693.v1.
6. Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *Ai & Society, 30*, 89–116.
7. Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion, 28*, 45–59.
8. Bisanzio, D., Kraemer, M. U., Bogoch, I. I., Brewer, T., Brownstein, J. S., & Reithinger, R. (2020). Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. *Geospatial health, 15*(1). https://doi.org/10.4081/gh.2020.882.
9. Blank, G., & Lutz, C. (2017). Representativeness of social media in great britain: Investigating Facebook, Linkedin, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist, 61*(7), 741–756.
10. Bonzanini, M. (2016). *Mastering social media mining with Python*. Packt Publishing Ltd. ISBN 1783552026, 9781783552023
11. Bosco, C., Patti, V., & Bolioli, A. (2013). Developing corpora for sentiment analysis: The case of irony and senti-tut. *IEEE intelligent systems, 28*(2), 55–63.
12. Buchmann, Katrin; Robison, Rosalyn A. V.; Foulds, Chris (2017). Transport sector decarbonisation - a social sciences and humanities annotated bibliography. Anglia Ruskin Research Online (ARRO). Report. https://hdl.handle.net/10779/aru.23783613.v1
13. Cosgrave, M. (2021). Digital humanities methods as a gateway to inter and transdisciplinary. *Global Intellectual History, 6*(1), 24–33.
14. Dadhich, Abhinav. Practical Computer Vision: Extract Insightful Information from Images Using TensorFlow, Keras, and OpenCV. Packt Publishing Ltd, 2018. ISBN 1788294769, 9781788294768
15. Das, S., Dutta, A., Medina, G., Minjares-Kyle, L., & Elgart, Z. (2019). Extracting patterns from Twitter to promote biking. *IATSS research, 43*(1), 51–59.
16. Del Vecchio, P., Mele, G., Ndou, V., & Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. *Information Processing & Management, 54*(5), 847–860.
17. Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment analysis of review datasets using naive bayes and k-nn classifier. *arXiv preprint* arXiv:1610.09982.
18. Dimovska, J., Angelovska, M., Gjorgjevikj, D., Madjarov, G. (2018). Sarcasm and Irony Detection in English Tweets. In: Kalajdziski, S., Ackovska, N. (eds) ICT Innovations 2018. Engineering and Life Sciences. ICT 2018. Communications in Computer and Information Science, vol 940. Springer, Cham. https://doi.org/10.1007/978-3-030-00825-3_11
19. Dormanesh, A., Majmundar, A., & Allem, J.-P. (2020). Follow-up investigation on the promotional practices of electric scooter companies: Content analysis of posts on Instagram and Twitter. *JMIR public health and surveillance, 6*(1), e16833.
20. European Commission. (1992). *GREEN PAPER on the impact of Transport on the Environment - A Community strategy for "sustainable mobility"*. https://op.europa.eu/en/publication-detail/-/publication/98dc7e2c-6a66-483a-875e-87648c1d75c8/language-en
21. Fersini, E., Pozzi, F. A., & Messina, E. (2015). *Detecting irony and sarcasm in microblogs: The role of expressive signals and ensemble classifiers*. 2015 IEEE international conference on data science and advanced analytics (DSAA). https://doi.org/10.1109/DSAA.2015.7344888.
22. Furlong, Mark (2010). 'Clear at a distance, jumbled up close': observation, immersion and reflection in the process that is creative research. Deakin University. Chapter. https://hdl.handle.net/10536/DRO/DU:30042512
23. Geels, F. W. (2011). The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental innovation and societal transitions, 1*(1), 24–40.

24. Geels, F. W. (2012). A socio-technical analysis of low-carbon transitions: Introducing the multi-level perspective into transport studies. *Journal of transport geography, 24*, 471–482.

25. Haas, T., Jürgens, I., & Brunnengräber, A. (2020). Die Corona-Pandemie als Transformationsbeschleuniger. Die Auswirkungen der Krise auf die Verkehrswende in Deutschland. *Forschungsjournal Soziale Bewegungen, 33*(4), 834–843.

26. Holden, E., Banister, D., Gössling, S., Gilpin, G., & Linnerud, K. (2020). Grand Narratives for sustainable mobility: A conceptual review. *Energy Research & Social Science, 65*, 101454.

27. Holden, E., Gilpin, G., & Banister, D. (2019). *Sustainable mobility at thirty. Sustainability, 11*(7), 1965.

28 Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we instagram: A first analysis of instagram photo content and user types. *Proceedings of the International AAAI Conference on Web and Social Media., 8*(1), 595–598.

29. Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the COVID-19 pandemic. PLoS One, 15(11), e0241957.

30. Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of higher education, 6*(5), 26–41.

31. Kühl, N., Goutier, M., Ensslen, A., & Jochem, P. (2019). Literature vs. Twitter: Empirical insights on customer needs in e-mobility. *Journal of cleaner production, 213*, 508–520.

32. Kumar, A., Sangwan, S.R., Nayyar, A. (2020). Multimedia Social Big Data: Mining. In: Tanwar, S., Tyagi, S., Kumar, N. (eds) Multimedia Big Data Computing for IoT Applications. Intelligent Systems Reference Library, vol 163. Springer, Singapore. https://doi.org/10.1007/978-981-13-8759-3_11

33. Roy Ka-Wei Lee, Tuan-Anh Hoang, and Ee-Peng Lim. 2017. On Analyzing User Topic-Specific Platform Preferences Across Multiple Social Media Sites. In Proceedings of the 26th International Conference on World Wide Web (WWW '17). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1351–1359. https://doi.org/10.1145/3038912.3052614

34. Murthy, D., Gross, A., & McGarry, M. (2016). Visual social media and big data. Interpreting Instagram images posted on Twitter. *Digital Culture & Society, 2*(2), 113–134.

35 Mustofa, R., & Prasetiyo, B. (2021). Sentiment analysis using lexicon-based method with naive bayes classifier algorithm on# newnormal hashtag in twitter. *Journal of Physics: Conference Series., 1918*(4), 042155.

36. Nam, M., Lee, E., & Shin, J. (2015). A method for user sentiment classification using Instagram hashtags. *Journal of Korea Multimedia Society, 18*(11), 1391–1399.

37. NLTK Project. (2015). *Twitter Samples*. https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/packages/corpora/twitter_samples.zip

38. Oliverio, J. (2018). A survey of social media, big data, data mining, and analytics. *Journal of Industrial Integration and Management, 3*(03), 1850003.

39. Qi, B., Costin, A., & Jia, M. (2020). A framework with efficient extraction and analysis of Twitter data for evaluating public opinions on transportation services. *Travel behaviour and society, 21*, 10–23.

40. Anna Izabel João Tostes Ribeiro, Thiago Henrique Silva, Fátima Duarte-Figueiredo, and Antonio A.F. Loureiro. 2014. Studying traffic conditions by analyzing foursquare and instagram data. In Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks (PE-WASUN '14). Association for Computing Machinery, New York, NY, USA, 17–24. https://doi.org/10.1145/2653481.2653491

41. Rogers, Richard. "Digital methods for cross-platform analysis." The SAGE handbook of social media (2017): 91-110. SAGE Publications Ltd, 2017 - 662 p. - ISBN: 9781473995802 - Permalink: http://digital.casalini.it/9781473995802 - Casalini id: 5018793

42. Salminen, J., Hopf, M., Chowdhury, S. A., Jung, S.-G., Almerekhi, H., & Jansen, B. J. (2020). Developing an online hate classifier for multiple social media platforms. *Human-centric Computing and information Sciences, 10*(1), 1–34.

43. Samah, K. A. (2021). Naïve Bayes Twitter Sentiment Analysis In Visualizing The Reputation Of Communication Service Providers: During Covid-19 Pandemic. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12*(5), 1753–1764.

44. Seyfi, M., & Soydaş, A. U. (2017). Instagram stories from the perspective of narrative transportation theory. *The Turkish online journal of design, art and communication, 7*(1), 47–60.

45. Singh, A., Halgamuge, M. N., & Moses, B. (2019). An Analysis of Demographic and Behavior Trends Using Social Media: Facebook, Twitter, and Instagram. Social Network Analytics, 87–108. https://doi.org/10.1016/B978-0-12-815458-8.00005-0

46. Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-centric Computing and information Sciences, 7*, 1–12.

47. Sinnott, R. O., Gong, Y., Chen, S., & Rimba, P. (2018). Urban Traffic Analysis Using Social Media Data on the Cloud. *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*. https://doi.org/10.1109/UCC-Companion.2018.00047.

48. Song, J., Kim, K. T., Lee, B., Kim, S., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. *KSII Transactions on Internet and Information Systems (TIIS), 11*(6), 2996–3011.

49. Sovacool, B. K. (2014). What are we doing here? Analyzing fifteen years of energy scholarship and proposing a social science research agenda. *Energy Research & Social Science, 1*, 1–29.

50. Sudira, H., Diar, A. L., & Ruldeviyani, Y. (2019). *Instagram sentiment analysis with naive bayes and KNN: exploring customer satisfaction of digital payment services in Indonesia*. 2019 International Workshop on Big Data and Information Security (IWBIS). https://doi.org/10.1109/IWBIS.2019.8935700.

51. Sujon, M., & Dai, F. (2021). Social Media Mining for Understanding Traffic Safety Culture in Washington State Using Twitter Data. *Journal of Computing in Civil Engineering, 35*(1), 04020059.

52. Taylor, S. J., & Bogdan, R. (1984). *Introduction to qualitative research methods: The search for meanings*. Wiley-Interscience. ISBN 0-471-88947-4

53. Tufekci, Zeynep. 2014. "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls". Proceedings of the International AAAI Conference on Web and Social Media 8 (1):505-14. https://doi.org/10.1609/icwsm.v8i1.14517.

54. Wisdom, V., & Gupta, R. (2016). *An introduction to twitter data analysis in python. Artigence Inc*. https://doi.org/10.13140/RG.2.2.12803.30243.

55. Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social media mining: An introduction*. Cambridge University Press.

56. Zhao, X., Ke, Y., Zuo, J., Xiong, W., & Wu, P. (2020). Evaluation of sustainable transport research in 2000–2019. *Journal of Cleaner Production, 256*, 120404.

Publisher's Note