

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Full Length Article

Public perception of electric vehicles on reddit over the past decade

Tao Ruan^{*}, Qin Lv

Department of Computer Science, University of Colorado Boulder, 430 UCB, Boulder, CO, 80309, USA



ARTICLE INFO

Keywords:

Reddit
Electric vehicles
Public perception
Online social network
Topic modeling
Sentiment analysis

ABSTRACT

Understanding public perception of electric vehicles (EVs) is imperative for increasing EV adoption, which can significantly reduce greenhouse gas (GHG) emissions, thereby mitigating climate change and global warming. While most existing research characterizes public perception of EV in the context of surveys, questionnaires, or interviews, our study leverages Reddit online social network (OSN) data to capture EV public perception at a much larger scale. We have collected 3,437,917 Reddit posts (including 274,979 submissions and 3,162,938 comments) between January 2011 and December 2020 relevant to EVs and analyzed them along several axes to understand how EVs are perceived by the public on Reddit through the following research questions: (1) What EV-related topics have been discussed by Reddit users? Whether/how Reddit users' interest in different topics has changed during 2011–2020? (2) What sentiment do Reddit users hold towards EVs? Whether public sentiment on Reddit has shifted over the past 10 years? (3) Whether/how do various Reddit communities (i.e., subreddits) have different perceptions of EVs? Our analysis evinces the potential of utilizing a large-scale OSN dataset for demonstrating a much wider spectrum of topics that the public is interested in than previous studies show, reveals fringe communities including r/conspiracy have many (controversial) discussions on EVs, especially on the environmental impacts of EVs, and one political community (r/The_Donald) has similar patterns with fringe communities in both sentiment and topic aspects. By answering these research questions, we aim to develop a more comprehensive understanding of the public perception of EVs in the past decade.

1. Introduction

Climate change and global warming are threatening the survival of humans and wild animals in multiple ways, due to their significant challenges in the ecosystem (van der Geest et al., 2019), food safety (Tirado et al., 2010), infectious disease management (Khasnis and Nettleman, 2005), and many other aspects. They are attributed to the human expansion of “greenhouse effect” by scientists (Oreskes, 2004; Karl et al., 2009; Pachauri et al., 2014) and the United Nations regards “carbon neutrality by 2050” as the world's most urgent mission (Guterres, 2020). Prior research indicates that the transportation sector contributes to approximately 20% of all carbon dioxide globally and road transportation accounts for the large majority of the emissions (Albuquerque et al., 2020). The proportion even rises to 29% for the United States as of 2020 (United States Environmental Protection Agency, 2020). Electric vehicles (EVs), including battery electric vehicles (BEVs) and plug-in hybrid electric vehicles (PHEVs), are often labeled “green” in that they can significantly reduce greenhouse gas (GHG) emissions (Thomas, 2012). Increasing EV adoption has become a preminent national public

priority as demonstrated by the Biden administration's objectives to pledge EV sales of between 40% and 50% by 2030 (Keith et al., 2021; The White House, 2021). However, motivating the general population to adopt EVs is by no means an easy task and previous studies indicate many obstacles need to be overcome before EVs will be widely adopted (Egbue and Long, 2012; Vassileva and Campillo, 2017). Therefore, understanding the public perception of EVs becomes imperative to reveal how people feel about the new technologies and what their main concerns are. Previous research on public perception of EVs mainly relies on questionnaires, interviews, or customer surveys (Ziefle et al., 2014; She et al., 2017; Bunce et al., 2014; Khandakar et al., 2020). Those methods have been proved to be valid and reliable in many other areas if designed carefully (Greenberg and Weiner, 2014), but they also have the following limitations:

- 1). Most of the questionnaires, interviews, and customer surveys are conducted in specific areas or are sent to specific groups of people (e.g., conventional vehicle owners or current electric vehicle owners). The sample size is typically limited to a few hundred or

^{*} Corresponding author.

E-mail addresses: tao.ruan@colorado.edu (T. Ruan), qin.lv@colorado.edu (Q. Lv).

<https://doi.org/10.1016/j.commtr.2022.100070>

Received 4 April 2022; Received in revised form 31 May 2022; Accepted 31 May 2022

Available online 18 June 2022

2772-4247/© 2022 The Author(s). Published by Elsevier Ltd on behalf of Tsinghua University Press. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

thousand at most and the results cannot represent the general public perception in our society.

- 2). The questionnaires, interviews, or customer surveys are typically conducted in a specific (short) time period, e.g., a couple of weeks or months. Therefore, those studies can only reflect people's responses towards EVs within a fairly limited time frame and fail to reveal the dynamic public perception over a longer time period.
- 3). As experimental studies, conclusions drawn from the methodologies mentioned above can suffer from "unsuspected or unassessed biases of the Hawthorne effect" (Cook, 1962), the phenomenon where study subjects behave differently because they know they are being observed by researchers (Salganik, 2019; Imai, 2018).

To overcome the limitations mentioned above, we investigate the general public perception of EVs through online social network (OSN) data in this paper. More specifically, we extract all EV-related discussions on the popular Reddit platform during 2011–2020, generating 274,979 Reddit submissions and 3,162,938 Reddit comments. To the best of our knowledge, this is the first and largest-scale EV discussion corpus in a 10-year period. Another advantageous characteristic of OSN data is non-reactivity, which means that users do not change their behaviors because they are either unaware that their data are being captured or so accustomed to this data collection (Salganik, 2019). With this motivation in mind, we identify and address the following research questions regarding public perception of electric vehicles on Reddit:

RQ1: What EV-related topics have been discussed by Reddit users and whether/how the users' interest in different topics has changed during 2011–2020?

RQ2: What sentiment do Reddit users hold towards EVs? Whether/how the public sentiment on Reddit has shifted over the past 10 years?

RQ3: Whether/how do various Reddit communities (i.e., subreddits) have different perceptions of EVs?

To answer these questions, with careful data preprocessing of the 3,162,938 Reddit comments, we conduct a comprehensive analysis of the topics, sentiment, temporal variations, and divergences among different subreddits. This work makes the following contributions:

- 1). We have constructed a pipeline to extract a large-scale OSN corpus related to EVs (274,979 Reddit submissions and 3,162,938 Reddit comments) and conducted the first research of its kind to explore public perception of EVs based on OSN data. The methodology we propose may be applied to other topics in transportation research (e.g., hydrogen vehicles).
- 2). We have identified a nearly exponentially increasing pattern in the public discussions of EVs on Reddit in the past 10 years. The public sentiment on EVs also become more stable due to the increasing volume of discussions.
- 3). We have revealed many EV-related topics that attract much public attention but are not discussed in previous studies, such as investment, other renewable energy, EV policies during elections, etc.
- 4). We have explored different communities' (i.e., subreddits) attention on diverse topics. We have found that many fringe communities, such as r/collapse, r/unpopularopinion, and r/conspiracy, also have a lot of discussions on EVs. To our surprise, r/The_Donald has a highly similar pattern with those communities, in both sentiment and topic aspects. Those subreddits have much interest in climate-related topics when discussing EVs, indicating that the climate influence of EVs can be a highly controversial topic in those communities.

2. Literature review

2.1. Research direction

2.1.1. Public perception of EVs

Recently, researchers have made significant progress in identifying influencing factors on the acceptance of EVs. Those studies typically utilize questionnaires (Schneiderei et al., 2015), interviews (Aksen et al., 2017; Delmonte et al., 2020), or surveys (Sintov et al., 2020) to analyze different types of attributes related to EV preference, such as personal social-economic attributes, financial attributes (e.g., EV purchase cost, fuel cost), technical attributes (e.g., range, charging time), infrastructure attributes (e.g., charging station availability), and policy attributes (e.g., rebate, subsidy) (Liao et al., 2017).

Liao et al. (2017) give a comprehensive literature review of research on consumer preferences for EVs and all the research works in the review are based on questionnaires, interviews, or surveys. They utilize models of attitude formation, such as the logit model (Horne et al., 2005; Potoglou and Kanaroglou, 2007; Hackbarth and Madlener, 2013; Valeri and Danielis, 2015; Helveston et al., 2015), hybrid choice model (Daziano, 2012; Jensen et al., 2013; Glerum et al., 2014; Kim et al., 2014), and latent class model (Hidrué et al., 2011; Bockarjova et al., 2014), to analyze the survey results and compare the importance of the attributes in terms of customer preference of EVs. These research works draw important conclusions and provide practical insights to EV manufacturers and policymakers but they have limitations. The survey-based research approach is commonly directed at a limited group of people in certain areas and the respondents range from hundreds to thousands of individuals. Moreover, surveys can only be performed in a fairly short time frame. Therefore, they are able to reflect people's attitudes in the given time period but they can hardly manifest the dynamic public perception in a long time. Our analysis overcomes these limitations by constructing and analyzing an extensive EV-related Reddit discourse corpus containing more than 3 million posts over a decade (from 2011 to 2020). Furthermore, as experimental studies, survey-based research suffers from the Hawthorne effect – respondents may change their behaviors when they realize they are being observed by researchers (Cook, 1962; Salganik, 2019; Imai, 2018). In contrast, Reddit posts and discussions are mostly spontaneous and our analysis can avoid the Hawthorne effect as most observational studies do.

2.2. Research methods

2.2.1. OSN analysis in energy and transportation

The advantages of online social network analysis discussed in the text above have attracted an increasing number of researchers to study social perception in energy and transportation areas.

Li et al. (2019) investigate the challenges and opportunities of using social media data (Twitter data) to understand the social perception of energy in depth. Rantala et al. (2020) conduct a social network analysis on a Facebook group to perform both quantitative and qualitative content exploration of open discussions on the reform of national-level energy policy in Finland. Chaniotakis et al. (2016) apply SWOT (strengths, weaknesses, opportunities, and threats) analysis and examine the applicability of social media data in transportation studies. Zayet et al. (2021) provide a systematic review of transportation-related research based on social media analysis during 2008–2018.

From the literature reviews, while there has been an increased interest in the use of social media data as important traffic and energy information sources (Lv et al., 2017), social media analysis on electric vehicle topics is rare (Carpenter et al., 2014; Zayet et al., 2021). Through this paper, we aim to bridge the gap in public responses research towards electric vehicle perception and adoption.

2.2.2. Topic modeling & sentiment analysis

Topic models, exemplified by latent Dirichlet allocation (LDA) (Blei

et al., 2003), are a useful tool to understand large corpus. Topic models can capture groups of words that usually appear in the same documents. Those words with the highest probabilities in one topic evince what this topic is about, and are therefore called keywords (Hu et al., 2014). Moreover, the topic modeling results can also represent each document in the corpus as a mixture of topics and thus a low-dimensional vector (i.e. topic weight vector) can be obtained for this document (Hu et al., 2014). This idea can be extended that we can use topic weight vectors to represent how similarly or dissimilarly different communities focus on the topics. In our paper, we perform the topic modeling using the wrapper for Mallet LDA (McCallum, 2002) in the Gensim toolkit (Rehurek and Sojka, 2010).

Sentiment analysis reveals the public perception from a different aspect as a computational treatment of people's opinions, sentiments, or subjectivity of text to explore the viewpoints of the authors on specific entities (Feldman, 2013; Medhat et al., 2014). Ribeiro et al. (2016) provide a comprehensive review of twenty-four popular sentiment analysis methods and present a benchmark comparison of them. In our analysis, we first apply several different methods, including Linguistic Inquiry and Word Count (LIWC) (Tausczik and Pennebaker, 2010), VADER (Valence Aware Dictionary for sEntiment Reasoning) (Hutto and Gilbert, 2014), AFINN (a list of English terms manually rated for valence with an integer between -5 (negative) and +5 (positive)) (Nielsen, 2011), on our dataset. We find despite the distinct mechanisms of these methods, they all evince consistent sentiment patterns over a long time period (10 years), which indicates they have identical performance in our large-scale dataset.

3. Materials and methodology

3.1. Reddit data collection and filtering

Our Reddit data was collected from Pushshift¹ (Baumgartner et al., 2020), which makes available all the submissions (i.e., RS dataset) and comments (i.e., RC dataset) posted on Reddit between June 2005 and December 2020. Pushshift maintains all the Reddit data in its database and releases monthly Reddit data. We, therefore, used the RS and RC over the past ten years (2011–2020). The Reddit datasets are from the whole platform and therefore only a small portion is EV-related discussions. We have carefully designed a filtering process to ensure adequate extraction of EV-related data.

Fig. 1 illustrates the workflow of our data filtering process. First, we traversed the RS (Reddit Submissions) dataset and used two sets of keywords to obtain all the EV-related submissions (274,979). The keywords are also displayed as **Keywords Filter** in Fig. 1: The first set is whether the original submission text contains 'EV' or 'EVs' and the second set is whether the lower-case submissions contain any of the keywords 'electric vehicle', 'electric car', 'electric truck', 'phev', 'plug-in hybrid', or 'plug-in ev'.

Based on the 'link_id' feature of those submissions, we retrieved the Subset A of EV-related Reddit comments by checking whether a comment shares the same 'link_id' as the EV-related submissions. Subset A includes all the comments under the 274,979 submissions. Meanwhile, we utilized the same **Keywords Filter** to get the Subset B of EV-related Reddit comments, whose submissions are not among the 274,979 submissions but the comment text body contains the keywords. Last, we checked whether the comments are posted in either r/electriccars or r/electricvehicles. Because these two subreddits focus specifically on EVs, we retained all the comments in these two subreddits as Subset C. We also kept all the submissions in these two subreddits and added them to EV-related Reddit submissions.

After merging Subset A, B, and C, we obtained the preliminary EV-related RC dataset. However, this preliminary collection still contains

"noisy" data. The main reason is that we used 'EV' and 'EVs' as our first set of keywords. 'EV' can have different meanings in other communities. For example, 'EV' means "Effort Value" in Pokémon, "Even strength" in Hockey, "Expect value" in poker and some other games, "Estradiol valerate" in the transgender community, as well as the abbreviation of some players in soccer, NBA, and other sports, etc. Therefore, further filtering is needed to remove these "noisy" data. We excluded the comments released in these communities through another set of subreddit lists: *hockey, stunfisk, Warframe, Barca, magicTCG, Dirtybomb, RWBY, FFBraveExvius, mtgfinance, vancouver, WonderTrade, MagicArena, live-sound, SVExchange, sto, swtor, TampaBayLightning, NintendoSwitch, path-of-exile, soccer, fantasyfootball, TheSilphRoad, easymons, gaming, evnova, spikes, lrcast, pcmasterrace, MTGO, VGC, qotsa, etheroll, senrankagura, DJs, AskMeFHRT, nba, cataclysmdda*. StarTrekTimelines, blackjack, Games, de, InsurgenceTrades, baseball, csgobetting, BreedingDittos, 3DS, MtF, battlebots, Beatmatch, 3dshacks, redsox, Stellaris, dankmemes, Chapo-TrapHouse, MHOC, DestinyTheGame, nfl, GiveMeTheVirus, hearthstone, dungeonddefenders. Besides the subreddits listed above, we also included other subreddits if their names include 'poke' or 'trans' so that we can exclude all the poker, Pokémon, and transgender-related communities.

After further filtering, we finally obtained 3,162,938 Reddit comments that are relevant to EVs. We randomly sampled the data, manually examined the text body, and found almost all the sampled posts are closely related to EVs. Fig. 2 shows the monthly number of Reddit submissions and comments between January 2011 and December 2020 after our filtering process.

Special patterns in 2020

From Fig. 2, we can observe a consistent pattern of significant increase in discussions between 2011 and 2019 for both Reddit comments and Reddit submissions, whereas 2020 is very different. We can see a dramatic decline in EV-related discussions on Reddit during the early months of 2020 and then an overwhelmingly rapid recovery afterward.

The different patterns we observe in 2020 are likely the result of two key events that occurred in 2020. The first is the COVID-19 (coronavirus disease-19) pandemic. Recent research (Ribeiro et al., 2021) has demonstrated how the COVID-19 pandemic has significantly impacted people's information-seeking behaviors online. People's attention may have shifted to other topics, especially at the beginning of the pandemic, causing a dramatic decrease in early 2020. The second is the 2020 U.S. presidential election. During the presidential election, both Trump and Biden promoted their support for electric vehicles, which was covered broadly by news media and therefore may also have triggered a multitude of discussions online in the second half of 2020.

Since 2020 is exceptionally different from previous years, in the following analysis, we always include a separate discussion about 2020 when presenting the analysis results.

3.2. Preprocessing

Before applying topic modeling on the Reddit posts, preprocessing is required. In our study, we follow a standard natural language processing pipeline to preprocess all the corpus data (Ruan et al., 2020). First, we removed all the mentions (@), hashtags(#), punctuation, and URLs using regular expressions. We also utilize the *simple_preprocess* function in the *gensim* python package for stripping tags, punctuation, multiple white-spaces, short words, digits, and removing stop words. In our analysis, all sentences are lower-cased, tokenized, and de-accented. Finally, a list of tokens are obtained for each Reddit post, and they are used for topic modeling. Meantime, all non-English tokens are removed based on the English dictionary.

3.3. Topic modeling and sentiment analysis

Topic modeling is a technique to identify the topics present in a large set of document collections. Latent Dirichlet Allocation (LDA) is an

¹ <https://pushshift.io>.

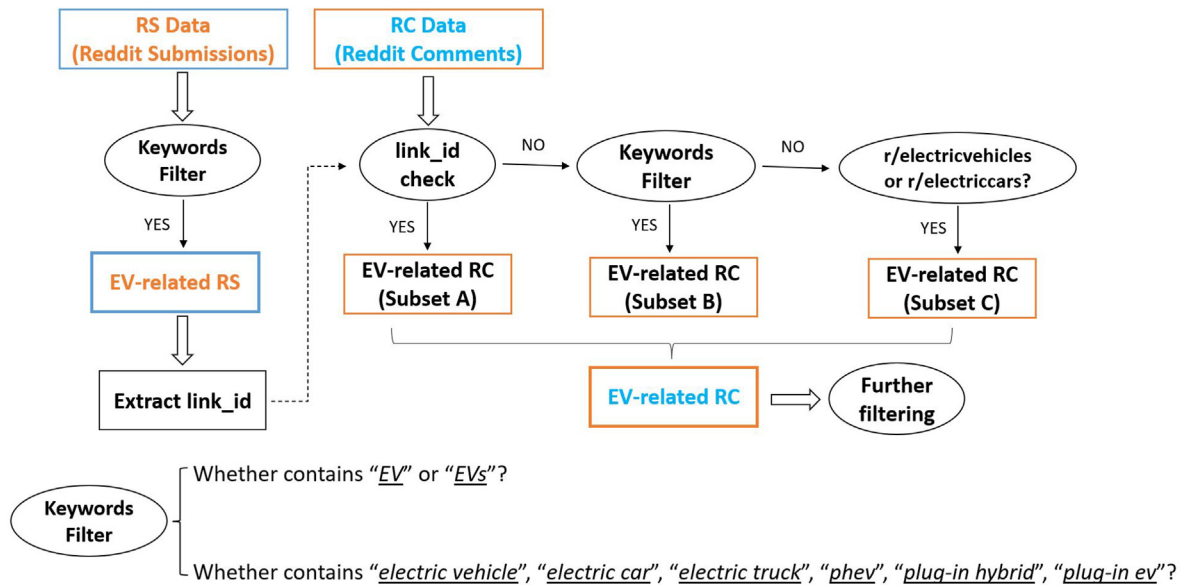


Fig. 1. Filtering EV-related posts from the Reddit RC/RS datasets. Note: RC is Reddit Comments and RS is Reddit Submissions.

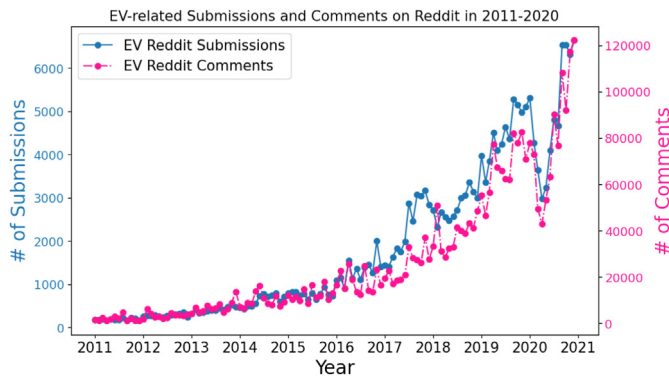


Fig. 2. Number of EV-related Reddit posts (submissions and comments, respectively) after filtering.

algorithm for text mining based on statistical (Bayesian) topic models and it has been widely used in text-based research. In this paper, we performed the LDA with the Mallet provided in the *Gensim* Python package. For LDA, the outputs include two matrices: the first matrix is the topic probability distributions over documents, represented by an $N \times K$ matrix; the other is the word probability distributions over topics, represented by a $K \times V$ matrix, where N is the number of documents, K is the number of topics and V is the size of vocabulary, namely the number of unique words in the whole corpus (Liu et al., 2016). The optimal number of topics can be determined using the coherence score, which measures the goodness of topic modeling results. A higher coherence score indicates that the topic result is more coherent (Mimno et al., 2011).

For sentiment analysis, we examined multiple sentiment scores introduced in the Related Work section and compared the temporal sentiment trend with different scores between 2011 and 2020. We found that, despite the various mechanisms of these sentiment scores, the overall trend is highly consistent, which implies that these methods perform similarly in our large-scale corpus. In the following analysis, we use the VADER score to measure public sentiment. More specifically, we utilize the *vaderSentiment* package in Python (Hutto and Gilbert, 2014) and calculate the VADER score for each post in our dataset. In order to obtain the overall monthly sentiment score, we compute the average VADER score of all posts in each month.

3.3.1. Combine topic modeling and sentiment analysis

As introduced above, LDA outputs the topic probability distributions over documents, which means each topic has a probability in each document. This provides us with the feasibility of combining topic modeling results with sentiment analysis. Following the methodology in the work by Yuan et al. (2020), we can define the initial topic weight as the sum of the probabilities of a topic in all documents (Eq. (1)). Then the topic weight can be computed as the ratio between the initial topic weight and the sum of all topics' initial topic weight (Eq. (2)).

$$\text{ToW}_i = \frac{\sum_{j=1}^{ND} P(T_i|D_j)}{\sum_{i=1}^{NT} \sum_{j=1}^{ND} P(T_i|D_j)} \quad (1)$$

$$\text{ToS}_i = \frac{\sum_{j=1}^{ND} P(T_i|D_j) \times \text{Sentiment}(D_j)}{\sum_{j=1}^{ND} P(T_i|D_j)} \quad (2)$$

Here, $P(T_i|D_j)$ is the probability of topic T_i in document D_j ; ND is the number of documents (i.e., Reddit Comments); NT is the number of topics; $\text{Sentiment}(D_j)$ is the sentiment of document D_j and in our analysis, we use VADER score to measure the sentiment of each document, which is between -1 (most extreme negative) and $+1$ (most extreme positive) (Hutto and Gilbert, 2014); ToW_i is the weight of topic T_i ; and ToS_i is the sentiment of topic T_i . As shown in the formulae, in general, a larger value of ToW_i indicates topic T_i attracts more public attention on Reddit and a larger value of ToS_i denotes Reddit users hold a more positive attitude towards topic T_i .

Fig. 3 is a flow chart illustrating the input and output of topic modeling, sentiment analysis, and their combination.

3.3.2. Topic popularity variation trends

Since our analysis is based on a long time range (10 years), it is important to explore whether people's interest in different topics (i.e., the popularity of topics) has changed over the past decade. We investigate two different approaches to measure the popularity of a certain topic: (1) we can assign each document the topic with the highest probability as the dominant topic and count the number of documents with this certain topic being the dominant topic, then the number of documents can be regarded as the popularity of this topic (Eq. (3)); (2) we can sum up the probability of each document belonging to a topic in a period of time as the topic weight and use the topic weight percentage in overall topic weight as the popularity of this topic (Eq. (4)).

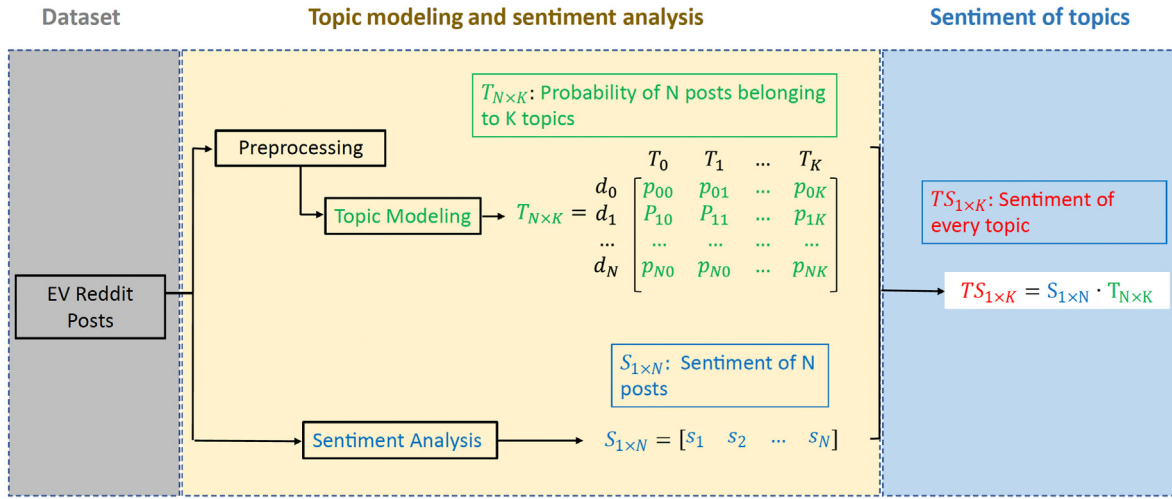


Fig. 3. Analysis flow chart (Topic modeling, sentiment analysis, and their combination.).

$$\text{Popularity_W}(ToP_i^k) = \frac{\sum_{j=1}^{ND_k} P(T_i|D_j)}{\sum_{i=1}^{NT} \sum_{j=1}^{ND_k} P(T_i|D_j)} \quad (3)$$

$$\text{Popularity_D}(ToP_i^k) = \frac{\sum_{j=1}^{ND_k} \mathbb{1}(\arg \max_T P(T|D_j) == T_i)}{ND_k} \quad (4)$$

where ND_k is the number of documents (i.e., Reddit comments) in the k-th month; ToP_i^k is the popularity of topic T_i in the k-th month. *Popularity_D* is the topic popularity defined by the number of dominant topics, and *Popularity_W* is the topic popularity defined by topic weight.

In order to test the evolutionary trends of the topic popularity, Mann-Kendall tests (Hamed and Rao, 1998), a non-parametric method used to identify a trend in a series, were performed on the temporal topic popularity. The null hypothesis for this test is that there is no monotonic trend in the series. The variation trend analysis provides a measure of long-run public perception of EVs, namely whether the public interest in diverse EV-related topics fades out or grows over time.

3.4. Subreddit analysis

Reddit has a unique mechanism: content on Reddit is organized in communities created by users, called “subreddits”, which indicate specific areas of interest. Users in the same subreddit typically share similar backgrounds or interests. Exploring how people from various areas talk about EVs differently also helps us obtain a better understanding of the public perception of EVs. In this work, we selected the top 50 subreddits with the most EV-related Reddit comments and performed a subreddit-level analysis based on the previous topic modeling and sentiment analysis results.

3.4.1. Hierarchical clustering of sentiment scores

Hierarchical clustering is a type of unsupervised machine learning algorithm that groups together data points with similar characteristics. We can use clustering to discover which subreddits have similar sentiments towards EVs. To perform hierarchical clustering, defining the distance between subreddits is essential. We utilized Wasserstein distance (del Barrio et al., 1999; Vallender, 1974) to calculate the sentiment score distributions’ divergence between two subreddits in our analysis. More specifically.

- Step 1:** Calculate the VADER score of each post in every subreddit;
- Step 2:** Calculate the Wasserstein distance of the score distributions between every two subreddits;

Step 3: Based on the pairwise subreddit distances, perform agglomerative clustering, a type of hierarchical clustering that iteratively merges similar subreddits into larger clusters.

After the hierarchical clustering is finished, a dendrogram is used to show the hierarchical relationship between objects (i.e., subreddits), which allows us to identify the best way to group subreddits into clusters.

3.4.2. Correlation matrix of topic distributions

Besides sentiment-based clustering, topic modeling results can reveal whether diverse communities focused on different topics during EV-related discussions. We examined the results of topic modeling for the comments in the 50 selected subreddits and calculated the topic weight for every topic. Then we computed the percentage of topic weight for each topic in every subreddit. In this way, we can generate a k-dimension topic weight distribution vector for each subreddit, where k is the number of topics.

Based on these k-dimension vectors, we can plot a heat map to indicate which topics attracted more attention from every subreddit. Meanwhile, the correlation matrix can be computed to indicate the pairwise similarity of topic weight distribution between every two subreddits, and thus the other heat-map can also be generated to illustrate how similar every two subreddits’ topic interests are during EV discussions.

4. Results

4.1. Topic modeling

We first generated the topic modeling results using Mallet on the preprocessed Reddit comments. Both coherence scores and manual checking are used to help decide the optimal number of topics (Grimmer et al., 2022). In this study, using 20 topics yields high coherence scores and the most explicit and comprehensive topics.

Fig. 4 shows the word clouds of the 20 topics identified. We see a wide range of words and topics being covered in the Reddit posts, indicating diverse EV-related interests of the online users. These topics are much more diverse than what previous studies have shown. For instance:

- 1). Potential customers talk about the price and brands of EVs (Topic 0: hybrid, petrol, affordable; Topic 5: price, expensive, buy; Topic 18: brand, ford, bolt), driving range, and charging time (Topic 8: range, charge, time).
- 2). Investors focus on the EV stock price (Topic 15: company, stock, market) and market performance (Topic 16: increase, rate, number).



Fig. 4. Word clouds for topic modeling results (Number of topics = 20).

3). Technophiles (those who are enthusiastic about new technology) discuss the EV speed, battery, and motor technologies (Topic 1: battery, motor, capacity; Topic 4: future, tech, design; Topic 14: speed, fast, engine).

4). Political users pay more attention to EV-related policies (Topic 12: tax, incentive, credit) and the presidential election (Topic 9: vote, government, policy).

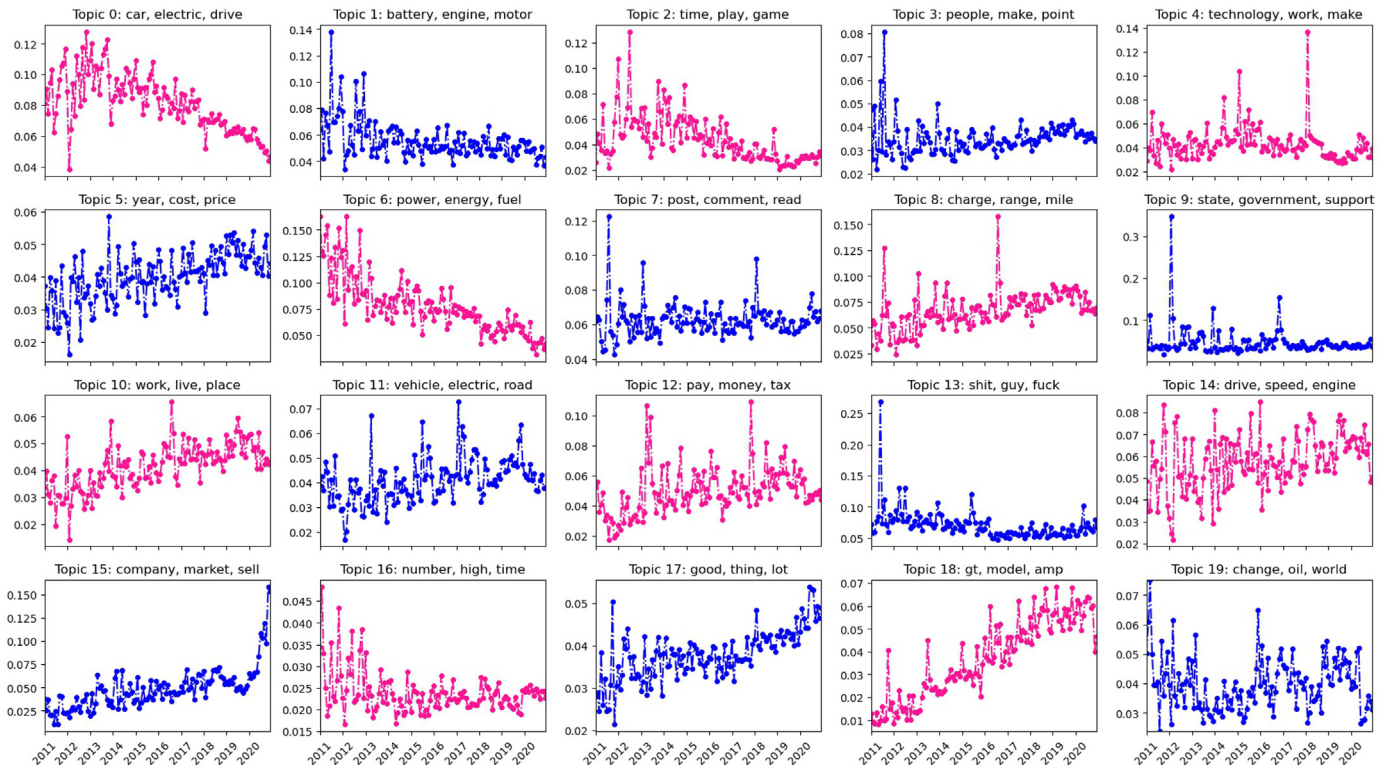


Fig. 5. Topic popularity variation trends during 2011–2020.

- 5). Environmentalists concentrate on EVs' impacts on climate change (Topic 19: climate, change, environment) and other renewable energy (Topic 6: emission, hydrogen, solar).

Topic popularity trends

In order to understand how people's interest in different topics has shifted in the past decade, we calculate the monthly topic popularity as defined in Section 3. In this paper, we present the topic weight as topic popularity but we also used the dominant topic and observed a consistent pattern.

Fig. 5 plots the monthly topic weight percentage for the 20 topics between 2011 and 2020. We only use the top three keywords to represent the topics in Fig. 5 but the topic order is the same as in Fig. 4. As we can see in Fig. 5, not all the topics share the same trend. Some topics, including Topic 18 (optional EV brands) and Topic 15 (EV stock performance) are attracting increasing attention from the public, while other issues such as Topic 6 (other renewable energy) have declined topic popularity on Reddit. Another noteworthy observation is that Topic 15 (EV stock performance) significantly increased in popularity in 2020. This is because the EV industry was one of the best performers in the stock market in 2020 and the EV stocks attracted substantial attention from investors. Meanwhile, Topic 18 is mainly about alternative EV brands such as Ford, Bolt, and this topic had steadily increased in popularity from 2011 to 2019 but decreased in 2020. This decrease in 2020 may be due to (1) most countries have implemented restrictions on mobility to prevent the spread of COVID (Oh et al., 2021) and thus caused a decline in the public enthusiasm for vehicle purchase; (2) the astonishing performance of Tesla on the stock market distracted people's attention from other EV brands. The trends of all the 20 topics are validated with Mann-Kendall tests (Hamed and Rao, 1998) introduced in the Materials and Methodology section, and they have consistent results with the visualization in Fig. 5.

4.2. Sentiment analysis

As described in Section 3, we examine people's sentiment toward EVs with monthly average sentiment scores in the past decade. Fig. 6 shows the VADER and AFFINE scores of EV-related Reddit comments in 2011–2020. We can observe that the two sentiment measures have consistent trends. We also investigated other sentiment scores, such as LIWC (Tausczik and Pennebaker, 2010) and textblob (Keen et al., 2020), and all of them exhibit the same patterns. This indicates that notwithstanding the distinct mechanisms of the sentiment scores, they have very similar outputs in our large-scale corpus. Without losing generality, we use VADER scores for further analysis.

It is worth noting that the emotion score becomes increasingly more stable in Fig. 6, which is because we have relatively few discussions in the early years and the average sentiment score can be fluctuating.

4.3. Combine topics and sentiment analysis

With the topic modeling results (topic distribution over each post) and sentiment analysis results (VADER score of each post), we apply Eqs. (1) and (2) to compute every topic's sentiment.

Fig. 7 illustrates the weighted sentiment of the 20 EV-related topics on Reddit. It is not surprising that Topic 13, which contains mainly negative curse words, has the lowest sentiment scores. However, we notice that Topic 19, which is about EV impacts on the climate change problem, has the second-lowest weighted sentiment scores. Our sub-reddit analysis in the next section provides further insights into this phenomenon – the climate change topic has been widely discussed in some fringe communities, such as r/unpopularopinion and r/conspiracy, indicating that whether EVs can help mitigate the climate change problem remains a controversial topic on Reddit and thus leads to many arguments.

Moreover, Fig. 7 shows that the technology (Topic 4: Technology, work, design), EV stock (Topic 15: Market, stock, sell), and renewable energy (Topic 6: Hydrogen, solar, energy) topics have the highest weighted sentiment scores, which indicate the public's optimism for these aspects of electric vehicles.

Besides the overall sentiment of each topic, we also perform an analysis of the polarization of public sentiment on each topic. Based on our literature survey, many different mechanisms have been used to measure the polarization of public opinions (Bramson et al., 2017; Bauer, 2019). As a matter of fact, the polarization measurement is still an active research topic up to now (Xing et al., 2022). In this research, we utilized aIn this research, we utilized one statistical measurement - coefficient of variance (CV), of the sentiment distribution, to reveal how polarized a topic is. The coefficient of variance is defined as population standard deviation divided by population mean, which can be used to describe the relative dispersion of data points around the mean value. A large CV value typically indicates more polarization. Besides coefficient of variance, we also used kurtosis, another commonly used statistical indicator for polarization measurement, as a reference.

Our polarization analysis indicated that political discussions (Topic 9: vote, government, policy) and environmental concerns (Topic 19: climate, change, environment) are the most polarized. In contrast, topics that discuss the price and brands of EVs (Topic 0: hybrid, petrol, affordable; Topic 5: price, expensive, buy; Topic 18: brand, ford, bolt), and feeling of EV (Topic 17: good, thing, lot) have a relatively consistent sentiment.

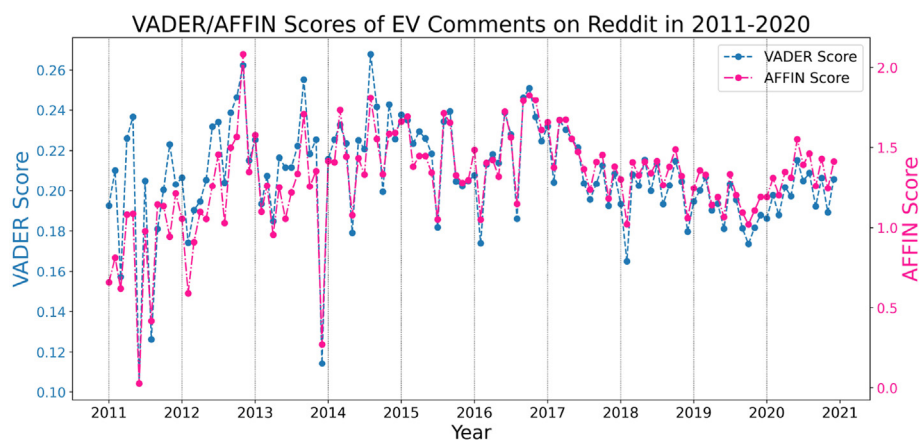


Fig. 6. Monthly VADER and AFFIN scores of EV Reddit comments over the past decade. AFFIN scores are added to show consistent trends between the different sentiment scoring mechanisms.

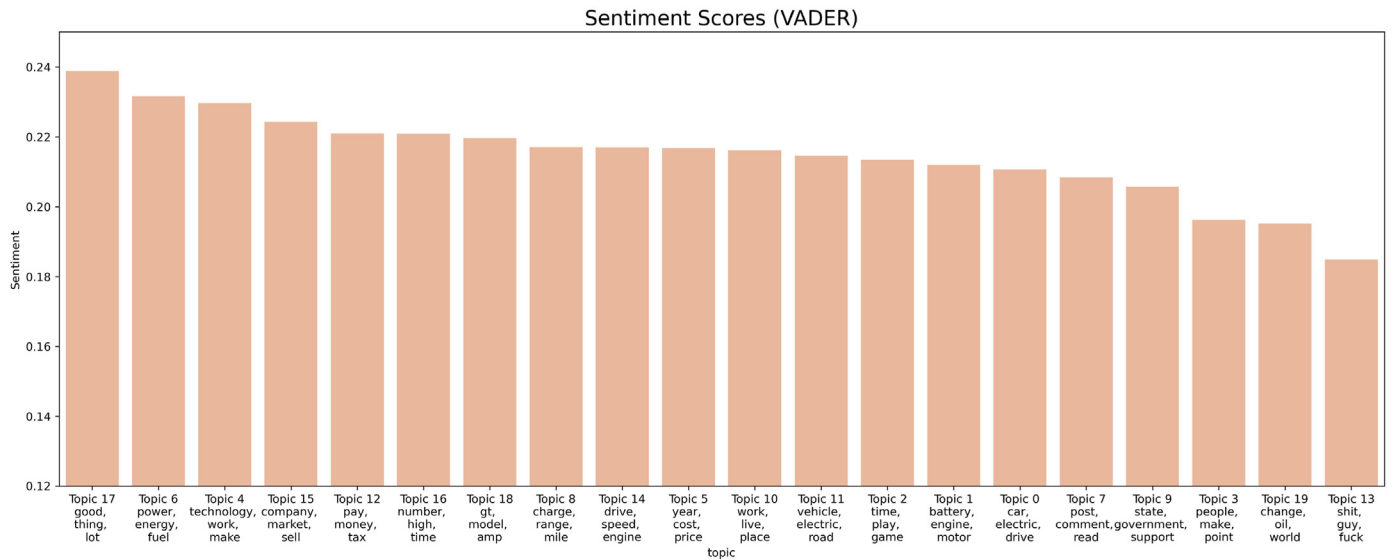


Fig. 7. Weighted sentiment of the 20 topics.

4.4. Subreddit analysis

Subreddit is a unique mechanism on Reddit. Posts on Reddit are organized in these user-created communities (i.e., “subreddits”), which are associated with certain areas of interest. This specific mechanism provides us with the possibility of studying public perception of EVs at the community level. Following the approaches described in Section 3, we delve into how different subreddits perceive EVs and whether certain subreddits have similar discussion patterns. We present the sentiment clustering and topic distribution correlation separately in the following

subsections.

4.4.1. Hierarchical sentiment clustering

Defining the Wasserstein distance between the sentiment score distributions in every two subreddits as the pairwise divergence, we perform hierarchical clustering to examine different communities’ sentiment toward EVs (Fig. 8).

We can easily observe many clusters with obvious similarities. For instance, r/BoltEV, r/leaf, and r/volt are close to each other because they are all common EV brands; r/europe, r/canada and r/australia belong to

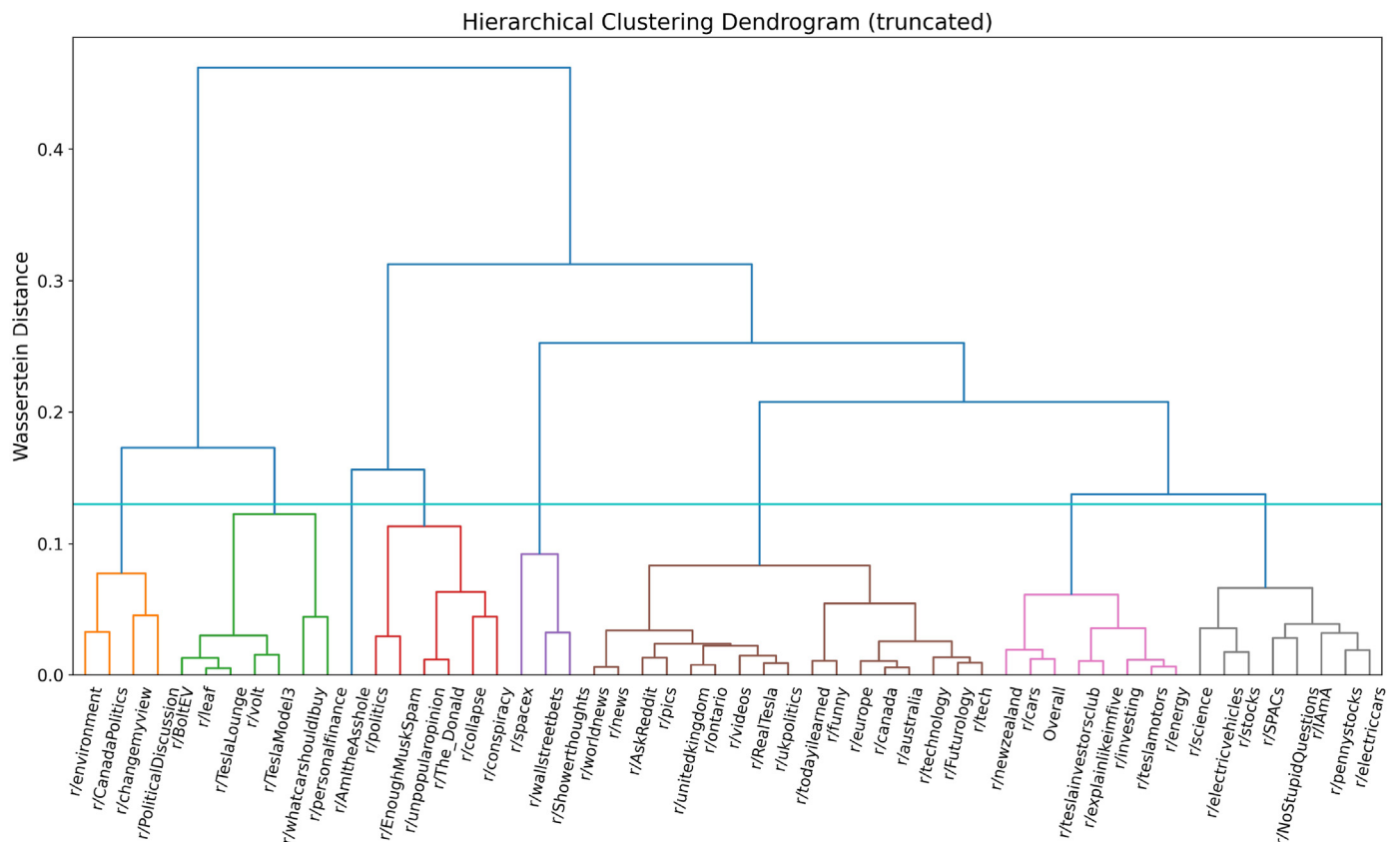


Fig. 8. Hierarchical clustering dendrogram based on pairwise Wasserstein distance of sentiment score distributions among top 50 subreddits.

the same cluster which indicates users inside these regional subreddits hold similar sentiments of EVs; r/tech, r/technology, and r/Futurology are communities interested in technology so their sentiment distributions are grouped together; potential customers tend to seek advice in r/personalfinance and r/whatcarshouldIbuy which also form a cluster. These clusters verify the effectiveness of our clustering method in capturing similarity of perception in different communities. It is notable that there is an interesting cluster in Fig. 8: r/The_Donald, r/collapse, r/conspiracy, and r/unpopularopinion are grouped together. r/The_Donald is a political community where more than 790 thousand users post memes, viral videos, and supportive messages about former U.S. President Donald Trump,² and the other three subreddits are fringe communities with alternative opinions or theories, while they have very similar sentiment distribution patterns.

4.4.2. Topic distribution correlation

Besides sentiment scores, the output of topic modeling can be used to examine different communities' interests as well. For every subreddit, we compute the topic weight for every topic and then calculate the percentage of each topic in this subreddit's overall weight. This percentage (i.e., topic distribution) reveals how each subreddit's users allocate their attention to different topics. After calculating the topic distributions for all subreddits, a correlation matrix can be obtained to measure how correlated every two subreddits are in terms of allocating their attention to different topics. In order to perform clustering based on the topic distribution, we define the topic distribution distance between two subreddits as $(1 - \text{correlation coefficient})$ and then apply hierarchical clustering on the distances.

Fig. 9 presents the correlation matrix based clustering of topic distributions. It can be clearly seen that the subreddits in the sentiment clustering also tend to have a high correlation in topic distributions and belong to the same clustering in the topic distribution clustering. In other words, certain subreddits not only hold similar sentiments towards EVs but also are interested in the same aspects of EVs.

5. Summary and impacts

Our work is the first to examine EV-related discourse on Reddit, a popular online social platform. Our analysis presents a comprehensive view on the public perceptions of electric vehicles on Reddit based on a large-scale corpus between 2011 and 2020. Besides capturing the full range of EV-related topics discussed on Reddit and sentiments, we have also investigated how the public attention on different topics changed over the course of the last decade, as well as a comparative study of different subreddits.

5.1. Impacts

Through the observational analysis of public perceptions of EVs on the Reddit platform, the main contributions of this work can be summarized as follows:

- 1). Our research reveals the feasibility of utilizing the spontaneous online social network (OSN) data to explore the public perception of electric vehicles, which complements traditional surveys, questionnaires, and interviews. Our approach leverages the advantages of OSN data (large-scale, longitudinal, and observational) and overcomes the potential shortcomings of traditional studies (limited-scale, cross-sectional, and experimental). These advantages allow us to obtain a more comprehensive view on public perceptions, explore the dynamic characteristics of public

responses and avoid the potential Hawthorne effect from traditional studies.

- 2). Our topic modeling analysis demonstrates a much wider spectrum of viewpoints that the public is interested in than that of previous studies, ranging from the investment value of the EV industry, other renewable energy-related to EV (e.g., hydrogen, solar), government policies, to public EV infrastructure, environmental impacts (e.g., climate change), etc. Our evolutionary analysis of topic popularity reveals that public interests in these topics have different trends.
- 3). Through the combination of topic modeling and sentiment analysis, we find that the environmental impact topics (e.g., climate change) have lower sentiment scores. The subreddit analysis also shows that the climate impact of EVs has plenty of discussions in fringe communities such as r/unpopularopinions, r/conspiracy, and r/collapse. This indicates that, despite the fact that environmental benefits of EV adoption have been expounded by many scientists and authoritative institutions, this topic remains controversial on online social networks. This observation also highlights a gap in accessible, trustworthy information resources online. Science and environment agencies need to put more effort into the popularization of the relevant science. Otherwise, people may turn to OSNs such as Reddit seeking information, which might be "polluted" by inaccurate information or even conspiracies. In addition, without authoritative engagement in these forums, misconceptions surrounding EVs may proliferate, fueling EV adoption hesitancy.
- 4). Our subreddit analysis based on hierarchical clustering of sentiment scores and correlation matrix of topic distribution propose new methods for examining the similarity of interest and sentiment on certain topics at a community level. In our EV-focused study, we manifest the effectiveness of the approaches and meanwhile gain quite a few interesting insights. For example, we find r/The_Donald has many similarities with the fringe communities such as r/collapse, r/unpopular-opinions, and r/conspiracy, in terms of both sentiment scores clustering or topic distribution correlation. Even though we cannot conclude Trump's supporters hold the same attitudes towards EV as the alternative-theory communities, our findings do indicate possible relation between political leaning and the fringe community preferences.
- 5). Our work can serve as a starting point for more analysis bringing about practical implications in future work. For example, our analysis revealed that various types of communities have active discussions on EVs, such as political subreddits and financial subreddits. Based on the different subreddits identified by our work, experts in different fields can look into the corresponding subreddits and further analyze the area in a more detailed manner. As an example, further analysis in the political subreddits can provide new insights on how the public responds to certain policies and therefore support practical policymaking to help promote EV adoption.
- 6). We found limited evidence of utilizing online social network data for studying public perception of emerging technologies in the transportation area, even less can be found for electric vehicles (Debnath et al., 2021). Our study can be regarded as the initial exploration encouraging the utilization of large-scale OSN datasets over a long time period to investigate the dynamic public perception of other topics in energy research, such as photovoltaic systems, hydrogen vehicles, etc.

5.2. Limitations and future work

There are several limitations in this study that need to be addressed in future work. First, the Reddit dataset provides sufficient temporal information but lacks geographical information. Public perceptions of EVs in different regions can vary due to the financial status, government

² This subreddit cannot be found on Reddit now since it was banned by Reddit acting against hate speech (Isaac, 2021).

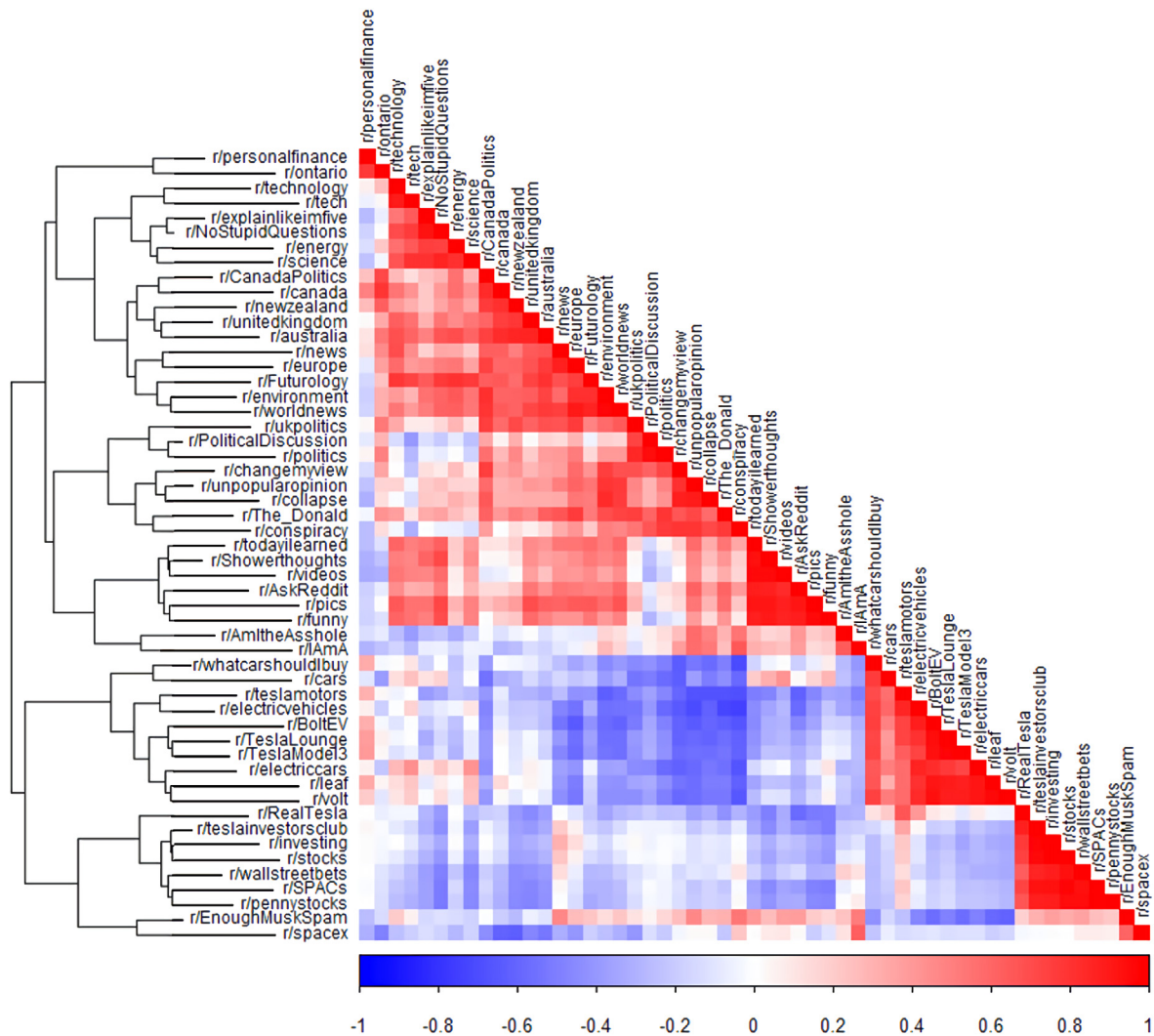


Fig. 9. Correlation matrix and hierarchical clustering based on topic distributions in each subreddit.

policies, political leanings, etc. Our study is unable to capture those important information. Second, Reddit data cannot represent the whole population. According to a telephone interview, 36 percent of respondents who indicated using Reddit were aged between 18 and 29 years old (Pew Research Center, 2021). Therefore, our study reveals the opinions of the relatively younger population. Third, the Reddit platform is anonymous and thus does not contain sufficient user information. Although we can gain some useful aggregated insights through the overall and subreddit analysis, the individuals' perception can also be valuable. For instance, it would be useful to understand how news media, government institutions, manufacturers, and individuals (e.g., experts, activists, celebrities) talk about EVs and how people respond to them. Last, we perform the polarization analysis of sentiment for each topic, however, our analysis is based on simple statistical measurements (i.e., coefficient of variance and Kurtosis) but sentiment polarization is a complex concept and also an active research topic. Therefore, the simple statistical measurement may not be sufficient.

As future work, those limitations can be addressed by the following approaches. First, we can combine OSN analysis with traditional survey-based investigations. Survey-based research can be manually designed for getting respondents' location, age, and other demographic information. The combination of OSN data and survey data can therefore complement each other. Second, we can use multi-source or cross-platform OSN datasets to mitigate the limitations. Some OSN data provides more user demographic information such as Twitter and Instagram. Based on

multiple OSN platforms, we can potentially obtain a more comprehensive understanding of the public perceptions of EVs. Third, more advanced methodologies should be considered for the public sentiment polarization analysis in future work. Lastly, dynamic topic modeling is a useful tool for revealing different topics of interest during different periods, especially over a long time. In our research, in order to combine the results of topic modeling and sentiment analysis, we do not use dynamic topic modeling. In future work, dynamic topic modeling can be used and a new methodology can be created to combine the dynamic topic modeling and sentiment analysis.

Replication and data sharing

The raw data used in the paper can be downloaded from the Pushshift website (<https://files.pushshift.io/reddit/>). The replication code and sample secondary datasets are available at <https://github.com/TaoRuan-Campus/EVReddit>. The complete secondary data can be easily reproduced with the steps in the paper.

Declarations of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the National Science Foundation (NSF) under Grant No. 1941524 for the NSF ASPIRE ERC. Thanks to Dr. Xiaodong Qian from Purdue University for proofreading the article. Also, thanks to Jason Baumgartner from PushShift for the data.

References

- Albuquerque, F.D., Maraqa, M.A., Chowdhury, R., Mauga, T., Alzard, M., 2020. Greenhouse gas emissions associated with road transport projects: current status, benchmarking, and assessment tools. *Transport. Res. Procedia* 48, 2018–2030.
- Axsen, J., Langman, B., Goldberg, S., 2017. Confusion of innovations: mainstream consumer perceptions and misperceptions of electric-drive vehicles and charging programs in Canada. *Energy Res. Social Sci.* 27, 163–173.
- Bauer, P.C., 2019. Conceptualizing and Measuring Polarization: A Review.
- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J., 2020. The pushshift reddit dataset. In: *Proceedings of the International AAAI Conference on Web and Social Media*, 14, pp. 830–839.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *JMLR* 3 (Jan), 993–1022.
- Bockarjova, M., Rietveld, P., Knockaert, J., Steg, L., 2014. Dynamic Consumer Heterogeneity in Electric Vehicle Adoption. Technical report.
- Bramson, A., Grim, P., Singer, D.J., Berger, W.J., Sack, G., Fisher, S., Flocken, C., Holman, B., 2017. Understanding polarization: meanings, measures, and model evaluation. *Philos. Sci.* 84 (1), 115–159.
- Bunce, L., Harris, M., Burgess, M., 2014. Charge up then charge out? drivers' perceptions and experiences of electric vehicles in the UK. *Transport. Res. Pol. Pract.* 59, 278–287.
- Carpenter, T., Golab, L., Syed, S.J., 2014. Is the grass greener? mining electric vehicle opinions. In: *Proceedings of the 5th International Conference on Future Energy Systems*, pp. 241–252.
- Chaniotakis, E., Antoniou, C., Pereira, F., 2016. Mapping social media for transportation studies. *IEEE Intell. Syst.* 31 (6), 64–70.
- Cook, D.L., 1962. The Hawthorne effect in educational research. *Phi Delta Kappan* 44 (3), 116–122.
- Daziano, R.A., 2012. Taking account of the role of safety on vehicle choice using a new generation of discrete choice models. *Saf. Sci.* 50 (1), 103–112.
- Debnath, R., Bardhan, R., Reiner, D.M., Miller, J., 2021. Political, economic, social, technological, legal and environmental dimensions of electric vehicle adoption in the United States: a social-media interaction analysis. *Renew. Sustain. Energy Rev.* 152, 111707.
- del Barrio, E., Giné, E., Matrán, C., 1999. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* 27 (2), 1009–1071.
- Delmonte, E., Kinnear, N., Jenkins, B., Skippon, S., 2020. What do consumers think of smart charging? Perceptions among actual and potential plug-in electric vehicle adopters in the United Kingdom. *Energy Res. Social Sci.* 60, 101318.
- Egbue, O., Long, S., 2012. Barriers to widespread adoption of electric vehicles: an analysis of consumer attitudes and perceptions. *Energy Pol.* 48, 717–729.
- Feldman, R., 2013. Techniques and applications for sentiment analysis. *Commun. ACM* 56 (4), 82–89.
- Glerum, A., Stankovikj, L., Thémans, M., Bierlaire, M., 2014. Forecasting the demand for electric vehicles: accounting for attitudes and perceptions. *Transport. Sci.* 48 (4), 483–499.
- Greenberg, M.R., Weiner, M.D., 2014. Keeping surveys valid, reliable, and useful: a tutorial. *Risk Anal.* 34 (8), 1362–1375.
- Grimmer, J., Roberts, M.E., Stewart, B.M., 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.
- Guterres, A., 2020. Carbon Neutrality by 2050: the World's Most Urgent Mission.
- Hackbarth, A., Madlener, R., 2013. Consumer preferences for alternative fuel vehicles: a discrete choice analysis. *Transport. Res. Transport Environ.* 25, 5–17.
- Hamed, K.H., Rao, A.R., 1998. A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol.* 204 (1–4), 182–196.
- Helveston, J.P., Liu, Y., Feit, E.M., Fuchs, E., Klampfl, E., Michalek, J.J., 2015. Will subsidies drive electric vehicle adoption? Measuring consumer preferences in the U.S. and China. *Transport. Res. Pol. Pract.* 73, 96–112.
- Hidru, M.K., Parsons, G.R., Kempton, W., Gardner, M.P., 2011. Willingness to pay for electric vehicles and their attributes. *Resour. Energy Econ.* 33 (3), 686–705.
- Horne, M., Jaccard, M., Tiedemann, K., 2005. Improving behavioral realism in hybrid energy-economy models using discrete choice studies of personal transportation decisions. *Energy Econ.* 27 (1), 59–77.
- Hu, Y., Boyd-Graber, J., Satinoff, B., Smith, A., 2014. Interactive topic modeling. *Mach. Learn.* 95 (3), 423–469.
- Hutto, C., Gilbert, E., 2014. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8.
- Imai, K., 2018. *Quantitative Social Science: an Introduction*. Princeton University Press.
- Isaac, M., 2021. Reddit, Acting against Hate Speech, Bans 'The Donald' Subreddit.
- Jensen, A.F., Cherchi, E., Mabit, S.L., 2013. On the stability of preferences and attitudes before and after experiencing an electric vehicle. *Transport. Res. Transport Environ.* 25, 24–32.
- Karl, T.R., Melillo, J.M., Peterson, T.C., Hassol, S.J., 2009. *Global Climate Change Impacts in the United States*. Cambridge University Press.
- Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., et al., 2020. Textblob: Simplified Text Processing. Accessed on 11.13.2021.
- Keith, L., Gabrielle, C., Jennifer, A.D., 2021. Biden Seeking Pledge for 40% of Car Sales to Be EV by 2030.
- Khandakar, A., Rizqullah, A., Ashraf Abdou Berbar, A., Rafi Ahmed, M., Iqbal, A., Chowdhury, M.E., Uz Zaman, S., 2020. A case study to identify the hindrances to widespread adoption of electric vehicles in Qatar. *Energies* 13 (15).
- Khasnis, A.A., Nettleman, M.D., 2005. Global warming and infectious disease. *Arch. Med. Res.* 36 (6), 689–696.
- Kim, J., Rasouli, S., Timmermans, H., 2014. Expanding scope of hybrid choice models allowing for mixture of social influences and latent attitudes: application to intended purchase of electric cars. *Transport. Res. Pol. Pract.* 69, 71–85.
- Li, R., Crowe, J., Leifer, D., Zou, L., Schoof, J., 2019. Beyond big data: social media challenges and opportunities for understanding social perception of energy. *Energy Res. Social Sci.* 56, 101217.
- Liao, F., Molin, E., van Wee, B., 2017. Consumer preferences for electric vehicles: a literature review. *Transport Rev.* 37 (3), 252–275.
- Liu, L., Tang, L., Dong, W., Yao, S., Zhou, W., 2016. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus* 5 (1), 1–22.
- Lv, Y., Chen, Y., Zhang, X., Duan, Y., Li, N.L., 2017. Social media based transportation research: the state of the work and the networking. *IEEE/CAA J Automatica Sinica* 4 (1), 19–26.
- McCallum, A.K., 2002. Mallet: a machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- Medhat, W., Hassan, A., Korashy, H., 2014. Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* 5 (4), 1093–1113.
- Mimmo, D., Wallach, H., Talley, E., Leenders, M., McCallum, A., 2011. Optimizing semantic coherence in topic models. In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 262–272.
- Nielsen, F.Å., 2011. A New Anew: Evaluation of a Word List for Sentiment Analysis in Microblogs. *arXiv preprint arXiv:1103.2903*.
- Oh, J., Lee, H.-Y., Khuong, Q.L., Markuns, J.F., Bullen, C., Barrios, O.E.A., Hwang, S.-s., Suh, Y.S., McCool, J., Kachur, S.P., et al., 2021. Mobility restrictions were associated with reductions in COVID-19 incidence early in the pandemic: evidence from a real-time evaluation in 34 countries. *Sci. Rep.* 11 (1), 1–17.
- Oreskes, N., 2004. The scientific consensus on climate change. *Science* 306 (5702), 1686–1686.
- Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., et al., 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ipcc.
- Pew Research Center, 2021. Percentage of U.S. Adults Who Use Reddit as of February 2021, by Age Group.
- Potoglou, D., Kanaroglou, P.S., 2007. Household demand and willingness to pay for clean vehicles. *Transport. Res. Transport Environ.* 12 (4), 264–274.
- Rantala, S., Toikka, A., Pulkka, A., Lyytimäki, J., 2020. Energetic voices on social media? Strategic Niche Management and Finnish Facebook debate on biogas and heat pumps. *Energy Res. Social Sci.* 62, 101362.
- Rehurek, R., Sojka, P., 2010. Software framework for topic modelling with large corpora. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Citeseer.
- Ribeiro, F.N., Araújo, M., Gonçalves, P., Gonçalves, M.A., Benevenuto, F., 2016. Sentibench—a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Sci.* 5 (1), 1–29.
- Ribeiro, M.H., Gligorić, K., Peyrard, M., Lemmerich, F., Strohmaier, M., West, R., 2021. Sudden attention shifts on Wikipedia during the COVID-19 crisis. In: *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, pp. 208–219.
- Ruan, T., Kong, Q., Zhang, Y., McBride, S.K., Lv, Q., 2020. An analysis of Twitter responses to the 2019 Ridgecrest earthquake sequence. In: *2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCLOUD/SocialCom/SustainCom)*. IEEE, pp. 810–818.
- Salganik, M.J., 2019. *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.
- Schneiderei, T., Franke, T., Günther, M., Krems, J.F., 2015. Does range matter? exploring perceptions of electric vehicles with and without a range extender among potential early adopters in Germany. *Energy Res. Social Sci.* 8, 198–206.
- She, Z.-Y., Sun, Q., Ma, J.-J., Xie, B.-C., 2017. What are the barriers to widespread adoption of battery electric vehicles? A survey of public perception in Tianjin, China. *Transport Pol.* 56, 29–40.
- Sintov, N.D., Abou-Ghalioum, V., White, L.V., 2020. The partisan politics of low-carbon transport: why democrats are more likely to adopt electric vehicles than Republicans in the United States. *Energy Res. Social Sci.* 68, 101576.
- Tausczik, Y.R., Pennebaker, J.W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lit. Semant.* 29 (1), 24–54.
- The White House, 2021. FACT SHEET: President Biden Announces Steps to Drive American Leadership Forward on Clean Cars and Trucks.
- Thomas, C.S., 2012. How green are electric vehicles? *Int. J. Hydrogen Energy* 37 (7), 6053–6062.
- Tirado, M.C., Clarke, R., Jaykus, L., McQuatters-Gollop, A., Frank, J., 2010. Climate change and food safety: a review. *Food Res. Int.* 43 (7), 1745–1765.
- United States Environmental Protection Agency, 2020. *Fast Facts on Transportation Greenhouse Gas Emissions*.

Valeri, E., Danielis, R., 2015. Simulating the market penetration of cars with alternative fuelpowertrain technologies in Italy. *Transport Pol.* 37, 44–56.

Vallender, S., 1974. Calculation of the Wasserstein distance between probability distributions on the line. *Theor. Probab. Appl.* 18 (4), 784–786.

van der Geest, K., de Sherbinin, A., Kienberger, S., Zommers, Z., Sitati, A., Roberts, E., James, R., 2019. The impacts of climate change on ecosystem services and resulting losses and damages to people and society. In: *Loss and Damage from Climate Change*. Springer, pp. 221–236.

Vassileva, I., Campillo, J., 2017. Adoption barriers for electric vehicles: experiences from early adopters in Sweden. *Energy* 120, 632–641.

Xing, Y., Wang, X., Qiu, C., Li, Y., He, W., 2022. Research on Opinion Polarization by Big Data Analytics Capabilities in Online Social Networks. *Technology in Society*, 101902.

Yuan, F., Li, M., Liu, R., 2020. Understanding the evolutions of public responses using social media: Hurricane Matthew case study. *Int. J. Disaster Risk Reduc.* 51, 101798.

Zayet, T.M., Ismail, M.A., Varathan, K.D., Noor, R.M., Chua, H.N., Lee, A., Low, Y.C., Singh, S.K.J., 2021. Investigating transportation research based on social media analysis: a systematic mapping review. *Scientometrics* 1–39.

Ziefle, M., Beul-Leusmann, S., Kasugai, K., Schwalm, M., 2014. Public perception and acceptance of electric vehicles: exploring users' perceived benefits and drawbacks. In: *International Conference of Design, User Experience, and Usability*. Springer, pp. 628–639.



Dr. Qian (Christine) Lv is a Professor and Co-Associate Chair for Graduate Education in the Department of Computer Science, University of Colorado Boulder. She received her Ph.D. degree in computer science from Princeton University. Lv's research focuses on full-stack data analytics, which integrates systems, algorithms, and applications for effective and efficient data analytics in ubiquitous computing and scientific discovery. Topics of interest include mobile/wearable/IoT sensing, multi-modal data fusion, anomaly detection, recommendation, and user behavior analysis. Her research is interdisciplinary in nature and interacts closely with a wide range of scientific domains including Earth sciences, transportation electrification, renewable and sustainable energy, as well as the information needs in people's daily life and cybersafety. Lv has received the IMWUT 2021 Distinguished Paper Award, SocialCom 2020 Best Paper Award, SenSys 2018 Best Paper Runner-up Award, 2017 Google Faculty Research Award, VLDB 2017 Ten Year Best Paper Award, ICTAI 2017 Best Student Paper Award, two Best Paper Award nominations, and Pervasive 2012 Computational Sustainability Award. Lv has served on the editorial board of PACM IMWUT as well as the technical program committee and organizing committee of many international conferences.



Tao Ruan is currently pursuing the Ph.D. degree in computer science with University of Colorado Boulder, Boulder, USA. His current research interests include computational social science, data mining and crisis informatics.