TRB Annual Meeting

Electric vehicle demand estimation and charging station allocation using urban informatics

--Manuscript Draft--

Full Title:	Electric vehicle demand estimation and charging station allocation using urban informatics
Abstract:	This paper performs a novel data-driven approach to optimize electric vehicle (EV) public charging. We translate the study area into a directed graph by partitioning it into discrete grids. A modified geographical PageRank (MGPR) model is developed to estimate EV charging demand, built upon trip origin-destination (OD) and social dimension features, and validated against real-world charging data. The results are fed into the capacitated maximal coverage location problem (CMCLP) model to optimize the spatial layout of public charging stations by maximizing their utilization. It is shown that MGPR can effectively quantify the EV charging demand with satisfactory accuracy. Optimized EV charging stations based on the CMCLP model can remedy the spatial mismatch between the EV demand and the existing charging station allocations. The developed methodological framework is highly generalizable and can be extended to other regions for EV charging demand estimation and optimal charging infrastructure siting.
Manuscript Classifications:	Data and Data Science; Information Systems and Technology AED30
Manuscript Number:	TRBAM-22-00608
Article Type:	Presentation
Order of Authors:	Zhiyan Yi
	Xiaoyue Cathy Liu
	Ran Wei
	Yirong Zhou
	Jianli Chen
Additional Information:	
Question	**Response**
The total word count limit is 7500 words including tables. Each table equals 250 words and must be included in your count. Papers exceeding the word limit may be rejected. My word count is:	7483

Electric vehicle demand estimation and charging station allocation using urban informatics

Zhiyan Yi
Department of Civil & Environmental Engineering, University of Utah
110 Central Campus Dr. RM 1650, Salt Lake City, Utah, 84112
Email: zhiyan.yi@utah.edu

Xiaoyue Cathy Liu, Corresponding Author
Department of Civil & Environmental Engineering, University of Utah
110 Central Campus Dr. RM 2137, Salt Lake City, Utah, 84112
Tel: 801-587-8858; Email: cathy.liu@utah.edu

Ran Wei
School of Public Policy
University of California, Riverside
Tel: (951) 827-9369; Email: ran.wei@ucr.edu

Yirong Zhou
Department of Civil & Environmental Engineering
University of Utah, Salt Lake City, Utah, 84112
Email: Yirong.Zhou@utah.edu

Jianli Chen
Department of Civil & Environmental Engineering, University of Utah
110 Central Campus Dr. RM 2032, Salt Lake City, Utah, 84112
Email: jianli.chen@utah.edu

Word count: 6,733 words + 3 tables (250 words per table) = 7,483 words

Submitted on [August 1st, 2021]

ABSTRACT
This paper performs a novel data-driven approach to optimize electric vehicle (EV) public charging. We translate the study area into a directed graph by partitioning it into discrete grids. A modified geographical PageRank (MGPR) model is developed to estimate EV charging demand, built upon trip origin-destination (OD) and social dimension features, and validated against real-world charging data. The results are fed into the capacitated maximal coverage location problem (CMCLP) model to optimize the spatial layout of public charging stations by maximizing their utilization. It is shown that MGPR can effectively quantify the EV charging demand with satisfactory accuracy. Optimized EV charging stations based on the CMCLP model can remedy the spatial mismatch between the EV demand and the existing charging station allocations. The developed methodological framework is highly generalizable and can be extended to other regions for EV charging demand estimation and optimal charging infrastructure siting.

Keywords: Electric vehicles, PageRank model, charging infrastructure optimization, spatiotemporal travel patterns.

INTRODUCTION

With the incentives and policy support from governmental agencies and EV manufactures, EV markets are progressively growing in the recent decade (*1*). EV sales reached to more than 2 million units globally in 2018 with an increase of 63% on a year-on-year basis (*2*). China and the Unites States are the two major EV markets accounting for over 20% of the sales worldwide (*3*). San Jose, San Francisco, and Los Angeles metropolitan areas, having some of the highest EV sales and market shares, where there are already more than a quarter-million EVs on the roads (*4*). The rapid growth of EV adoption results in the increase of charging demand as well. EV charging events can be mainly divided into home charging, workplace charging, and public charging depending on the locations (*5*). In 2010, ECOtality and Idaho National Laboratory conducted EV charging units analysis (*6*). According to their result on 2,903 private EV owners in United States, 80% of charging events are conducted at the participants' home, and over 70% of the vehicles have charged at locations other than home, such as shopping malls, restaurants, and work offices. Therefore, effectively satisfying EV users' public charging needs is crucial. In fact, a recent study shows that public charging infrastructure is a key to the growth of EV market (*4*). Another reason for promoting public charging infrastructure is that although large portion of EV adopters in the United States and Europe have home charging facilities, countries like China still have low penetration of home charging due to fewer single-family dwellings. It is estimated that in China public charging (as compared against home charging option) will increase from 55% to 80% by 2030 (*7*). Yet, two major barriers exist for most countries in terms of public charging infrastructure expansion. First, some ill-chosen charging stations can be significantly underutilized due to their inconvenience of access (e.g. distance). Distance plays an important role for EV users when choosing charging stations because of range anxiety (*8*). Second, insufficient charging networks in certain regions could fail to meet the charging demand (*9*). To this end, how to optimally allocate the public charging stations to improve the charging coverage and effectively exploit their utilization are the main challenges for siting public electric vehicle supply equipment (EVSE).

Solving EVSE allocation problem generally involves two steps: estimate spatial distribution of charging demand; and apply mathematical modeling to obtain optimal locations with specific objectives. When estimating charging demand, a common approach is to model the EV battery usage, and simulate EVs' energy consumption under different traffic scenarios (*10–12*). Such simulation-based approaches would be difficult to scale to large urban context due to computational expense. In remedy to that, data-driven methodologies have gained more interests in recent years with the proliferation of urban informatics and mobility data. For instance, geographical features including point of interest (POI), traffic flow, and population density are extracted as inputs to spatial statistical models to infer public charging demand (*13*, *14*). Traffic data such as GPS trajectory are available for exploring driver' travel patterns and further identifying hot spots of large public charging demands. The general procedure for conducting such type of analysis involves formulating it as a discrete problem by partitioning the study area into sub-regions (or cells), extracting drivers' travel patterns from travel mobility data, and finally inferring public charging demand for each cell (*14-16*). Yet, given that charging events usually occur at the trips' destinations, it would be far-fetched to associate transient locations along one's trip to the charging demand. Another issue with the existing charging demand estimation is the lack of real-world data to either construct or validate the proposed models (*13-14*). Public charging infrastructures are typically managed by governmental agencies or private companies, and the charging event data are usually not publicly accessible.

To this end, this paper aims to develop an innovative approach leveraging PageRank algorithm, graph theory, geographical features, and trajectory data to quantify the spatial distribution of EV charging demand. PageRank algorithm is uniquely suited for this problem in that it quantifies the importance of a web page via its linkage to other pages. Travel behavior to certain extent resembles people's internet browsing process (treated as a random walk). While PageRank algorithm is able to identify the important nodes within the graph topology, we develop a modified geographical PageRank (MGPR), which is capable of incorporating geographical features to estimate the EV charging demand within a region. The model validation is further built upon an automated dynamic crawling pipeline for retrieving and storing the public charging information. The estimated EV charging demand is then fused into a capacitated maximal coverage location problem (CMCLP) model to optimize the EVSE distribution by maximizing the utilization of charging stations. The framework is beneficial to transportation agencies and could provide insightful guidance for future public EVSE installation.

LITERATURE REVIEW

EV charging demand estimation

A myriad of studies have utilized GPS trajectory data to explore trip purposes and spatiotemporal travel patterns to infer EV charging demand. Hu et al. (*17*) used drivers' travel activities to evaluate the feasibility of replacing the gasoline yellow taxi with BEVs in New York City. GPS data of 13,587 taxis spanning the entire year of 2013 were analyzed to extract spatiotemporal driving patterns, travel demand, dwelling and other features. Their proposed BEV feasibility model indicated that only 8% of current taxis can be freely charged under the constraints of mileage range and pickup activity. Similarly, Tu et al. (*18*) employed optimization algorithm to optimize the location of electric taxi in Shenzhen, China. Dynamic pickup demands were estimated using trip data in combination with the corresponding transportation network information first, and then a spatial-temporal demand coverage location model is applied to maximize the taxi service coverage while minimizing the charging wait time. Their results indicate that downtown area, airport, and railway stations have intensive charging demand for electric taxis. The aforementioned studies offer insights on exploring the charging demand of electric taxis based on GPS data and trip activity. However, travel patterns and charging demand can be drastically different between taxis and private vehicles. Kontou et al. (*16*) explored the relationship between charging demand and people's daily activities for private vehicles. They identified places with high trip destination densities and prioritized those regions for charging infrastructure installation. The result showed that if top 10% most frequently visited grid cells have installed charging stations in the Puget Sound region, then EV users will be able to access public charging on 71% of their trips. This study suggested that charging probability is highly associated with people's daily travel activities and trip destinations. For EV drivers, they prefer to leave their EVs charging at nearby stations while conducting other activities, e.g., working, shopping, etc. For people who plan to purchase EV in the future, they are less willing to compromise their daily routines to go to distant charging stations. Meanwhile, Vazifeh (*15*) reconstructed trajectory using cellular data to track individual movement patterns in Boston area. A discrete optimization model is formulated by minimizing the total number of charging stations and the average travel distance on those routes.

Apart from GPS trajectory data, urban informatics such as POIs are utilized to analyze the charging demand. Wagner et al. (*13*) built a linear regression model using POIs to fit the usage data from more than 32,000 charging sessions in Amsterdam. Results indicated that POI imposes significant influence on the charging behavior of EV users. Likewise, Dong et al. (*14*) applied spatial features including traffic flows, population density, and POI to model the charging demand based on the distribution of current charging stations in London using Bayesian spatial log-Gaussian Cox process model. Statistical analysis showed that transport, retail, and commercial POIs significantly influence the charging demand in urban-scale region.

PageRank model and its application

PageRank is one of the most widely used web pages ranking algorithms, developed by Google (*19*). PageRank model formulates the internet as a huge directed graph, where each website represents a node and the hyperlinks are the edges connecting those nodes. Each node is assigned a PageRank value, denoting the importance of the website. PageRank models users' internet browsing behavior as a random walk process. The underlying assumption of PageRank is that more popular web pages are likely to be linked from other web pages, and their importance tends to propagate via hyperlinks. Nodes that are more frequently visited will receive higher PageRank scores and are subsequently deemed more important. PageRank is proved to be extremely efficient and simple enough to solve complex graph problems. Yet, a few strategies could be applied to improve its performance. The original PageRank algorithm assumes that the transition probability from one node to its all linked nodes is equal. However, that is not always the case in reality. For example, it is likely that a user jumps to a more popular website over the less popular ones. To fix this issue, Xing and Ghorbani (*20*) proposed a weighted PageRank algorithm. The core concept of this extended model is to assign higher transition probability to more popular pages instead of distributing equally. Another deficiency of the original PageRank is that it does not consider the content of web pages. In many situations, users jump to other web pages with similar content. Haveliwala (*21*) proposed a topic-sensitive PageRank algorithm by clustering the web pages into a set of topics and biased the original PageRank with those topics. Such innovative idea makes it possible for the PageRank model to incorporate more features for augmented model performance.

Although PageRank was originally used for ranking websites, it is quite effective to capture a variety of relations among vertices of graphs (*22*). For instance, PageRank model has been employed to infer traffic states in urban region. Kim et al. (*23*) explored the traffic congestion at 57 intersections in Cheongju city. Specifically, they generated a network graph to connect intersections by roads, and then applied PageRank to extract intrinsic relationship of traffic conditions across intersections. The result indicated that the intersections with higher PageRank scores generally have higher traffic density and thus are prone to congestion. Wang et al. (*24*) studied the traffic states in urban area of Beijing by partitioning the area into 62 by 65 grids and constructing the network using 12,000 taxi GPS trajectories. The traffic volume between two adjacent grids are used as the weight of the link. It is found that there exists a positive correlation between the PageRank value and congestion index for most regions, and the PageRank value can therefore predict the upcoming congestion. Besides the traffic states inference, PageRank model has been applied to analyze other geographical-related problems, such as revealing urban structure through roads connectivity (*25*).

To the best of our knowledge, PageRank model has not been utilized for EV charging demand estimation to date. Yet, with the proliferation of big data (e.g. GPS trajectory, POI) as reviewed in Section 2.1, PageRank is well suited for modeling EV charging demand from a graph-theory perspective, as EV users travel from an origin and a destination can be treated as Markov process, and the charging demand is highly correlated with the features of trip destinations (*14*). A modified PageRank model not only can consider the EV users' travel habits for demand estimation, but also can incorporate social dimension features (e.g. POI, land use) to improve model accuracy.

Optimization of EVSE location

Public charging infrastructure deployment problem can be deemed as optimally siting EVSE on a landscape. A variety of optimization algorithms have been employed to attempt this from multiple angles. Among them, maximal coverage location problem (MCLP) is a classic model to optimally assign facilities (*26*). Dong et al. (*14*) applied a standard MCLP to maximize the coverage of EV charging demand by assigning a fixed amount of charging stations. The model did not consider the constraint of charging stations' capacity, yet in reality charging stations are constrained by energy load. The CMCLP model is subsequently proposed to optimize EVSE location by taking into account the capacity constraints (*27*, *28*). A more sophisticated approach to optimizing EVSE location is to formulate it as a multi-objective optimization problem. Wang and Wang (*29*) proposed a mixed integer programming model to site refueling stations to serve intercity and intra-city travel with the goals of minimizing siting cost and maximizing population coverage. Vazifeh et al. (*15*) treated this as a set covering problem with dual objectives of minimum number of charging stations required and minimum average distance for drivers to the nearest accessible charging stations. Extended upon that, Kınay et al. (*30*) developed a full cover modeling framework with a novel objective function, which optimizes the charging station locations and determines the optimal OD routes so that the total en-route recharging is minimal for each trip. Yet most of multi-objective optimization problems are computationally infeasible due to large amount of intricate constraints. Such optimization problems mostly require heuristic algorithms to obtain near-optimal solutions.

To this end, the CMCLP model appears to be computationally efficient and suited for the EV charging infrastructure allocation problem. The flexibility of having the capacity constraint is uniquely aligned with the EV charging problem, as agencies when siting the charging stations, have to consider the energy load capacity for the specific area. Further, CMCLP is not a binary allocation, in that the model allows partial charging demand assignment. In case where one charging station reaches its capacity, the rest of the charging demand can be allocated to another eligible station nearby.

METHODOLOGY

In this section, a proposed MGPR model is presented, followed by the formulation of CMCLP optimization model. PageRank is uniquely suited for this study, as we partition the entire study area into grid cells. We further extract the spatial correlation built off of trips' origins and destinations (OD) and associate that with potential charging demand. The problem as such can be treated as a directed graph, where each grid or cell within the study area is treated as a node and can be characterized by the PageRank score, describes how appealing that cell is to the EV users.

Modified geographical PageRank (MGPR) model

Given a weighed directed graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{W})$, where $\boldsymbol{V}$ represents the set of nodes, $\boldsymbol{E}$ represents the set of edges, and $\boldsymbol{W}$ is the set of weight corresponding to each edge. The simplified version of PageRank model is defined as follows:

$$\boldsymbol{R}_{t+1} = \boldsymbol{M}\boldsymbol{R}_t \tag{1}$$

where $\boldsymbol{R}_t$ is the vector indicating the PageRank value of each node at step *t*. Formally, $\boldsymbol{R}_t = [PR(v_1), \dots, PR(v_n)]^T$, and *n* is the total number of nodes. $\boldsymbol{M}$ is the stochastic matrix that describes the transition probability from one node to another. For each node *j*, the transition probability has the following two properties:

$$M_{ij} \geq 0 \quad (2)$$

$$\sum_{i=1}^{n} M_{ij} = 1 \quad (3)$$

After infinite steps of long walk, the PageRank value for each node will converge to a stationary probability denoted by the following equation:

$$\boldsymbol{MR} = \boldsymbol{R} \quad (4)$$

To guarantee convergence, the graph is required to be a strongly connected graph, and an aperiodic one. However, not all network graphs can meet the aforementioned requirements. To address this problem, the simplified PageRank incorporates a random term to make the graph strongly connected and aperiodic. Eq. (4) is therefore modified as follows:

$$\boldsymbol{R} = d\boldsymbol{MR} + \frac{(1-d)}{n}\boldsymbol{1} \quad (5)$$

where the second term allows each node has a certain probability to transfer to all other nodes, and *d* is the damping factor that controls the tradeoff between the first and the second terms.

In the original PageRank model, the network graph does not consider the weight of links. Instead, it assumes all links have equal transition probabilities. Specifically, for node *j*, M_{ij} shares equal transition probability for each incoming node *i*, if there is an edge between nodes *i* and *j*, and otherwise 0. As noted by Kontou et al. (*16*), trip destination density can be used as a surrogate to measure potential EV charging demand. Therefore, in the MGPR, instead of assuming equal transition probability across nodes, we use the trip counts (derived from vehicle trajectories) to construct the transition matrix. The transitional probability M_{ij} is thus re-defined as:

$$M_{ij} = \frac{w_{ij}}{\sum_{i=1}^{n} w_{ij}} \quad (6)$$

where w_{ij} is the trip count from node *i* to node *j*. Another flaw of the original PageRank is the inability to incorporate other information such as web content. Inspired by the topic-sensitive PageRank algorithm (*21*), we adopted similar concept to our proposed model such that social dimension that might influence the EV charging demand could be incorporated. The core idea is to add a third term in the PageRank model. Different from the second term in Eq. (5) which has equal transition probability to all nodes, the third term will direct the drivers to other nodes with varying probabilities. Grid cells that are more favorable to charging activities would receive higher transitional probabilities. For instance, if a grid cell contains large number of commercial buildings, EV drivers are more likely to charge in that cell during the day (while they work). The proposed MGPR model is presented below:

$$\boldsymbol{R} = \alpha_1\boldsymbol{MR} + \frac{\alpha_2}{n}\boldsymbol{1} + \alpha_3\boldsymbol{G} \quad (7)$$

where $\boldsymbol{G}$ represents the social dimension term, α_1, α_2, and α_3 are the weights of PageRank term, random transferring term, and social dimension term, respectively. The sum of α_1, α_2, and α_3 should equal to 1. In this study, POI data, land-use type information, and socio-economic factors are used to describe urban and geographical features. The social dimension term $\boldsymbol{G}$ is therefore defined as:

$$\boldsymbol{G} = \frac{1}{3}(\boldsymbol{G}_{POI} + \boldsymbol{G}_{land-use} + \boldsymbol{G}_{socio-economics}) \quad (8)$$

Detailed information and definition with respect to the social dimension features will be introduced in the *Data* and *Results* sections.

CMCLP optimization model

In this study, we formulate the optimal public EVSE allocation problem as CMCLP following similar ideas from (*27*, *28*). The mathematical formulation is defined as follows:

Objective function:

$$Maximize\ \sum_i \sum_{j \in N_i} Z_{ij} \tag{9}$$

Subject to:

$$\sum_j Z_{ij} \leq a_i, \forall i \tag{10}$$

$$\sum_j X_j = p \tag{11}$$

$$\sum_i Z_{ij} \leq c_j X_j, \forall j \tag{12}$$

$$X_j = \{0,1\}, \forall j \tag{13}$$

where

i is the index of grid cells

j is the index of candidate grid cells that can be assigned with new charging stations

Z_{ij} indicates the amount of charging demand that can be covered in grid *i* by the neighboring charging stations *j*

$X_j = \begin{cases} 1; if\ charging\ station\ j\ is\ sited \\ 0; otherwise \end{cases}$

a_i is the estimated charging demand in each grid *i*

N_i is the set of potential neighboring charging stations for grid cell *i*

c is the capacity of each charging station

p is the total number of charging stations

The objective function Eq. (9) is maximizing the coverage of charging demand for all grids. Constraint (10) guarantees the charging energy provided by neighboring charging stations is under the estimated charging demand for grid *i*. Constraint (11) defines the total number of charging stations that are planned to be installed in the study area. Constraint (12) defines that the allowed charging energy each charging station serves to the neighboring area cannot be greater than its capacity. Constraint (13) ensures X_j is a binary variable. According to Prianka (*31*), the willingness of people walking from parking location to their activity place would become extremely low when the distance is over 3,000 feet (0.91 km). For this reason, we define that the neighboring charging stations for a grid cell refer to charging stations located in its adjacent 8-directional grid cells. The CMCLP is solved using a commercial optimization solver Gurobi in this study.

DATA PROCESSING AND ANALYTICS

Data

OD data

As noted earlier, in order to apply the MGPR to inferring charging demand, the problem is treated as a directed graph, where each grid cell within the study area is treated as a node and the edge is characterized by the number of trip ODs. The OD data are obtained from probe vehicle trajectories provided by Inrix. The trajectories were extracted from a portion of vehicle stream using probe sensors, such as cell phone and automated vehicle location (AVL). The raw data contains 2.5 million trips distributed in the State of Utah during September of 2018. We further filtered the trips that are enclosed within the Salt Lake City metropolitan area, as it is the boundary of this study. As described in Section 3, we employ a grid-based approach to partition the region and modeling the EV charging demand. Such level of granularity can remedy the GPS reading errors, while providing sufficient resolution for EVSE planning purpose. Note that determining the size of grid cells is an empirical process. If the size is set too large, it would be difficult to pinpoint the optimal locations of charging stations with fine granularity. On the contrary, small grid cell size may lead to failure in capturing hot spots of high charging demand. Previous studies (*14*, *16*, *18*) suggest the appropriate size of grid cells set as 1km by 1km. After grid segmentation, there are 756,303 trip OD pairs in total within the region. The maximum number of origin count in a grid cell is 9,281, and the maximum number of destination count in a grid cell is 9,268.

POI data

POI data can effectively represent urban context and infer people's trip purposes (*13*). In this study, we use Google Place API to extract POIs in our study area. There are 62,673 POIs in total obtained from Google Place API with 103 different labels. In fact, many labels share similar denotation. For instance, both hospital and doctor refer to health-related POI. For simplicity and practical concerns, we further classify the 103 labels into 11 categories. The detailed information of classified POI data is shown in Table 1.

TABLE 1 Description of POI Data

POI ID	Category	Label Examples	Total Number
1	Business	office, personal business	23,472
2	Health	hospital, health, doctor	8,982
3	Finance	agency, finance building	6,691
4	Retail	supermarket, grocery store	10,066
5	Restaurant	restaurant, food delivery	2,181
6	Transportation	bus station, train station	3,140
7	Education	school, university	1,290
8	NGO	church, government building	1,591
9	Entertainments	park, salon, bar ,zoo	2,422
10	Service	post office, gas station, laundry	2,427
11	Hotel	hotel, lodging	410

Among the eleven categories, a few do not have apparent association with public charging behaviors. To reduce the noise that might be incurred, Health, NGO, and Service POIs are eliminated for further analysis.

Social dimension features

Socioeconomics describes the relationship between social behavior and economics, while land-use information reveals the human use of land. These two types of geographical-based features are highly associated with people's parking behaviors and subsequently could impact the potential EV charging demand (*9*). For this reason, we incorporate socioeconomic and land use features in our model to infer the public charging need. These two datasets are obtained from Wasatch Front Regional Council (WFRC), the metropolitan planning organization that synthesizes a variety of data sources for transportation planning in the region. Population data is obtained at traffic analysis zone (TAZ) level. Land-use is categorized into agriculture, commercial area, residential area, recreation, and transportation. For simplicity, we reclassified the land-use into commercial vs. non-commercial region only, since most public charging facilities are inclined to be installed in commercial places.

Real-world public charging data

The real-world charging data is crawled from ChargePoint, an online application that assist EV users navigate and review nearby charging sites. ChargePoint operates the largest online network of independently owned EV charging stations, operating in 14 countries worldwide. Directly accessing the utilization information of public EV charging stations is challenging, since it requires the authorization from owners of EVSE. Alternatively, such information can be accessed via Google Map Service. We search all public charging stations in the study area first, and then retrieve the real-time data (e.g. the number of in-use port) constantly for each charging station. Meanwhile, associated features (e.g. power, location) for each charging station are collected. There are 126 public charging stations with 576 charging ports recorded by ChargePoint in the Salt Lake City metropolitan area. Among them, 109 charging stations (516 ports) broadcast real-time utilization information (i.e. number of in-use port at current time point), indicating charging status. A sample of collected features of public charging stations is displayed in Table 2.

TABLE 2 A sample of Stationary Features of Five Public Charging Stations

Station ID	Number of Port	Address	Latitude	Longitude	Power of Port (kW)	Status
0	2	425S Orchard Dr, North SLC	40.83**	-111.91**	7.2	Available
1	4	2280 Rose Park Ln, SLC	40.77**	-111.94**	7.2	Unknown
2	2	210N 1950W, SLC	40.77**	-111.94**	7.2	Available
3	8	168 N 1950 W, SLC	40.77**	-111.94**	7.2	Available
4	6	195 N 1950 W, SLC	40.77**	-111.95**	7.2	Available

In order to obtain the real-time charging station utilization, we applied *beautifulsoup* package in Python, which is a package for crawling on HTML. The program is further deployed on Amazon Web Service (AWS) Elastic Compute Cloud (EC2) and Simple Storage Service (S3) for dynamic crawling. The dynamic crawling framework and schemas of database are displayed in Figure. 1(a) and (b), separately. Three virtual machines were rented on EC2 to cover all charging stations in the study area, and the crawling was triggered every 10-minute for each charging station. The data collection spanned from Nov 5th, 2020 to Dec 12th, 2020. The dataset contains 656,179 records for the 109 charging stations.

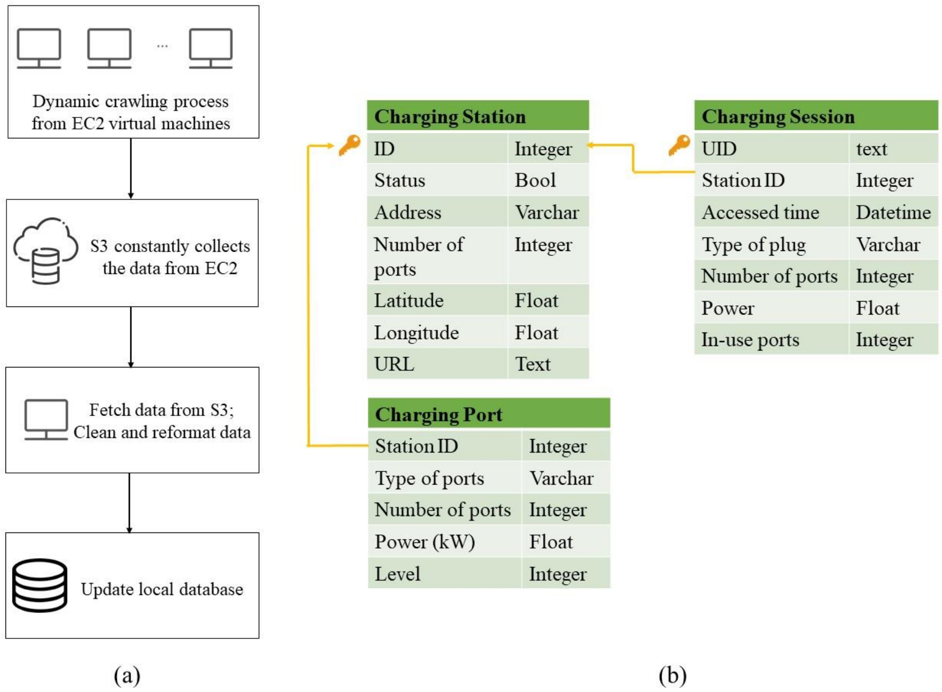


Figure 1 (a) Framework of dynamic crawling; and (b) the SQL schema for the database of EV charging information

There are three tables in our database to store the public charging data, as shown in Figure. 1(b). *Charging Station* and *Charging Port* tables record the associated features for charging stations and corresponding charging ports. *Charging Session* table documents the dynamic crawling records. The *in-use ports* feature reflects the number of ports being occupied at a charging station at a specific time point. We further aggregated the crawled records to obtain the charging energy at each station. In a nutshell, once the crawler detects that the charging ports in a station are in use, the energy consumption at that time point is calculated as the total number of in-use ports multiplied by the corresponding power of the port and 0.167 hour (the crawling interval). The accumulative energy consumption is then summed up across entire crawling period as the total charging energy consumption.

Spatiotemporal Analysis

We examine the spatiotemporal distribution of the charging energy consumption based on the data collected. Figure. 2 indicates the accumulative energy consumption at each charging station within the study period.

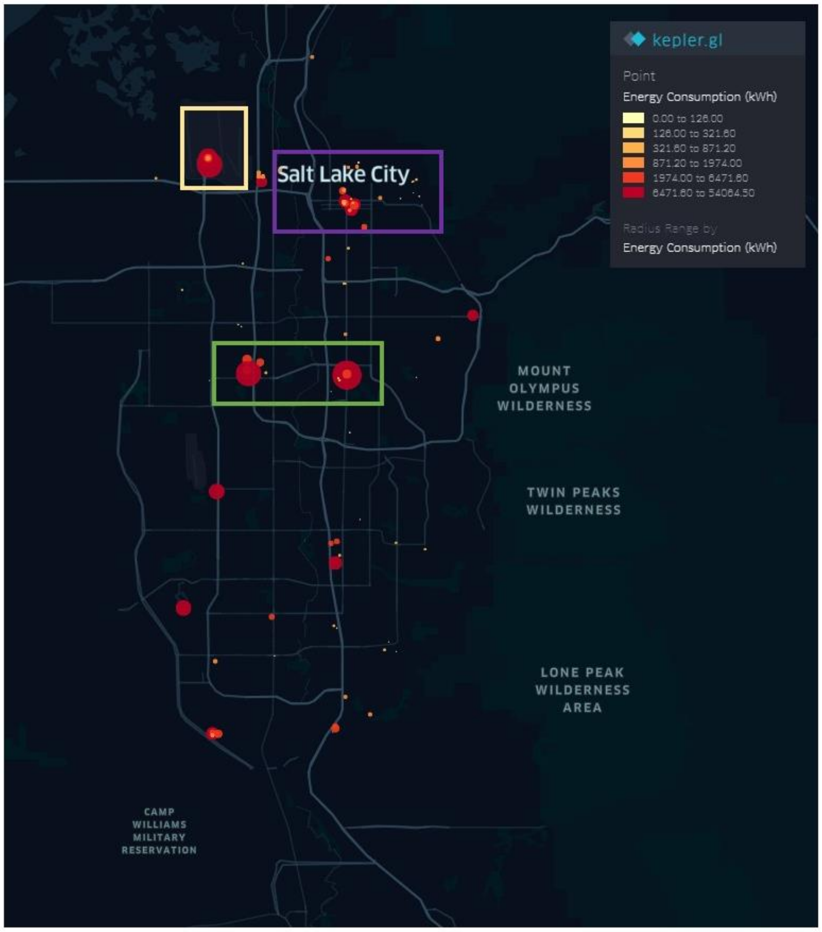


Figure 2 Total energy consumption of each public charging stations across entire period

In Figure. 2, two charging stations highlighted by green show very high energy consumption. These two charging stations are equipped with Level 3 charging ports (CSS and CHAdeMO) which have much higher power than level 2 ports. The energy consumptions are therefore larger. A cluster of charging stations with high demand exist around the Salt Lake City downtown area (marked by purple). Although none of them exhibit extremely large energy consumption, the average usage frequency is relatively high compared to stations in other areas. It is worthy to note that the Salt Lake City international airport (marked by yellow) indicates high energy consumption as well. EV users might charge their vehicles while waiting to pick up someone or travel to other places while leaving their vehicles charged at the airport.

We further aggregate the charging station utilization data by different time resolutions to examine the temporal patterns. Figure. 3(a) shows that most of the charging activities occur during the day, between 7:00 am to 7:00 pm. Figure. 3(b) indicates that the average charging demand on Monday and Tuesday is relatively low, while the highest charging demand is observed on Sunday. There is a significant difference between charging patterns across the day-of-week. The average number of daily in-use port is 145 Monday through Thursday, and 185 Friday through Sunday.

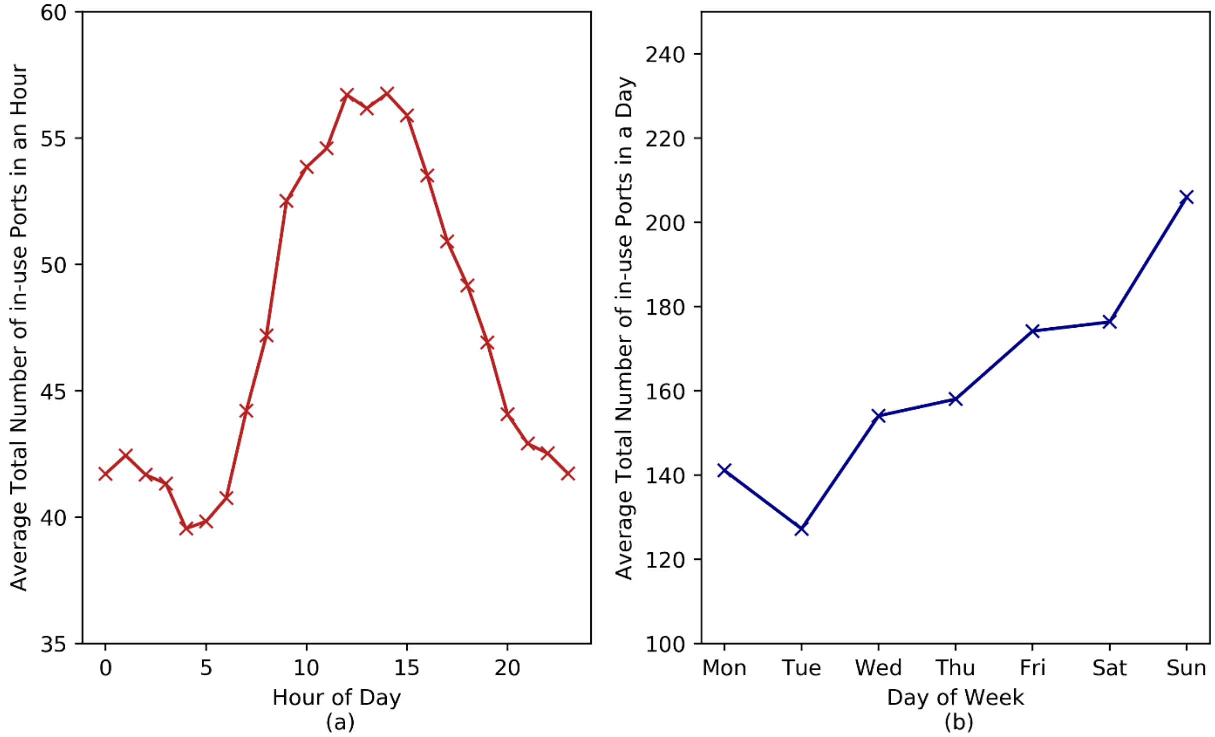


Figure 3 (a) The average total number of in-use ports in each hour of day; and (b) the average total number of in-use ports in each day of week

Subsequently, we explore the difference of charging patterns by day-of-week. Specifically, for each station, we calculate the difference between the average daily charging energy consumption from Friday to Sunday and that value from Monday to Thursday. In fact, charging behaviors during Monday to Thursday are more likely to be linked with work trips, while charging behaviors during Friday to Sunday are more likely to be linked with non-work trips. We further overlay the POI data to infer the nature of trip purposes around the charging stations. For POI data, we combine the finance, business, and education into one group to infer work trips. Retail, entertainment, and restaurant POIs are grouped together to infer the non-work trips. The charging differences by day-of-week (i.e. average daily charging energy consumption from Friday to Sunday minus that value from Monday to Thursday) are presented under different POI types in Figure. 4 (a) and (b), respectively.

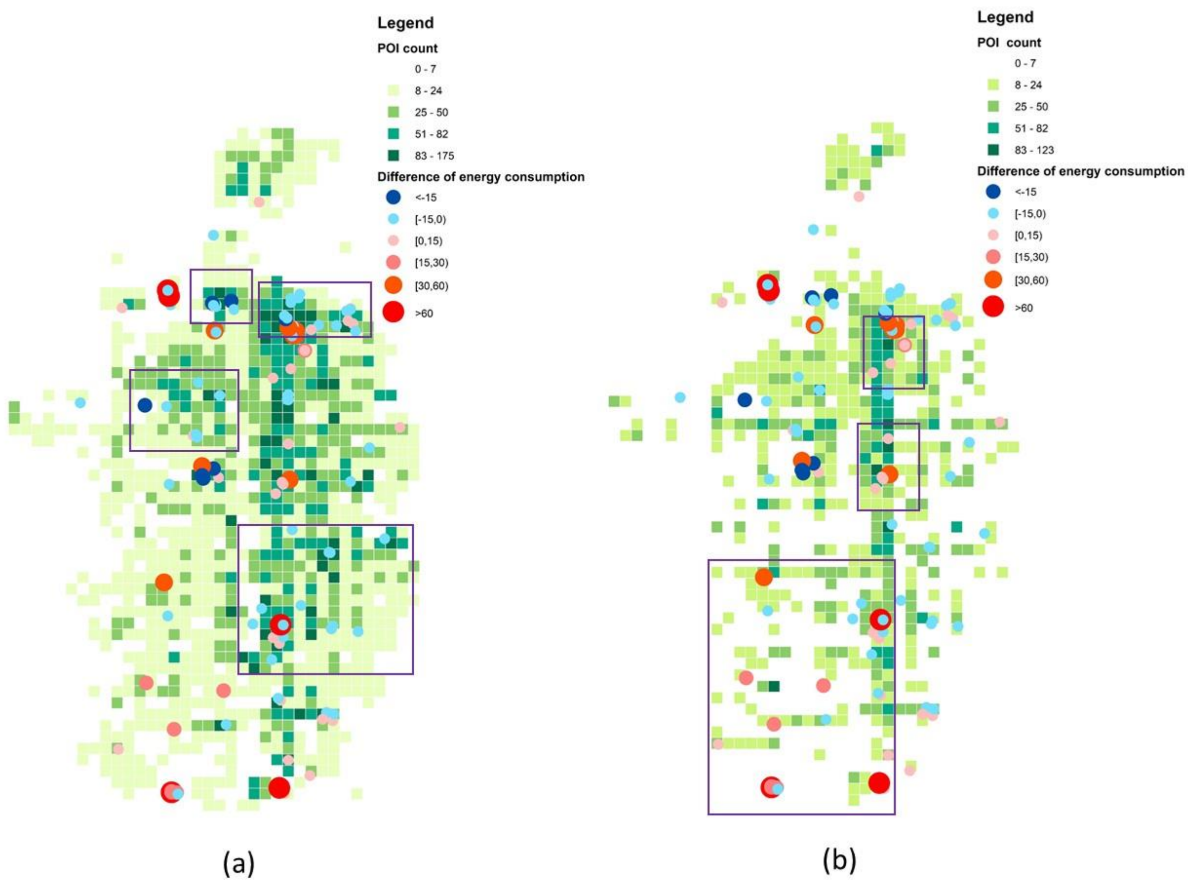


Figure 4 (a) The work-related POIs (financial buildings, business, and education) distribution and public charging stations' charging patterns; (b) the non-work-related POIs (entertainment places, retails, and restaurants) distribution and public charging stations' charging patterns

In Figure. 4(a), it is found that charging stations that are more frequently used during weekdays are located around regions with large number of POIs associated with workplace. In contrast, charging stations that are more frequently used during weekends are mostly scattered in remote regions away from downtown as shown in Figure. 4(b). POIs such as parks, grocery stores are identified in neighboring regions. This distinction of charging patterns with respect to day-of-week validates the aforementioned hypothesis. Note that the Salt Lake City downtown area is mixed land use with both commercial buildings and recreational places. As a result, there is a mix of usage both on weekdays and weekends.

RESULT AND ANALYSIS

PageRank model

In our proposed MGPR model, the OD matrix $\boldsymbol{M}$ is defined in Eq. (6). The social dimension term $\boldsymbol{G}$ includes $\boldsymbol{G}_{POI}$, $\boldsymbol{G}_{socio-economics}$, and $\boldsymbol{G}_{land-use}$. $\boldsymbol{G}_{POI}$ for grid cell i is calculated as the number of POIs in the grid cell i divided by the total number of POIs in study area. Similarly, population density for grid cell i is calculated to represent $\boldsymbol{G}_{socio-economics}$ value for grid cell i. Lastly, dummy variable is used to indicate whether grid cell i is a commercial area. The dummy value is then normalized to represent $\boldsymbol{G}_{land-use}$ value for grid cell i. The coefficients α_1, α_2, and α_3 control the OD matrix, random effect, and social dimension, respectively. For example, large α_1 will cause OD matrix dominates PageRank score. In this study, the parameter values are determined empirically. α_2 is set as 0.05 since random effect is expected to be marginal. Meanwhile, it is found that when α_1 ranges between 0.5 and 0.7, no much variation on results is detected. The optimal result is observed when α_1, α_2, and α_3 are set as 0.6, 0.05, and 0.35, separately. Once the coefficients are determined, PageRank value is determined via an iterative process. We keep updating the PageRank vector $\boldsymbol{R}_t$ in Eq. (7) at each step t until it reaches convergence. The total number of iteration steps is set as 500 for the MGPR. Numeric result indicates that when the step reaches 20, the PageRank value almost converges.

To test the effectiveness of our proposed model, the original PageRank model and weighted PageRank model are also developed for comparison purpose. For original PageRank and weighted PageRank models, we adopt 0.85 for the dampening factor d in Eq. (5) as seen in (*19*, *20*). The computation of PageRank values for original PageRank and weighted PageRank models follows similar process. In order to quantify model performances, the following two metrics are used:

Weighted total energy (WTE): the weighted total energy is defined as follows:

$$WTE = \sum_{i \in S} R_i * d_i \tag{14}$$

where S is the set of grid cells with charging facilities; R_i is the normalized PageRank value of the i^{th} grid cell, and d_i is the actual charging energy of the i^{th} grid cell. It measures the ability of the model to capture paramount charging demand regions. If the normalized PageRank values of grid cells with large charging energy are high, WTE would be large as well.

Mean absolute rank difference (MARD): MARD is calculated by averaging the absolute difference of the rank of PageRank value and the rank of actual charging energy for each grid cell equipped with charging facilities. It is expressed as follows:

$$MARD = \frac{1}{N} \sum_{i \in S} |Rank\_R_i - Rank\_d_i| \tag{15}$$

where N is the total number of grid cells with EVSE; $Rank\_R_i$ is the ranking of PageRank value for the i^{th} grid cell, and $Rank\_d_i$ is the ranking of actual charging energy consumption for the i^{th} grid cell. The smaller MARD is, the better the result is. MARD value reflects how well the model can distinguish high and low charging demand regions. Additionally, a random scenario, where each grid has equal PageRank value (1/N), is added to benchmark the PageRank models. The results of all models are shown in Table 3.

TABLE 3 The WTE and MARD Values for Original PageRank, Weighted PageRank, and MGPR Models

Metrics	Random Scenario	PageRank	Weighted PageRank	MGPR
WTE	92.0	444.1	550.2	630.0
MARD	NA	18.8	18.2	11.8

In Table 3, it is observed that WTE has a substantial increase over the random scenario by using original PageRank model. Such result indicates that trip count can imply EV charging demand effectively in urbanized area. Compared to original PageRank model, the weighted PageRank increases WTE value by 106.1, suggesting that trip density of each OD grid pair should be taken into consideration. Moreover, the MGPR further augments the model performance with

WTE value of 630. One possible explanation is that social dimension features such as the number and type of POIs reflect trip purposes that are associated with public charging behaviors. Note that in random scenario, all grid cells share same PageRank value and same rank, hence the MARD value cannot be computed. As for the original PageRank and weighted PageRank model, the MARDs are relatively close and much higher than MARD for MGPR. The significant reduction of MARD by incorporating social dimension features illustrates that for some of regions with high trip density their charging demands are not necessarily high, e.g., residential neighborhoods. To better visualize the results geographically, we present the grid plot for each model, separately, in Figure. 5.

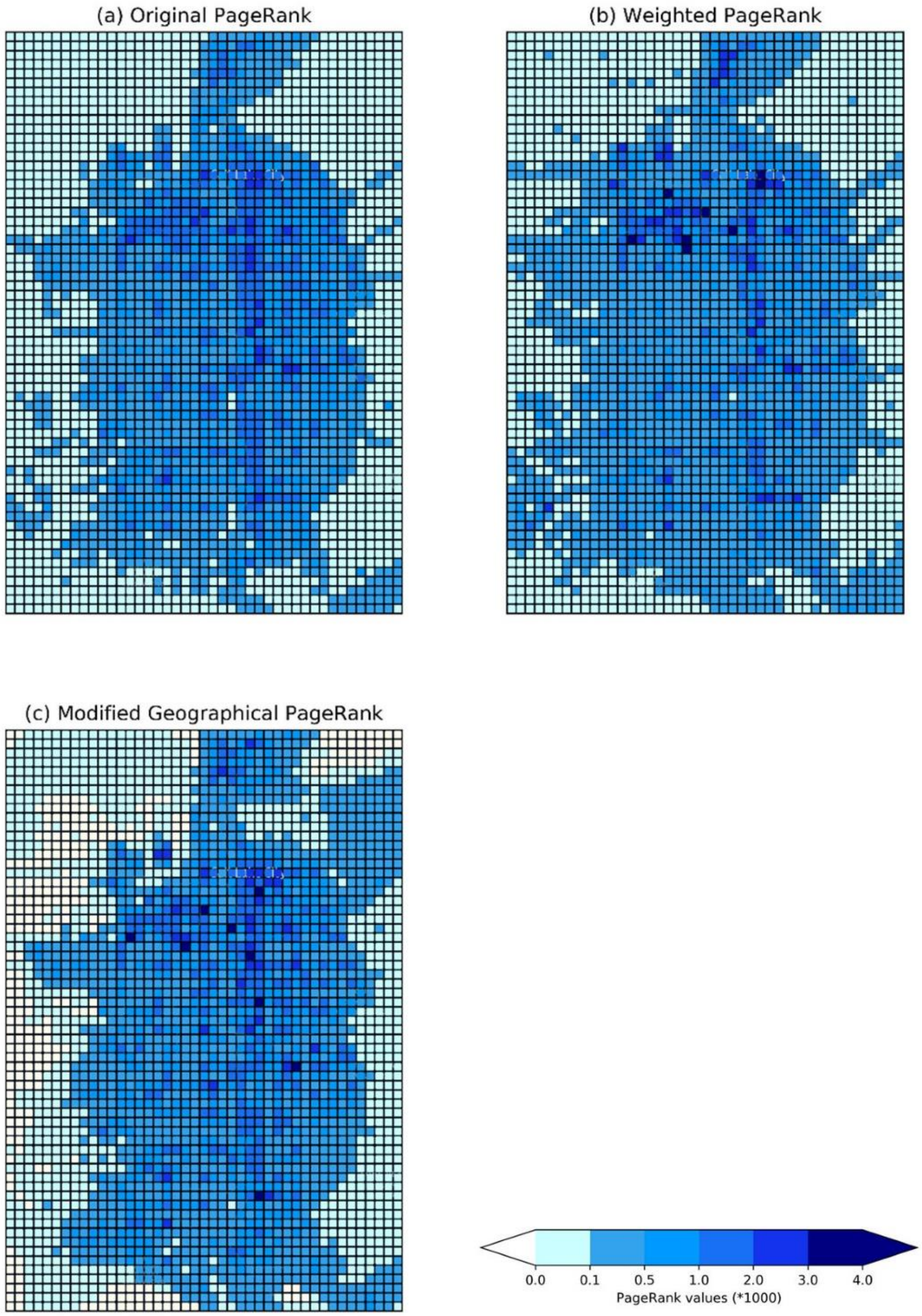


Figure 5 (a) PageRank values distribution of original PageRank model; (b) weighted PageRank model; and (c) MGPR model

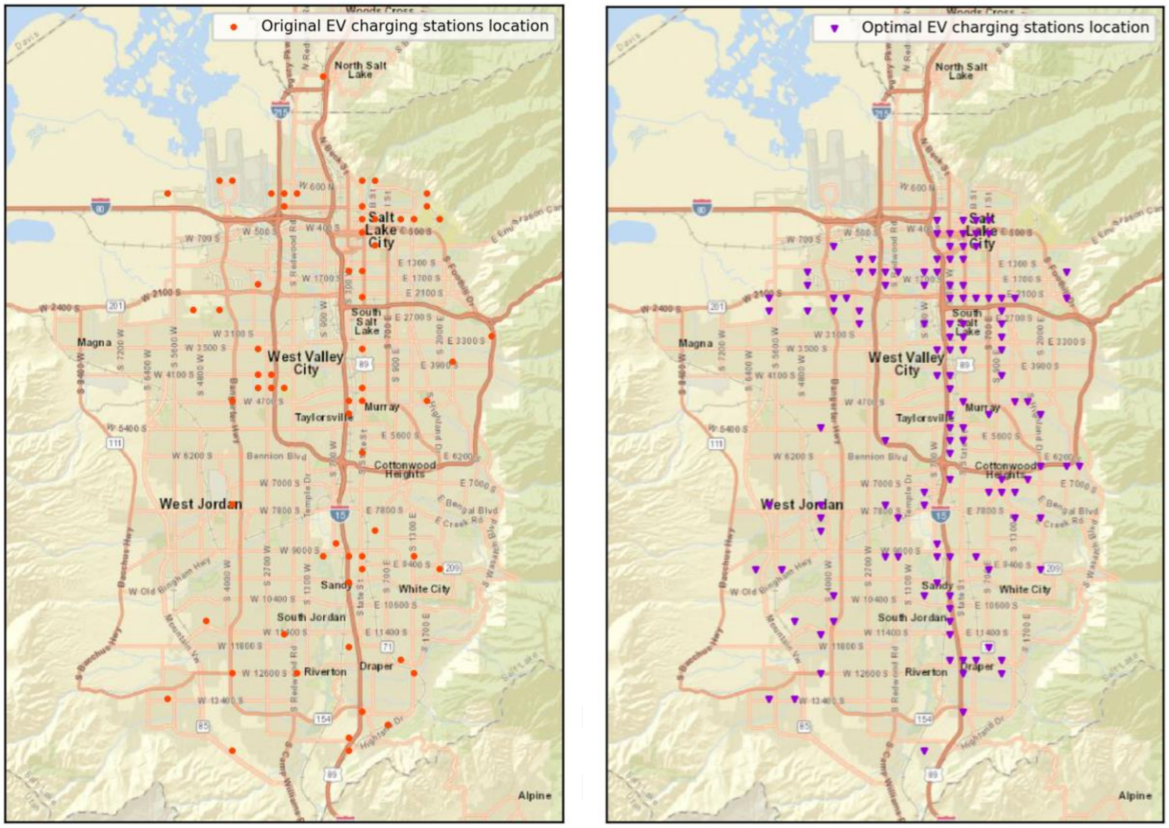
The original PageRank and weighted PageRank indicate similar public charging demand distributions as illustrated in Figure. 5(a) and (b). However, some cells receive higher PageRank values by employing weighted PageRank. Those cells have higher trip density, and subsequently more appealing to charging activities. In Figure. 5(c), it is observed that the PageRank values of large portion of cells located in the outskirts decrease to 0. We also notice that the PageRank values for some cells increase on the contrary, which is attributable to the POIs, population density, and land-use type. In summary, MGPR presents a more distinct pattern compared to original PageRank, and Weighted PageRank.

Optimizing public charging stations

As indicated by WTE and MARD in Section 5.1, the MGPR model can effectively quantify EV charging demand using PageRank score. With the real charging energy data collected, one can build a model to map the PageRank score to the charging energy consumption for each grid. In this study, 63 grid cells that contains the ground truth EV charging energy consumption are utilized to build a regression model. Specifically, the regression model is defined as follows:

$$y = \max(0, \beta_1 x_1^2 + \beta_2 x_2 + \beta_0) \tag{16}$$

where y is the daily estimated charging energy (kwh) in a grid cell; x_1 and x_2 denote the normalized PageRank value and the number of charging ports in that grid cell, respectively; β_0, β_1 and β_2 are corresponding coefficients. The least square method is applied to obtain the estimated coefficients, where the values of β_0, β_1 and β_2 are -107.73, $1.25*10^7$, and 15.46, respectively. The value of R^2 being 0.78 indicates a satisfactory model fitness. We then use this model to estimate the charging demand at each grid cell and Figure 6 (a) shows its spatial distribution and the demand coverage of existing 126 charging stations. The total potential charging demand in Salt Lake City metropolitan area is estimated as $11.89*10^4$ kWh, and the existing 126 charging stations can only supply $1.58*10^4$ kWh energy, which is 13.3% of the total public charging demand in the Salt Lake City metropolitan area.



(a) (b)

Figure 6 (a) Existing allocation of public EV charging stations; and (b) optimal allocation of public EV charging stations

Given the estimated charging demand across the region, we can further explore the optimal siting of public charging stations by employing the CMCLP optimization model. While there are 2,816 grid cells, not all grid cells are suitable for siting new public charging stations due to their unique land-use types. We assume that the new public charging stations can be only installed in grids with public parking lots. The public parking lots data is available from WFRC, and there are 463 grid cells in total with public parking lots as shown in Figure. 7. Meanwhile, for those grid cells that are currently not equipped with charging stations, the number of charging ports x_2 is unavailable. Without loss of generality, we assume each grid cell is allowed to install at most one public charging station. Each charging station is equipped with four 7.2kW charging ports, and the capacity constraint for each port is set as 57.6 kWh.

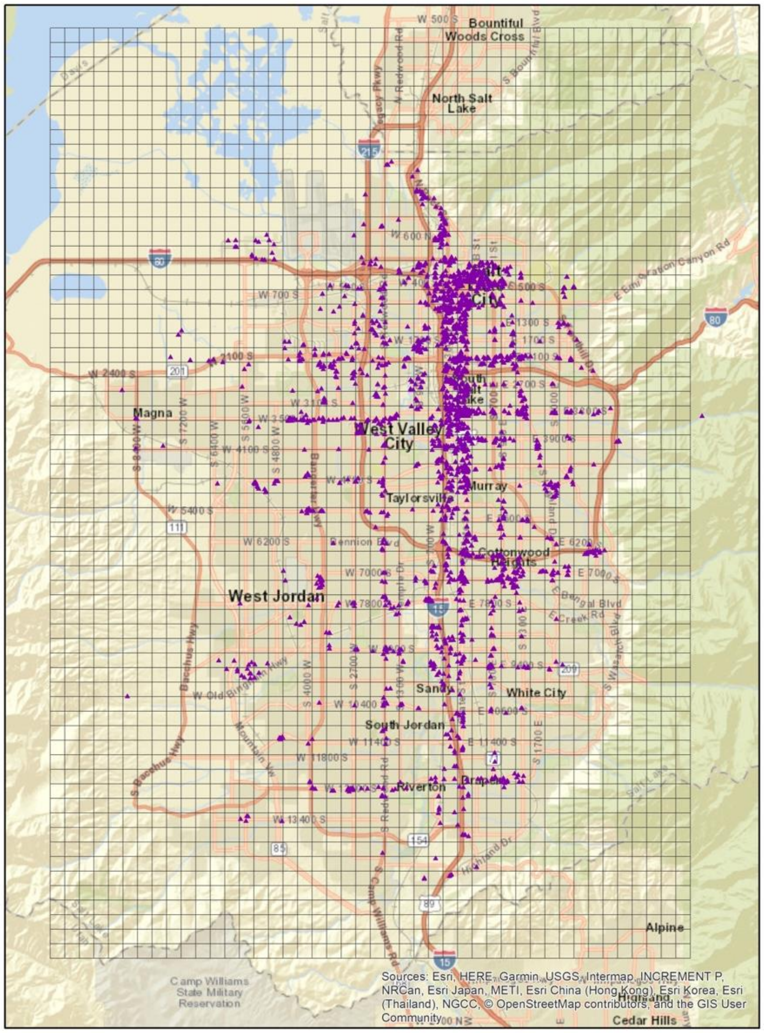


Figure 7 The distribution of public parking lots in the Salt Lake City metropolitan area

The CMCLP is solved for various potential values of p (number of charging stations to site) using Gurobi to explore the tradeoff between the cost of building charging stations and the service coverage of charging demand. Figure. 8 shows the tradeoff curve, where Scenario 1 assumes that there is no existing public EV charging station prior to optimization. As shown from the curve, clearly the coverage of charging demand increases linearly as the number of charging stations increases from 1 to 300. However, the growth rate decreases after p = 300, indicating that some of the additional charging stations are not fully utilized. The maximum service coverage is reached when the number of sited charging stations is 370 (p=370). No more charging stations can lead to increased service coverage. It is also important to note that 126 sited charging stations can provide service for 24.4% of charging demand in Salt Lake City metropolitan area. This is a significant demand coverage increase compared with 13.3% service coverage provided by the existing 126 charging stations, suggesting that these newly planned sites can serve EVs more effectively. These newly planned 126 sites are shown in Figure. 6 (b). It is observed that the optimally allocated charging stations are clustered in Salt Lake City downtown (northeastern), indicating large charging demand in that area. In fact, a large portion of current charging stations are deployed in downtown already. Such result suggests that most of charging stations in this region are effectively exploited. Yet there are areas (e.g. West Valley) where a congregated number of EVSEs are currently present that might need to be reallocated. One possible explanation is that those EVSEs are located in residential neighborhoods where people can just charge their EVs at home instead. Another observation is that many optimal charging stations are distributed along the freeways, such as I-15. Chen et al. (2016) mentioned that freeways generally have excessive vehicle flow and correspondingly high quick-charging demand. Future deployment of public charging stations can be considered to be located in the close vicinity of freeway entrance or exit.

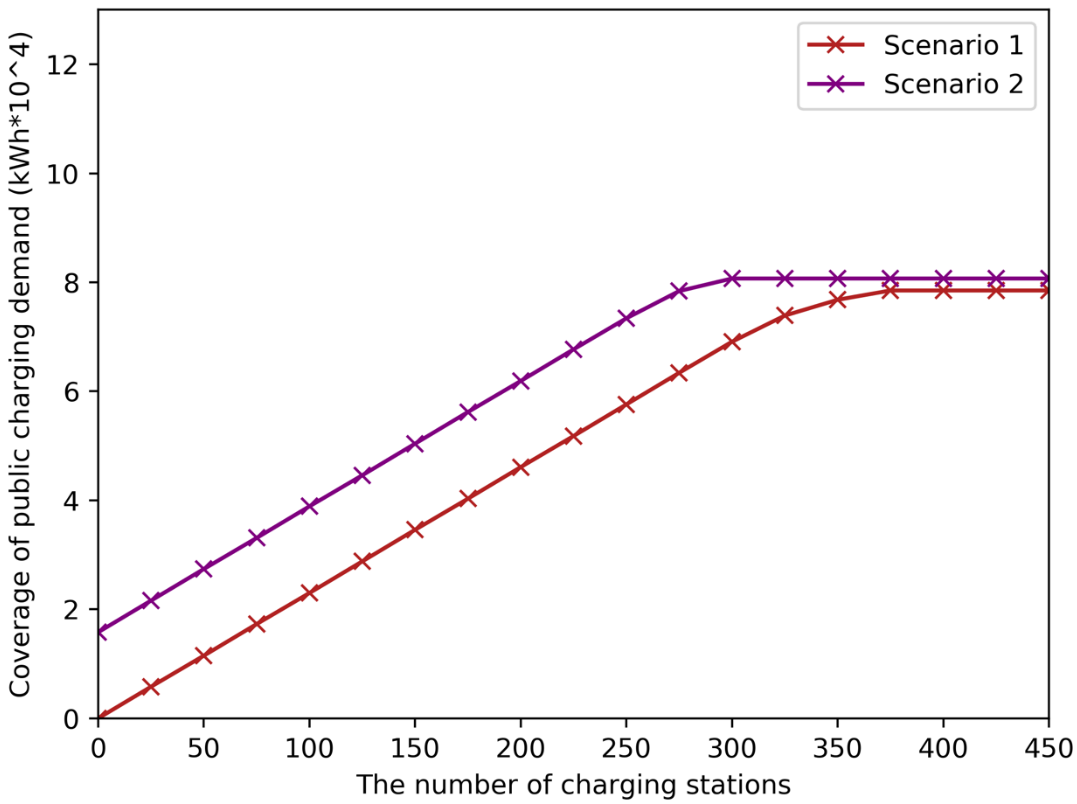


Figure 8 Total charging demand coverage under different number of public charging stations

In addition, we explore where to site new charging stations to maximize the service coverage of charging demand given the existing 126 charging stations. Such tradeoff curve between demand coverage and number of sited charging stations is displayed in Figure. 8 (Scenario 2). The demand coverage reaches its peak at p=300, providing service to 67.9% of charging demand in Salt Lake City Metropolitan Area.

CONCLUSION

In this paper, we present a methodological framework for EV charging demand estimation and EVSE re-allocation using advanced graph-theory based approach and mathematical modeling. First, we developed a MGPR model based on the classic web ranking algorithm-PageRank-to explore potential charging demand by translating the problem into a directed graph and utilizing trip ODs and social dimension features. Second, an optimization model CMCLP is employed to optimize the locations of EVSEs by maximizing the utilization of charging stations. In addition, a crawling pipeline is created to retrieve real-world public charging data for modeling purpose. Such pipeline framework can be widely generalizable to other cities covered by ChargePoint.

The Salt Lake City metropolitan area is chosen to demonstrate the effectiveness of the framework. Real-world data were obtained from 109 public charging stations from Nov 5th, 2020 to Dec 12th, 2020. Both WTE and MARD values indicate satisfactory performance of MGPR model, which proves that trip ODs and social dimension features can effectively infer the public charging demand. Once charging demand distribution is obtained based on the PageRank score integrated regression model, CMCLP is employed to optimize the EVSE locations. It is found that most of existing charging stations located at Salt Lake City downtown are effectively exploited; while some charging stations located in West Valley City are underused. Meanwhile, it is observed that there are mismatches between the currently deployed charging infrastructures and charging demand. More charging stations are encouraged to be sited along the interstate highways for future planning. We further examine the public charging demand coverage rate with the increase of EVSEs. The numerical results indicate that the currently installed charging stations can only cover 13.3% of the total charging demand in Salt Lake City metropolitan area. With 300 more newly sited charging stations at public parking lots, the coverage rate can reach to 67.9%. In the CMCLP model, considering the small amount of Level 3 charging stations (34 out of 576), we simplified the scenarios by assuming all charging stations are Level 2. However, several cities such as Los Angeles have much higher proportion of Level 3 charging. Future work therefore includes incorporating additional constraints to distinguish Level 2 and Level 3 charging stations.

AUTHORS CONTRIBUTION

The authors confirm contribution to the paper as follows: study conception and design: ZY, XCL, RW; data collection: ZY, XCL, RW; analysis and interpretation of results: ZY, XCL, RW; draft manuscript preparation: ZY, XCL, RW, YZ. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

1. Huo, H., H. Cai, Q. Zhang, F. Liu, and K. He. Life-Cycle Assessment of Greenhouse Gas and Air Emissions of Electric Vehicles: A Comparison between China and the U.S. *Atmospheric Environment*, Vol. 108, 2015, pp. 107–116. https://doi.org/10.1016/j.atmosenv.2015.02.073.
2. McKinsey. Expanding Electric - Vehicle Adoption despite Early Growing Pains. *McKinsey & Company*, No. August, 2019.
3. Gersdorf, T., P. Hertzke, P. Schaufuss, and S. Schenk. McKinsey Electric Vehicle Index: Europe Cushions a Global Plunge in EV Sales. *McKinsey & Company*, Vol. July, No. July, 2020, pp. 1–12.
4. Hall, D., and N. Lutsey. Emerging Best Pratices for Electric Vehicle Charging Infrastructure. *ICCT White Paper*, No. October, 2017.
5. Grote, M., J. Preston, T. Cherrett, and N. Tuck. Locating Residential On-Street Electric Vehicle Charging Infrastructure: A Practical Methodology. *Transportation Research Part D: Transport and Environment*, Vol. 74, No. April, 2019, pp. 15–27. https://doi.org/10.1016/j.trd.2019.07.017.
6. Smart, J., and S. Schey. Battery Electric Vehicle Driving and Charging Behavior Observed Early in the EV Project. *SAE Technical Papers*, Vol. 1, No. 1, 2012, pp. 27–33. https://doi.org/10.4271/2012-01-0199.
7. Engel, H., R. Hensley, S. Knupfer, and S. Sahdev. Charging Ahead : Electric- Vehicle Infrastructure. *McKinsey & Company*, No. Exhibit 1, 2018, pp. 1–8.
8. Xu, M., H. Yang, and S. Wang. Mitigate the Range Anxiety: Siting Battery Charging Stations for Electric Vehicle Drivers. *Transportation Research Part C: Emerging Technologies*, Vol. 114, No. January, 2020, pp. 164–188. https://doi.org/10.1016/j.trc.2020.02.001.
9. Csiszár, C., B. Csonka, D. Földes, E. Wirth, and T. Lovas. Urban Public Charging Station Locating Method for Electric Vehicles Based on Land Use Approach. *Journal of Transport Geography*, Vol. 74, No. June 2018, 2019, pp. 173–180. https://doi.org/10.1016/j.jtrangeo.2018.11.016.
10. Qian, K., C. Zhou, M. Allan, and Y. Yuan. Load Model for Prediction of Electric Vehicle Charging Demand. 2010.
11. Chaudhari, K., N. K. Kandasamy, A. Krishnan, A. Ukil, and H. B. Gooi. Agent-Based Aggregated Behavior Modeling for Electric Vehicle Charging Load. *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 2, 2018, pp. 856–868.
12. Leou, R.-C., C.-L. Su, and C.-N. Lu. Stochastic Analyses of Electric Vehicle Charging Impacts on Distribution Network. *IEEE Transactions on Power Systems*, Vol. 29, No. 3, 2013, pp. 1055–1063.
13. Wagner, S., M. Götzinger, and D. Neumann. Optimal Location of Charging Stations in Smart Cities: A Point of Interest Based Approach. *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design*, Vol. 3, 2013, pp. 2838–2855.
14. Dong, G., J. Ma, R. Wei, and J. Haycox. Electric Vehicle Charging Point Placement Optimisation by Exploiting Spatial Statistics and Maximal Coverage Location Models. *Transportation Research Part D: Transport and Environment*, Vol. 67, No. November 2018, 2019, pp. 77–88. https://doi.org/10.1016/j.trd.2018.11.005.

15. Vazifeh, M. M., H. Zhang, P. Santi, and C. Ratti. Optimizing the Deployment of Electric Vehicle Charging Stations Using Pervasive Mobility Data. *Transportation Research Part A: Policy and Practice*, Vol. 121, No. September 2018, 2019, pp. 75–91. https://doi.org/10.1016/j.tra.2019.01.002.
16. Kontou, E., C. Liu, F. Xie, X. Wu, and Z. Lin. Understanding the Linkage between Electric Vehicle Charging Network Coverage and Charging Opportunity Using GPS Travel Data. *Transportation Research Part C: Emerging Technologies*, Vol. 98, No. November 2017, 2019, pp. 1–13. https://doi.org/10.1016/j.trc.2018.11.008.
17. Hu, L., J. Dong, Z. Lin, and J. Yang. Analyzing Battery Electric Vehicle Feasibility from Taxi Travel Patterns: The Case Study of New York City, USA. *Transportation Research Part C: Emerging Technologies*, Vol. 87, No. April 2017, 2018, pp. 91–104. https://doi.org/10.1016/j.trc.2017.12.017.
18. Tu, W., Q. Li, Z. Fang, S. lung Shaw, B. Zhou, and X. Chang. Optimizing the Locations of Electric Taxi Charging Stations: A Spatial–Temporal Demand Coverage Approach. *Transportation Research Part C: Emerging Technologies*, Vol. 65, No. 3688, 2016, pp. 172–189. https://doi.org/10.1016/j.trc.2015.10.004.
19. Page, L., S. Brin, R. Motwani, and T. Winograd. *The PageRank Citation Ranking: Bringing Order to the Web*. 1999.
20. Xing, W., and A. Ghorbani. Weighted PageRank Algorithm. *Proceedings - Second Annual Conference on Communication Networks and Services Research*, 2004, pp. 305–314. https://doi.org/10.1109/dnsr.2004.1344743.
21. Haveliwala, T. H. Topic-Sensitive Pagerank: A Context-Sensitive Ranking Algorithm for Web Search. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No. 4, 2003, pp. 784–796. https://doi.org/10.1109/TKDE.2003.1208999.
22. Chung, F., and W. Zhao. PageRank and Random Walks on Graphs. *Bolyai Society Mathematical Studies*, Vol. 20, 2010, pp. 43–62. https://doi.org/10.1007/978-3-642-13580-4_3.
23. Kim, Y. Y., K. H. Lee, H. A. Kim, C. H. Choi, C. H. Shin, and W. S. Cho. Analysis on the Transportation Point in Cheongju City Using Pagerank Algorithm. *ACM International Conference Proceeding Series*, Vol. 20-23-Octo, 2015, pp. 165–169. https://doi.org/10.1145/2837060.2837087.
24. Wang, M., S. Yang, Y. Sun, and J. Gao. Discovering Urban Mobility Patterns with PageRank Based Traffic Modeling and Prediction. *Physica A: Statistical Mechanics and its Applications*, Vol. 485, 2017, pp. 23–34. https://doi.org/10.1016/j.physa.2017.04.155.
25. Agryzkov, T., J. L. Oliver, L. Tortosa, and J. F. Vicent. An Algorithm for Ranking the Nodes of an Urban Network Based on the Concept of PageRank Vector. *Applied Mathematics and Computation*, Vol. 219, No. 4, 2012, pp. 2186–2193. https://doi.org/10.1016/j.amc.2012.08.064.
26. Church, R., and C. ReVelle. The Maximal Covering Location Problem. No. 32, 1974, pp. 101–118.
27. Current, J. R., and J. E. Storbeck. Capacitated Covering Models. *Environment and planning B: planning and Design*, Vol. 15, No. 2, 1988, pp. 153–163.
28. Chung, C. H., D. A. Schilling, and R. Carbone. The Capacitated Maximal Covering Problem: A Heuristic. No. 1983, 1983, pp. 1423–1428.
29. Wang, Y. W., and C. R. Wang. Locating Passenger Vehicle Refueling Stations. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 46, No. 5, 2010,

pp. 791–801. https://doi.org/10.1016/j.tre.2009.12.001.

30. Kınay, Ö. B., F. Gzara, and S. A. Alumur. Full Cover Charging Station Location Problem with Routing. *Transportation Research Part B: Methodological*, Vol. 144, 2021, pp. 1–22. https://doi.org/10.1016/j.trb.2020.12.001.

31. Prianka, B., and U. T. Division. In the Quest for Solutions That Would Satisfy These Conflicting Needs , Most Local Authorities Have Preoccupied Themselves with Deriving Meth- Ods That Would Alleviate Traffic Congestion . For Example , Parking Control , Traffic Management Schemes ,' " Pe. Vol. I, No. 4, 1985, pp. 365–376.