

## Article

# Planning of Electric Taxi Charging Stations Based on Travel Data Characteristics

Yan Wang <sup>1,\*</sup> , Shan Gao <sup>1</sup>, Hongyan Chu <sup>2</sup> and Xuefei Wang <sup>3</sup>

<sup>1</sup> Department of Electrical Engineering, Southeast University, Nanjing 210096, China; shangao@seu.edu.cn

<sup>2</sup> Department of Energy and Mechanical Engineering, Nanjing Normal University, Nanjing 210042, China; njnuchuhongyan@163.com

<sup>3</sup> State Grid NANJING Power Supply Company, State Grid Jiangsu Electric Vehicle Service Company, Ltd., Nanjing Branch, Nanjing 320105, China; wangxuefei19680307@163.com

\* Correspondence: wyddseu@163.com; Tel.: +86-181-6808-0707

**Abstract:** In view of the practical application requirements for the rapid expansion of electric taxis (ETs) and the reasonable planning of charging stations, this paper presents a method for mining latent semantic correlation of large data by the trajectory of ETs and the planning of charging stations with optimal cost. Firstly, the vector space modeling method of ET trajectory data is studied, and the semantic similarity of the trajectory data matrix is evaluated. Secondly, the hidden characteristics of the mass trajectory data are extracted by matrix decomposition. Then, the latent semantic correlation characteristics of trajectory data are mined. Finally, the fast clustering of ETs is realized by the spectral clustering method. On this basis, with the objective of minimizing the annual construction and maintenance costs of charging stations, the optimal planning scheme of charging stations for ETs is given. In this paper, the spectrum clustering processing method of the potential semantic correlation of the big data of the driving track of ETs can be combined with the operation and maintenance costs of the charging station, and the convenience of charging for ET users is also considered. This provides decision support information for the reasonable planning of charging stations.

**Keywords:** electric taxi; charging station; big data; clustering; optimal planning



check for updates

**Citation:** Wang, Y.; Gao, S.; Chu, H.; Wang, X. Planning of Electric Taxi Charging Stations Based on Travel Data Characteristics. *Electronics* **2021**, *10*, 1947. <https://doi.org/10.3390/electronics10161947>

Academic Editor: Alexey Vinel

Received: 8 July 2021

Accepted: 8 August 2021

Published: 12 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the continuous development of science and technology and the concern of environmental protection, renewable energy power generation [1,2] and electric vehicles (EVs) have been rapidly developed and attracted more and more attention [3,4]. The rapid development of EVs also brings new challenges to the operation of charging stations. However, the temporal and spatial distribution characteristics of EV users and the convenience of charging should be considered. Therefore, problems related to economic benefits and user convenience must be solved in the research of charging station planning. Therefore, the study of reasonable charging station planning can make it more convenient for the daily charging of EVs, which has become the focus of research in recent years [5–7]. This can not only accelerate the rapid popularization of EVs but also be part of the efforts to develop smart cities in the coming years [8,9].

The planning of EV charging stations is affected by many factors. The study of this problem should consider the construction cost, traffic flow and user's charging demand. Many researchers put forward different research methods. In Reference [10], a new charging spot model is designed, namely a single output multiple cables charging spot. Reference [11] focuses on an EV charging station planning strategy, and it considers the construction cost and drivers' satisfaction. Studies [12] considered land cost, construction cost and other key indicators affecting the planning of charging stations and established the charging station layout model. In Reference [13], economic, environmental and social factors are taken as the evaluation indexes of planning and uncertainty, and qualitative and

quantitative factors are also considered in the planning. In [13], a Bayesian network model that considers not only quantitative factors but also qualitative ones was proposed. The developed Bayesian network model was validated through a sensitivity analysis approach. Finally, different propagation analyses have been performed to make special types of reasoning. In order to predict the actual demand of electric vehicle charging power supply, Arias, Kim and Bae [14] considered the real-time closed-circuit television data from an actual urban road network. In [14], the Markov chain traffic model and invisible state transfer method were used to determine the arrival rate of electric vehicles at charging stations near the urban road network. Then, the information obtained in the first step was used to determine the charging power demand of the public fast charging station [14]. Taking the charging demand of electric vehicles as an example, the charging demand of urban public fast charging stations was predicted. Some EV charging stations are planned by an artificial intelligence method or stochastic programming [15,16]. In the research [17], the location and scale of EV stations are determined based on the optimal economic benefits. Benefits change with time, location and capacity. In [17], a planning model method considering net present value and life cycle cost is proposed to determine the site and size of the charging stations. This model integrated the distribution network constraint, the user constraint and the traffic flow captured constraint. Islam, Hussain and Mohamed [18] converted traffic network information into EV charging demand and then planned charging stations. Their conversion method was relatively simple, and the accuracy of the analysis results needs to be further improved. Wang et al. [19] proposed a multi-objective, multistage collaborative planning model for a coupled EV charging station infrastructure and power distribution network. The planning model aims to minimize the investment and operation costs of the distribution system while maximizing the annually captured traffic flow. The uncertainties of EV charging loads for three different types of charging stations were studied: slow charging mode, battery swapping mode and fast charging mode [19]. Guo et al. [20] constructed a charging path navigation architecture integrating power grid information and traffic information. This pushes the capacity information of charging stations to EV users through an intelligent transportation system to guide EVs to charge quickly, but the authors do not conduct an in-depth analysis on the planning of charging stations. In order to improve the electrification level of smart cities, Reference [21] optimizes a combination of EV chargers. The uncertainty of EV load is also considered in [21], and the EV load is modeled as a stochastic process, and, then, the particle swarm optimization algorithm is used to solve the nonlinear stochastic problem. In Mu et al. [22], the start–stop matrix is introduced to represent the travel traffic flow, and a spatiotemporal distribution model integrating the transportation system and the distribution network system is established. The impact of the EV charging station on the operation of the distribution network is analyzed by the Monte Carlo simulation method. To a certain extent, the relationship between transportation system operation and charging station planning is established. In Gong et al. [23], the location problem of charging stations for public ETs is studied. The authors take into consideration the factors affecting charging station locations, including mileage, public ET distribution and passenger distribution.

Although these studies have different thoughts on charging station planning, they seldom study the problem from the needs of the users. As known, the charging time of ETs takes a long time, and the minimum effective charging time is usually not less than half an hour. If an ET is parked in a place for more than half an hour and the charging station is nearby, charging can be arranged in this time period, which can be convenient for ET users. It can greatly reduce the time cost and opportunity cost of the user, thus greatly improving the charging efficiency and convenience of the ET. On this basis, by considering the investment cost and operation and maintenance cost of the charging stations, the planning will be very conducive to the economic operation of the charging stations, as well as to the popularization of ETs and the development of the whole ET industry. Considering this idea, this paper studies the charging station planning problem by spectral clustering and economic optimization based on the large data of the travel characteristics of the

users of ETs from the perspective of the users, the economics of charging stations and the popularization and development of ETs. It has become one of the research hotspots at home and abroad to study the correlation between driving behavior and urban settings through vehicle trajectory data mining and then to optimize the construction planning of urban related facilities [24,25]. Massive trajectories of ETs contain not only abundant information about traffic patterns and driving behavior but also valuable information that can provide services for charging station planning. Therefore, decision support information for the planning of charging stations of ETs can be provided by acquiring a large amount of trajectory data.

The main research contents of this paper are as follows: Based on a large amount of ET driving trajectory data and the extraction of valuable information that can reflect the characteristics of the vehicle, the information can reflect the latent semantics of ET travel. Then, the study area passed by the ETs is divided by spectral clustering, and the most appropriate place is selected from the results of spectral clustering to realize charging station planning. Finally, the economics of building charging stations is evaluated. In the research process, both the demand of ET users and the economics of charging station operation are fully considered.

The rest of the paper is organized as follows: Section 2 describes the methodology used to perform charging station planning. Section 3 describes some basic definitions and the model analysis used in this paper, while Section 4 gives the research method for extracting valuable information from the driving trajectory data of ETs and analyzes how to carry out spectral clustering. The general steps of calculating and analyzing charging station planning and the evaluation of the economics of charging stations are presented in Section 5. Section 6 proposes the calculation and analyzes the results with the driving data of actual ETs, and the effectiveness of the proposed method is verified, while Section 7 concludes this paper.

## 2. Review Methodology

We now explain the methodology used for the literature review. This review focuses on papers that were published by the IEEE Xplore Digital Library, Elsevier database and Wiley database; as described in the Introduction Section, studies from earlier years are mainly focused on the electrical engineering aspect of the research area. The keyword that the search was based on was applied to the title of a paper, the abstract and keywords of the paper. The core search terms were “EVs”, “ETs”, “charging station planning”, “big data” and “clustering”. For example, in the IEEE Xplore Digital Library, search results contain keywords: charging station planning (581 results), big data (51,024 results), EV (59,868 results) and clustering (111,321 results). These papers contain more information. We select 116 highly relevant and valuable papers for careful research and analyze some of the methods of ET charging station planning.

The research on ET charging station planning has developed rapidly, and there are many methods. It can be roughly divided into two categories: the first is to consider the ET charging station planning separately [12–14,17], such as the location and capacity of the charging station, and the second is to consider the joint planning of the ET charging station and distribution network [18–22]. The research on these two kinds of problems is described in detail as follows.

For the first kind of problem, only the planning of charging stations is considered separately. Generally, the environment, service distance and other factors will be considered; then the linear or nonlinear optimization objective function will be established; and, finally, the optimization algorithm will be used to solve it. Although research on these problems can obtain a planning scheme, it only considers one-sided interests, considers the interests of ET users less and does not consider comprehensive factors.

The second kind of problem considers two factors: charging station planning and distribution network operation. A hybrid optimization model is established to minimize the total cost, and then the optimization algorithm is used to solve it. Some also put forward

the joint optimization model of distribution network development and the location and capacity of charging stations. Some also consider the impact of distributed generation on charging station planning. Most of these studies are based on the determined distribution network. Most of them plan the charging station considering the interests of distribution network operation without considering the interests of ET users at the same time.

### 3. ET Trajectory Data and Modeling Analysis

#### 3.1. Definition

**Definition 1 (Planning Space).** *The area surrounded by longitude ( $Lg_{min}$ ,  $Lg_{max}$ ) and latitude ( $La_{min}$ ,  $La_{max}$ ) of the whole area to be planned is called computing space  $Q$ , namely*

$$Q = \int_{Lg_{min}}^{Lg_{max}} (La_{max} - La_{min}) dLg \quad (1)$$

**Definition 2 (Grid space).** *Planning space  $Q$  is divided into equal  $M$  columns and  $N$  rows according to the longitude priority principle. ( $M \times N$ ) grids are formed, which are called grid space  $G$ . Each grid is called spatial grid, which is called grid  $g_i$  for short, namely*

$$G = \bigcup_{i=1}^{M \times N} g_i \quad (2)$$

where  $m$  and  $n$  are the column number and row number of the grid  $g_i$ , respectively;  $i$  is the spatial index number; and  $i = (n - 1) \times M + m$ ; thus,  $g_{m,n}$  can be expressed by  $g_i$ .

From Definitions 1 and 2, it can be known that the planning space  $Q$  and the grid space  $G$  have  $Q \equiv G$ , and for any  $i, j$  ( $i \neq j$ ), they have  $g_i \cap g_j = \Phi$ . Given any position  $p(x, y)$  in the planning space  $Q$ , a unique  $g_i$  in the grid space  $G$  exists, making  $p \in g_i$ .

**Definition 3 (ET trajectory).** *Trajectory is a time series consisting of space, time and other information collected from ET  $i$  during driving, namely*

$$D_i = \{d_1, d_2, \dots, d_k\} \quad (3)$$

where  $d_i = (Lg_i, La_i, t_i)$  is spatiotemporal data containing longitude, latitude and time information.

#### 3.2. Vector Space Modeling of the ET Trajectory

When a car passes through one location, it will inevitably pass through another location. In this part, the high-precision original driving data are transformed into a finite-precision equivalent space grid sequence, and the vector space model is built. The irregular and unstructured ET driving data are transformed into a unified structured data matrix, which is able to provide structured data support for classifying the intrinsic characteristics of the trajectory data. An ET's trajectory can be identified by the ET ID number, and a "grid-ET" matrix  $K$  is constructed in which the row of the matrix corresponds to the grid number ( $g_i$ ) of the road network space, and the column of the matrix corresponds to the ET number ( $w_i$ ) of the trajectory; thus, the structured processing of the large-scale traffic trajectory can be realized.

Assuming that there are  $n$  ETs, each ET is a column of the matrix; the number of grids in the planning space is  $m$ ; and each grid is a row of the matrix, which constitutes a "grid ET" matrix  $K$  of order  $m$  rows and  $n$  columns. In matrix  $K$ , according to the different

assignments of matrix elements  $k(i, j)$  and their physical meanings, the traffic trajectory can be mined in different aspects. Matrix  $K$  is represented as follows:

$$K = \begin{bmatrix} & w_1 & w_2 & \cdots & w_n \\ g_1 & k(1,1) & k(1,1) & \cdots & k(1,n) \\ g_2 & k(2,1) & k(2,2) & \cdots & k(2,n) \\ \vdots & \vdots & \vdots & \ddots & \cdots \\ g_m & k(m,1) & k(m,1) & \cdots & k(m,n) \end{bmatrix} \quad (4)$$

**Definition 4 (Grid heat matrix).** *If the element of matrix  $K$  is assigned to the number of times that the ET travels in the grid for more than half an hour (if it stays in a grid for less than half an hour, it will not be able to effectively charge here, so it will be regarded as invalid to pass through), then the grid heat matrix  $F$  can be obtained. Matrix  $F$  can be expressed as follows:*

$$F = \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{m1} & \cdots & f_{mn} \end{pmatrix} \quad (5)$$

where  $f_{ij}$  is defined as the number of valid passes when the ET  $w_j$  stays in the grid  $g_i$  for more than half an hour. If it passes in the grid but for less than half an hour, the number of times is zero, which is regarded as an invalid pass.

The driving heat of ETs in the grid is described as follows: if there are more effective times of ETs passing through a grid in unit time, the grid heat is higher; otherwise, if a grid passes by very few vehicles, it means that the grid heat is lower.

According to the definition, matrix  $F$  is a high-dimensional sparse matrix. By analyzing the semantic spatial correlation of matrix  $F$ , the spatial grid can be clustered automatically based on the driving heat of ETs. Therefore, this paper studies the spectral clustering algorithm based on matrix  $F$ , which can support the latent semantic correlation mining of large trajectory data of ETs, and carries out the research and validation of charging station planning.

## 4. Spectral Clustering Analysis with Trajectory Big Data

### 4.1. Trajectory Space Analysis

The essence of trajectory data clustering is to classify objects according to the similarity of their trajectory data. Therefore, how to evaluate the distance or similarity between trajectory data is one of the key issues in clustering processing.

The trajectory data of ETs are a series of time data composed of spatial position information, including both time and space information. The similarity of trajectory data is their proximity in time and space dimensions. In order to mine the correlation among the trajectories of ETs and find their commonness, a method of similarity measurement of trajectory data combining Euclidean distance in semantic space and its Gauss kernel function is applied to each cluster [26]. Firstly, the grid thermal matrix is decomposed into the semantic space of large-scale trajectory data by singular value decomposition, and, then, its similarity is measured by the Gauss kernel function according to the Euclidean distance of the latent semantic information of the trajectory data.

Singular value decomposition is an effective method for extracting context semantic information [25]. By singular value decomposition of matrix  $F$ , a semantic space consisting of several left singular orthogonal vectors is generated, which implies users' interest in various locations, and the degree of interest is derived from the inner product of the eigenvectors of the space and the corresponding singular values.

According to the singular value decomposition theorem, set  $F$  is a real matrix of order  $m \times n$ , and  $n$ -order orthogonal matrices  $M = \{m_1, m_2, \dots, m_s\}$  and  $N = \{n_1, n_2, \dots, n_s\}$  exist, so that

$$F = M * \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_s) * N^t. \quad (6)$$

where  $\lambda_i$  ( $i = 1, 2, \dots, s$ ) is a singular value of matrix  $F$ , and, in descending order, the column vector of  $M$  is the eigenvector of  $FF'$ ; the column vector of  $N$  is the eigenvector of  $F'F$ ; and  $s$  is the rank of matrix  $F$ .

By singular value decomposition of matrix  $F$ , the semantic space reflecting the context information between spatial grids is obtained, that is, the left singular vector matrix. The  $l$  component signals in front of the semantic space reflect the general picture of the original signal, which is represented as follows:

$$F_l = M_l * \text{diag}(\lambda_1, \dots, \lambda_l) * N_l^t. \quad (7)$$

The left singular vector  $M_l$  is the reduced dimension semantic subspace. The selection of dimension  $l$  affects the performance of the algorithm. If the descent speed of the  $l + 1$  singular value decreases significantly compared with the first  $l$ , then this  $l$  value is appropriate [27]. This paper selects according to this principle. In the dimension reduction subspace, the urban grid semantic space mainly represents the common information between grids. Therefore, the product of the column vector and its corresponding singular value of the reduced dimension semantic subspace constitutes the weighted semantic subspace of potential upper and lower semantic information in the spatial grid, which is expressed as follows:

$$\overline{M}_l = M_l * \text{diag}(\lambda_1, \dots, \lambda_l). \quad (8)$$

In urban spatial grid weighted semantic space, any grid has a latent semantic sequence of equal dimensions  $\{m_1, m_2, \dots, m_l\}$  and contains most of the latent semantic information between grids. Therefore, Euclidean distance can be used to evaluate the semantic similarity between any two grids. That is

$$d_{ab} = \sqrt{\sum_{i=1}^l (a[i] - b[i])^2}. \quad (9)$$

where  $d_{ab}$  is the distance between grid  $a$  and  $b$  in semantic space,  $a[i]$  ( $i = 1, 2, \dots, l$ ) is a latent semantic sequence of grid  $a$ .

#### 4.2. Application of Dimension Reduction Technology

With the increase in the number of ETs, the dimension of matrix  $F$  will become higher rapidly, which will cause dimension disaster and bring great challenges to the calculation. In this part, we give a data compression method based on JL transformation theory, which can map the data in high-dimensional space to a low-dimensional space, and keep all the important features of the original data unchanged.

**Lemma 1 (Johnson–Lindenstrauss lemma [28]).** Set  $\{u_1, u_2, \dots, u_r\}$  in  $u$ -dimensional space; for any  $\zeta \in (0, 1)$ , let integer  $h$  satisfy  $h > O\left(\frac{\ln(r)}{\zeta^2}\right)$ ; then, a linear mapping exists  $C: R^u \rightarrow R^h$ , such that for all  $i, j \in \{1, 2, \dots, r\}$ , the following holds:

$$\frac{(1 - \zeta) \|u_i - u_j\|^2}{(1 + \zeta) \|u_i - u_j\|^2} \leq \|Cu_i - Cu_j\|^2 \leq \|u_i - u_j\|^2. \quad (10)$$

Lemma 1 ensures that the data existing in the  $u$  dimensional space can be embedded in an  $h$  dimensional space so that the relative distance between data pairs can be maintained under an arbitrarily small constant factor  $\zeta$ .

In general, there are many methods to construct the random projection matrix  $C_{Jxu}$  satisfying Lemma 1. The most commonly used method is the Gaussian random projection matrix in which each element obeys Gaussian normal distribution, but the calculation of the Gaussian matrix is more complex [29]. Li proposed that a very sparse random projection matrix can be used [30]. Each element in the matrix obeys the following distribution:

$$c_{ij} = \begin{cases} \sqrt{v}, & \text{with probability } \frac{1}{2v} \\ 0, & \text{with probability } 1 - \frac{1}{v} \\ -\sqrt{v}, & \text{with probability } \frac{1}{2v} \end{cases}. \quad (11)$$

where  $v$  is a positive number.

Compared with the Gaussian random projection matrix, the sparse random projection matrix is faster. In this paper, the sparse random projection matrix is used. Through the random projection method, the heat matrix  $F$  with high dimension can be projected into a low-dimensional matrix, and the time complexity of singular value decomposition can be reduced rapidly; thus, the efficiency of singular value decomposition can be greatly improved.

## 5. Spectral Clustering and Economic Analysis of Charging Stations

### 5.1. Spectral Clustering Analysis of Charging Stations

The correlation between area grids can be represented by their semantic similarity maps. Given spatial data sets  $\{g_1, g_2, \dots, g_n\}$  and the semantic similarity relationship  $e_{i,j}$  between any two grids  $g_i$  and  $g_j$ , each grid in the planning space can be regarded as the vertex of the graph, and, then, a fully connected semantic similarity graph can be constructed. Thus, the Laplace matrix of the graph can be constructed for correlation analysis and clustering processing. The process of the spectral clustering algorithm for charging station planning is as follows:

(1). The semantic space of the trajectory of ETs is constructed. A large-scale trajectory data matrix  $F$  is established based on the trajectory data of ETs, and, then, singular value decomposition of the matrix is performed. Furthermore, the first  $k$  vectors in the trajectory semantic space are selected to represent most of the common information in the original space. The weighted semantic subspace  $\bar{M}_l$ , which reflects the potential semantic information of the spatial grid, is constructed by the inner product operation.

(2). The semantic distance matrix of the planning spatial grid is generated. The Euclidean distance is used to evaluate the semantic distance relationship among spatial grids, and the symmetric semantic distance matrix  $A$  of spatial grids is generated. The matrix  $A$  is expressed as follows:

$$A = \begin{pmatrix} & g_1 & g_2 & \cdots & g_n \\ g_1 & 0 & & & \\ g_2 & d_{2,1} & 0 & & \\ \vdots & \cdots & \cdots & \ddots & \\ g_n & d_{n,1} & d_{n,2} & \cdots & 0 \end{pmatrix}. \quad (12)$$

where  $d_{i,j}$  is the Euclidean distance between grid  $g_i$  and  $g_j$ .

(3). The semantic similarity matrix of the urban spatial grid is constructed. According to the semantic distance matrix  $A$ , this paper uses the Gauss kernel function to measure the similarity [31] and constructs the symmetric spatial grid semantic similarity matrix  $E$ . The matrix  $E$  is as follows:

$$E = [e_{ij}]_{n \times n}. \quad (13)$$

where  $e_{ij} = \exp\left(\frac{-(d_{ij})^2}{2\gamma^2}\right)$ ,  $i, j = 1, 2, \dots, n$ ;  $d_{ij}$  is the Euclidean distance in the semantic space. The larger the value is, the smaller the similarity between the spatial grids is.  $\gamma$  is the bandwidth parameter of the kernel function. It describes the speed at which the semantic

similarity of the grid decreases. The larger  $\gamma$  is, the wider the bandwidth of the kernel function is, the smoother the function curve is, and the slower the similarity decreases.

(4). The Laplacian matrix is constructed to characterize the spectral characteristics of spatial grids. According to the graph theory [32], two graphs are said to be spectrally similar if their Laplacian matrices have the same feature set. This paper studies the properties of graphs by analyzing the characteristic polynomials, eigenvalues and eigenvectors; constructs the semantic similarity matrix  $E$ ; and constructs the Laplace matrix  $J$  of undirected graphs. The matrix  $J$  is expressed as follows:

$$J = I - Z^{-1/2}EZ^{-1/2}. \quad (14)$$

where  $I$  is the unit matrix,  $Z$  is the diagonal matrix.  $Z$  represents the degree of association between each vertex and its connected vertices in the graph. It also has

$$z_{ij} = \sum_{j=1}^n e_{ij}. \quad (15)$$

(5). Grid spectral clustering is carried out based on Laplacian matrix characteristics. According to the spectral clustering algorithm, the eigenvalues and eigenvectors of Laplace matrix  $J$  are calculated, and the eigenvectors corresponding to the first  $q$  minimum eigenvalues are extracted. Matrix  $U$  is constructed. Through the  $k$ -means clustering algorithm, the data composed of the row vectors of matrix  $U$  are aggregated into  $t$ -th class.

### 5.2. Economic Analysis of Charging Station

Considering the economics of charging stations, this paper starts from the investment and construction cost and the operation and maintenance costs. The economics of charging stations mainly include annual construction cost, annual operation and maintenance costs and annual depreciation cost. The construction cost mainly includes the investment cost of the distribution transformer, land cost, charger and the cost of a new feeder and power corridor in the distribution network. The operation and maintenance costs mainly include human resource cost, equipment maintenance, repair cost and network loss. Therefore, the annual total cost of a charging station can be expressed as follows:

$$H_e = (B_e + b) \left\{ \sum_{i \in e} (r_f + r_T n_i + r_G s_i + r_C k_i) \right\} + r_m. \quad (16)$$

where  $B_e$  is the annual equivalent investment recovery factor of the charging station,  $B_e = z(1+z)^t / [(1+z)^t - 1]$ . Among them,  $z$  is the discount rate;  $t$  is the service life;  $b$  is the annual conversion coefficient;  $r_f$  are the cost of new feeder and power corridors;  $r_T$  and  $n_i$  are the unit price and quantity of the transformers, respectively;  $r_G$  and  $s_i$  are the land price and area of the charging station, respectively;  $r_m$  are the cost of operation and maintenance;  $r_C$  and  $k_i$  are the unit price and quantity of the chargers, respectively.

### 5.3. Planning Process

Based on the application of spectral clustering analysis, the operation economics of charging stations are comprehensively considered. The detailed planning analysis process is expressed as follows:

Step 1. ET driving data set is established.

Step 2. According to the clustering algorithm, the spatial semantic clustering analysis results of the ETs are given; that is, the priority of the candidate charging station location is determined.

Step 3. In the same clustering, the economic evaluation of each candidate charging station is carried out to determine the economic optimal priority ranking of candidate stations.

Step 4. According to the results of the cluster analysis and economic evaluation, the final result of the planning of the charging station is determined.

## 6. Numerical Calculation and Analysis

The experimental operating environment of this paper is the Windows 10 (64 bit) operating system. The hardware configuration of the experimental workstation is the following: the CPU is Intel i5-4210, the main frequency is 1.7 GHz, dual core and 8 GB memory; the GPU is NVIDIA P100; and the algorithm is written in MATLAB language. MATLAB software has powerful functions. Through programming, the collected driving information of ETs can be effectively extracted, and the corresponding semantic matrix can be obtained to complete clustering.

### 6.1. Planning Area and Data Preprocessing

The ET trajectory data used in the numerical calculation part of this paper are the real trip trajectory data set of Hangzhou city [33]. This data set contains 552 ETs, which is the data of nearly one month from June to July in 2019. The details are shown in Table 1.

**Table 1.** Hangzhou city taxi travel trajectory data set.

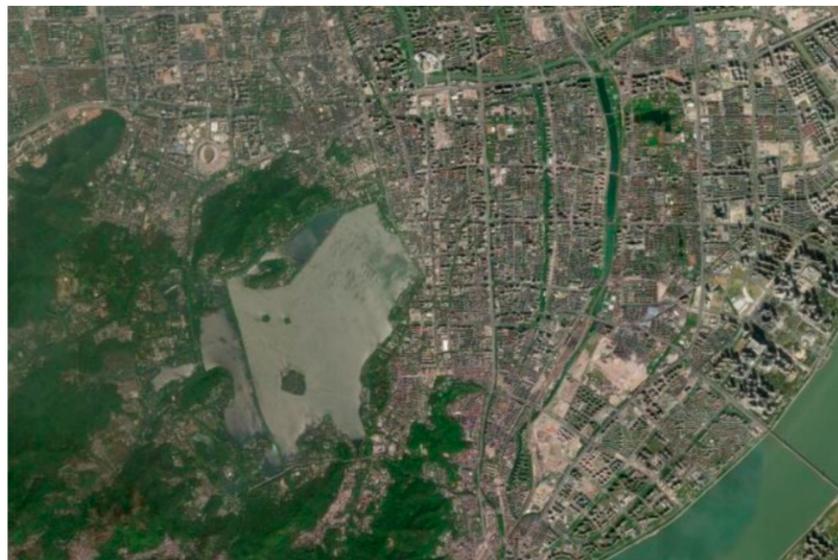
Data Set	EV Trajectory Data
Location point	241,403
Number of trajectory	380

The appearance track of the ET contains the following fields:

ET ID: each taxi has a unique number;

GPS location: there is certain location information in the extraction time.

The urban area map is shown in Figure 1 (due to the limited space, only part of the map is represented).



**Figure 1.** Planning area map.

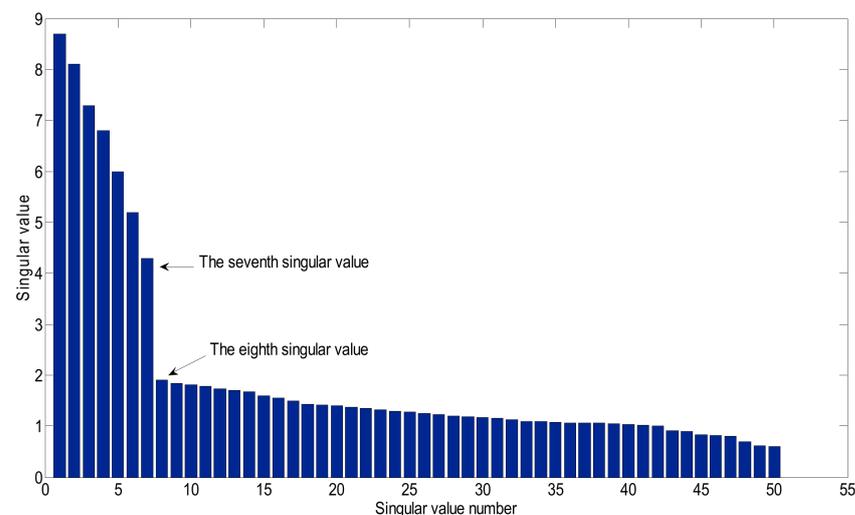
The area of the urban map studied in Figure 1 is gridded. The range of the studied area is (120.065817, 120.235709) in longitude and (30.179788, 30.337624) in latitude. This part divides the urban area into a single grid with a size of 0.003 degrees in longitude and 0.003 degrees in latitude.

### 6.2. Analysis of Charging Station Planning Results

In this part, the calculation and analysis is made based on the ET driving data in the planning space, and the priority of the charging station planning position is ranked, which provides scientific reference for the actual needs of the project.

### 6.2.1. Clustering Result of Driving Data

The matrix  $K$  is obtained from the driving data of ET in Section 6.1. The grid heat matrix  $F$  is constructed based on the experimental data, and, then, the matrix element value is the number of times the ET passes through a grid. The ET trajectory data are sparse; thus, matrix  $F$  is a highly sparse matrix. The sparse random projection transform is used to reduce the dimension of matrix  $F$ , and the time of singular value decomposition is 25.3 s. Figure 2 is a curve of the singular value. From the trend of the singular value change in Figure 2, it can be concluded that after the seventh, the range of the singular value change decreases obviously and tends to be stable. Therefore, this paper sets the dimension of semantic subspace to 7 and carries out a similarity analysis between the Laplace maps to achieve semantic spectral clustering.



**Figure 2.** Singular value in semantic space.

The calculation process is described as follows: after obtaining the grid heat matrix  $F$  and the “grid-ET” matrix  $K$ , calculate the semantic similarity of any grid with Formula (9); then, obtain the semantic distance matrix and similarity matrix shown in Formula (11) and Formula (12), respectively; further, calculate the Laplace matrix  $J$  with Formula (13); and, finally, conduct the similarity analysis between the Laplace maps. Through the  $k$ -means clustering algorithm, the semantic spectrum clustering is realized. According to the clustering results, the urban grid is divided into different categories; that is, the urban planning spatial areas are divided into different clusters. Figure 3 shows the number of times ETs pass through all grid sets in different clusters. Figure 4 shows the number of ET passes in a single grid in different clusters. Table 2 shows the actual physical meaning of the planning space corresponding to different clusters. Among them, station includes bus stations and railway stations. The market shopping center includes commodity centralized circulation places, supermarkets, large markets, etc.

From Figures 3 and 4, the clustering results show the different heat levels of the ET and the strong similarity among the grids of the trajectory. The calculation results in Table 2 also show that most of the clustering results of the proposed algorithm have practical physical significance, except that it is difficult to carry out qualitative analysis due to the small amount of data in the individual categories. It also realizes the mining and extraction of the semantic information contained in the driving of electric taxis in different grids. In the clustering calculation, some grids with large thermal disparity are clustered into the same category because of the correlation of the trajectories. In the calculation of Table 2, the number of ETs passing through the station is the highest, but the time that the ETs stay for is not long, and the number of those staying more than half an hour is less. Therefore, the station does not appear in cluster 1. The clustering results of different grids are consistent

with the actual heat of the physical locations, which also shows that the clustering results conform to the daily driving habits of ETs.

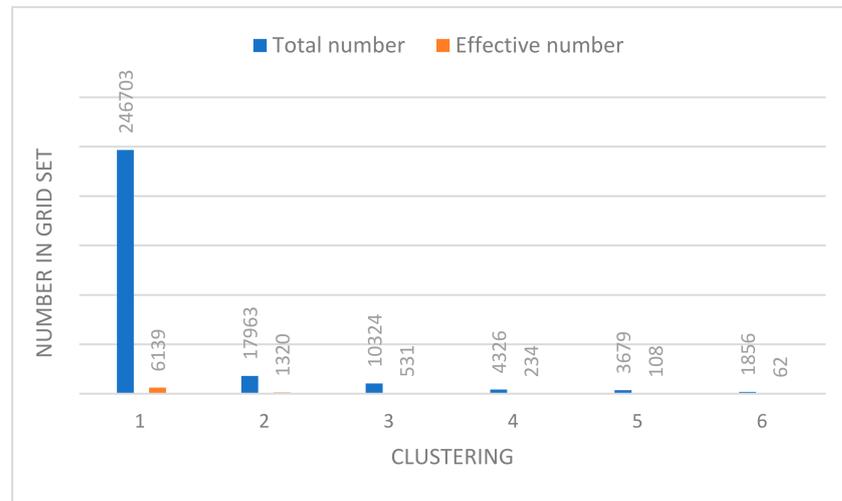


Figure 3. Total number of ETs passing in different clustering.



Figure 4. Number of passes of a single grid in different clustering.

Table 2. Physical location in different clustering.

Clustering	Physical Location
1	market, dwelling, hospital, hotel
2	dwelling, station, bar
3	hotel, park
4	school, park
5	general region
6	other

According to the planning space and clustering results, 41 grids are included in cluster 1. Some results of cluster 1 are shown in Table 3 (only partial data results are listed due to space limitation).

Table 3. Clustering results in planning area.

Clustering	Grid
Cluster 1	grid 362, grid 898, grid 1873, grid 1875, grid 1978, grid 1985, grid 2043

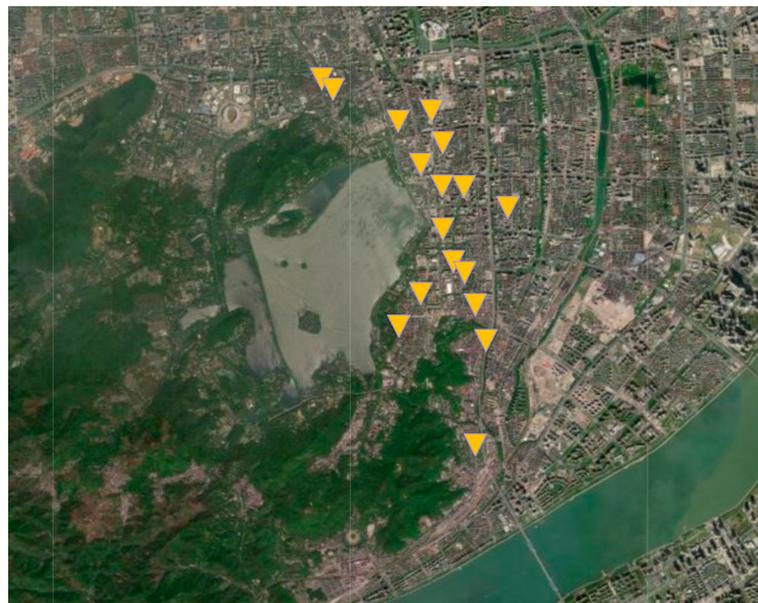
In addition, in order to reflect the potential semantic information of this planning method, Table 4 only uses the number of grids with effective times to count. Some grids with more times are listed in Table 4. In Table 4, grid 898, grid 1875 and grid 1978 are in cluster 1, while other grids are not in cluster 1.

**Table 4.** Part of the grid where the ET passes more times.

Grid	Physical Locations	Effective Pass Times
898	dwelling	135
1875	hotel	173
1074	general area	97
1978	hospital	201
2267	a company	142

Through the comparative analysis of Tables 3 and 4, the grid is simply clustered according to the number of effective passes, and the grid clustering does not completely correspond to the actual hot physical locations. If only clustering by effective pass times, the meaning of the clustering results is not clear enough, which will cause the loss of ET trajectory information and cannot reflect the semantic information of the upper and lower grid. It is difficult to achieve effective clustering and the planning of ET charging stations.

The partial grid description of cluster 1 is shown in Figure 5 (only partial data results are listed due to space limitation). In Figure 5, “▼” is the grid of cluster 1.



**Figure 5.** Partial results of cluster 1 in spatial grid. Orange symbol indicates the planning locations of the charging stations.

The clustering results in Figure 5 reflect the characteristics of the real space. The clustering results are consistent with the characteristics of the real physical location in the space. It shows that the proposed method can effectively mine the potential relevance of the semantic space and achieve the correct clustering of driving characteristics. The planning of charging stations for ETs will be preferred in cluster 1. A part of the statistical data of ET trajectories in different grids is shown in Figure 6 (only partial data results are listed due to space limitation).

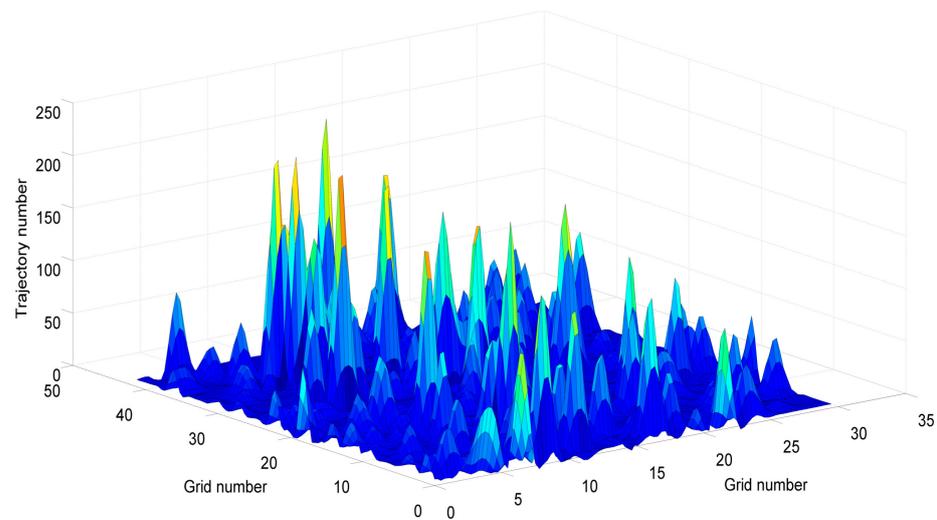


Figure 6. Part of statistics of electric taxi trajectory.

By comparing the results of cluster 1 with the actual driving track data in Figure 6, it is found that the clustering results are roughly consistent with the driving heat of the actual urban space, but the clustering results are not completely consistent with the track data; that is, those with more effective passes are not necessarily in cluster 1. As shown in grid 1074 in Figure 6, the number of effective passes of ETs is 97, which does not appear in cluster 1. This is because the trajectory of ETs in the grid is not stable every day. If it only use the statistical results of the ET trajectory to plan the charging station, it will cause the loss of track semantic space context information, which obviously cannot achieve effective clustering. In the semantic spectrum clustering given in this paper, some grids with large heat differences are automatically aggregated in the same category because of the correlation of the driving trajectory. The clustering results are not entirely based on the statistical characteristics of driving heat, which fully combines the strong correlation of ET trajectories among different grids. In this paper, the clustering results contain more information than the single trajectory data statistics.

### 6.2.2. Economic Evaluation of Charging Stations

According to the calculation results, priority is given to the locations of the charging stations in cluster 1. In the same clustering, by considering the economics to achieve economic evaluation of the charging station, the best economic factor is the location of the charging station. Because of different geographical locations, distribution network settings and other reasons, there are great differences in the economics of charging station construction in different locations. The economic evaluation of different grids is shown in Table 5.

Table 5. Annual total cost analysis of charging station.

Grid	Cost of Transformer/Million	New Feeder Cost/Million	Cost of Transformer/Million	New Feeder Cost/Million	Cost of Transformer/Million	New Feeder Cost/Million
362	0.38	0.19	5.3	0.21	0.26	2.08
898	0.39	0.21	4.1	0.22	0.21	1.72
1873	0.40	0.23	5.8	0.26	0.30	2.32
1875	0.42	0.27	3.7	0.27	0.28	1.70
1975	0.46	0.26	5.2	0.22	0.31	2.12
1978	0.50	0.20	5.9	0.28	0.26	2.38
2043	0.42	0.18	4.0	0.2	0.27	1.69

The calculation results from Table 5 show that the planning and construction costs of charging stations are different. In grid 2043, the construction cost is the smallest and the

most economical, while in grid 1978, the construction cost is higher. The calculation results also show that the total cost of different regions have great influence on charging station construction. On the basis of meeting the user's travel convenience, the economics should also be the key factor considered.

## 7. Conclusions

The reasonable location planning of ET charging stations plays a vital role in the development of ETs. Based on the travel characteristics of ETs, this paper gives the charging station planning method, constructs the charging station location model and considers the economic factors. Then, the planning method is calculated and analyzed using the actual data of ET travel. The results showed that the charging station planning method given in this paper comprehensively considered the travel habits, economy and other factors of ETs; can accurately be evaluated by the rationality of the spatial distribution of charging station resources; and, further, balanced the interests of charging stations and users. The numerical results also show that the planned location of the charging station is basically consistent with the actual popular physical location. The proposed method shows a good data processing ability and can effectively mine the potential and hidden semantic information of the data. For ETs with different calculation scales, the proposed method also effectively mined and processed the data and had good calculation advantages.

The analysis method in this paper also has some limitations. The feature meaning expressed by the clustering results is relatively single, and it cannot deal with clustering with multiple feature meanings.

Due to space constraints, this paper has some shortcomings: the ETs analyzed do not consider the different effects of the use attributes of other vehicles in the charging station planning. Because the use attributes of EVs are quite different, the reasons for staying somewhere are quite different. From a practical point of view, the collected driving track of EVs will have different effects on the planning of charging stations. This paper also does not consider the different capacities of charging stations. The different capacities of charging stations will affect the charging quantity of EVs.

In future research, the authors will fully consider the impact of the use attributes of different vehicles (such as family vehicles, buses and exclusive vehicles) on charging station planning and analyze it from the heat of the route, so as to complete the planning more in-line with the actual needs of the project. If the use attributes of EVs are considered in the study, the form and the trajectory of EVs need to be divided and classified in detail according to use attributes, which is one of the difficulties in planning. In addition, the capacity of EV charging stations will also be considered in future research. If the capacity of different charging stations is different, the charging scale and charging efficiency of EVs will be different, which will affect the results of charging station planning. These factors will contribute to more complex and in-depth research content.

**Author Contributions:** Conceptualization, methodology, investigation, simulations and experimental validation were carried out by Y.W. and S.G.; Y.W. and H.C. carried out the original draft preparation; S.G. and X.W. contributed to the writing of the paper; and H.C. and X.W. were in charge of the project supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hao, C.; Fangxing, L.; Yurong, W. Wind Power Forecasting based on outlier smooth transition autoregressive GARCH Model. *J. Mod. Power Syst. Clean Energy* **2018**, *6*, 532–539.
2. Zhaoxing, M.; Hongjie, J.; Liping, X. Study on influence of various wind farm correlations on voltage stability with multidimensional data streams method. *J. Chin. Inst. Eng.* **2019**, *42*, 738–747.

3. Yu, L.; Zhao, T.; Chen, Q.; Zhang, J. Centralized bi-level spatial-temporal coordination charging strategy for area electric vehicles. *CSEE J. Power Energy Syst.* **2015**, *1*, 74–83. [[CrossRef](#)]
4. Zhipeng, L.; Fushuan, W.; Ledwich, G. Optimal planning of electric-vehicle charging stations in distribution systems. *IEEE Trans. Power Deliv.* **2013**, *28*, 102–110.
5. Sadeghi-Barzani, P.; Rajabi-Ghahnavieh, A.; Kazemi-Karegar, H. Optimal fast charging station placing and sizing. *Appl. Energy* **2014**, *125*, 289–299. [[CrossRef](#)]
6. Cheng, S.; Feng, Y.; Wang, X. Application of lagrange relaxation to decentralized optimization of dispatching a charging station for electric vehicles. *Electronics* **2019**, *8*, 288. [[CrossRef](#)]
7. Huang, H.; Balasubramaniam, S.; Todeschini, G.; Santoso, S. A photovoltaic-fed DC-bus islanded electric vehicles charging system based on a hybrid control scheme. *Electronics* **2021**, *10*, 1142. [[CrossRef](#)]
8. Kim, H.; Mokdad, L.; Ben-Othman, J.; Dong, Z. Designing UAV Surveillance Frameworks for Smart City and Extensive Ocean with Differential Perspectives. *IEEE Commun. Mag.* **2018**, *56*, 98–104. [[CrossRef](#)]
9. Kim, H.; Ben-Othman, J.; Mokdad, L. UDiPP: A Framework for Differential Privacy Preserving Movements of Unmanned Aerial Vehicles in Smart Cities. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3933–3943. [[CrossRef](#)]
10. Zhang, H.; Hu, Z.; Xu, Z.; Song, Y. Optimal planning of PEV charging station with single output multiple cables charging spots. *IEEE Trans. Smart Grid* **2017**, *8*, 2119–2128. [[CrossRef](#)]
11. Liu, J.; Peper, J.; Lin, G.; Zhou, Y.; Awasthi, S.; Li, Y.; Rehtanz, C. A planning strategy considering multiple factors for electric vehicle charging stations along German motorways. *Electr. Power Energy Syst.* **2021**, *124*, 1–12. [[CrossRef](#)]
12. Ren, X.; Zhang, H.; Hu, R.; Qiu, Y. Location of electric vehicle charging stations: A perspective using the grey decision-making mode. *Energy* **2019**, *173*, 548–553. [[CrossRef](#)]
13. Hosseini, S.; Sarder, M.D. Development of a Bayesian network model for optimal site selection of electric vehicle charging station. *Int. J. Electr. Power Energy Syst.* **2019**, *105*, 110–122. [[CrossRef](#)]
14. Arias, B.; Arias Myungchun, K.; Sungwoo, B. Prediction of electric vehicle charging-power demand in realistic urban traffic networks. *Appl. Energy* **2017**, *195*, 738–753. [[CrossRef](#)]
15. Xiaoqing, H.; Jie, C.; Hang, Y.; Yijia, C.; Weide, G.; Bicheng, H. Economic planning approach for electric vehicle charging stations integrating traffic and power grid constraints. *IET Gener. Transm. Distrib.* **2018**, *12*, 3925–3934.
16. Hongcai, Z.; Zechun, H.; Zhiwei, X.; Yonghua, S. An integrated planning framework for different typers of PEV charging facilities in urban area. *IEEE Trans. Smart Grid* **2016**, *7*, 2273–2284.
17. Hamid, S.; Hasan, D.; Hadi, R.; Gharehpetian Gevork, B. Cost-based optimal siting and sizing of electric vehicle charging stations considering demand response programmes. *IET Gener. Transm. Distrib.* **2018**, *12*, 1712–1720.
18. Mainul, I.M.; Shareef, H.; Mohamed, A. Optimal location and sizing of fast charging stations for electric vehicles by incorporating traffic and power networks. *IET Intell. Transp. Syst.* **2018**, *12*, 947–957.
19. Shu, W.; Zhaoyang, D.; Fengji, L.; Ke, M.; Yongxi, Z. Stochastic collaborative planning of electric vehicle charging stations and power distribution system. *IEEE Trans. Ind. Inform.* **2018**, *14*, 321–331.
20. Guo, Q.; Xin, S.; Sun, H.; Li, Z.; Zhang, B. Rapid-charging navigation of electric vehicles based on real-time power systems and traffic data. *IEEE Trans. Smart Grid* **2014**, *5*, 1969–1979. [[CrossRef](#)]
21. Zeb, M.Z.; Imran, K.; Khattak, A.; Janjua, A.K.; Pal, A.; Nadeem, M.; Zhang, J.; Khan, S. Optimal placement of electric vehicle charging stations in the active distribution network. *IEEE Access* **2020**, *8*, 68124–68134. [[CrossRef](#)]
22. Mu, Y.; Wu, J.; Jenkins, N.; Jia, H.; Wang, C. A spatial-temporal model for grid impact analysis of plug-in electric vehicles. *Appl. Energy* **2014**, *114*, 456–465. [[CrossRef](#)]
23. Gong, D.; Tang, M.; Buchmeister, B.; Zhang, H. Solving location problem for electric vehicle charging stations-A sharing charging model. *IEEE Access* **2019**, *7*, 138391–138402. [[CrossRef](#)]
24. Guibin, W.; Zhao, X.; Fushuan, W.; Po, W.K. Traffic constrained multi-objective planning of electric vehicle charging stations. *IEEE Trans Power Deliv.* **2013**, *28*, 2363–2372.
25. Lv-chao, L.; Xin-hua, J.; Fu-min, Z.; Wen-wu, H.; Huai, Q. A spectral clustering method for big trajectory data mining with latent semantic correlation. *Acta Electron. Sin.* **2015**, *43*, 956–964.
26. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
27. Yang, Q.; Lihong, W.; Shuanhu, W.; Yibin, S. Analysis of similatrity between DNA sequence sets based on laplace matrix. *J. Beijing Jiaotong Univ.* **2010**, *33*, 137–140.
28. Johnson, W.B.; Lindenstrauss, J. Extension of Lipschitz mapping into a Hilbert space. *Contemp. Math.* **1984**, *26*, 189–206.
29. Dasgupta, S.; Gupta, A. An elementary proof of the Johnson-Lindenstrauss lemma. *Random Struct. Algorithms* **1999**, *22*, 1–5.
30. Li, P.; Hastie, T.J.; Church, K.W. Very sparse random projections. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA, USA, 20–23 August 2006; pp. 287–296.
31. Wei, G.; Ding-Xuan, Z. Convergence of spectral clustering with a general similarity function. *Sci. Sin. Math.* **2012**, *42*, 985–994.
32. Merris, R. Laplacian matrices of graphs: A survey. *Linear Algebra Appl.* **1994**, *197*, 143–176. [[CrossRef](#)]
33. Wangsheng, Z.; Shijian, L.; Gang, P. Mining the semantics of origin-destination flows using taxi traces. In Proceeding of the 4th International Workshop on Location Social Networks, Pennsylvania, PA, USA, 8–10 September 2012; pp. 46–50.