Research on urban public bicycle dispatching optimization method

Lin Fei [1] , **Yang Yang** [Corresp., 1] , **Wang Shihua** [1] , **Xu Yudi** [1] , **Ma Hong** [1]

[1] Hangzhou Dianzi University, Hangzhou, China

Corresponding Author: Yang Yang
Email address: 162050103@hdu.edu.cn

Unreasonable public bicycle dispatching area division seriously affects the operational efficiency of the public bicycle system. To solve this problem, this paper innovatively proposes an improved community discovery algorithm based on multi-objective optimization (***CDoMO***). The data set is preprocessed into a lease/return relationship, thereby it calculated a similarity matrix, and the community discovery algorithm ***Fast Unfolding*** is executed on the matrix to obtain a scheduling scheme. For the results obtained by the algorithm, the workload indicators (scheduled distance, number of sites, and number of scheduling bicycles) should be adjusted to maximize the overall benefits, and the entire process is continuously optimized by a multi-objective optimization algorithm ***NSGA2***. The experimental results show that compared with the clustering algorithm and the community discovery algorithm, the method can shorten the estimated scheduling distance by 20%-50%, and can effectively balance the scheduling workload of each area. The method can provide theoretical support for the public bicycle dispatching department, and improve the efficiency of public bicycle dispatching system.

* Corresponding author. *E-mail address*: *yangyang_hdu@163.com*

Research on Urban Public Bicycle Dispatching Optimization Method

Fei Lin[a],Yang Yang[a*],Ritai Yu[a],Shihua Wang[a],Yudi Xu[a],Hong Ma[a]

[a]*School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018 ,China*

Abstract

Unreasonable public bicycle dispatching area division seriously affects the operational efficiency of the public bicycle system. To solve this problem, this paper innovatively proposes an improved community discovery algorithm based on multi-objective optimization (***CDoMO***). The data set is preprocessed into a lease/return relationship, thereby it calculated a similarity matrix, and the community discovery algorithm ***Fast Unfolding*** is executed on the matrix to obtain a scheduling scheme. For the results obtained by the algorithm, the workload indicators (scheduled distance, number of sites, and number of scheduling bicycles) should be adjusted to maximize the overall benefits, and the entire process is continuously optimized by a multi-objective optimization algorithm ***NSGA2***. The experimental results show that compared with the clustering algorithm and the community discovery algorithm, the method can shorten the estimated scheduling distance by 20%-50%, and can effectively balance the scheduling workload of each area. The method can provide theoretical support for the public bicycle dispatching department, and improve the efficiency of public bicycle dispatching system.

Keywords: public bicycle system; multi-objective optimization; community discovery algorithm; regional scheduling workload; elite strategy.

1. Introduction

With the progress of urbanization, people's awareness of low carbon life and health is increasing. The public bicycle system can provide a green and healthy way to travel, and gradually become an important part of the public transport system. However, the study of the division of public bicycle dispatching area is still in the primary stage. The division of the public bicycle scheduling area has two purposes: decomposing the scheduling between large-scale sites, and reducing the computational complexity of path planning.

At present, the mainstream regional division method is based on the urban administrative area, and each area is an independent scheduling area. However, the boundaries of residents' travel are not as clear as the administrative areas. With the development of the city, the links between the areas are more closely related, so the division based on urban administrative areas is lack of scientific basis. Because the size and population density of each administrative area are different, the number of sites in each area varies greatly. The administrative area is large in size and concentrated in population. Therefore, there are more sites, public bicycle turnover is high, and dispatching workload is large; but if there are fewer sites, public bicycle turnover is low, and dispatching workload is small. Above all, lack of a scientific planning method often leads to higher scheduling capital costs.

Aiming at the problems, this paper proposes an improved community discovery algorithm based on multi-objective optimization. By using this innovative algorithm, the results show that the algorithm brings three major benefits: it can effectively shorten the public bicycle scheduling distance, improve the scheduling efficiency, and effectively balance the workload of regional scheduling.

2. Related work

The division of the public bicycle dispatching area involves operational research, and researchers have made significant contribution. Public bicycles and buses, as well as cargo transport vehicles are public transport, and their operations have similarities. Therefore, they can learn dispatching methods from each other. T. Tulabandhula [1] proposed a passenger monitoring system for dispatching vehicles in a public transportation network, it monitors passengers at the station, vehicle scheduling information and processed hardware equipment. G. Q. Pan [2] designed a heuristic simulated annealing hybrid search algorithm for large-scale VRP distributed problems. Firstly, based on the actual road network of GIS, the mathematical model is established. Secondly, the large-scale VRP path planning problem is studied. Forma [3] considers the spatial nature of public bicycle rentals, and the original inventory factor of bicycles. Then the paper establishes a regional maximum diameter distance constraint model. Finally, the best classification results are obtained by heuristic algorithms to minimize the overall inventory cost. Schuijbroek [4] applied the maximum algebra algorithm to the division of the public bicycle scheduling area, and the paper established the corresponding partition mathematical model. The goal of zoning is to minimize the maximum completion time based on a reasonable level of service. Kloimllner [5] decomposes the problem of public bicycles into two sub-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

problems: scheduling area partitioning and scheduling path planning. Then create an integer programming model to achieve as few bicycle rental points as possible.

In addition, other researchers chose to use clustering algorithms. H. Z. Dong [6] used the rental rules between public bicycle stations, the space of public bicycle stations, and the non-spatial attributes of public bicycle stations, as well as the self-flow characteristics, using association rules to classify sites with strong correlation into the same category. Finally, various types of site space enclosed areas serve as the scheduling area for public bicycles. J. Zhang [7] proposed a public bicycle scheduling area division scheme based on the improved *K-means* clustering algorithm. In the data analysis, the algorithm effectively estimates the *k* central sites at the initial central site. After the *K-means* clustering algorithm is divided, the edge sites are clustered and adjusted again according to the scheduling requirements. C. Wang [8] integrates the spatial relationship between the sites, and the lease relationship of the bicycle, establishes the similarity matrix of the site, and proposes the parameters of the regional coupling, quantifies the degree of connection between the regions, and finally uses the clustering algorithm to obtain the corresponding result. W. C. Yu [9] established a dynamic regional scheduling model, for large-scale public bicycle scheduling problems, and proposed a multi-stage re-optimized dynamic clustering algorithm, integrates optimal division, task balance between regions and regions. Within the balance of demand, three factors are progressively clustered, and in the process of solving, the abnormal sites are continuously split to gradually improve the clustering results. J. M. Liu [10] has studied the public bicycle dispatching area, found that there are often abnormal sites in the division, and he proposed a *K-Center* algorithm, adaptively limits the capacity of the rental site. Austwick M. Z. [11] analyzed the spatial attributes and community structure of public bicycles, and the paper used the community discovery algorithm to analyze the community structure of public bicycles in Washington, London and Boston, and verified the existence of community structure in the public bicycle network.

The main method of scheduling area division is model method [12] and clustering algorithm [13]. The model method requires abstract research, and there are many constraints and it is not easy to solve. Clustering algorithm is very difficult to determine the number of clusters, and it is difficult to evaluate. Moreover, the scheduling workload has no evaluation criteria, and it does not consider whether the workload is balanced. Therefore, this paper proposes a new method to solve the problem.

3. Scheduling area division model design

This part establishes the division model of public bicycle scheduling area, including the description of the model, and the assumptions of some conditions, and some interpretations of the parameters. Finally, this chapter will propose a lease/return point demand forecasting model, the data obtained from this model can help this paper verify whether ***CDoMO***'s estimated total dispatch distance is the shortest.

3.1. Problem description

At present, the clustering algorithm is mainly used to solve the problem of scheduling area division. The data set abbreviated to ***DS*** is preprocessed using a data preprocessing program. Turn a data set into a lease/return relationship abbreviated to ***LRR*** between sites. Then, through the similarity calculation between the sites, the similarity matrix abbreviated to ***SM*** is generated [14]. Conversion from ***DS*** to ***SM***, as shown in ***Equation (3.1)***, where R_{ij} represents the similarity between site *i* and site *j*, Q_{ij} represents the number of bicycles rented from the site *i* and returned to the site *j*, Q_{ji} represents the number of bicycles rented from the site *j* and returned to the site *i*.

$$\boldsymbol{DS} \xrightarrow{\boldsymbol{c_1}} \boldsymbol{LRR} = \begin{array}{c} \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1j} \\ Q_{21} & Q_{22} & \ddots & Q_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{i1} & Q_{i2} & \cdots & Q_{ij} \end{bmatrix} \\ \begin{bmatrix} Q_{11} & Q_{12} & \cdots & Q_{1i} \\ Q_{12} & Q_{22} & \ddots & Q_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{j1} & Q_{j2} & \cdots & Q_{ji} \end{bmatrix} \end{array} \xrightarrow{\boldsymbol{c_2}} \boldsymbol{SM} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1j} \\ R_{21} & R_{22} & \ddots & R_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ R_{i1} & R_{i2} & \cdots & R_{ij} \end{bmatrix} \tag{3.1}$$

The conversion process represented by $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ is as follows ***equation (3.2)***, ***equation (3.3)***, ***M*** represents the time range, which is based on the number of days:

$$\boldsymbol{c_1}{:}progressing\ program \tag{3.2}$$

$$\boldsymbol{c_2}{:}R_{ij} = \frac{Q_{ij} + Q_{ji}}{\boldsymbol{M}} \tag{3.3}$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

$$\boldsymbol{SM} = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1j} \\ R_{21} & R_{22} & \ddots & R_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ R_{i1} & R_{i2} & \cdots & R_{ij} \end{bmatrix} \xrightarrow{\boldsymbol{CA}} \boldsymbol{DR} = \{R_1, R_2, \cdots, R_n\} \tag{3.4}$$

Finally, the clustering algorithm abbreviated to ***CA*** is used for dividing, $\boldsymbol{R_n}$ stands for dividing into ***n*** independent scheduling areas is shown in ***equation (3.4)***. If the division result abbreviated to ***DR*** conforms to the lease/return law abbreviated to ***LRL***, the user can actively complete a part of the scheduling work to reduce the scheduling workload. However, in the actual scheduling area division, in order to obtain the highest comprehensive benefits, the regional division should not only conform to the law, but also achieve the balance of scheduling workload as much as possible [15]. The regional scheduling workload is mainly determined by the distance within the area and the number of stations in the area. $\boldsymbol{Z_1}$ and $\boldsymbol{Z_2}$ should be as small as possible if the regional workload is balanced. This balance problem can be transformed into multi-objective optimization problem. The objective function $\boldsymbol{f}$ is shown in ***equation (3.5)***:

$$\boldsymbol{R} = \{R_1, R_2, \cdots, R_n\} \xrightarrow{\boldsymbol{MOO}} \min \boldsymbol{f} = [Z_1, Z_2]^T \tag{3.5}$$

$\boldsymbol{Z_1}$: Variance of the dispatch distance $\boldsymbol{Z_2}$: variance of the number of sites

MOO: Multi-objective optimization

Calculation of $\boldsymbol{Z_1}$ in the following ***equation (3.6)***, ***n*** represents the number of areas, $\boldsymbol{D_i}$ represents the estimated dispatch distance of area ***i***, and $\boldsymbol{\overline{D}}$ represents the average of the estimated dispatch distances:

$$Z_1 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2 \tag{3.6}$$

Calculation of $\boldsymbol{Z_2}$ in the following ***equation (3.7)***, ***n*** represents the number of areas, $\boldsymbol{N_i}$ represents the number of internal sites in area ***i***, and $\boldsymbol{\overline{N}}$ represents the average number of internal sites:

$$Z_2 = \frac{1}{n-1}\sum_{i=1}^{n}(N_i - \overline{N})^2 \tag{3.7}$$

s.t.

$$S = [(S_i - P) \cup (S_j - P)] \cup P \tag{3.8}$$

Equation (3.8) indicates that each site must be divided into an area. $\boldsymbol{S_i}$ and $\boldsymbol{S_j}$ represent the partition set. ***P*** represents the parking lot sites collection:

$$[[S_i - P] \cap [S_j - P]] = \emptyset \ (i \neq j) \tag{3.9}$$

Equation (3.9) indicates that a site can only be divided into an area:

$$S_i \cap P \neq \emptyset, S_j \cap P \neq \emptyset \tag{3.10}$$

Equation (3.10) indicates that each scheduling area contains at least one dispatch center. There are two optimization goals for this issue:

- **Minimize** the variance between the estimated dispatch distance between each area;
- **Minimize** the variance between the numbers of sites in each area.

3.2. Model assumptions and parameter description

The scheduling area dividing process is complicated, and the abstract model involves many parameters. In order to make the model as close as possible to the actual division, before the model is established, some assumptions about the scheduling area dividing process are assumed:

- The scheduling distance of each area can be estimated theoretically, the estimated scheduling distance is approximately equal to the actual scheduling distance;
- Dispatching vehicles are not limited by driving time and mileage;
- Only one dispatching vehicle in each area is responsible for bicycle dispatch;
- Model of the dispatching vehicle is consistent with all parameters;
- The dispatching vehicle departs from the dispatching center, completes the dispatching task, and then returns to the original dispatching center, regardless of vehicle failure, and other unexpected factors.

Based on the problem description and model assumptions, the parameters and variables of the model in ***Table 3.1*** are defined.

3.3. Leasing demand forecasting model

After the scheduling area is divided, in order to calculate the estimated total distance of the scheduling, it is necessary to ensure that the demand for the lease/return site is known, so it is necessary to predict the scheduling demand for the lease/return site in the future. This section will be divided into ***24-time*** periods in hours per day named t, $t \in \{0,1,\cdots,23\}$. A Meteorology Similarity Weighted K-Nearest-Neighbour (***MSWK***) method is introduced to predict the number of least and returned bicycle at the site.

3.4.1. Leasing number forecast model

MSWK is an improved method for predicting lease/return bicycle quantity based on ***KNN*** algorithm. Analysed the amount of leasing in a similar time period to predict future leasing. Weather, temperature, humidity, winds speed, and visibility are measured in ***5*** indicators.

In the measurement of the similarity of weather, the weather is split into ***5*** levels and assigned corresponding values. The exact values are shown in ***Table 3.2***.

The quantified weather conditions at p and q for two days t is denoted by $W_{D_p}{}^t$ and $W_{D_q}{}^t$, respectively, and the weather similarities for t in p and q are defined as follows ***equation (3.11)***:

$$\lambda_1 = \frac{1}{2\pi\sigma_1} e^{-\frac{\left(W_{D_p}{}^t - W_{D_q}{}^t\right)^2}{\sigma_1^2}} \tag{3.11}$$

The temperatures of the p and q two days t periods are denoted by $F_{D_p}{}^t$ and $F_{D_q}{}^t$. The temperature similarities of the t time periods in p and q are defined as follows ***equation (3.12)***:

$$\lambda_2 = \frac{1}{2\pi\sigma_2} e^{-\frac{\left(F_{D_p}{}^t - F_{D_q}{}^t\right)^2}{\sigma_2^2}} \tag{3.12}$$

The three dimensions of humidity, wind speed, and visibility are represented by a ***3-D*** Gaussian kernel function, and $H_{D_p}{}^t$, $S_{D_p}{}^t$, $V_{D_p}{}^t$ represents the humidity, wind speed, and visibility of the t time period in p, respectively. The humidity, wind speed, and visibility similarity of p and q periods in t are defined as follows ***equation (3.13)***:

$$\lambda_3 = \frac{1}{2\pi\sigma} e^{-\left(\frac{\left(H_{D_p}{}^t - H_{D_q}{}^t\right)^2}{\sigma_3^2} + \frac{\left(S_{D_p}{}^t - S_{D_q}{}^t\right)^2}{\sigma_4^2} + \frac{\left(V_{D^t_p} - V_{D_q}{}^t\right)^2}{\sigma_5^2}\right)} \tag{3.13}$$

In order to simplify the calculation, the temperature, humidity, wind speed, and visibility are normalized and all $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$ are set to 1, thereby simplifying the calculation; finally, by weighting the above three similarity indexes, p and q can be obtained. The overall similarity indicator at time t as follows ***equation (3.14)***:

$$M\left(D_p{}^t, D_q{}^t; a\right) = \delta_w\left(D_p{}^t, D_q{}^t\right) \sum_{i=1}^{3} a_i \lambda_i \tag{3.14}$$

Where $\delta_w\left(D_p{}^t, D_q{}^t\right)$ is a judgment function, when both p and q are working days or all non-working days, $\delta_w\left(D_p{}^t, D_q{}^t\right) = 1$, otherwise $\delta_w\left(D_p{}^t, D_q{}^t\right) = 0$.If you want to predict the amount of rent in the t time period in q, select the most similar ***K*** days and use the ***MSWK*** algorithm to calculate the predicted value. The specific ***equation (3.15)*** is as follows:

$$s_i \cdot pd\left(D_q{}^t; a\right) = \frac{\sum_{p=1}^{K} M\left(D_p{}^t, D_q{}^t; a\right) s_i \cdot pd\left(D_p{}^t\right)}{\sum_{p=1}^{K} M\left(D_p{}^t, D_q{}^t; a\right)} \tag{3.15}$$

3.4.2. *returning number forecast model*

After a user rents a bicycle, they often return the bicycle to an adjacent site within a certain period of time [16]. Therefore, there is a need for prediction data of the number of bicycles based on neighbouring sites, which is used to predict the number of bicycles returned to the site. Bicycles rented from site i during time t may be returned to site j adjacent to i during time t or $t+1$. For the forecast of the number of return bicycles within the lease time t period, it is

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

necessary to first estimate the number of bicycles rented from the site *i* and at the site *j* within the time period *t*. The specific ***equation (3.16)*** is as follows:

$$e_{ij}{}^{t} = s_i \cdot pd(t)\frac{e_{ij} \cdot f}{s_i \cdot pd} \tag{3.16}$$

Among them, $s_i \cdot pd(t)$ is the predicted value of bicycle rental quantity from site *i* in time period *t*, $e_{ij} \cdot f$ is historical record of bicycle rental from site *i* and is still at site *j*. $s_i \cdot pd$ is historical total bicycle rental record from site *i*. Through the analysis of historical data, it is found that the user's riding time law can be fitted by the 2-Gaussian function. Therefore, the riding time $D_{ij}(t)$ between rental sites *i* and *j* can be estimated by ***equation (3.17)***:

$$D_{ij}(t) = g_1(t;\mu_1,\sigma_1) + g_2(t;\mu_2,\sigma_2) \tag{3.17}$$

Assume that the user's return time is evenly distributed, and the user's behaviour of returning the leased bicycle is completed within the *t* time period or *t+1*-time period. During the time periods *t* and *t+1*, the user t_1 rents a bicycle from the site *i* at the moment, and the probability of returning the ticket at the site *j* at t_2 is as follows ***equation (3.18)***, ***equation (3.19)***:

$$P_{ij}{}^{t} = \frac{1}{|t|}\int_0^{|t|}\int_0^{|t|-t_1{}'} dt_1{}' dt_2 D_{ij}(t_2) \tag{3.18}$$

$$P_{ij}{}^{t+1} = \frac{1}{|t|}\int_0^{|t|}\int_{|t|-t_1{}'}^{+\infty} dt_1{}' dt_2 D_{ij}(t_2) \tag{3.19}$$

Finally, considering the traffic patterns and the corresponding probabilities of the adjacent sites, the formula for predicting the number of return bicycles within the sites is obtained as follows ***equation (3.20)***:

$$s_i \cdot dd(t) = \sum_{j \neq i} e_{ij}{}^{t} P_{ij}{}^{t} + e_{ij}{}^{t-1} P_{ij}{}^{t+1} \tag{3.20}$$

So far, the demand ΔN of the site *i* at the time *t* in the future will be calculated by combining the demand for rental and return of the rental site *i* at the time *t* in the future. The specific formula is as follows ***equation (3.21)***:

$$\Delta N = s_i \cdot dd(t) - s_i \cdot pd(t) \tag{3.21}$$

If ΔN is less than zero, it means that the site *i* will not be able to meet the user's bicycle rental demand at the time *t* in the future, and it is necessary to dispatch the bicycle through dispatch [17]. If ΔN is greater than the number of parking spots at the leased site, it means that the site *i* at the time *t* in the future cannot satisfy the user's demand for returning the car. It is necessary to reduce the number of bicycles by scheduling.

4. Community discovery algorithm based on multi-objective optimization

Community discovery algorithm based on multi-objective optimization, which combines quantitative indicators of regional scheduling workloads, community discovery algorithms [18], and multi-objective optimization algorithms [19]. Firstly, the ***Fast Unfolding*** community discovery algorithm [20] is performed based on the similarity matrix of the site. Secondly, the workload adjusts the results of the community discovery algorithm. Throughout the process, the results are continuously optimized through a multi-objective optimization algorithm.

*4.1. **CDoMO** scheduling workload analysis*

Scheduling workload is an indicator to measure the workload of a dispatch line. The scheduling itself involves many fields, so there is no uniform standard [21]. The generalized scheduling workload is mainly determined by the scheduling distance, the delivery volume and the number of service outlets. The three parameters are weighted and integrated, and the workload of the dispatching line can be quantified. Suppose $\boldsymbol{W}$ is the generalized scheduling workload, $\boldsymbol{D}$ is the driving distance(*km*), N is the number of outlets(*pieces*), $\boldsymbol{S}$ is the delivery amount(*pieces*), and $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$,$\boldsymbol{\rho_3}$ is the driving distance weight, the delivery amount weight, and the service outlet quantity weight as follows ***equation (4.1)***:

$$W = \rho_1 \cdot D + \rho_2 \cdot N + \rho_3 \cdot S \tag{4.1}$$

This paper combines generalized scheduling workload with public bicycles, and then obtains a quantitative formula for regional scheduling workload, W_i is the scheduling workload of area i , and $\boldsymbol{D_i}$ is the scheduling distance of area i, which is calculated by the maximum generation star algorithm. $\boldsymbol{N_i}$ is the number of stations in area i, $\boldsymbol{S_i}$ is the number of stations in area i, and $\boldsymbol{\rho_1}$,$\boldsymbol{\rho_2}$,$\boldsymbol{\rho_3}$ is the corresponding weight coefficient as follows ***equation (4.2)***:

$$W_i = \rho_1 \cdot D_i + \rho_2 \cdot N_i + \rho_3 \cdot S_i \tag{4.2}$$

Since the regional scheduling is based on all stations in the entire area, and in the scheduling area division stage, the waiting scheduling sites and scheduling quantities of each area are unknown, so in this paper, the impact of $\boldsymbol{S_i}$ on

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

the scheduling workload is ignored, that is, let $\boldsymbol{\rho_3} = 0$. So, the ***equation (4.3)*** can be simplified to:

$$W_i = \rho_1 \cdot D_i + \rho_2 \cdot N_i \tag{4.3}$$

In the quantitative formula of scheduling workload, the weight coefficient cannot be determined manually, but when the scheduling workload balance is satisfied, the estimated scheduling distance variance in each area, and the variance of the number of stations in each area should be as small as possible, so the scheduling balance the problem can be turned into a multi-objective optimization problem. The objective function is$\min f = [Z_1, Z_2]^T$. ***NSGA2*** is the most popular multi-objective genetic algorithm. ***NSGA2*** first genetically manipulates the population *P* to obtain the population *Q*; then the populations are combined and then combined with non-inferior sorting and crowding distance sorting, and then a new population is established. Repeat the above process, until the termination condition is met. The detailed process is as follows:

(1) Randomly generate the initial population P_0, then sort the populations non-inferiorly, and assign a non-dominant value to each individual; then perform the operations of selection, crossover, and mutation on the initial population P_0 to obtain a new population Q_0, set to $i=0$.

(2) Combine the populations of the father and offspring, then form a new population $R_i = P_i \cup Q_i$, and then sort the population R_i non-inferiorly to obtain the non-inferior layer *F1, F2,⋯*.

(3) Perform replication, crossover, and mutation operators on population P_{i+1} to form population Q_{i+1}.

(4) If the termination condition holds, then it ends; otherwise, $i=i+1$, go to step (2).

The main process diagram of ***NSGA2*** is shown in ***Figure 4.1***:

This paper uses the ***NSGA2*** multi-objective optimization algorithm to resolve the scheduling area partition model [22]. The length of the chromosome in ***NSGA2*** is 2 [23], which corresponds to the value of the weight parameter in the area scheduling workload. Each individual corresponds to a scheduling workload formula, based on schedule the workload adjustment community found the results of the division. ***Figure 4.2*** shows the restricted flow of ***NSGA2*** algorithm.

*4.2. **CDoMO** algorithm design*

Community discovery algorithm built on multi-objective integrates quantitative indicators of regional scheduling workloads, community discovery algorithms and multi-objective optimization algorithms. Firstly, the ***Fast Unfolding*** community discovery algorithm is implemented based on the similarity matrix of the least sites; secondly, the workload index is used to adjust the results of the community discovery algorithm. The entire process continuously optimizes the results from a multi-objective optimization algorithm.

Table 4.1 displays the detailed algorithm calculation.

5. Experiment and analysis

The rest of the paper is part of the experiment and analysis. The experimental section was divided into two groups, which were experiments using New York public bicycle data and Chicago public bicycle data. In the analysis section, the two groups of experiments use ***K-means*** clustering algorithm, and ***Fast Unfolding*** community discovery algorithm as comparisons, it compares the three aspects of the number of rental sites, the variance of the number of scheduled bicycles, and the estimated total distance of scheduling. The comparative data show that the algorithm is effective against both sets of experiments.

5.1. New York public bicycle

5.1.1. Data set introduction

New York Public Bicycle [24] is a people-benefit project launched by the New York City Government. ***Figure 5.3*** displays the spatial distribution of rental sites. Blue represents Manhattan, with 250 rental sites; Green represents Brooklyn, with 77 sites. Each Citi Bicycle rental site has GPS location information, so it is not difficult to locate the rental site. The system records the user's data onto each cycle. The package contains the location and time data onto the start and the end of the site, the entire riding process, the bicycle ID, and the user's gender and birth date. This experiment will use the May 2016 rent-return dataset of New York public bicycles to conduct an experiment, a total of 96, and 1986 rent-return data. The dataset contains 16 fields, and the 9 fields related to this experiment are shown in the following ***Table 5.1***.

This paper uses the pre-processing program to process the leased data, it turned into the lease-return relationship between the least sites [25]. It also generates a similarity matrix based on the rent-return relationship. The similarity calculation formula for the least sites is as follows ***equation (4.4)***:

$$R_{ij} = \frac{Q_{ij} + Q_{ji}}{M} \tag{4.4}$$

Among them, R_{ij} represents the similarity between site *i* and site *j*; Q_{ij} represents the number of times to rent a bicycle from site i and site j to return the bicycle; Q_{ji} represents the number of times of renting a bicycle from site i and returning it at site j; M represents the time range in days. In this experiment, the data set was a total of 31 days in May 2016, so M=31.The corresponding abstract network can be generated through the lease-return relationship as ***Figure 5.4*** shows. As shown in ***Table 5.1***, due to the dense population, dense sites, and prosperous business, the sites in Manhattan are more closely linked, and Brooklyn is a river is separated from Manhattan, so the connection between the two regional sites is sparse except for the leases along the river.

5.1.2. Experimental result

In the experiment, we first used the Gephi visualization network analysis platform to analyse the community structure in the data [26]. The Gephi platform uses the integrated Fast Unfolding algorithm, it divides the public bicycle abstraction network according to the rules of public bicycle rental. The Fast Unfolding algorithm mainly includes two phases. The first phase is known as Modularity Optimization. The main part is to divide each node into the community, its neighbourhood nodes are located, so that the value of the module degree becomes larger; the second phase is called community. Aggregation is mainly to aggregate the communities divided in the first step into one site, that is, to rebuild the network based on the community structure generated in the previous step. Repeat the above process, until the structure of the network no longer changes as ***Figure 5.1*** shows.

After the ***Fast Unfolding*** algorithm for the New York public bicycle rental site in this paper， the result are shown in ***Figure 5.2*** shows the internal community structure of the abstract network of New York's public bicycles, where the dots represent sites, where the sites of different communities are represented by different colours, and the lines represent the relationships between the sites; Obviously, six communities have more close contact with leases within the same community, and the links between different societies are relatively sparse. The results of the ***Fast Unfolding*** community discovery algorithm are mapped to map on New York, as shown in ***Figure 5.5*** Manhattan is a densely populated administrative district, and the vast majority of public bicycles in the area ride on the inside, so the Manhattan District is divided into five areas according to the law of rent. Brooklyn is structured in a district. Although the division results are relatively reasonable, there are still many abnormal sites. These abnormal sites are far away from their respective areas; the number of sites of each area is uniform.

CDoMO is based on community discovery algorithm, considering the regional scheduling workload factors. The regional scheduling workload is determined by estimated dispatch distance and the number of regional least sites. If the community finds out that there are abnormal sites, it will cause regional forecasting scheduling distance become larger, so that the variance between the scheduling distances will become larger. If there is a major difference in the number of sites between areas, the variance between the numbers of sites will increase. The goal of ***CDoMO*** is to optimize the variance of the distance between the regional scheduling, and optimize the variance of the number of sites. In the optimization process, the division results can be adjusted to make it more reasonable. The division process does not take into consideration the deficiencies in the workload balance in each scheduling area. After the community discovery algorithm based on multi-objective optimization solves the division model of the public bicycle scheduling area, the experimental results shown in ***Figure 5.6*** are obtained. By comparison between ***Figure 5.4***, the result shows that the sites along the Williamsburg Bridge and the riverside along Manhattan is divided into the same dispatch area, which is more n line with the rules of public bicycle rental and resolving the anomaly [27]. The difference between the number of sites and regional sites is too large [28].

In order to maintain the consistency of the experiment, the value of k in the classical clustering algorithm ***K-means*** algorithm is set to 6 [29], and then the clustering is based on the same data set; the space area enclosed by the sites in each class as the scheduling area In order to achieve regional division, the results of the regional division based on the clustering algorithm are shown in ***Figure 5.7***, it can be found that when the clustering number ***k*=6**, the clustering algorithm achieves a poor regional division. The number of sites in the class represented by the red is very large, while the number of classes represented by purple and beige is very small, and the number of sites to vary greatly from the types. In addition, the boundaries of each scheduling area are unclear and overlapped [30].

5.1.3. Algorithm performance comparison results

Built on the overall experimental results of the above three methods, it is found that the multi-objective optimization-based community discovery algorithm proposed to this paper can make the division of the areas consistent with the rules and make the regional scheduling workload as balanced as possible. In addition to the analysis of the overall distribution of provincial division space, the paper also compares and analyses the three dimensions of the regional rental site variance, the regional dispatch distance variance, and the estimated total dispatch distance. ***Figure 5.8*** compares the variance between the numbers of sites. The data show that the variance between the ***CDoMO*** compared to the ***K-means*** algorithm is reduced by 63.31%, and the variance of the ***Fast Unfolding*** algorithm is reduced by 32.32%. ***Figure 5.9*** compares the variance of the number of bicycles dispatched in the area. The data show that the variance of the ***CDoMO*** algorithm compared to the ***K-means*** algorithm is reduced by 88.06%, and the variance of the ***Fast Unfolding*** algorithm is reduced by 38.14%. ***Figure 5.10*** compares the estimated total scheduled distances. The data show that the variance of the ***CDoMO*** algorithm is 55.17% compared with the ***K-means*** algorithm and 27.54% compared to the ***Fast Unfolding*** algorithm. When scheduling and partitioning based on multi-objective optimization algorithm, the estimated scheduling distance can be shortened, and the estimated scheduling distance is positively related to the actual scheduling distance, so the actual scheduling distance will also be shortened; in addition, the scheduling work of each area will also be made. Relatively balanced. ***Figure 5.11*** is a comparative display of experimental results.

5.2. Chicago Public Bicycle

5.2.1. Data set introduction

First of all, the data set cited in this paper is Chicago public bicycle data [31]. The starting site is 2015-1-1, and the deadline is 2015-6-30. There are two quarters and six months of data, a total of 759,789 data records. This paper did some data pre-processing: Trips that did not include a start or end date were removed from the original table. Then, in order to ensure that the information of the data set more abundant, this paper decided to use the data set, distance information of each pair of source address and destination address. Finally, we utilize certain data pre-processing methods to remove weather and other data because it can be considered as an ideal condition. The dataset contains 12 fields, and the 10 fields related to this experiment are presented in the following ***Table 5.2***.

5.2.2. Experimental result

Results of the ***Fast Unfolding*** community discovery algorithm can be mapped to Chicago map as the picture shows [32]. In contrast, the division results are more uniform and reasonable, but there are too many abnormal sites in the middle. These abnormal sites are a long way from where they should have existed. The number of rental sites in a divided area is not particularly uniform, as showed in ***Figure 5.12***.

Based on ***CDoMO***, in the optimization process, the division results are dynamically adjusted in time. Therefore, in this case, the division result is more reasonable, and the problem of scheduling balance, this algorithm obviously adds more consideration. It not only addresses the problem of abnormal sites, but also solves the problem of differences in the number of regional sites at the same time. As showed in ***Figure 5.13***. In order to make the experiment consistent, we set the value of ***k*** in the ***K-means*** algorithm as **5**, and then we clustered the uniform data set. The results of the clustering are presented in the figure. This paper believes that the results obtained by the clustering algorithm are very poor, because the red sites represent a particularly large number of sites. The yellow site represents a particularly small number of rental sites. This shows that the various types of leases, the number of differences is too large, in addition, this algorithm also led to the border is not clear, and there is some inevitable overlap. In the actual scheduling work, this situation is not allowed, as showed in ***Figure 5.14***.

5.2.3. Algorithm performance comparison results

This paper will describe the quantified experimental results of the three methods, it compares the differences between them. It is easy to see that the algorithm proposed in this paper is optimal, compared to the other two algorithms. ***Figure 5.15*** compares the difference between the numbers of sites. Compared to the ***K-means*** clustering algorithm and the ***Fast Unfolding*** community discovery algorithm, the variance of the ***CDoMO*** algorithm is reduced by 66.98% and 22.57%. ***Figure 5.16*** in this paper compares the number of scheduled bicycles, it finds that the ***CDoMO*** algorithm set out in the present paper is an optimal algorithm. Similarly, opposed to the ***K-means*** clustering algorithm and the ***Fast Unfolding*** community finding algorithm, the variance is reduced by 83.77% and 48.72% shown in ***Figure 5.17***. ***Figure 5.18*** in this paper compares the estimated total distance of scheduling with the other two algorithms, and the conclusion shows that the distance is decreased by 50.82% and 22.08%.

Then we can conclude that the ***CDoMO*** algorithm proposed in this paper: It effectively reduces the number of sites; it effectively reduced the variance in the number of bicycles dispatched; it effectively reduced the estimated total distance for scheduling.

6. Conclusion

In order to solve the problem of regional division of public bicycles, this paper proposes ***CDoMO***. The algorithm fully considers the special law of public bicycle lease/return, and in order to balance the scheduling workload between areas, the regional scheduling workload index is proposed. This problem is identified as a multi-objective optimization problem with two objective functions: minimize the variance between the estimated dispatch distances between each area; minimize the variance between the numbers of sites in each area. The regional scheduling workload can adjust the results of the community discovery algorithm in real time and dynamically. In the end, the results obtained can meet the special rules of public bicycle lease/return, and balance the workload between the areas. The experimental results show that the ***CDoMO*** can effectively shorten the scheduling distance of public bicycle system, effectively improve the scheduling efficiency, and make the workload of each scheduling area relatively balanced. The next step is to have a more appropriate solution if you limit the travel time and mileage of the scheduling vehicle.

Data Availability Statement

The [New York public bicycle data] data used to support the findings of this study have been deposited in the [Citi

Bike Trip Histories] repository ([https://www.citibikenyc.com]).

The [Chicago public bicycle data] data used to support the findings of this study have been deposited in the [Divvy Data] repository ([https://www.divvybikes.com]).

Acknowledgment

This study was supported by the National Natural Science Foundation of China (No. 61602141). The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper. The authors declare that there are no conflicts of interest regarding the publication of this paper.

Reference

1. Tulabandhula, T., & Bodas, T. P. (2018). Method and system for dispatching of vehicles in a public transportation network.
2. Pan, G. Q., Jun-Yi, H. U., & Min, H. (2015). Gis-based logistics distribution area division and its vrp algorithm. *Journal of Dalian Maritime University*, 41(1), 83-90.
3. Forma, I. A., Raviv, T., & Tzur, M. (2015). A 3-step math heuristic for the static repositioning problem in bicycle-sharing systems. *Transportation Research Part B*, 71, 230-247.
4. Schuijbroek, J., Hampshire, R. C., & Hoeve, W. J. V. (2016). Inventory rebalancing and vehicle routing in bicycle sharing systems. *European Journal of Operational Research*, 257(3), 992-1004.
5. Kloimüllner, C., Papazek, P., Hu, B., & Raidl, G. R. (2015). A Cluster-First Route-Second Approach for Balancing Bicycle Sharing Systems. International Conference on Computer Aided Systems Theory (pp.439-446). *Springer International Publishing*.
6. Dong, H., Shi, C., Chen, N., & Liu, D. (2013). Clustering division of public bicycle scheduling regional based on association rules. *Bulletin of Science & Technology*.
7. J Zhang, Y Liang, WJ Wei. (2017). Regional division of public bicycle stations based on improved k-means algorithm. *Information Communication* (4), 42-44.
8. C Wang. (2016). Research on Partition Scheduling Model of Public Bicycle System. (*Doctoral dissertation*, Hangzhou Dianzi University).
9. WC Yu. (2015). Study on intelligent scheduling optimization algorithm of urban public bicycles system. (*Doctoral dissertation*, Shanghai Jiao Tong University).
10. Liu, J., Sun, L., Chen, W., & Xiong, H. (2016). Rebalancing Bicycle Sharing Systems: A Multi-Source Data Smart Optimization. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp.1005-1014). ACM.
11. Zaltz, A. M., O'Brien, O., Strano, E., & Viana, M. (2013). The structure of spatial networks and communities in bicycle sharing systems. *Plos One*, 8(9), e74685.
12. Tolić, I. H., Martinović, G., & Crnjacmilić, D. (2018). Optimization methods in modern transportation systems.
13. Doreswamy, & Ghoneim, O. A. (2018). Traffic jams detection and congestion avoidance in smart city using parallel k-means clustering algorithm.
14. Yan, Q., Ding, Y., Zhang, J. J., Xia, Y., & Zheng, C. H. (2018). A discriminated similarity matrix construction based on sparse subspace clustering algorithm for hyper spectral imagery. *Cognitive Systems Research*.
15. JIANG Yanping, KONG Decai, YUAN Duoning. Two-sided stable matching decision-making method with ordinal interval preference [J]. *Systems Engineering - Theory & Practice*, 2017, 37(8): 2152-2161.
16. Shpak, M., Ni, Y., Lu, J., & Müller, P. (2017). Variance in estimated pairwise genetic distance under high versus low coverage sequencing: the contribution of linkage disequilibrium. *Theoretical Population Biology*, 117.
17. Schamel, J., & Job, H. (2017). National parks and demographic change – modelling the effects of ageing hikers on mountain landscape intra-area accessibility. *Landscape & Urban Planning*, 163, 32-43.
18. X Feng, FX Zhu, and SC Liu. Label Propagation Community Discovery Algorithm Based on DeepWalk Model [J]. *Computer Engineering*, 2018, 44(3): 220-225,232.
19. Shivach, P., Nautiyal, L., & Ram, M. (2018). Applying Multi-Objective Optimization Algorithms to Mechanical Engineering. In M. Ram, & J. Davim (Eds.), *Soft Computing Techniques and Applications in Mechanical Engineering* (pp. 287-301). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-3035-0.ch014
20. Mori, T., & Saito, S. (2016). Molecular mechanism behind the fast folding/unfolding transitions of villain headpiece subdomain: hierarchy and heterogeneity. *Journal of Physical Chemistry B*.
21. Zhi-yuan Sun, Yue Li, Wen-cong Qu, and Tanveer Muhammad, "Unified Framework for Optimal Routing Choice under Guidance Information," *Complexity*, vol. 2018, Article ID 9764204, 11 pages, 2018. https://doi.org/10.1155/2018/9764204.
22. GAO, X., Tian, Y., Sun, B., GAO, X., Tian, Y., & Sun, B. (2018). Multi-objective optimization design of bidirectional flow passage components using rsm and nsga ii: a case study of inlet/outlet diffusion segment in pumped storage power station. *Renewable Energy*, 115.
23. Wu, T. (2014). Research of logistics distribution path planning based on improved nsga-ii. *Journal of Information & Computational Science*, 11(7), 2143-2153.
24. Basch, C. H., Ethan, D., Zybert, P., Afzaal, S., Spillane, M., & Basch, C. E. (2015). Public bicycle sharing in New York City: helmet use behaviour patterns at 25 city bicycles stations. *Journal of Community Health*, 40(3), 530.

25. Shixiong Jiang, Wei Guan, Zhengbing He, and Liu Yang, "Exploring the Intermodal Relationship between Taxi and Subway in Beijing, China," *Journal of Advanced Transportation*, vol. 2018, Article ID 3981845, 14 pages, 2018. https://doi.org/10.1155/2018/3981845.
26. Chao Guo, Cheng Gong, Haitao Xu, and Zhiyong Yao, "Queuing Game Theory Based Optimal Routing Scheme for Heterogeneous Users over Space Information Networks," *Mathematical Problems in Engineering*, vol. 2017, Article ID 8064989, 9 pages, 2017. https://doi.org/10.1155/2017/8064989.
27. Xiaowei Hu, Shi An, and Jian Wang, "Taxi Driver's Operation Behavior and Passengers' Demand Analysis Based on GPS Data," *Journal of Advanced Transportation*, vol. 2018, Article ID 6197549, 11 pages, 2018. https://doi.org/10.1155/2018/6197549.
28. Zhang, W., Ge, Y., Yuan, B., & Xu, H. (2011). Busy Line Analysis with Improved Association Rules Mining Algorithm for Hangzhou Public Free-Bicycle System. *Intelligent Systems* (Vol.1, pp.139-142). IEEE.
29. Manju, V. N., & Fred, A. L. (2018). Ac coefficient and k-means cuckoo optimisation algorithm-based segmentation and compression of compound images. *Iet Image Processing*, 12(2), 218-225.
30. Lin, W., & Song, Y. (2017). A comparative study on x-ray diffraction mineral quantitative analysis of two methods in sediments. *Journal of Earth Environment.*
31. Zhen, Z., Xiangqian, L. I., Dept, S., & Univ, S. N. (2016). Development and enlightenment of public bicycle sport in Chicago. *Journal of Sports Adult Education.*
32. Ye, P., Chu, C., & Xu, L. (2015). IC Card-Based Data Mining Characteristics of Urban Public Bicycles. *International Conference on Transportation Engineering* (pp.2124-2132).

Figure 1

the main process of NSGA2 algorithm

the main process of NSGA2 algorithm

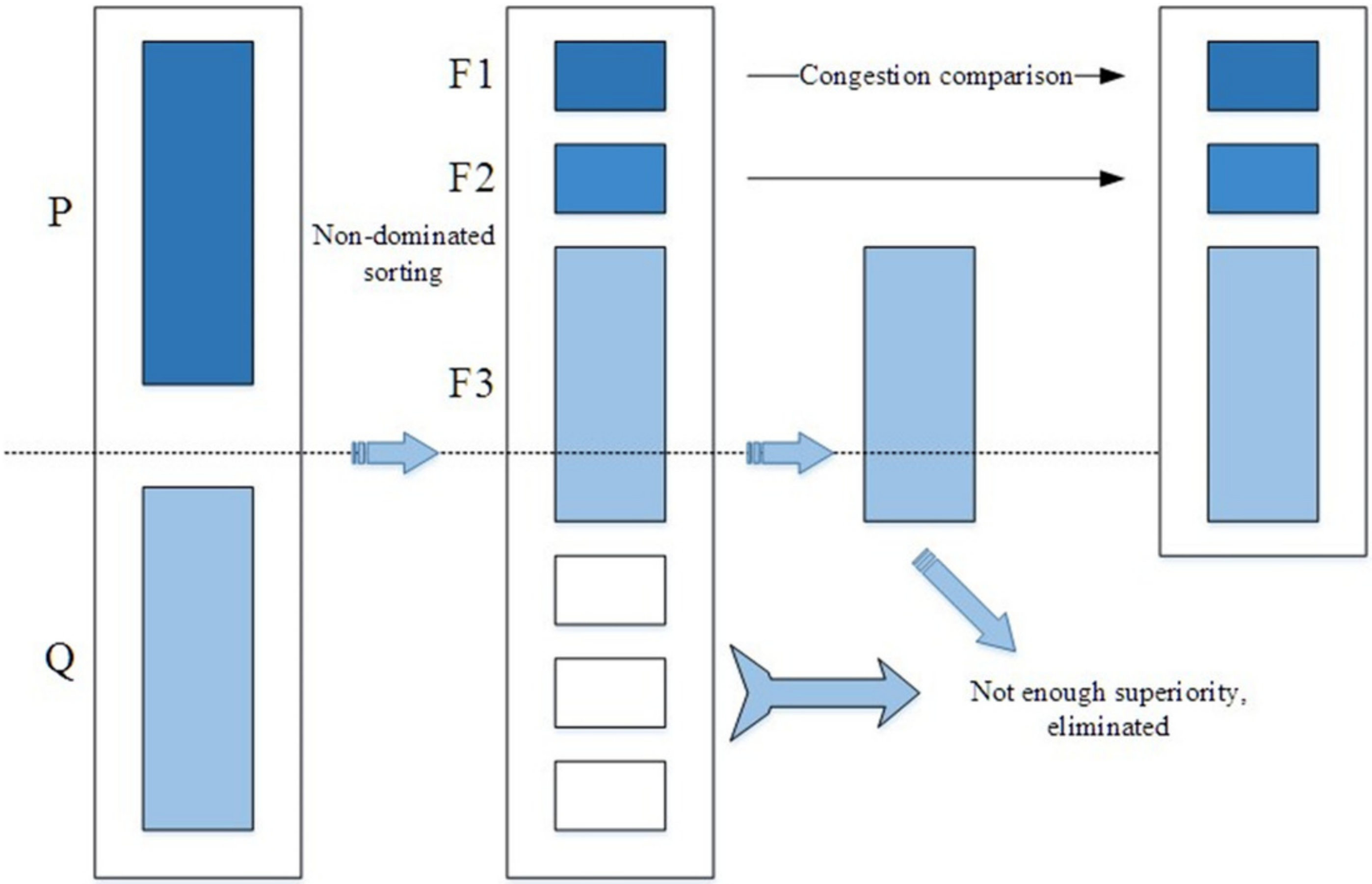


Figure 2

specific_flow_of_NSGA2_algorithm

specific_flow_of_NSGA2_algorithm

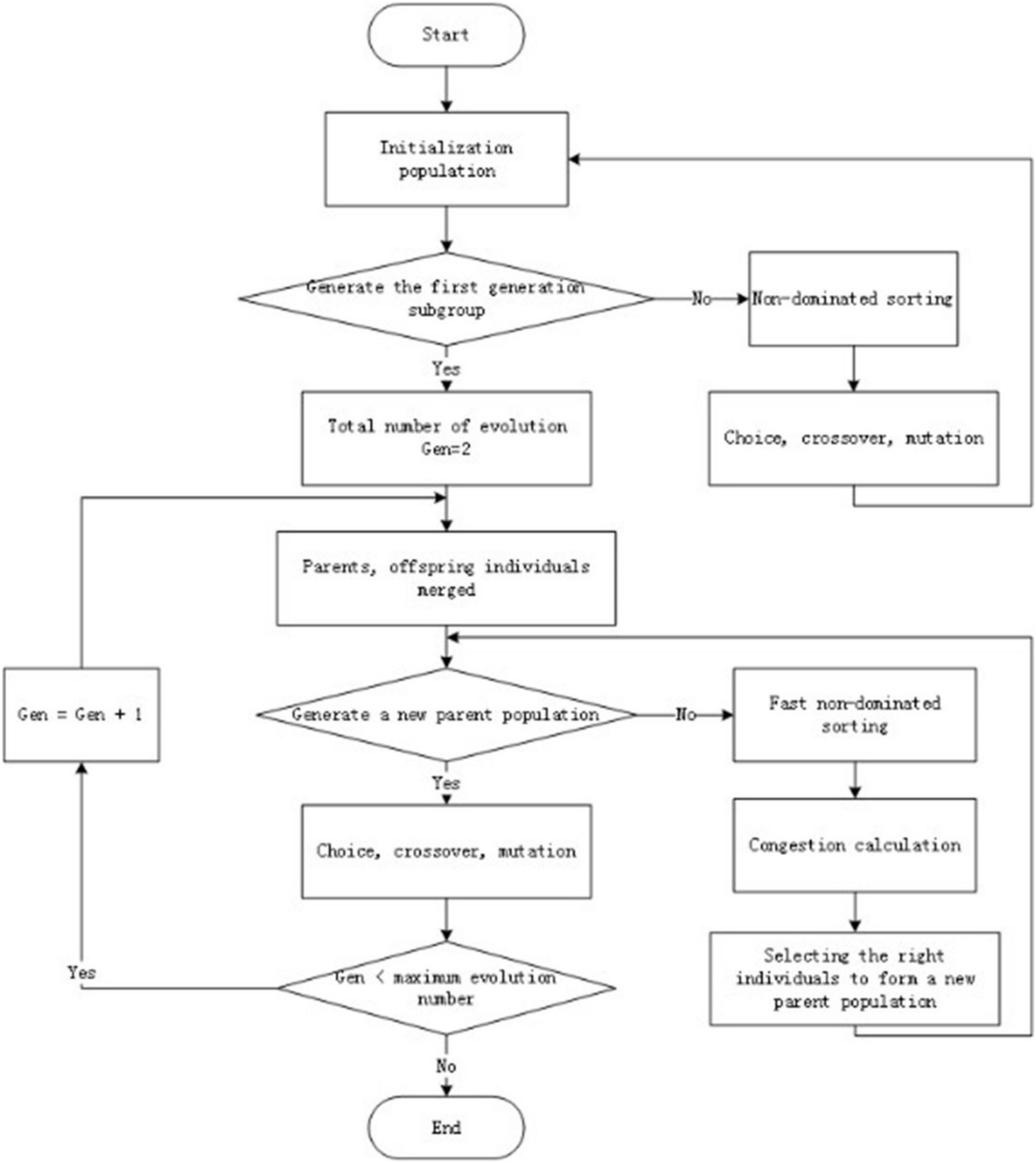

Start
Initialization population
Generate the first generation subgroup
No
Non-dominated sorting
Choice, crossover, mutation
Yes
Total number of evolution Gen=2
Parents, offspring individuals merged
Generate a new parent population
No
Fast non-dominated sorting
Congestion calculation
Selecting the right individuals to form a new parent population
Yes
Choice, crossover, mutation
Gen = Gen + 1
Gen < maximum evolution number
Yes
No
End

Figure 3

Fast_unfolding_community_discovery_algorithms

Fast_unfolding_community_discovery_algorithms

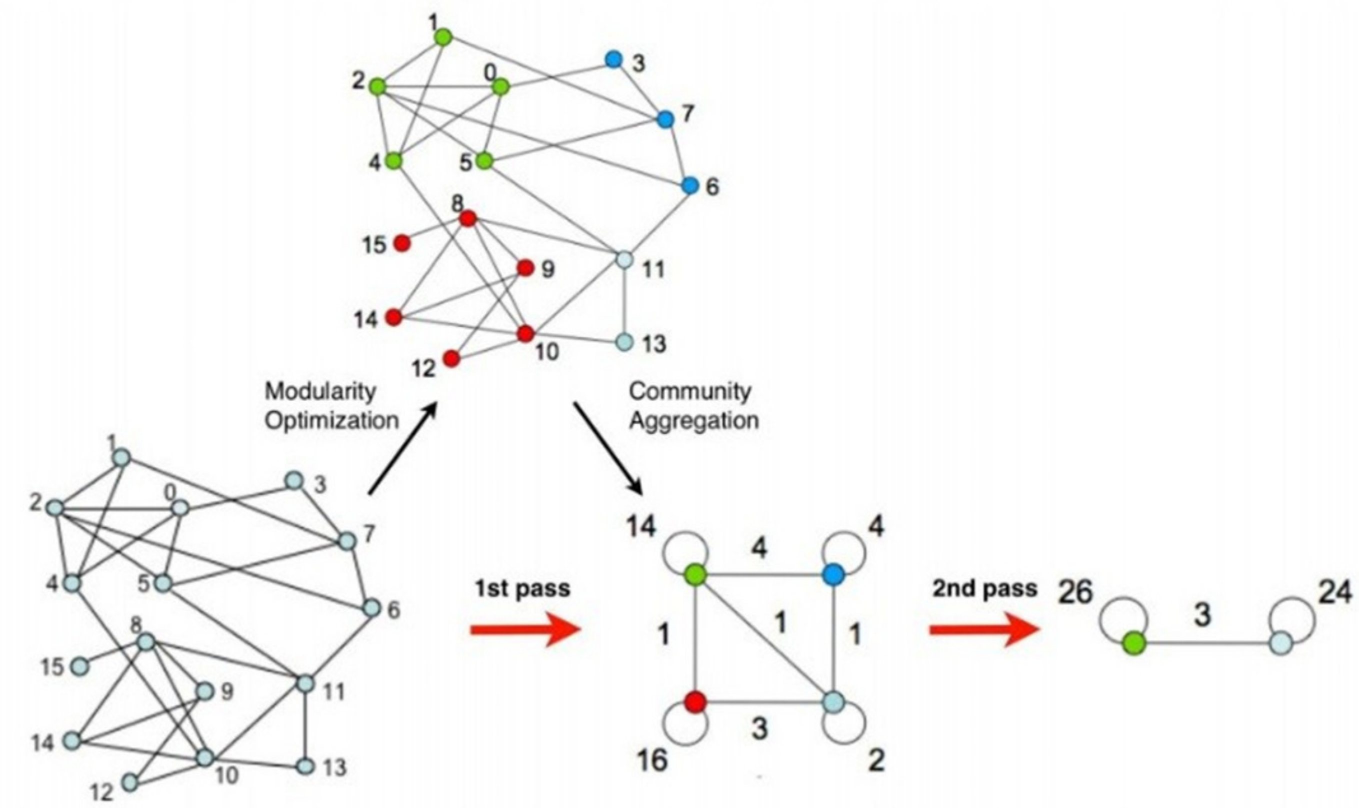


Figure 4

Internal_community_Structure_of_the_abstract_network

Internal_community_Structure_of_the_abstract_network

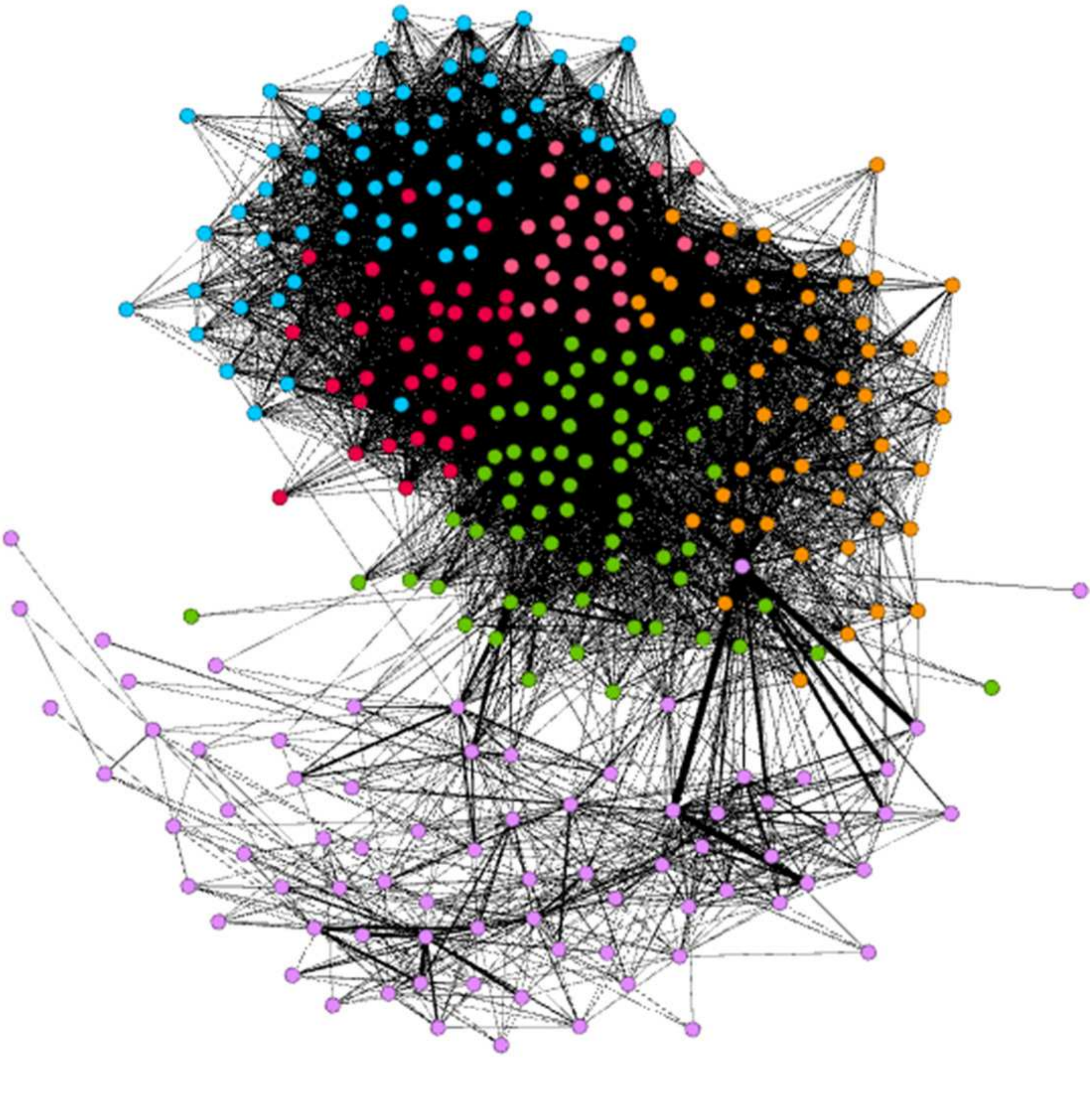


Figure 5

Spatial distribution

Spatial distribution

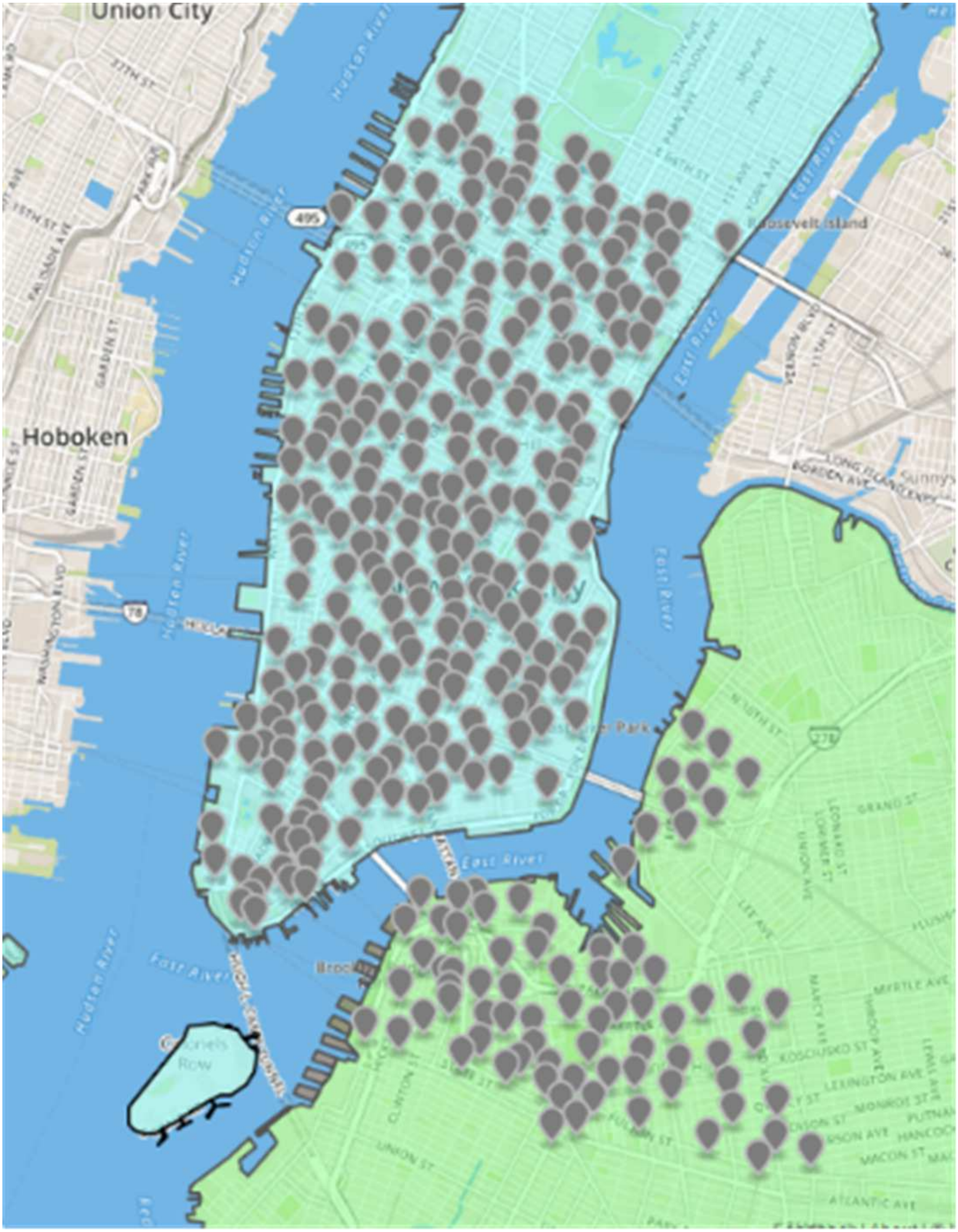

Union City
Hoboken
Hudson River
East River
Roosevelt Island
495
78
278

Figure 6

Corresponding abstract network

Corresponding abstract network

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Figure 7

Results of the fast unfolding community discovery algorithm

Results of the fast unfolding community discovery algorithm

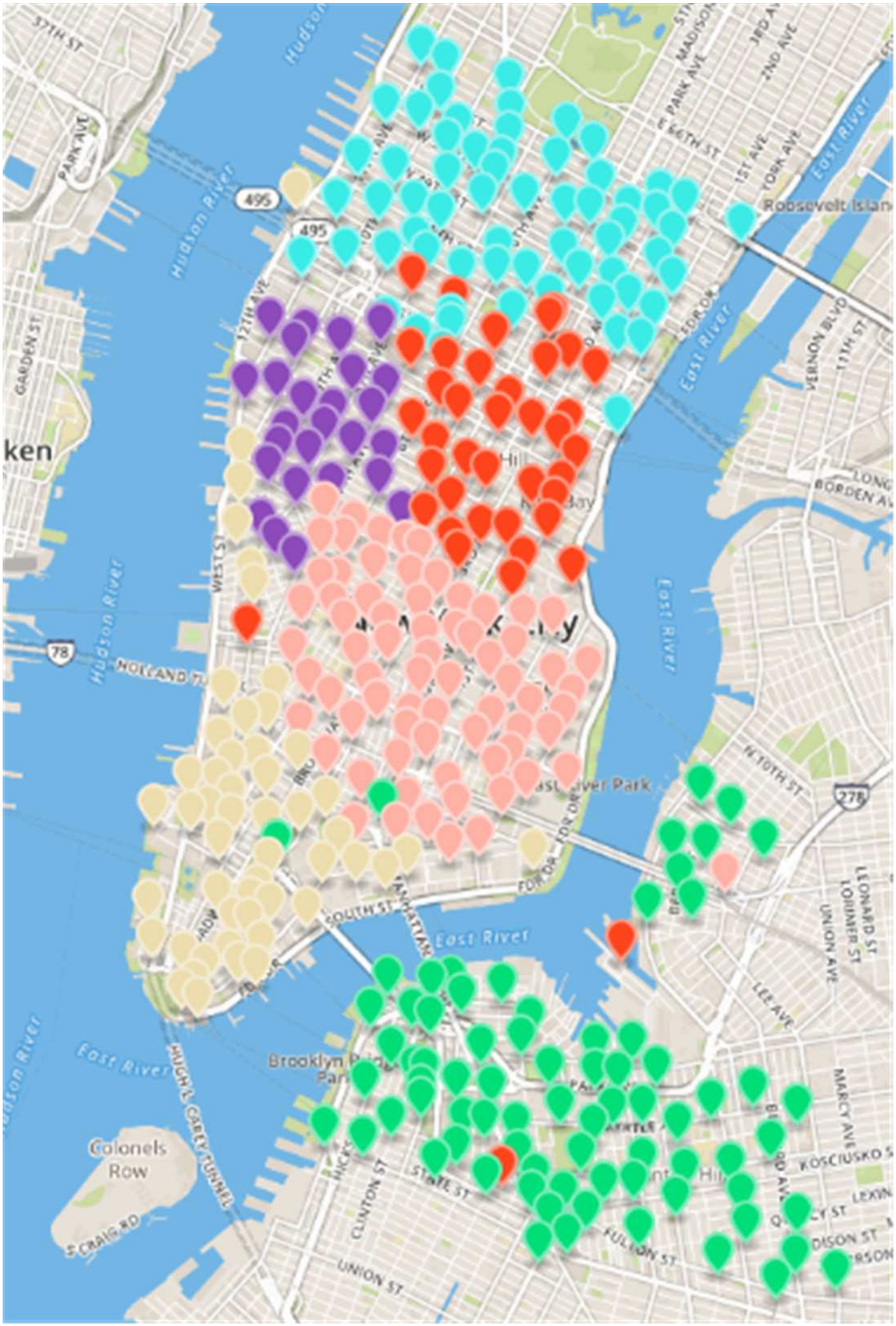

Hudson River
East River
Roosevelt Island
Holland Tunnel
Colonels Row
Brooklyn Bridge Park
East River Park
South St
State St
Fulton St
Union St
Clinton St
Lee Ave
Union Ave
Marcy Ave
Kosciusko St

Figure 8

Results of the multi objective optimization algorithm

Results of the multi objective optimization algorithm

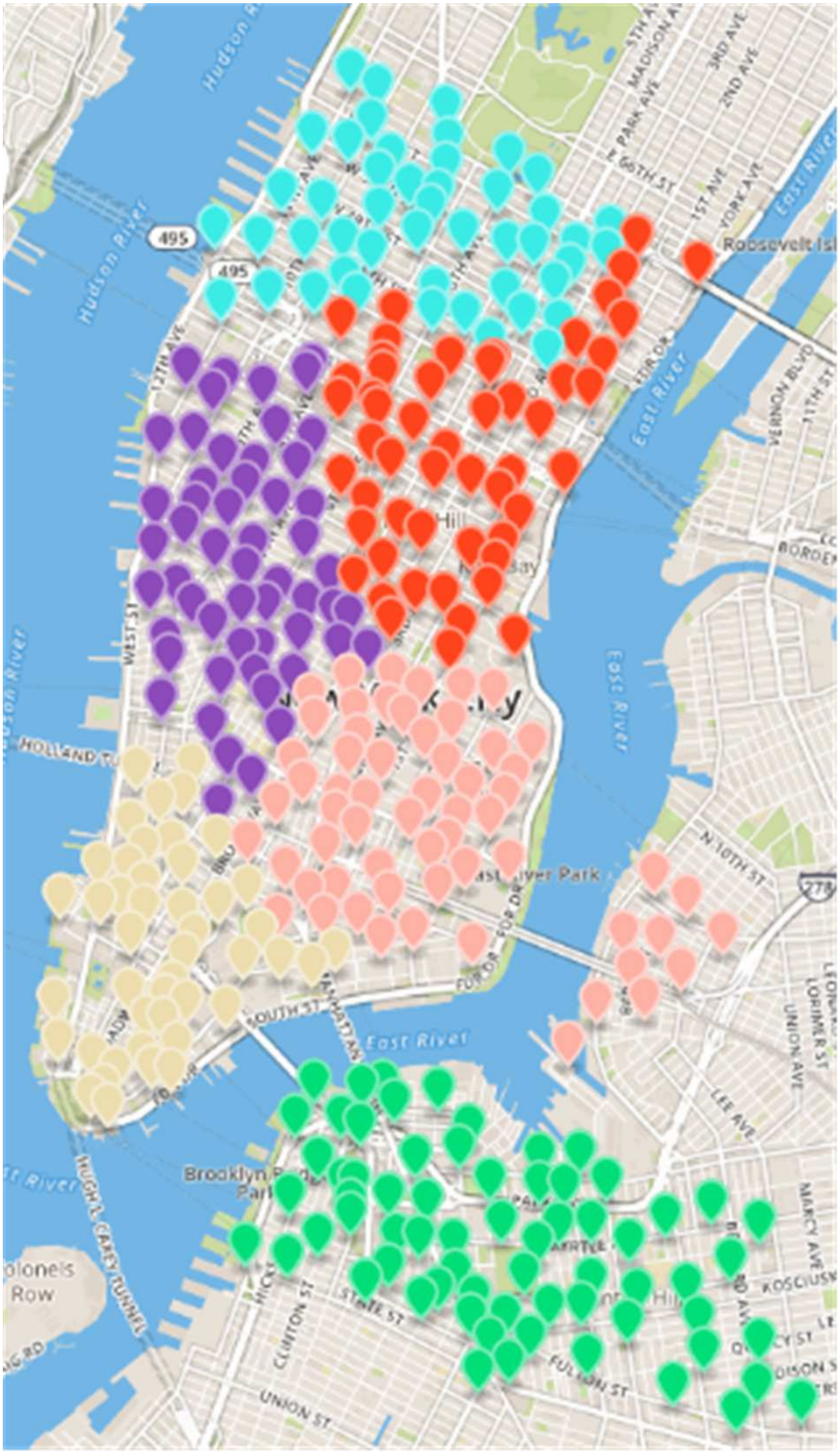

Hudson River
495
East River
PARK AVE
MADISON AVE
3RD AVE
2ND AVE
1ST AVE
YORK AVE
VERNON BLVD
12TH AVE
WEST ST
N 10TH ST
278
SOUTH ST
UNION AVE
LORIMER ST
LEE AVE
MARCY AVE
CLINTON ST
STATE ST
FULTON ST
UNION ST

Figure 9

Results of the clustering algorithm

Results of the clustering algorithm

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

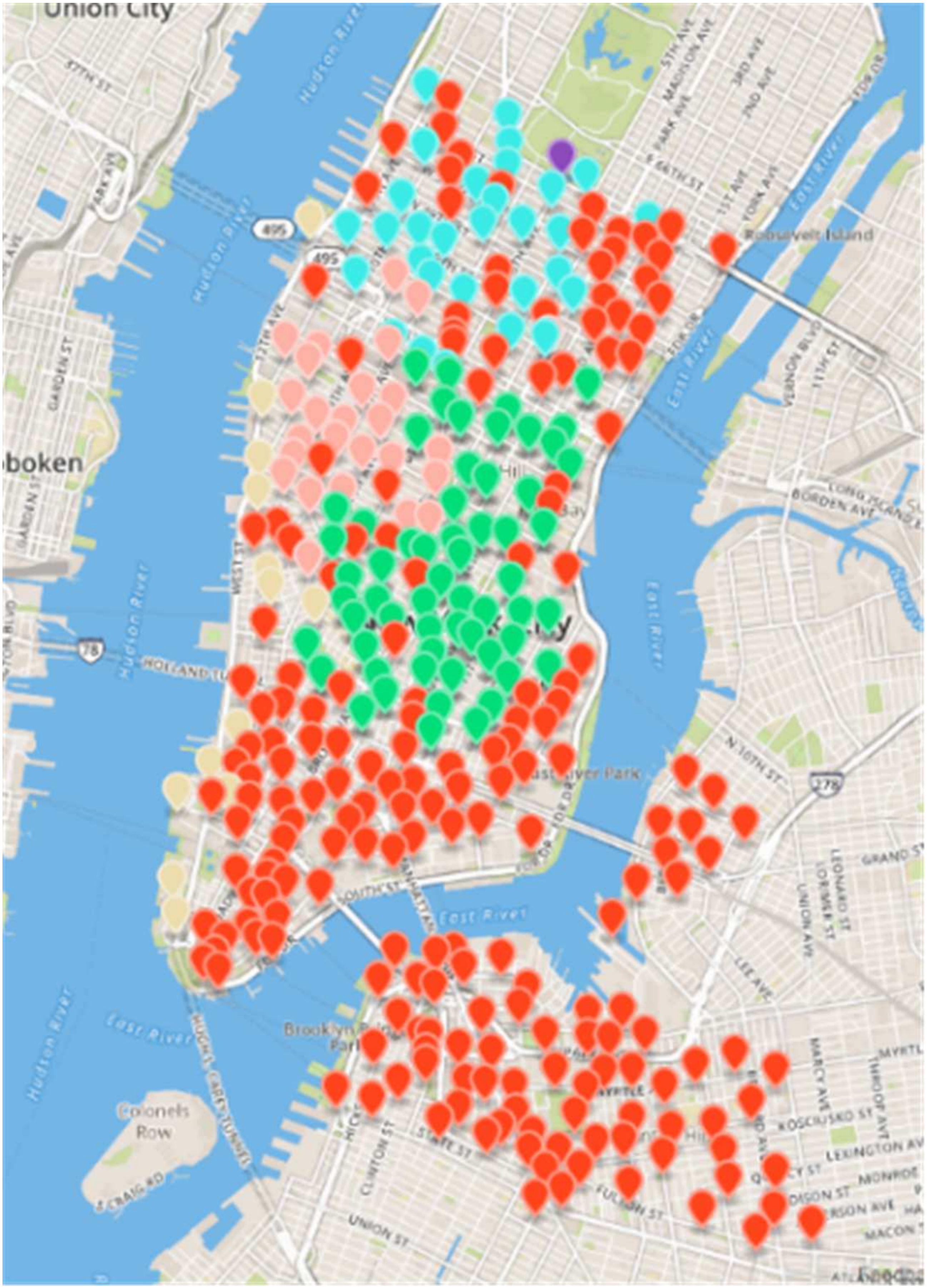

Union City
Hudson River
East River
Roosevelt Island
495
78
278
WEST ST
HOLLAND TU
SOUTH ST
East River Park
Colonels Row
FULTON ST
STATE ST
CLINTON ST
UNION ST
MARCY AVE
LEXINGTON AV
MONROE
GRAND ST
LEE AVE
UNION AVE
KOSCIUSKO ST

Figure 10

Variance_of_the_number_of_rental_sites

Variance_of_the_number_of_rental_sites

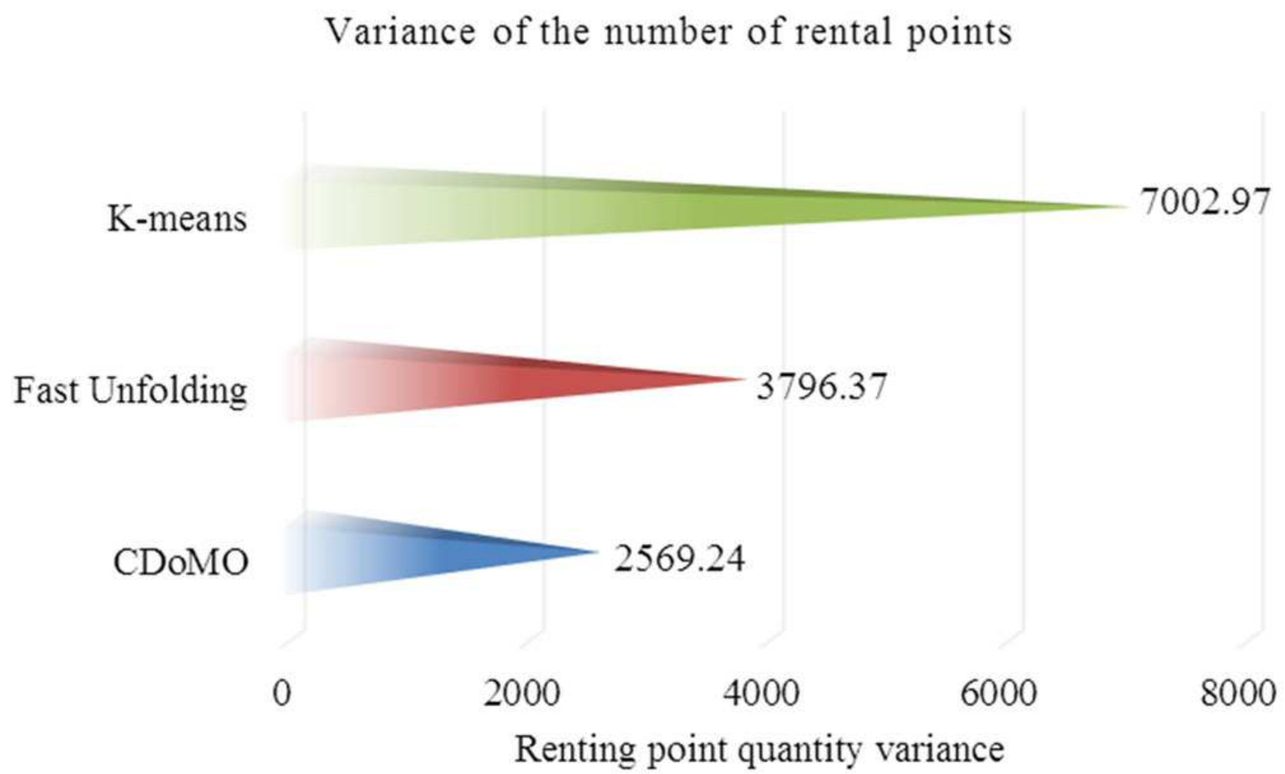
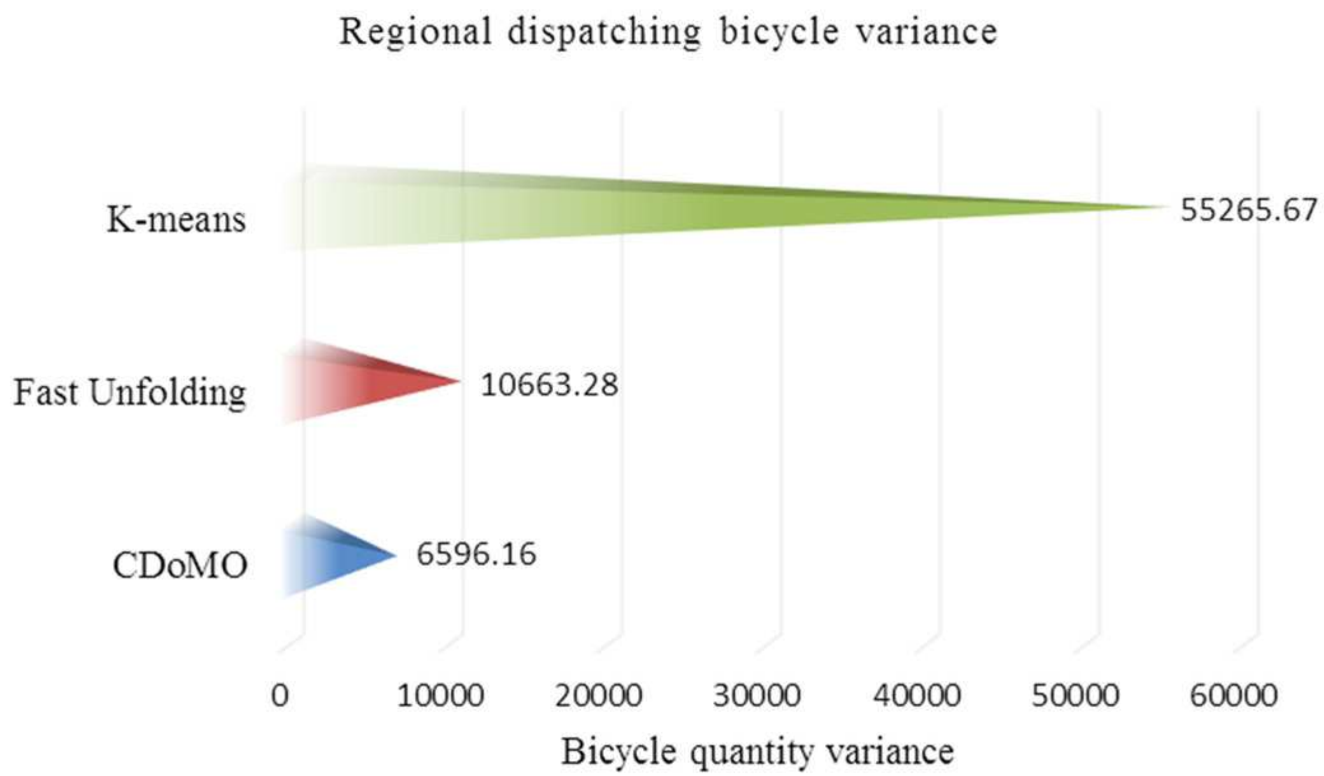


Figure 11

Regional_dispatching_bicycle_variance

Regional_dispatching_bicycle_variance

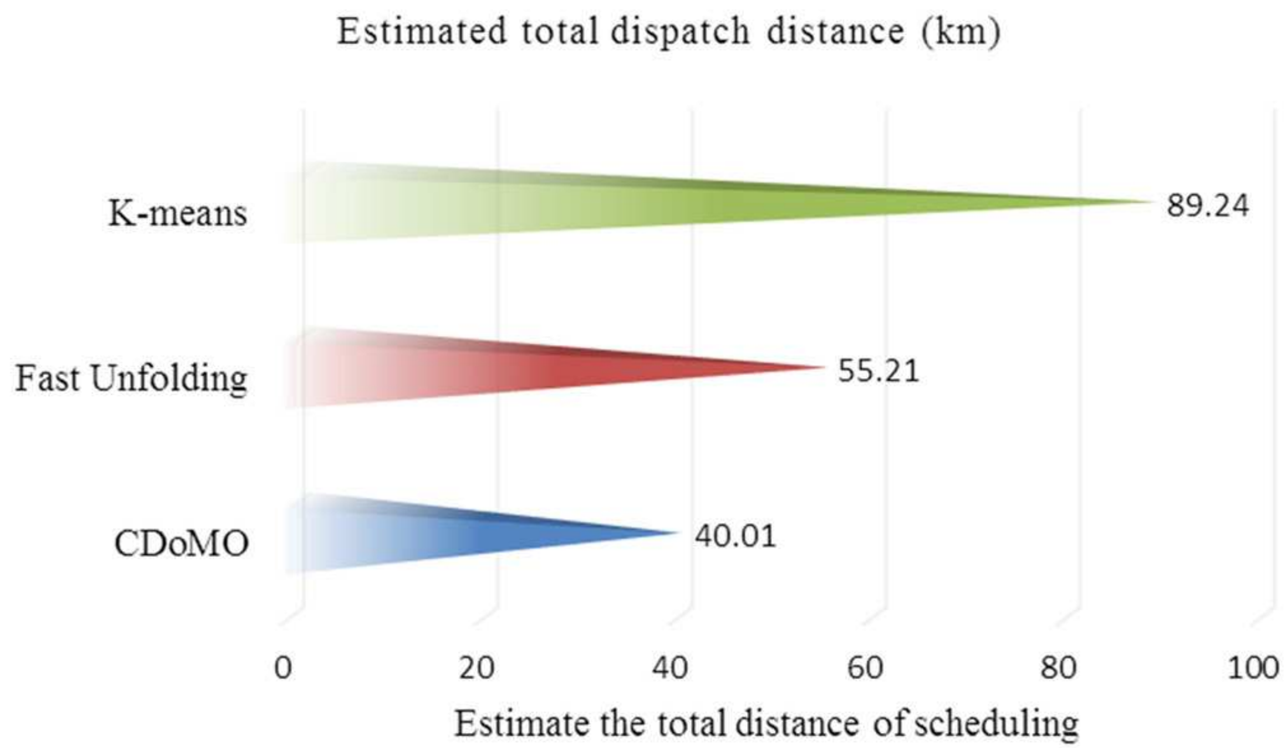


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Figure 12

Estimated total dispatch distance

Estimated total dispatch distance



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Figure 13

Comparison_of_regional_division_experiment_results

Comparison_of_regional_division_experiment_results

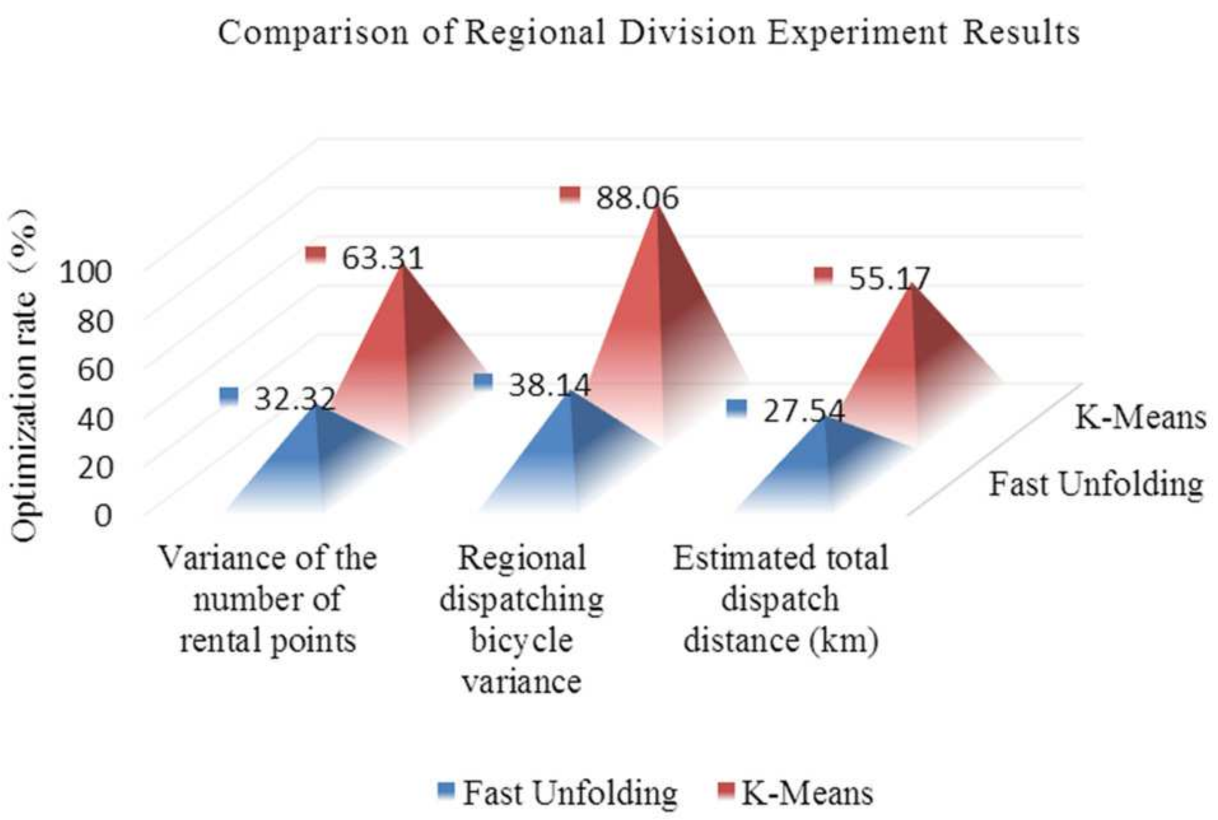


Figure 14

Results of the fast unfolding community discovery algorithm

Results of the fast unfolding community discovery algorithm

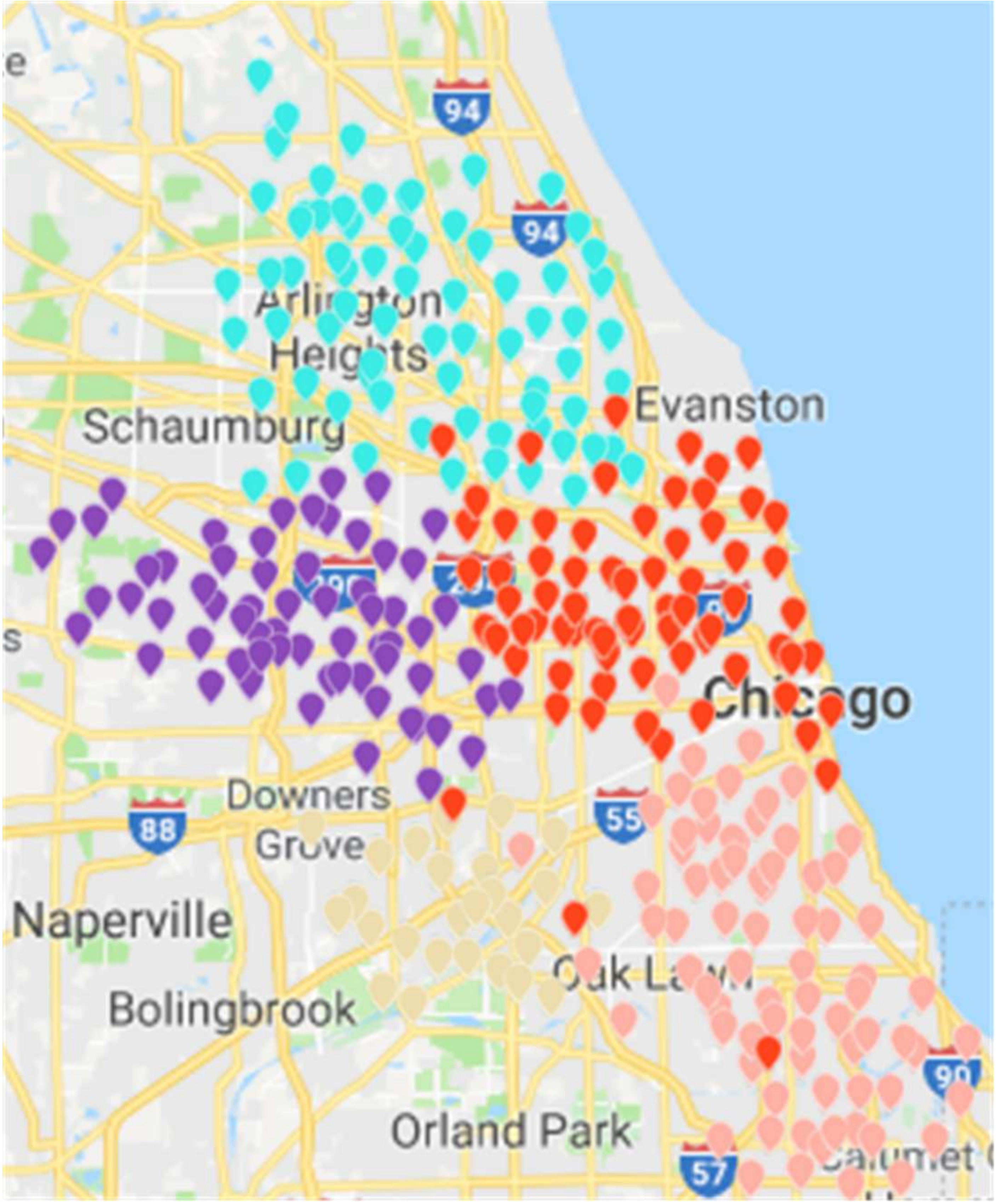
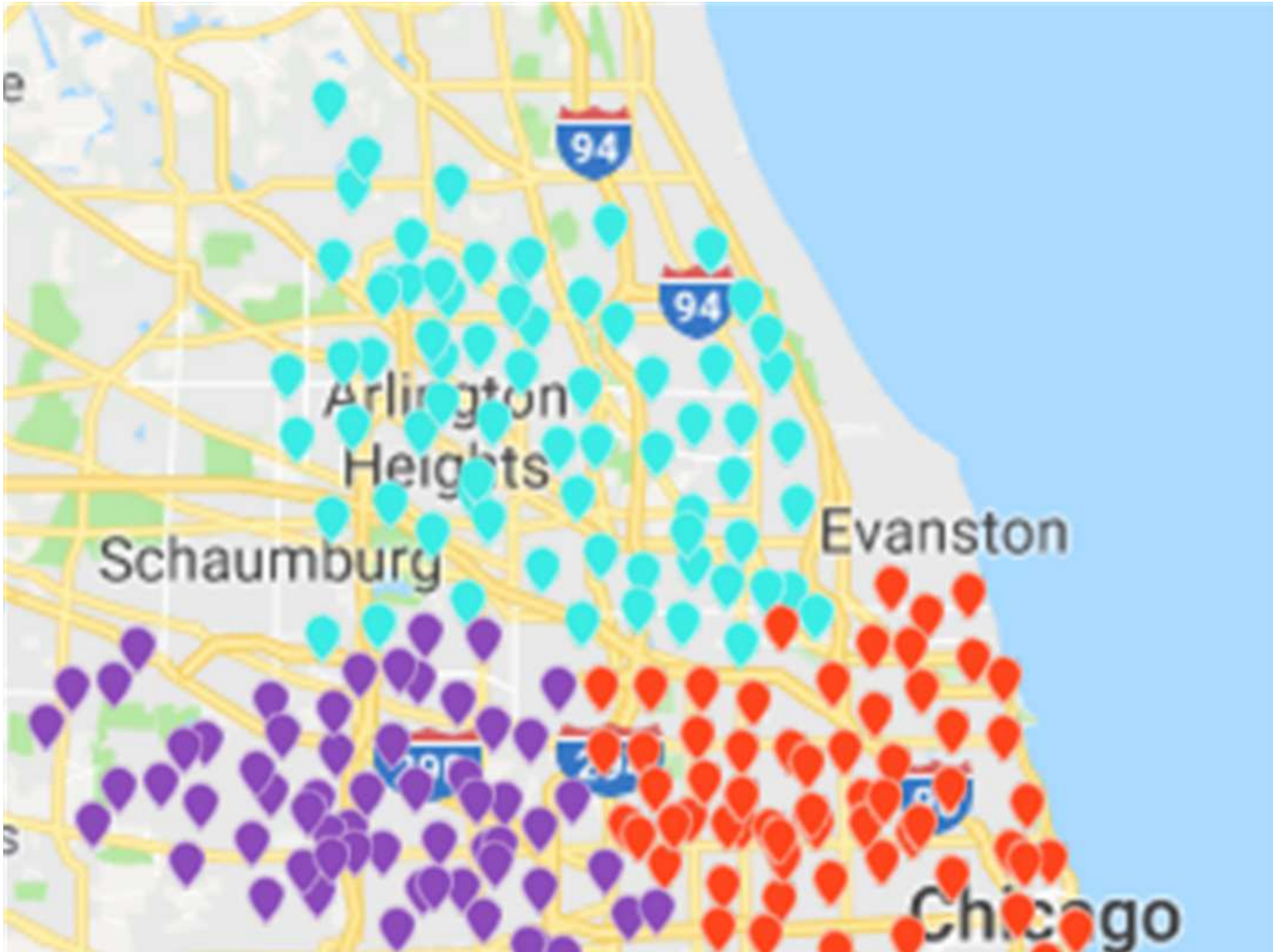

Arlington Heights
Schaumburg
Evanston
Chicago
Downers Grove
Naperville
Bolingbrook
Orland Park
94
88
55
57
90

Figure 15

Results of the multi-objective optimization algorithm

Results of the multi-objective optimization algorithm



Arlington Heights
Schaumburg
Evanston
Chicago
Downers Grove
Naperville
Bolingbrook
Oak Lawn
Orland Park
94
94
290
88
55
90
57

Figure 16

Results of the clustering algorithm

Results of the clustering algorithm

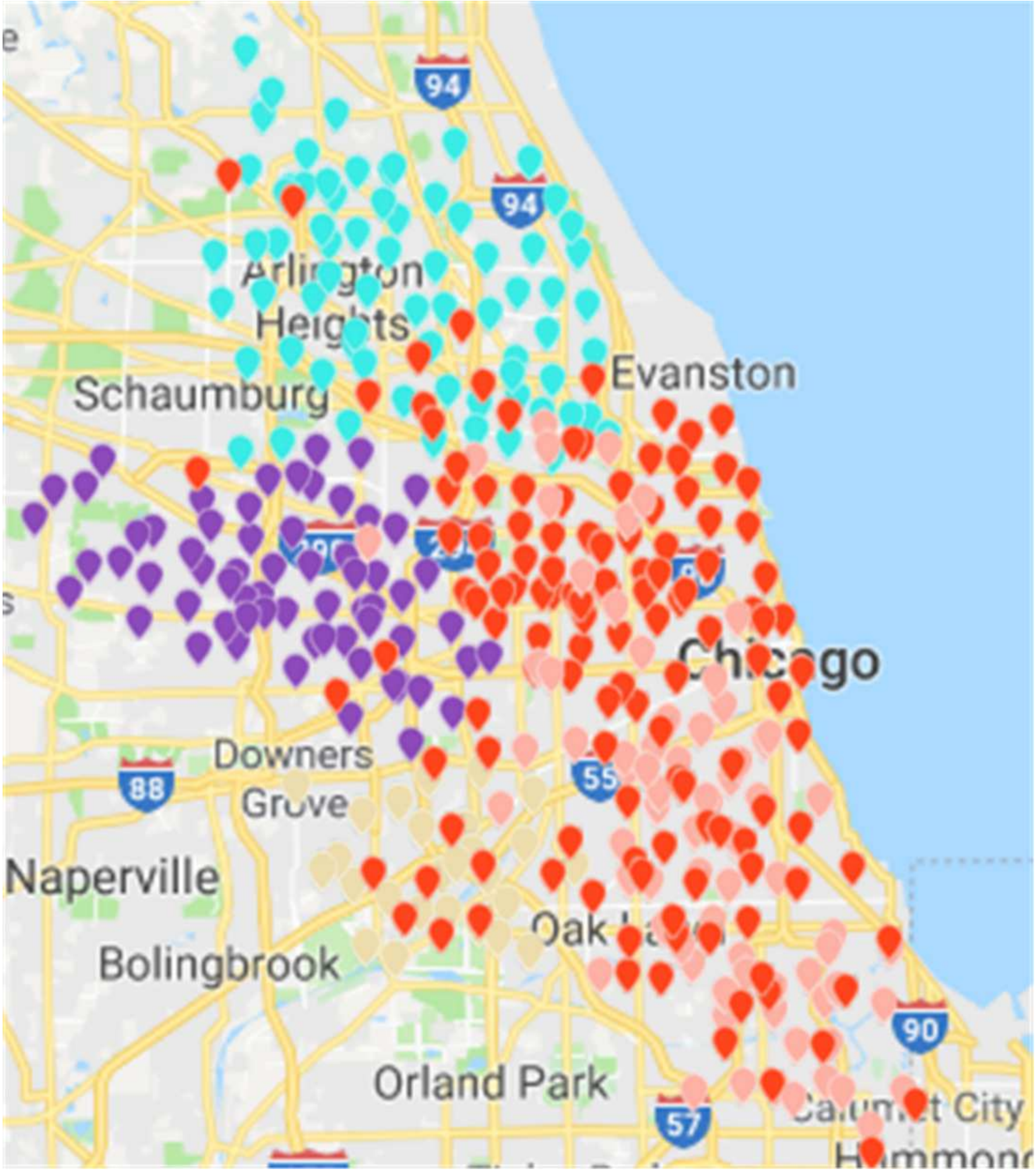

Arlington Heights
Schaumburg
Evanston
Chicago
Downers Grove
Naperville
Oak Lawn
Bolingbrook
Orland Park
Calumet City
94
290
55
88
57
90

Figure 17

Variance of the number of rental sites

Variance of the number of rental sites

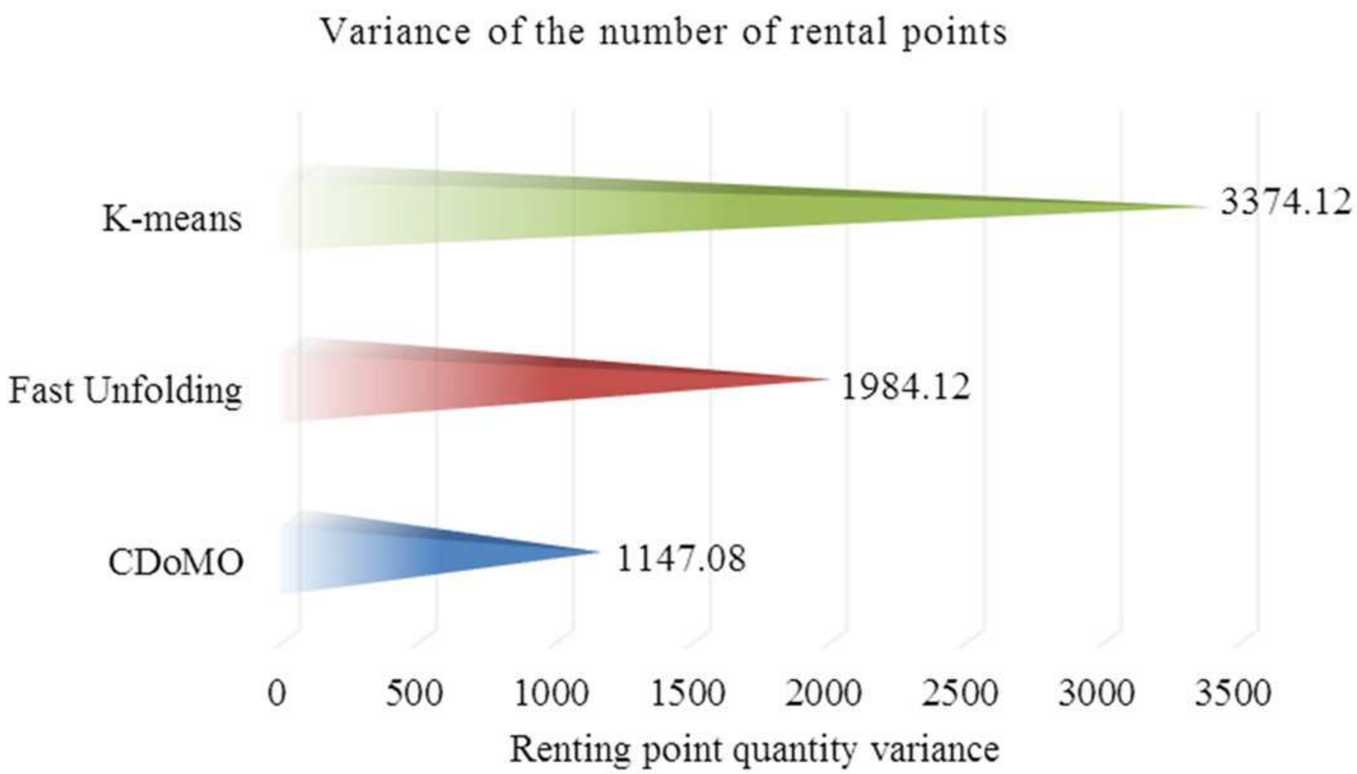
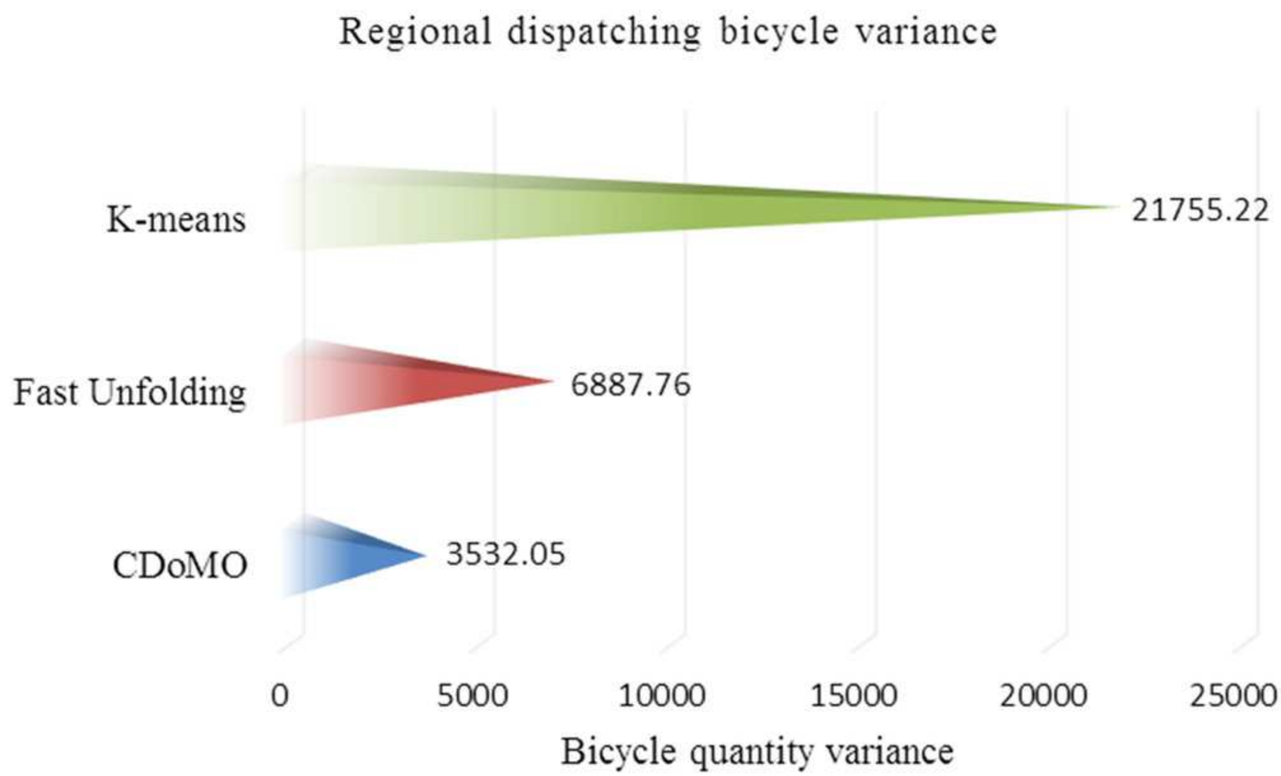


Figure 18

Regional dispatching bicycle variance

Regional dispatching bicycle variance



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Figure 19

Estimated total dispatch distance

Estimated total dispatch distance

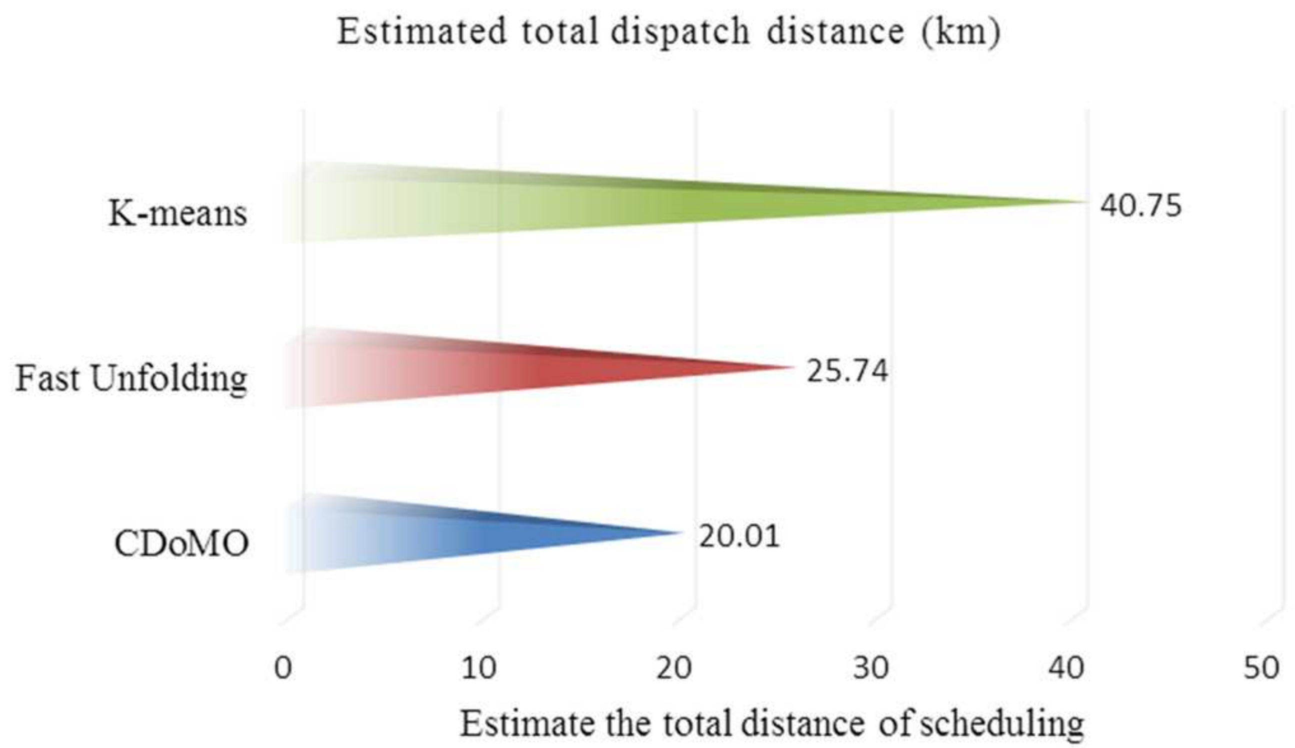


Figure 20

Comparison of regional division experiment results

Comparison of regional division experiment results

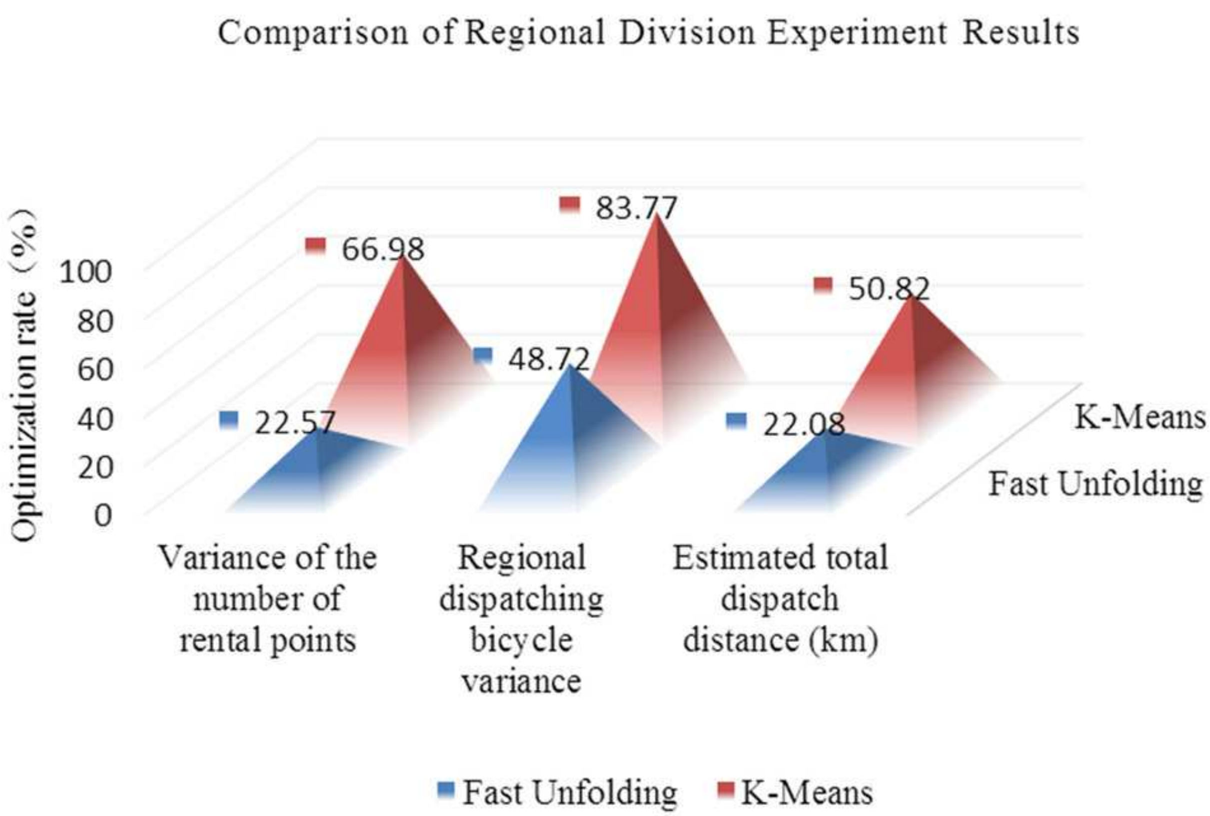


Table 1(on next page)

parameters and variables of the model

Based on the problem description and model assumptions, the parameters and variables of the model in ***Table 3.1*** are defined.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Table 3.1. Model-related parameters/variables description table

Parameters/variables	Parameter/variable meaning
n	The number of areas
i , j	Area number
D_i	The estimated scheduling distance of the area ***i***
D_j	The estimated scheduling distance of the area ***j***
$\overline{D}$	Regional estimated dispatch distance average
N_i	The number of sites in area ***i***
N_j	The number of sites in area ***j***
$\overline{N}$	Average number of sites within the area
S	Collection of sites
S_i	Site division set for area ***i***
S_j	Site division set for area ***j***
P	parking lot site collection

Table 2(on next page)

exact values

In the measurement of the similarity of weather, the weather is split into ***5*** levels and assigned corresponding values. The exact values are shown in ***Table 3.2***.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Table 3.2. Quantification of Weather Conditions

weather	value
heavy snow, heavy rain	1
snow, light snow, moderate rain, light rain	0.75
foggy	0.5
sunny and cloudy	0.25

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Table 3 (on next page)

detailed algorithm calculation

Table 4.1 displays the detailed algorithm calculation.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Table 4.1. Detailed explanation of algorithm flow

Algorithm:	Community discovery algorithm based on multi-objective optimization
Input:	Site similarity matrix *X*, population number *popsize*, maximum number of iterations *MaxGen.*
Output:	Optimal regional division results $\rho_1{}^*$, and workload index parameters $\rho_2{}^*$.
1.	Initialize the historical optimal solution f^* and its workload index parameter $\rho_1{}^*$,$\rho_2{}^*$.
2.	Perform a pass phrase of the Fast Unfolding community discovery algorithm, and obtain the results of the preliminary zoning division as *R*.
3.	Calculate the estimated distance D_i of each area in *R*, number of regional sites N_i.
4.	Individual genes in the population as weight coefficients ρ_1,ρ_2.Finally, the scheduling workload of each area is calculated by the formula$W_i = \rho_1 \cdot D_i + \rho_2 \cdot N_i$. The variance of the regional workload is denoted as *V*.
5.	For each rental site *i*, try to put *i* into other communities and calculate the incremental ΔV of the adjustment workload, the entire process records the maximum ΔV_{max} and the corresponding community *k*. If $\Delta V_{max} < 0$, site node *i* does not adjust; if $\Delta V_{max} > 0$, node *i* is adjusted to community *k*. Traverse all the site until all the site are adjusted and the result is recorded as R^*.
6.	Define the variance function f_1 of the regional site, and define the regional dispatch distance variance function f_2, they are two objective functions to perform fast non-dominated sorting on the results, the records of the optimal solution in the contemporary population as f', and its corresponding scheduling workload parameters are denoted as $\rho_1{}'$,$\rho_2{}'$. If $f' > f^*$ after comparison, letting$\rho_1{}^* = \rho_1{}'$,$\rho_2{}^* = \rho_2{}'$.
7.	Determine whether the number of program iterations exceeds the maximum number of iterations *MaxGen*. If it exceeds, the optimal regional division results, and workload index parameters $\rho_1{}^*$,$\rho_2{}^*$ are output; otherwise, a new population is generated through elite strategy selection, which can ensure that certain elite individuals will not be discarded during the evolution process, thereby improving the accuracy of the optimization results, and expanding the sampling space. And gene crossover and mutation processes and the execution continue from 1.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Table 4(on next page)

dataset contains 16 fields, and the 9 fields

The dataset contains 16 fields, and the 9 fields related to this experiment are shown in the following ***Table 5.1***.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27562v1 | CC BY 4.0 Open Access | rec: 3 Mar 2019, publ: 3 Mar 2019

Table 5.1. The meaning of the field

No.	Fields	meaning
1	start time	Starting time
2	stop time	End Time
3	start_station_id	Bicycle rental site ID
4	start_station_name	Name of bicycle rental site
4	start_station_longitude	Longitude of rental bicycle rental site
5	start_station_latitude	Latitude of rental bicycle rental
6	end_station_id	Return bicycle rental ID
7	end_station_name	The name of the bicycle rental site
8	end_station_longitude	The longitude of the bicycle rental site
9	end_station_latitude	The longitude of the bicycle rental site

Table 5(on next page)

dataset contains 12 fields, and the 10 fields

The dataset contains 12 fields, and the 10 fields related to this experiment are presented in the following ***Table 5.2***.

Table 5.2. The meaning of the field

No.	Fields	meaning
1	start time	day and time trip started, in CST
2	stop time	day and time trip ended, in CST
3	from_station_id	ID of station where trip originated
4	from_station_name	name of station where trip originated
5	from_station_longitude	Longitude of rental bicycle rental site
6	from_station_latitude	Latitude of rental bicycle rental
7	to_station_id	ID of station where trip terminated
8	to_station_name	name of station where trip terminated
9	to_station_longitude	The longitude of the bicycle rental site
10	to_station_latitude	The longitude of the bicycle rental site