

Article

Research on Big Data Digging of Hot Topics about Recycled Water Use on Micro-Blog Based on Particle Swarm Optimization

Hanliang Fu ^{1,*} , Zhaoxing Li ^{2,*}, Zhijian Liu ³  and Zelin Wang ⁴ 

¹ School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China

² School of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

³ Department of Power Engineering, North China Electric Power University, Baoding 071003, China; zhijianliu@ncepu.edu.cn

⁴ School of Humanities and Social Sciences, Xi'an Jiaotong University, Xi'an 710049, China; zelin.wang@foxmail.com

* Correspondence: fuhanliang@xauat.edu.cn (H.F.); abde0309@126.com (Z.L.)

Received: 31 May 2018; Accepted: 12 July 2018; Published: 16 July 2018



Abstract: The public's acceptance level of recycled water use is a key factor that affects the popularization of this technology; therefore, it is critical to know the public's attitude in order to make guiding policies effectively and scientifically. To examine the major focuses and hot topics among the public about recycled water use, one of the major platforms for social opinion in China, the micro blog, is used as a source to obtain data related to the topic. Through the "follow-be followed" and "forward-dialogue" behaviors, a network of discussion of recycled water use among micro-blog users has been constructed. Improved particle swarm optimization has been used to allow deep digging for key words. Ultimately, key words about the topic of have been clustered into three categories, namely, the popularization status of recycled water use, the main application, and the public's attitude. The conclusion accurately describes the concerns of Chinese citizens regarding recycled water use, and has important significance for the popularization of this technology.

Keywords: recycled water use; micro-blog; big data; particle swarm optimization

1. Introduction

With the continued development of the social economy, population growth, and improvements in people's requirements in terms of life quality, the consumption of natural water resources for human activities has increased significantly. In the past 100 years, the demand of for water has increased by nearly 8 times [1]. Human activities have become the most important factor affecting the natural water cycling system, and its influence has exceeded the limit of ecological bearing capabilities, thus causing irreversible damage [2,3]. In urban areas with relatively dense industries and population, the supply-demand conflict of water resources is extremely significant, making water resources an important factor that restricts the development of the social economy in the 21st century [4]. Therefore, recycled water, an alternative to natural water resources, has become an important approach to solve the problem of water shortages.

At present, the main alternatives include sea water desalination, rainfall accumulation, and recycled water [5]. Recycled water can be obtained by wastewater treatment. It is a stable resource and is not affected by season. Meanwhile, it is more energy efficient and environmentally sound in terms of its production, compared to desalination [6]. Moreover, the production of recycled water can be used to relieve pollution in local water environments [7,8]. However, the public is disgusted by

recycled water, because waste water is the original source, so the level of public acceptance of recycled water is far lower than for other water resources [9,10].

Recycled water use in modern society dates back to the middle of 19th century. However, before 1990, wastewater treatment technology was backward, and recycled water was usually used for agricultural irrigation in less developed areas. However, in the following twenty years, with the rapid development of science and technology, the application scope of recycled water has expanded. In recent years, because of the emergence of wastewater treatment technologies like membrane technology [11], the recycled water obtained with deep treatment could meet the standards [12,13] of drinking water. Wastewater treatment technology is not the major difficulty that restricts the popularization of recycled water use. Rather, the public's resistance toward its use is a new issue [14]. Therefore, it is extremely important to know the public's opinions about recycled water, and to prevent these kind of issues from occurring [15,16]. The micro-blog is one of the main social communication tools for Chinese netizens; it is the major source and transmission approach of social opinion. One can have a good understanding of hot topics about recycled water use through the digging of micro-blog big data related to recycled water use, so as to confirm the main concerns of Chinese netizens [17,18]. However, it is not easy to find the key concerns among massive amounts of micro-blog data. Therefore, a type of algorithm that is easy to operate and which has high accuracy and fast rate of convergence is required to search this hot topic; namely, the particle swarm optimization algorithm.

The Particle Swarm Optimization algorithm was an evolutionary algorithm that was developed by Kennedy and Eberhart in 1995. By observing the foraging behaviors of birds, it was found that the flocks always changed their direction separately, only to get together again suddenly. So, the behaviors of bird flocks are not predictable, but are consistent overall. Each bird in the flock keeps the most suitable distance. A micro-blog is a very popular social network in China at present. Micro-blog users can post information on the internet and exchange with others. Each user can keep in touch with others with "follow up-followed" and "forward-dialogue". As a complicated social network consists of these relationships, a micro-blog is very similar to bird flocks, in the sense that both change from time to time. Therefore, using particle swarm optimization algorithm to dig out big data from a micro-blog is a wise choice. Using related theories of complicated networks, key words were used as nodes, while the follow up behaviors are the edges, which conform to a complicated network among users. With data digging, sensitive key data information was acquired. Then, those data were used to analyze the topological structure of a complicated relational network. Improved particle swarm optimization was used to dig out sensitive communities within a complicated network, so as to describe the hot topics about recycled water use among Chinese netizens.

2. Theoretical Background

2.1. Complexes Network

Networks consist of different types of complicated systems; in the real world, complex collaboration networks [19] and the world-wide web [20] are two typical examples which help us to communicate with each other in more convenient way.

We are living in a complicated world with all kinds of network systems, so analysis of the properties of these networks can help us to better know these systems. The small-world [21] and the scale-free properties are two thousand of the most significant characteristics of networks. In recent years, many fields have paid attention to the property of network community structures [22]. A graph consists of nodes and edges which are used to show a network community. A network community can be seen as a subset of a graph. Usually, communities need to meet the requirement that nodes should share similar features, although different communities have their own node properties. One must use a microscopic perspective to understand the functionality of a complicated system based on the knowledge of the network community structure. Many different methods have been proposed in the past decades to detect the community structure of networks.

Detection on the network community is connected to clustering problems. Clustering is an issue of optimization. Evolutionary algorithms (EAs) can be regarded as a class of intelligent optimization algorithms based on some biological and ethological principles. Those algorithms have been shown to be effective in solving many optimization problems. As a result, it is very common that EAs be used to solve network-clustering problems [23]. On the basis of this opinion, a lot of single-objective and multi-objective EA-based network-clustering methods have been proposed. Pizzuti [24] came up with the idea of using a genetic algorithm to discover network communities based on a community evaluation criterion called the *community score*. Gong et al. [25] proposed a mimetic algorithm-based approach that mentions a hill-climbing-based local search tactic to improve algorithm performance in searching data. *Local search* refers to the capability to approach the optimal solution infinitely, while *comprehensive convergence ability* refers to the ability to find the general location of the optimal solution. Both local search ability and comprehensive search ability are indispensable. In this paper, the local search ability of the particles and search efficiency were improved through the algorithm. In terms of network-clustering problem, previous researchers have come up with multi-objective optimization models.

Pizzuti [26] narrated a multi-objective genetic algorithm that can be used to reveal network communities. Recently, Gong et al. [27] proposed a multi-objective discrete PSO to solve related problems about network clustering. Moreover, many relevant researches and tests have proven the effectiveness of the proposed algorithm.

2.2. Particle Swarm Optimization

Eberhart and Kennedy introduced the Particle Swarm Optimization (PSO) algorithm in 1995. This algorithm is a population-based stochastic optimization technique that can be used to solve optimization problems. The inspiration for the design of PSO came from some social behaviors, like fish schooling and bird flocking [28]. Nowadays, PSO has become one of the most commonly-used optimization techniques to solve continuous optimization problems.

PSO is used to solve problems more efficiently by clustering individuals. Each individual can be regarded as one particle. Every particle can learn from other, nearby particles in order to adjust the velocity itself. In PSO, the particle's position vector can simulate a candidate solution to optimize the problem. After a new velocity is acquired, the position vector of the particle will be updated. That means that a new solution can be generated in this way, given that pop is the size of the particle swarm while n is the dimension of the search space.

Based on some tests on the PSO, it is obvious that it is very effective in solving problems about optimization. Different types of creative work have been done in order to improve the algorithm's performance [29]. Compared with other evolutionary algorithms, PSO does not have too many parameters. PSO has a concise framework that can be realized easily. However, PSO was supposed to be used in providing solutions for optimization-related problems. Because of problems occurring in applications, people have taken great effort to use PSO to solve many real questions in applications.

3. Method

In this paper, a single-objective PSO algorithm is proposed to solve problems about network-clustering. This algorithm can improve the quality of popular community significantly, as well as its modularity. Girvan and Newman [30] introduced the concept of modularity in their divisive hierarchical clustering algorithm. They used modularity as a standard to prevent network divisions from becoming subnetworks. Different kinds of methods have been developed to improve modularity, including the simulated annealing-based method, the external optimization-based approach, and the spectral optimization-based technique. However, people are not satisfied with the performance of these methods. As a result, the optimization of modularity becomes a NP-hard problem. In this paper, a particle turbulence operation is developed in order to improve the proposed algorithm. A local search strategy is used to improve algorithm use. Based on relevant real-world experiments and

operations, it turns out that the proposed algorithm is very effective in solving problems, and is much more efficient than some other, state-of-the-art approaches.

3.1. Network Clustering

Usually, a set of nodes and edges can be used in a graph to represent a network. Network clustering is used to divide network communities into different classes based on certain rules and principles. Every community is a cluster.

Given a network represented by a graph denoted by $G = (V, E)$, where V is the set of nodes and E is the set of edges, let A be the adjacency matrix of G and a_{ij} the ij th element of A . Given that S is a subgraph (i.e., $S \in G$), let $d_{i \text{ in}} = \sum_{j \in S} a_{ij}$ and $d_{i \text{ out}} = \sum_{j \in S} a_{ij}$ be respectively the internal and external degrees of node i . In this paper, S can be regarded as a community in the strong sense, if

$$\forall i \in S, d_i^{\text{in}} > d_i^{\text{out}} \quad (1)$$

And S is a community in a weak sense if

$$\sum_{i \in S} d_i^{\text{in}} > \sum_{i \in S} d_i^{\text{out}} \quad (2)$$

However, the definition mentioned above only describes the conditions that a network community shall have. In order to solve the network-clustering problem, a quantitative index shall be used to measure the network partition. In this aspect, a modularity index was introduced by Girvan and Newman [30]. Modularity (normally denoted by Q) can be written as:

$$Q = \frac{1}{2m} \sum \left(A_{ij} - \frac{k_i k_j}{2m} \right) \sigma(i, j) \quad (3)$$

where m is the number of edges and $\sigma(i, j) = 1$ if nodes i and j are in the same community, otherwise $\sigma(i, j) = 0$. Normally, by assumption, we take it that the larger the value of Q , the more accurate the partition.

3.2. Improved Particle Swarm Optimization

PSO has been shown to be very promising for solving many optimization problems. Each individual can be regarded as one "particle", and this can be used to simulate a possible solution in order to get the best solution. Based on flying experience and the knowledge acquired from other particles, it can update its own flying velocity and position accordingly.

Given a particle swarm size m , and a particle dimension n , let $V_i = \{v_1, v_2, \dots, v_n\}$ and $X_i = \{x_1, x_2, \dots, x_n\}$ be the velocity and position vectors, respectively, of the i th ($i = 1, 2, \dots, m$) particle. The updating process for each particle in the basic form can be formulated by the following equations:

$$V_i \leftarrow \omega V_i + c_1 r_1 (P_i - X_i) + c_2 r_2 (G - X_i) \quad (4)$$

$$X_i \leftarrow X_i + V_i \quad (5)$$

where $P_i = \{p_1, p_2, \dots, p_n\}$ and $G = \{g_1, g_2, \dots, g_n\}$ are the personal best position of the i th particle and the global best position of the swarm, respectively. The parameters r_1 and r_2 are random numbers between 0 and 1, while c_1 and c_2 are acceleration coefficients termed *cognitive* and *social* components.

The rapid loss of population diversity is one of the most important causes of premature convergence of algorithm, while the multi population strategy is an important approach to improve population diversity and the global search of the algorithm. In this paper, a multi-population coordination mechanism is introduced, and a multi-objective particle swarm optimization (MOPSO) is proposed based on object segmentation. The multi-population coordination mechanism is a coordination mechanism for multiple populations; its function is to solve multi-objective optimization problems through multiple populations, to introduce external archives mechanism to store the non-inferior solutions, and to change the speed updating formula of particles correspondingly. The particle flight is not only affected by the population particles, but also by the non-inferior solutions in the external archives. This mechanism makes it possible for every object to find its optimization solution and coordinate conflicts among solutions.

This algorithm confirms the quantity of the particle swarm based on the quantity of object functions. Using single object optimization technology, each particle swarm can find the non-dominated extreme point for the corresponding single objective function. All non-inferior extreme points obtained from the particle swarm will form a major particle swarm, and operate partial MOPSO, which fixes the distribution gap among the non-dominated extreme points obtained from the particle swarm, so as to generate an optimal Pareto solution set with even distribution and wide coverage.

In this paper, a master-slave strategy of multi-population co-evolution is proposed so as to offer a MOPSO algorithm of multi-population coloration based on object segmentation. This algorithm adopts a master-slave evolution model and confirms the particle swarm quantity according to the object function quantity: one particle swarm corresponds to one object function. This will then be used to search the non-dominated solution of a single object function. A partial search algorithm operates on the major particle swarm to fix the existing gap among the optimal solutions obtained from the particle swarms, so a Pareto front with good distribution can be acquired. For an extreme point of object function, if it includes the optimal solution of Pareto, then this extreme point can be regarded as non-dominated. Furthermore, as with most MOPSO algorithms, the major particle swarm also uses an external reserve set to keep all non-dominated solutions obtained in the algorithm's evolution.

Figure 1 shows the parallel model of the aforementioned algorithm. This model includes M subordinate particle swarms and one principle particle swarm; one subordinate particle swarm corresponds to a single object function. Every subordinate particle swarm could operate the improved particle swarm optimization (PSO) independently, and then search the non-dominated extreme point of corresponding single object function. The principle particle swarm will operate a partial MOPSO to generate a Pareto front with good distribution. During every iteration process, all acquired non-dominated extreme points will be transferred to the principle particle swarm by the subordinate particle swarms, and will then be kept in the reserve set. After the reserve set accepts these non-dominated extreme points, the elements that are distributed in sparse area will be chosen as the guiders, i.e., to guide the particles to develop these potential areas emphatically.

In order to find all non-dominated extreme points of a single object function rapidly, an improved particle swarm optimization based on clustering is proposed in this paper. This algorithm can improve the subordinate particle swarms and locate multiple non-dominated extreme points by introducing two new evolutionary operators, namely, clustering technology based on Pareto, and a mature escape strategy. Taking the M_{th} subordinate particle swarm $S_m(t)$ as an example, the description of two improved algorithms can be shown as below:

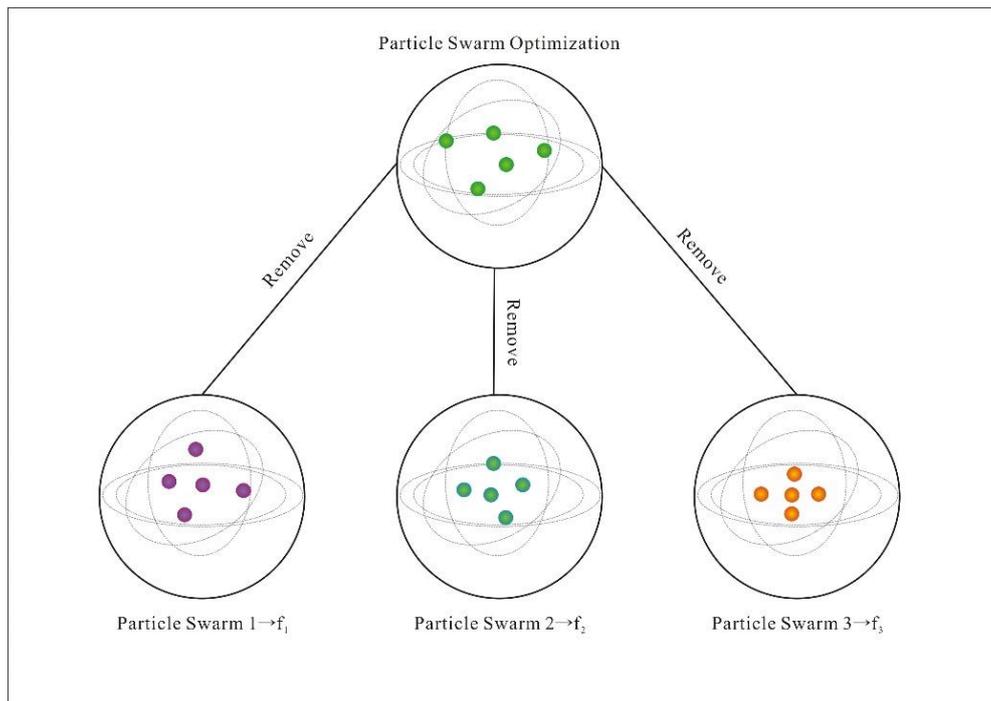


Figure 1. Parallel optimization model of algorithm.

In this paper, a clustering technology dominated by Pareto has been proposed. This technology can divide the particles in the subordinate sets $S_{(1)}$ into multiple categories, so different categories could search different extreme points at the same time. To be specific, this clustering technology mainly includes confirmation of the class seed and its classes. The Algorithm 1 is described below:

Algorithm 1: Class seed determination algorithm

Input: Lsorted

Output: PS

1: Begin

2: $PS = \emptyset$

3: While not reaching the end of Lsorted;

4: Get best unprocessed $\rho \in$ Lsorted;

5: flag = 0;

6: For all $sp \in PS$

7: If $d(s_p, p) \leq r_s$;

8: flag = 1;

9: break;

10: Endif

11: Endfor

12: If flag = 0 and p is non-dominated with respect to the PS

13: Let $PS \leftarrow PS \cup \{p\}$

14: Endif

15: Endwhile

16: End

In this algorithm, the set Lsorted is the input. All particles in the subordinate particle swarms are in descending order based on the f_m value of elements in Lsorted. Firstly, one null set PS is initialized to reserve all acquired class seeds; next, the distance between the particles in Lsorted and discovered class seed in PS is inspected. If the distance between one particle and a random class seed in the PS

is greater than the radius r_s , and cannot be dominated by the class seed in PS, then this particle can reserved in the set PS as a new class seed. The class radius in this paper is:

$$r_s = k_0 \sqrt{\sum_{i=1}^n (ub(i) - lb(i))^2} \quad (6)$$

where $ub(i)$ and $lb(i)$ are the upper and lower bound of the i th decision variable, K_0 is the radius coefficient among the value range $[0, 1]$.

Multiple class seeds with good distribution technology can be confirmed through the clustering technology mentioned above, and the distance between the particles in $S_m(t)$ and PS has been inspected one by one. If the distance between the particle and the class seed is smaller than the radius, then this particle will be kept in the corresponding classes of class seed. Lastly, all particles that haven't been classified successfully will be gathered, forming pseudo classes. Based on the two improved operators mentioned, improved particle swarm optimization can be obtained.

3.2.1. General Framework

A new PSO algorithm for use in network clustering has been proposed in this paper. The principles that are used to define the particle status are introduced in this paper. Moreover, in order to make improvements in exploring the effect of this algorithm, turbulence operations were developed. The introduction of a local search strategy would be useful in enhancing use. Thus, detailed information of the algorithm will be discussed in this section.

The proposed PSO-based network-clustering algorithm (Algorithm 2) works as follows:

Algorithm 2: PSO-based network-clustering algorithm

Input:

pop: particle swarm size;
 g_{\max} : number of iterations;
 ω : inertia weight, learning factors c_1 and c_2 ;
 A : network adjacency matrix.

Output: the network community structure.

- 1: **(Step 1)** Initialization
 - 2: **(Step 1.1)** Initialize the population, i.e., initialize the position vectors $X_1, X_2, \dots, X_{\text{pop}}$ and initialize the velocity vectors $V_1, V_2, \dots, V_{\text{pop}}$;
 - 3: **(Step 1.2)** Initialize the personal best position vectors $P_1, P_2, \dots, P_{\text{pop}}$ and the global best position vector G ;
 - 4: **(Step 2)** Update: For $i = 1, \dots, \text{pop}$, do
 - 5: **(Step 2.1)** Update the particle velocity vectors according to Equation (7);
 - 6: **(Step 2.2)** Update the particle position vectors according to Equation (8);
 - 7: **(Step 2.3)** Turbulence operation on the position vectors;
 - 8: **(Step 3)** Local search: implement the local search procedure on the global best vector G ;
 - 9: **(Step 4)** Stopping criterion: If the stopping criterion is satisfied, then stop and output. Otherwise, go to Step 2).
-

3.2.2. Particle Representation

A solution to solve optimization problem can be represented by a particle. The position vector of particles is also used to solve network-clustering problem. The string coding system is used to solve problems related to optimization. In this system, the position vector can be regarded as one integer permutation.

Figure 2 shows the particle representation scheme in detail. Every element of the position vector is regarded as one integer, which is regarded as the cluster label of the corresponding node. If nodes share the same label, they belong to the same cluster.

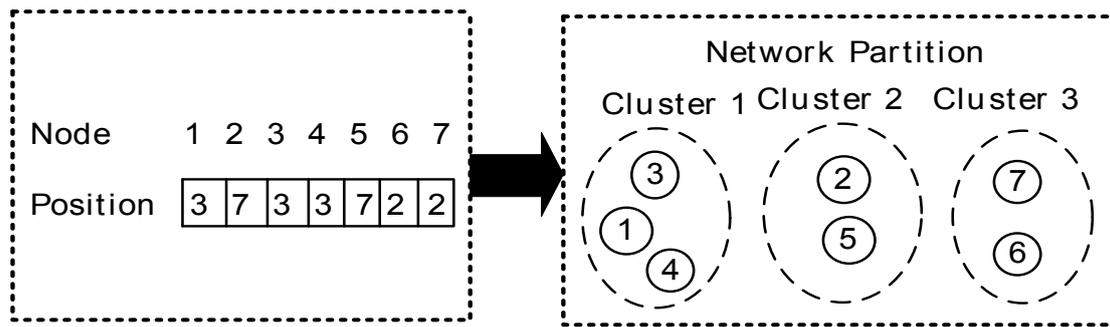


Figure 2. Graphical illustration of the particle representation scheme unique interest.

It is easy to use and obtain the adopted particle representation scheme. It is not necessary to define the number of network communities, as this is determined automatically already.

3.2.3. Particle Update Rules

Because the particle velocity and position vectors are all integer coded, the update rules in Equations (4) and (5) do not fit the requirements of the recommendation problem. In this paper, we can define the update rules in a discrete form again. The redefined rules are the following:

$$\mathbf{V}_i \leftarrow \text{sig}(\omega \mathbf{V}_i + c_1 r_1 (\mathbf{P}_i \oplus \mathbf{X}_i) + c_2 r_2 (\mathbf{G} \oplus \mathbf{X}_i)) \quad (7)$$

$$\mathbf{X}_i \leftarrow \mathbf{X}_i \ominus \mathbf{V}_i \quad (8)$$

where the operator \oplus is the XOR operation, and the $\text{sig}(x)$ function works as follows:

$$\text{sig}(x) = \begin{cases} 1 & \text{if } \text{rand}(0,1) < 1/(1 + e^{-x}) \\ 0 & \text{if } \text{rand}(0,1) \geq 1/(1 + e^{-x}) \end{cases} \quad (9)$$

The $\text{sig}(x)$ function maps the velocity vector into a binary code sequence; as a result, the operator \ominus can be seen as the key operation of the newly defined particle update rules. In this paper, in consideration of the network topology and the features of the optimization objective function, we can define the operation of \ominus in the following way:

$$\mathbf{X}_i \ominus \mathbf{V}_i = \{x_{i1}', x_{i2}', \dots, x_{in}'\} \quad (10)$$

In this equation, if an element of V_i is zero, $x_{ij}' = x_{ij}$; otherwise, x_{ij}' is calculated as follows:

$$x_{ij}' = \text{argmax} \Delta Q(x_{ij} \leftarrow x_{ik} | k \in N_j) \quad (11)$$

where N_j is the neighbor set of node j .

From Figure 3, we can see the operations of the updated rules clearly. The topology information can be fully used because of the updated rules of the particle status. It turns out that the proposed algorithm can be used to feasibly complete community discovery tasks.

In Equation (11), the value x_{ik} that generates the largest modularity increment is turned, while the x_{ij} element is repulsed by the value x_{ik} . After defining the position update rules, a large value can be obtained for this object function because the community identifier is found to be effective in increasing the modularity. Therefore, if multiple x_{ik} can lead to significant modularity increases, we will choose one of the modularities at random.

After updating the particle position, a turbulence operator was designed to preserve the population diversity. If most people or one person share the same feature, they can be regarded as one cluster. The designed turbulence operation can be used in the following way:

Firstly, it is important to know whether it can work with the possibility of 0.5. Secondly, we will try to get a dominated identifier from other, nearby particles in order to update the position element afterwards. If more than one dominated identifier is found, we just pick one. Algorithm 3 has shown the pseudocode of the operation.

Algorithm 3: Pseudocode of the local search operation

```

1: for  $i = 0; i < \text{vertex}; i++$ 
2:   (a)  $\text{NL}[] = \text{find}(\text{node}[i].\text{NeighborLabel})$ ;
3:   (b)  $\text{DiffL}[] = \text{DifferentLabel}(\text{NL})$ ;
4:   (c) for  $j = 0; j < \text{size}(\text{DiffL}); j++$ 
5:     (c)i.  $\text{DQ}[] = \text{deltaQ}(\text{node}[i].\text{label} \leftarrow \text{DiffL}[j])$ ;
6:   (d) end for
7:   (e) if  $\text{max}(\text{DQ}[]) > 0$ 
8:     (c)i.  $\text{node}[i].\text{label} = \text{DiffL}[\text{max}]$ ;
9:     (c)ii. break;
10:  (f) end if
11: end for
    
```

The global best particle is regarded as a leader, which makes the local search step only able to lead the swarm in certain regions as a guider.

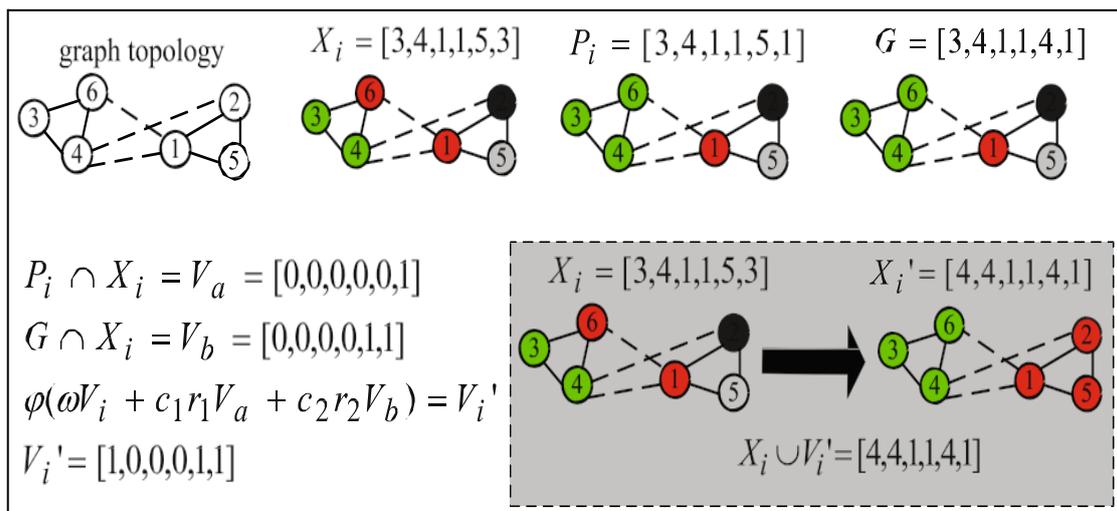


Figure 3. A graphical illustration of the operations of the update rules.

4. Results and Analysis

This study takes big data from micro-blog topics about recycled water use as a data source, and 18 high-frequency keywords are taken as nodes of complex networks about recycled water use; the behavior of “follow-followed by” and “forward-dialogue” among micro-blog users are taken as clues. With PSO, twelve key words among micro-blog topics are clustered into 3 communities, which are shown in Figure 4.

NMI is an index that is used to measure the proximity of two set elements. With this index, one can know the difference between the obtained community structure with the algorithm, and the actual structure.

Assuming *R* is the actual community classification of the network, *D* is the community classification of the network obtained from algorithm. *M* is a matrix, and the element m_{ij} in *M*

is the node quantity of the intersection between the i th community set and the j th community set. The mutual information computation formal between R and D is the following:

$$NMI = \frac{-2\sum_{i=1}^{KR} \sum_{j=1}^{KD} m_{ij} \log\left(\frac{m_{ij}N}{m_i m_j}\right)}{\sum_{i=1}^{KR} m_i \log\left(\frac{m_i}{N}\right) + \sum_{j=1}^{KD} m_j \log\left(\frac{m_j}{N}\right)} \quad (12)$$

KR and KD represent the quantity of R community and D community respectively, while m_i , (m_j) represents the sum of elements of the i th row (j line) in the matrix; N is the quantity of network nodes.

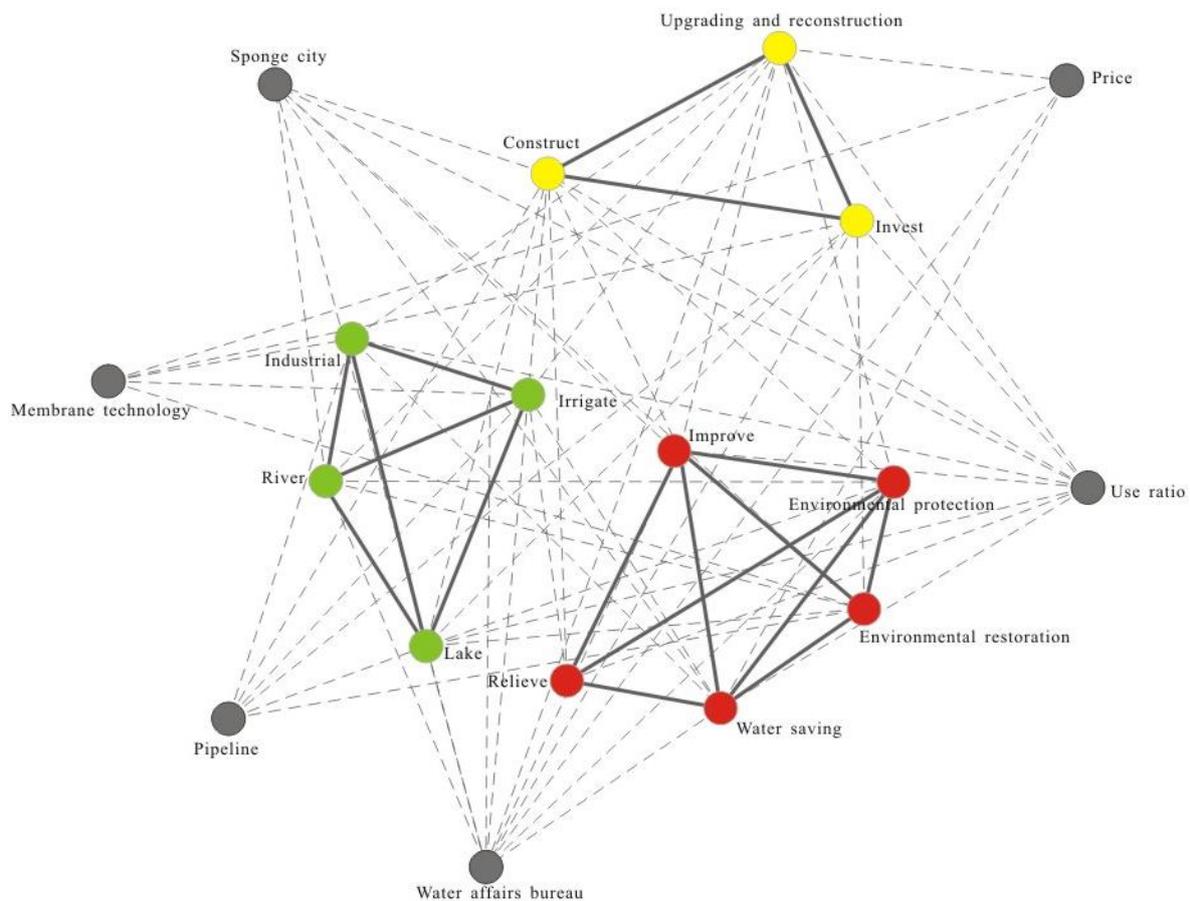


Figure 4. Detected communities structure of the complex network regarding to the topic of recycled water use on micro-blog.

When the actual network does not exist, we will compare the functional values obtained from each algorithm, because we used the algorithm to find the corresponding network community classification when the modularity density is at its maximal value; the greater the obtained functional value, the more comprehensive search ability of the algorithm.

In order to inspect whether the results shown in Figure 4 are credible, traditional NMI indicators were used to measure the advantages and disadvantages of the obtained network community structure obtained with the algorithm. Independent operation of the algorithm was performed 30 times. The obtained network community structure obtained from each algorithm operation test was recorded, and the NMI value was calculated. Because the topological structure of a simulative network is controlled by three parameters, the results obtained in the algorithm with all parameter combinations are recorded in the statistics, and the statistical results are shown in Figure 5.

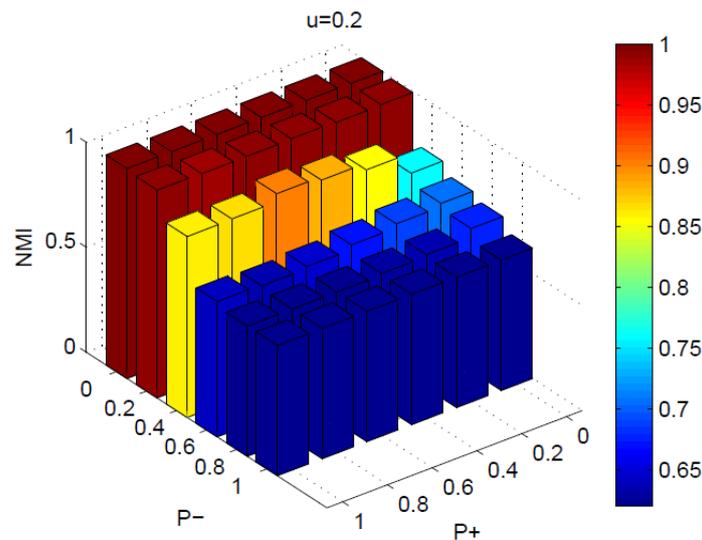


Figure 5. Average NMI values for the complex network regarding to the topic of recycled water use on micro-blog.

Parameter μ determines the proportion of the quantity of internal edges in the community and the edges among communities. The edge among communities will be denser if the parameter μ is greater; then the community structure will be vaguer. When $\mu = 0.2$, a clear community outcome can be obtained. Figure 5 shows a three-dimension display of the statistic results obtained after having performed 30 independent operations of the algorithm. From the results show in Figure 5, it can be seen that the network community has clear discrimination, which verifies the fact that PSO is effective in classifying the community structure about recycled water use on the micro-blog. Therefore, a conclusion could be drawn that the result shown in Figure 4 could responsibly reflect hot topics about recycled water use on a micro-blog.

As shown in Figure 4, the key words appearing in micro-blog topics about recycled water use can be clustered into three communities (shown in Figure 4), consisting of yellow points, green points and red points.

The community of yellow points consists of key words like “Upgrading and reconstruction”, “Construct”, and “Invest”. Considering micro-blog topics and the development situation of recycled water use in China, key words such as “construct” and “invest” could reflect the development tendency of recycled water use industry in China at present. Meanwhile, the key words “upgrading and reconstruction” reflect the popularization of the industry of recycled water use in China, so as to make improvements on the processing technologies of municipal wastewater treatment, or to add in-depth processing technology to current wastewater treatment procedures as a major processing technology.

The community in red points consists of key words like “improve”, “Environmental protection”, “Environmental restoration”, “Water saving” and “Relieve”. This series of key words describes the positive effect of recycled water use on the environment, and reflects the positive image of recycled water use in the minds of Chinese netizens. It can be seen that recycled water use has a good foundation in terms of social opinion and popular will.

The community in green points consists of key words like “industrial”, “irrigate”, “lake”, and “river”. These key words indicate the main usage of recycled water in China at present. Currently, the popularization of recycled water use is at an initial stage of development. The construction of recycled water facilities and the public’s acceptance of recycled water use determine the popularity; at present, recycled water is mainly used as industrial cooling water, agriculture irrigation water, river and lake water replenishing, and therefore, does not come into contact with human bodies directly.

According to the other key words in the topics of recycled water use on the micro-blog, conclusions can be made: the use ratio of recycled water, the price of recycled water, the pipeline to transport

recycled water, and emerging membrane technology that is used in the wastewater treatment are major concerns of micro-blog netizens. Meanwhile, the Water Affairs Bureau also shows up in discussions of recycled water use, which shows its important role in the popularization of the technology. Recycled water use could improve the water environment. Therefore, the popularization of recycled water use has great significance for *sponge city*. *Sponge city* appears in the key words that show up with high frequency in the topic about recycled water use, which also verifies the close connection between recycled water use and *Sponge city*.

5. Discussion

The core object of this research is to find the hot topic of the public's concern about recycled water use, and cluster the key words appearing in the micro-blog topics about recycled water use. The research conclusions reflect the concerns of Chinese netizens; we obtained the following inspiration from these research conclusions:

In recent years, recycled water use and wastewater treatment plants have been built rapidly in China. For example, the capacity and efficiency of wastewater treatment plants have increased from $0.4 \times 10^6 \text{ m}^3$ and 14.9%, respectively, in 1991 to $3.8 \times 10^6 \text{ m}^3$ and 89.3% in 2013 [31]. The conclusions of micro-blog big data digging conforms to the actual situation; key words that were contained in the yellow points community like "Upgrading and reconstruction" and "Construct" and "Invest" reflect the rapid growth of recycled water use in China.

At present, Chinese netizens are very positive towards the reclaimed water use. As shown in the red points community, Chinese citizens are mainly concerned about the improvement of environment resulting from reclaimed water use. Numerous research has been undertaken in different countries like Australia and the United States, indicating that residents felt disgusted about the reclaimed water use, i.e., the so called "yuck factor" [32]. Despite this finding, this is not the concern of Chinese netizens. This may have something to do with the low popularity of reclaimed water facilities, and the fact that reclaimed water use hasn't affected the vital interests of most netizens. However, based on the experience of popularizing the use of reclaimed water in many countries, the public's aversion and objection to it is an important reason for the popularization of reclaimed water use. Therefore, preventing this from happening and guiding Chinese citizens to keep a positive attitude shall be taken into serious consideration.

At present, the approaches of reclaimed water use in China are very restricted. Reclaimed water use in modern society began in the middle of the 19th century. With nearly two centuries of development, current technology can be used to produce reclaimed water with any water quality standard. Reclaimed water has been widely used in groundwater recharge, non-potable uses, and even potable uses [33], as shown in Figure 4: for the usage of reclaimed water, the Chinese netizens concern about industrial uses, irrigating, lake and river recharge, as well as the usage with low degree of contact with human body. Therefore, it can be seen that the usage of reclaimed water is very limited, so the water management department must consider the issue of expanding safe and reliable usage of reclaimed water that can be accepted by the public.

Pipeline construction is an important factor that affects the popularization of reclaimed water use in China. The reused reclaimed water in China at present is produced by wastewater treatment plants, and then transported to water users through pipelines [34]. Meanwhile, the construction expense of pipelines is high; as one of the most important factors restricting the popularization of reclaimed water use in China, the construction of pipelines has drawn much attention from Chinese netizens.

New types of wastewater treatment technologies comprising membrane technology have provided a new direction for popularizing the use of reclaimed water in China. The emergence of new types of wastewater treatment technologies makes it possible to produce large-scale and automatic reclaimed water production equipment, and it resolves the technical obstacle of adopting distributed reclaimed water use modes. Therefore, using small-scale reclaimed water equipment in

populated areas that are far away from municipal pipelines is a new tendency for the popularization of reclaimed water use in China.

Price is a factor that has a direct effect on the public's acceptance level vis a vis reclaimed water use. As a substitute for running water, the public's acceptance degree is affected by both price, and price relations between recycled water and tap water prices. Therefore, the pricing of reclaimed water is more complicated than common commodities. Meanwhile, as a major concern of Chinese netizens, price also shows the public's concern about the reclaimed water price.

Urban construction in Haimian provides a new opportunity for reclaimed water use. Keeping the purification of water is one of the important goals to promote the urban construction of Mianyang. The reclaimed water use can not only increase the volume of wastewater treatment, but also reduce the discharge of pollution in natural water systems. Moreover, it also reduces the exploitation of water resources by human activities. Therefore, reclaimed water use plays an important role in urban construction in Mianyang. In contrast, the urban construction of Mianyang also provides a new opportunity for the popularization of reclaimed water use in China.

Government-oriented popularization of reclaimed water use is a major tendency among hot topics like use-ratio and the water affair bureau. By setting goals for the use-ratio for reclaimed water use, we can construct facilities to make reclaimed water use compulsory. Undoubtedly, at the first stage of the popularization, this solution could speed up the progress for the popularization of reclaimed water use. However, if guiding policy of urban residents' behavior cannot be changed in time to improve public attitudes towards reclaimed water use, and willingness to pay for it, the facilities that are constructed may amount to nothing.

The deficiencies of this study are that most topics about recycled water in micro-blogs are lead by government organizations and enterprises, and there are very limited discussions made by ordinary citizens. As a result, it was difficult for the author to find the real opinions of the public regarding to the recycled water by the described means. In future studies, we shall obtain data about ordinary citizens' attitude towards recycled water use.

Author Contributions: Conceptualization, H.F.; Methodology, Z.L. (Zhaoxing Li); Software, Z.L. (Zhaoxing Li); Validation, Z.W.; Formal Analysis, Z.L. (Zhaoxing Li); Data Curation, Z.L. (Zhaoxing Li); Writing-Original Draft Preparation, H.F. and Z.L. (Zhaoxing Li); Writing-Review & Editing, Z.L. (Zhijian Liu); Visualization, Z.L. (Zhijian Liu) and Z.W.; Supervision, H.F.; Project Administration, Z.L. (Zhaoxing Li); Funding Acquisition, H.F.

Funding: Chinese Postdoctoral Science Foundation "Research on the influencing factors and driving strategies of urban residents' recycled water reuse behaviors" (2018M633480).

Acknowledgments: This work is supported by the Chinese Postdoctoral Science Foundation "Research on the influencing factors and driving strategies of urban residents' recycled water reuse behaviors" (2018M633480).

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

References

1. Veldkamp, T.I.E.; Wada, Y.; Aerts, J.C.J.H.; Döll, P.; Gosling, S.N.; Liu, J.; Masaki, Y.; Oki, T.; Ostberg, S.; Pokhrel, Y.; et al. Water scarcity hotspots travel downstream due to human interventions in the 20th and 21st century. *Nat. Commun.* **2017**, *8*, 15697. [[CrossRef](#)] [[PubMed](#)]
2. Lu, S.; Zhang, X.; Bao, H.; Skitmore, M. Review of social water cycle research in a changing environment. *Renew. Sustain. Energy Rev.* **2016**, *63*, 132–140. [[CrossRef](#)]
3. Zeng, Y.; Xie, Z.; Zou, J. Hydrologic and climatic responses to global anthropogenic groundwater extraction. *J. Clim.* **2017**, *30*, 71–90. [[CrossRef](#)]
4. Liu, K.; Fu, H.; Chen, H. Research on the Influencing Mechanism of Traditional Cultural Values on Citizens' Behavior Regarding the Reuse of Recycled Water. *Sustainability* **2018**, *10*, 165. [[CrossRef](#)]
5. Fielding, K.S.; Dolnicar, S.; Schultz, T. Public acceptance of recycled water. *Int. J. Water Resour. Dev.* **2018**, *1*, 1–36. [[CrossRef](#)]
6. Kajenthira, A.; Siddiqi, A.; Anadon, L.D. A new case for promoting wastewater reuse in Saudi Arabia: Bringing energy into the water equation. *J. Environ. Manag.* **2012**, *102*, 184–192. [[CrossRef](#)] [[PubMed](#)]

7. Fu, H.; Liu, X. Research on the Phenomenon of Chinese Residents' Spiritual Contagion for the Reuse of Recycled Water Based on SC-IAT. *Water* **2017**, *9*, 846.
8. Smith, H.M.; Brouwer, S.; Jeffrey, P.; Frijns, J. Public responses to water reuse—Understanding the evidence. *J. Environ. Manag.* **2017**, *207*, 43–50. [[CrossRef](#)] [[PubMed](#)]
9. Hurlimann, A.; Dolnicar, S. Public acceptance and perceptions of alternative water sources: A comparative study in nine locations. *Int. J. Water Resour. Dev.* **2016**, *32*, 650–673. [[CrossRef](#)]
10. Wang, S.; Leviston, Z.; Hurlstone, M.; Lawrence, C.; Walker, I. Emotions predict policy support: Why it matters how people feel about climate change. *Glob. Environ. Chang.* **2018**, *50*, 25–40. [[CrossRef](#)]
11. Li, C.; Wang, L.; Wang, X.; Kong, M.; Zhang, Q.; Li, G. Synthesis of PVDF-g-PSSA proton exchange membrane by ozone-induced graft copolymerization and its application in microbial fuel cells. *J. Membr. Sci.* **2017**, *527*, 35–42. [[CrossRef](#)]
12. Qian, N.; Leong, C. A game theoretic approach to implementation of recycled drinking water. *Desalin. Water Treat.* **2016**, *57*, 1–9. [[CrossRef](#)]
13. Kosovac, A.; Hurlimann, A.; Davidson, B. Water Experts' Perception of Risk for New and Unfamiliar Water Projects. *Water* **2017**, *9*, 976. [[CrossRef](#)]
14. West, C.; Kenway, S.; Hassal, M.; Yuan, Z. Why do residential recycled water schemes fail? A comprehensive review of risk factors and impact on objectives. *Water Res.* **2016**, *102*, 271–281. [[CrossRef](#)] [[PubMed](#)]
15. Dolnicar, S.; Hurlimann, A.; Grün, B. Branding Water. *Water Res.* **2014**, *57*, 325–338. [[CrossRef](#)] [[PubMed](#)]
16. Fu, H.; Liu, X. A Study on the Impact of Environmental Education on Individuals' Behaviors Concerning Recycled Water Reuse. *Eurasia J. Math. Sci. Technol. Educ.* **2017**, *13*, 6715–6724. [[CrossRef](#)]
17. Chen, W.; Bai, Y.; Zhang, W.; Lyu, S.; Jiao, W. Perceptions of Different Stakeholders on Reclaimed Water Reuse: The Case of Beijing, China. *Sustainability* **2015**, *7*, 9696–9710. [[CrossRef](#)]
18. Gu, Q.; Chen, Y.; Pody, R.; Cheng, R.; Zheng, X.; Zhang, Z. Public perception and acceptability toward reclaimed water in Tianjin. *Resour. Conserv. Recycl.* **2015**, *104*, 291–299. [[CrossRef](#)]
19. Tian, G.; Zhang, H.; Feng, Y.; Jia, H.; Zhang, C.; Jiang, Z.; Li, L.; Li, P. Operation patterns analysis of automotive components remanufacturing industry development in China. *J. Clean Prod.* **2017**, *64*, 1363–1375. [[CrossRef](#)]
20. Liu, Z.; Wu, D.; Yu, H.; Ma, W.; Gin, G. Field measurement and numerical simulation of combined solar heating operation modes for domestic buildings based on the Qinghai-Tibetan plateau case. *Energy Build.* **2018**, *167*, 312–321. [[CrossRef](#)]
21. Watts, D.J.; Strogatz, S.H. Collective dynamics of “small-world” networks. *Nature* **2012**, *393*, 440–442. [[CrossRef](#)] [[PubMed](#)]
22. Zeng, Y.; Xie, Z.; Yu, Y.; Liu, S.; Wang, L.; Zou, J.; Qin, P.; Jia, P. Effects of anthropogenic water regulation and groundwater lateral flow on land processes. *J. Adv. Model. Earth Syst.* **2016**, *8*, 1106–1131. [[CrossRef](#)]
23. Zechman, E.M.; Giacomoni, M.H.; Shafiee, M.E. An evolutionary algorithm approach to generate distinct sets of non-dominated solutions for wicked problems. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1442–1457. [[CrossRef](#)]
24. Li, Z.; He, L.; Li, Y. A novel multiobjective particle swarm optimization algorithm for signed network community detection. *Appl. Intell.* **2016**, *44*, 621–633. [[CrossRef](#)]
25. Gong, M.; Fu, B.; Jiao, L.; Du, H. Memetic algorithm for community detection in networks. *Phys. Rev. E* **2011**, *84*, 056101. [[CrossRef](#)] [[PubMed](#)]
26. Liu, Z.; Cheng, K.; Li, H.; Cao, G.; Wu, D.; Shi, Y. Exploring the potential relationship between indoor air quality and the concentration of airborne culturable fungi: A combined experimental and neural network modeling study. *Environ. Sci. Pollut. Res.* **2018**, *25*, 3510–3517. [[CrossRef](#)] [[PubMed](#)]
27. Liu, Z.; Li, H.; Cao, G. Quick Estimation Model for the Concentration of Indoor Airborne Culturable Bacteria: An Application of Machine Learning. *Int. J. Environ. Res. Public Health* **2017**, *14*, 857. [[CrossRef](#)] [[PubMed](#)]
28. Trelea, I.C. The particle swarm optimization algorithm: Convergence analysis and parameter selection. *Inf. Process. Lett.* **2016**, *85*, 317–325. [[CrossRef](#)]
29. Gong, M.; Cai, Q.; Chen, X.; Ma, L. Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition. *IEEE Trans. Evol. Comput.* **2014**, *18*, 82–97. [[CrossRef](#)]
30. Girvan, M.; Newman, M.E.J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, Q.; Yang, W.; Ngo, H.; Guo, W.; Jin, P.; Dzkapasu, M.; Yang, S.; Wang, Q.; Wang, X.; Ao, D. Current status of urban wastewater treatment plants in China. *Environ. Int.* **2016**, *92–93*, 11–22. [[CrossRef](#)] [[PubMed](#)]

32. Leong, C. The Role of Emotions in Drinking Recycled Water. *Water* **2016**, *8*, 548. [[CrossRef](#)]
33. Chen, Z.; Ngo, H.; Guo, W. A Critical Review on the End Uses of Recycled Water. *Crit. Rev. Environ. Sci. Technol.* **2013**, *43*, 1446–1516. [[CrossRef](#)]
34. Liu, X.J.; Guo, Z.Y.; Ding, C.; Fu, H. Industrial Structure Adjustment of Shaanxi Province under the Limited Water Resources Condition. *Front. Eng. Manag.* **2014**, *1*, 378–384.



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).