Research Article

Constructing a Method for an Evaluation Index System Based on Graph Distance Classification and Principal Component Analysis

Keyou Shi, Yong Liu [ID], Zhijun Zhang [ID], Qing Yu [ID], and Qiucai Zhang [ID]

School of Resource Environmental and Safety Engineering, University of South China, Hengyang 421001, China

Correspondence should be addressed to Yong Liu; liuyong81668@163.com and Qiucai Zhang; zqiucai@163.com

Based on the importance of having an evaluation index system, a new method that combines PCA with graph distance classification is presented to make up the deficiencies of principal component analysis in the process of index screening, and this method is applied in the construction of an evaluation index system for the environmental quality of decommissioning uranium tailing. The seepage indexes were classified into six classes using graph distance classification, which selects the representative elements, including pH, $\sum \alpha$, ^{210}Pb, ^{210}Po, F$^-$, and NO^{3-}. All of the representative elements were analyzed by PCA while determining the seepage indexes, including pH, U, Ra, $\sum \alpha$, NH$_4$-N, and F$^-$, and establishing an index system for environmental quality evaluation that consists of two primary indexes (seepage and radiation environment) and 12 secondary indexes. The results showed that the model had ensured that the sifted indexes had a significant effect on the evaluation result and avoided the deletion of some important indexes and that it had stronger applicability and maneuverability.

1. Introduction

The selection of an evaluation index and whether the index system is reasonable or not have an essential effect on the evaluation results. Thus, how to select an index in a complicated and enormous index system is a problem in the construction of an evaluation index system for the environmental quality of decommissioning uranium tailing. If all of the indexes are evaluated, then the computing effort of the information processing will be greatly increased because of the excessive indexes. However, if only a few indexes are selected individually, then a large amount of the original data's information could be lost, which causes inaccuracy of the evaluation results.

Principal component analysis (PCA), as an evaluation index screening method, could not only substitute a small number comprehensive variables for the original multidimensional variable and obviously simplify the data structure under the precondition of minimizing the loss of the original data's information but also avoid subjectivity and

arbitrariness. In the application, PCA is widely applied in index selection, evaluation, and prediction, with strong applicability [1–5]. At the same time, there are also some defects. For example, the contribution rate of the discard component on the data analysis could be similar to the chosen principal component, or there is a high level of correlation between the major elements of the chosen principal component and the elements of the discard component, which is prone to loose the important indexes and affect the evaluation results.

To make up for the deficiency of PCA, a new classification method that combines the correlation coefficient [6] with the shortest path theory [7, 8], namely, graph distance classification (GDC), was presented. Combining GDC with PCA, GDC was used to classify the indexes, and the representative elements selected from the classification were analyzed by PCA; this approach can simplify the data and reduce the redundant indexes to simplify the evaluation index system. This method was applied to simplify the evaluation index system for the environmental quality of

decommissioning uranium tailing, which can reduce the computing effort of the information processing and ensure the rationality of the index system.

2. Construction of the Method of an Evaluation Index System

The establishment of an evaluation index system based on graph distance classification and principal components analysis can be divided into two stages: index classification and index screening, as shown in Figure 1. The first stage is the selection of the initial indexes, and then, the indexes are classified and selected by graph distance classification and principal components analysis. Eventually, the simplified index system is determined. In the graph distance classification method, the highly correlated elements are ascribed to the same class. Then, the representative elements that are selected from the classification are analyzed by PCA, which can reduce the information processing workload. According to the screening results of the PCA, the selected indexes and the same class indexes serve as important indexes when constructing the index system, which can avoid the loss of important indexes.

3. Index Classification Model Based on a Graph Distance Classification

3.1. Data Preprocessing and Standardization. Because of subjective and objective reasons, there are some cases in which data are missing in the sample data. Thus, the polynomial interpolation method is adopted. At the same time, the dimension of the unit value indexes often varies, and to eliminate the influence of the dimension, the standardization method of data [5] is often used to make the indexes dimensionless.

Let $x_i (i = 1, 2, \ldots, n)$ be the observed value of a certain index, and let x_i' be the standardized data of the index. The formulas of the standardization are

$$x_i' = \frac{x_i - \overline{x}}{s},$$

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2},$$

and $\overline{x}$ and s represent the mean and standard deviation of the observed values, respectively.

3.2. Determination of the Distance between the Indexes Based on the Correlation Coefficient. Based on the correlation coefficient, the reciprocal of the correlation coefficient is used as the distance between the indexes. The higher the correlation coefficient between the indexes is, the shorter the distance is.

The formula for the correlation coefficient is

$$r_{ij} = \frac{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)(x_{kj} - \overline{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)^2 (x_{kj} - \overline{x}_j)^2}}, \tag{2}$$

$$i = 1, 2, \ldots, n, \ \ j = 1, 2, \ldots, m,$$

where r_{ij} is the correlation coefficient between index i and index j. Here, x_{ki} and x_{kj} represent the observed values of index i and index j of the evaluation object k, respectively, and $\overline{x}_i$ and $\overline{x}_j$ represent the average values of index i and index j, respectively.

Let the reciprocal of the correlation coefficient be the distance between the indexes. In other words, we have

$$d_{ij} = \begin{cases} \dfrac{1}{r_{ij}}, & i \neq j, \\ \\ 0, & i = j. \end{cases} \tag{3}$$

3.3. Calculation of the Shortest Path Based on the Floyd Algorithm. The complete weighted graph G is determined according to the distance between the indexes. The initial index is used as the vertex of graph G, and the shortest path between the vertexes is the shortest distance between the indexes.

Let each index correspond to one vertex v in graph G. Suppose that graph G is a complete simple graph with the vertex set $V = \{v_1, v_2, \ldots, v_k\}$ and edge set $E = \{(v_i, v_j) \mid \forall v_i \in V, \forall v_j \in V, v_i \neq v_j\}$ $(1 \le i, j \le k)$.

Let the determining weights of each edge $(v_i, v_j) (i \neq j)$ be

$$w_{ij} = d_{ij}. \tag{4}$$

The graph G is said to be a complete weighted graph if each edge has determining weights w_{ij}. Calculate the shortest distance between any two vertexes based on the shortest path algorithm. Let d_{ij}' be the shortest distance between vertexes $v_i (i = 1, 2, \ldots, k)$ and $v_j (j = 1, 2, \ldots, k)$. Calculate the shortest path between any two vertexes by using the Floyd algorithm. The steps of the Floyd algorithm are as follows:

(i) Input the weight matrix $W = [w_{ij}]_{n \times n}$ of the complete weighted graph.

(ii) For vertex v_i and vertex v_j, in the adjacency matrix, when we have $e_{ik} + e_{kj} < e_{ij}$, the data must be updated, using $e_{ik} + e_{kj}$ instead of e_{ij}. Repeat the steps until the shortest path is found and determine the shortest path matrix $D[d_{ij}']_{n \times n} (d_{ij}' = \min[e(v_i, v_j)])$. Here, e_{ij} is the distance between vertex v_i and vertex v_j, namely, the determining weights of the edge (v_i, v_j). Additionally, e_{ik} and e_{kj} represent the determining weights of edge (v_i, v_k) and edge (v_k, v_j), respectively.

3.4. Index Classification. Index classification, in essence, is a partition of vertex sets in graph G. In other words, we have
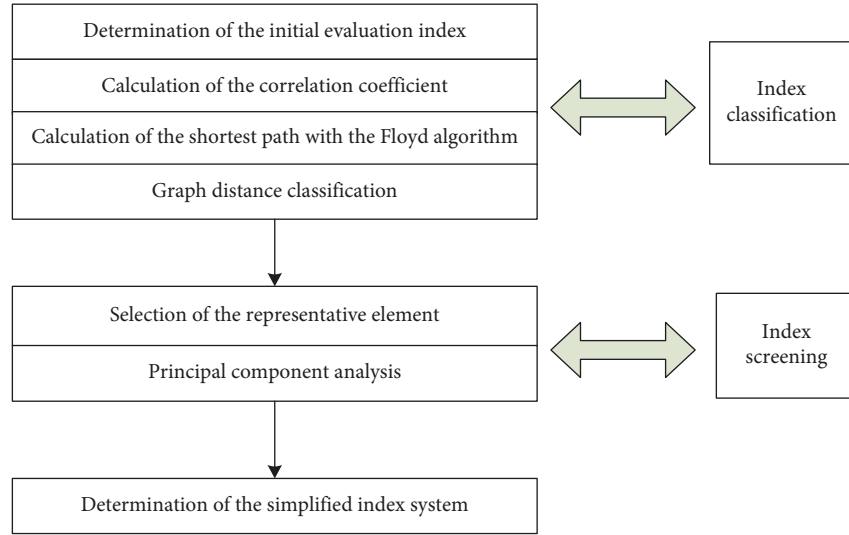
FIGURE 1: Flow chart for the construction of the index system.

$$V = V_1 \cup V_2 \cup \ldots, \cup V_t,$$
$$V_i \cap V_j = \Phi \, (i, j = 1, 2, \ldots, t), \tag{5}$$
$$V_i = \{u \mid \forall u \in V, \forall v \in V \, \& \, d \, (u, v) \le \rho\} \cup \{v\},$$

where ρ is the distance parameter, which is determined according to the actual situation.

4. Model of Index Screening Based on Principal Component Analysis

4.1. Selection of the Representative Element. According to the shortest distance relationship graph of all of the classes, the sum of the distance between one index and the other indexes can be calculated. The smaller the distance is, the closer the relationship is. The minimum distance index will be used as the representative element to be analyzed by the PCA. The representative elements for all of the classes are shown in Table 1.

4.2. Principal Component Analysis. The basic model of principal component analysis [9] is as follows:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p, \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p, \\ \quad \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p, \end{cases} \tag{6}$$

where x_i is the index i, z_j is the principal component j, and l_{ij} is the principal component load of index i in principal component j.

The concrete steps of selecting the index based on PCA are as follows:

(i) Calculate the correlation matrix $R = (r_{ij})_{p \times p} (r_{ii} = 1)$ of the standardized data of the index.

(ii) Calculate the eigenvalues and eigenvectors of the correlation matrix R, variance contribution rate, cumulative contribution rate, and factor load of the principal components.

(iii) Select the principal component and determine the number of principal components according to the eigenvalues or cumulative contribution rate.

(iv) Screen the index according to the absolute value of the factor load of the principal component. The larger the absolute value of the factor load is, the more significant the influence of the index on the evaluation results is. Such an index should be retained.

5. Application Example

The construction of an environmental quality evaluation index system of decommissioning uranium tailing is taken as an example to realize the process of making and modifying the index system. Graph distance classification was used to classify the indexes, and the principal component analysis was used to select the indexes.

5.1. Selection of the Initial Evaluation Index. According to the evaluation purpose and considering the availability and integrity of the existing monitoring data, the construction of an environmental evaluation index system focuses on the pollution angle and selecting the seepage index and radiation index as the primary indexes. The seepage indexes include pH, $\sum \alpha$, $\sum \beta$, U, Ra, ^{230}Th, $\sum$Th, ^{210}Po, ^{210}Pb, Mn, NH$_4$-N, F$^-$, SO$_4{}^{2-}$, NO$_3{}^-$, Zn, and Cd. The radiation indexes include the radon concentration, radon exhalation rate, α aerosol, γ, surface α, and surface β.

5.2. Index Classification. This paper takes the seepage indexes of decommissioning uranium tailing as the research object, and the water monitoring data of six monoliths (A–F) of a decommissioning uranium tailing is targeted as the sample data. The original data originate from the environmental monitoring report of a decommissioning

TABLE 1: Standardized data of the seepage indexes.

Number	Indexes	Dam section					
		A	B	C	D	E	F
1	pH	−0.89367	−1.02547	0.71194	−0.45632	0.10684	1.55668
2	U	−1.25434	0.07486	1.7085	0.2027	−0.67672	−0.055
3	Ra	−0.75681	−0.65197	1.85449	−0.38958	−0.48067	0.42452
4	$\sum \alpha$	−0.98028	0.09968	1.8813	−0.18001	−0.65595	−0.16473
5	$\sum \beta$	−0.57751	1.30033	−0.57767	1.27698	−0.70983	−0.71231
6	^{230}Th	−1.02349	1.45878	−0.88318	0.92995	−0.06296	−0.41911
7	$\sum$Th	−0.97383	1.58721	−0.49956	0.8853	−0.49956	−0.49956
8	^{210}Po	−0.69878	−0.02688	−1.15568	−0.48377	1.45132	0.91379
9	^{210}Pb	−0.69117	1.34045	−0.71211	1.23573	−0.62834	−0.54456
10	NH$_4$-N	−0.63823	−0.31411	2.00895	−0.11289	−0.46543	−0.47828
11	Mn	−0.41453	1.10042	−0.50982	1.36143	−0.36757	−1.16993
12	F$^-$	−1.4763	0.08122	0.10938	−0.56603	0.30652	1.5452
13	SO$_4{}^{2-}$	−0.61445	−0.48732	−0.15679	−0.26697	−0.48732	2.01286
14	NO$_3{}^-$	−0.41206	−0.41348	−0.40372	−0.41126	−0.40069	2.04122
15	Zn	−0.81908	0.26535	−1.00956	−0.74023	1.29041	1.01311
16	Cd	−0.64282	−0.64282	−0.64282	−0.64282	1.42955	1.14172

uranium tailing. The standardized data of the seepage indexes are shown in Table 1.

The standardized data of the seepage indexes were substituted in formula (2) to calculate the correlation coefficient, which is shown in Table 2. The distances between the indexes were the reciprocals of the correlation coefficients.

The vertexes of graph G are determined according to the number of indexes. Let each index be a vertex v of graph G, and the vertex set is $V = \{v_1, v_2, \ldots, v_{16}\}$. Let d_{ij} be the weights w_{ij} of each edge (v_i, v_j), and the weight matrix $W = [w_{ij}]_{n \times n}$ is produced, where W is

$$
\begin{bmatrix}
0 & 2.49 & 1.45 & 2.67 & 1.67 & 2.18 & 2.13 & 3.07 & 1.82 & 3.16 & 1.39 & 1.25 & 1.21 & 1.31 & 2.79 & 1.67 \\
2.49 & 0 & 1.17 & 1.03 & 8.13 & 35.7 & 4.85 & 2.24 & 10.9 & 1.1 & 15.6 & 3.05 & 7.94 & 38.5 & 2.85 & 3.33 \\
1.45 & 1.17 & 0 & 1.12 & 2.58 & 2.15 & 3.25 & 2.91 & 2.4 & 1.12 & 2.27 & 2.51 & 2.97 & 4.76 & 3.77 & 20.8 \\
2.67 & 1.03 & 1.12 & 0 & 125 & 8.78 & 11.9 & 1.99 & 20 & 1.05 & 19.6 & 3.82 & 16.9 & 12.7 & 2.56 & 3.02 \\
1.67 & 8.13 & 2.58 & 125 & 0 & 1.1 & 1.05 & 3.95 & 1.01 & 7.41 & 1.04 & 4.37 & 3.09 & 2.84 & 4.22 & 1.82 \\
2.18 & 35.7 & 2.15 & 8.78 & 1.1 & 0 & 1.04 & 6.49 & 1.06 & 3.43 & 1.17 & 14.7 & 4.98 & 4.81 & 5.43 & 5.68 \\
2.13 & 4.85 & 3.25 & 11.9 & 1.05 & 1.04 & 0 & 12.7 & 1.03 & 9.8 & 1.14 & 62.5 & 4.57 & 4.03 & 55.6 & 2.59 \\
3.07 & 2.24 & 2.91 & 1.99 & 3.95 & 6.49 & 12.7 & 0 & 6.8 & 1.71 & 3.55 & 1.65 & 2.65 & 2.22 & 1.02 & 1.08 \\
1.82 & 10.9 & 2.4 & 20 & 1.01 & 1.06 & 1.03 & 6.8 & 0 & 5.03 & 1.07 & 7.35 & 4.05 & 3.7 & 7.81 & 2.19 \\
3.16 & 1.1 & 1.12 & 1.05 & 7.41 & 3.44 & 9.8 & 1.71 & 5.03 & 0 & 9.43 & 15.6 & 10.5 & 4.31 & 1.93 & 2.75 \\
1.39 & 15.6 & 2.27 & 19.6 & 1.04 & 1.17 & 1.14 & 3.55 & 1.07 & 9.43 & 0 & 2.44 & 1.82 & 1.74 & 3.29 & 1.75 \\
1.25 & 3.05 & 2.51 & 3.83 & 4.37 & 14.7 & 62.5 & 1.65 & 7.35 & 15.6 & 2.44 & 0 & 1.26 & 1.32 & 1.43 & 1.47 \\
1.21 & 7.94 & 2.97 & 16.9 & 3.09 & 4.98 & 4.57 & 2.65 & 4.05 & 10.5 & 1.82 & 1.26 & 0 & 1.01 & 2.34 & 1.94 \\
1.31 & 38.5 & 4.77 & 12.6 & 2.84 & 4.81 & 4.03 & 2.22 & 3.7 & 4.31 & 1.73 & 1.32 & 1.01 & 0 & 2.01 & 1.78 \\
2.79 & 2.85 & 3.77 & 2.56 & 4.22 & 5.43 & 55.6 & 1.02 & 7.81 & 1.93 & 3.29 & 1.43 & 2.34 & 2.01 & 0 & 1.11 \\
1.67 & 3.33 & 20.8 & 3.02 & 1.82 & 5.68 & 2.59 & 1.08 & 2.19 & 2.75 & 1.75 & 1.47 & 1.94 & 1.78 & 1.11 & 0
\end{bmatrix}. \tag{7}
$$

Calculate the shortest path by the Floyd algorithm, using MATLAB programming, which is the shortest distance d'_{ij} between each pair of indexes, as shown in Table 3.

Given $\rho = 1.2$, according to the shortest distance (Table 3), the vertex set is divided into six categories.

TABLE 2: Correlation coefficient matrix. Table 2 is reproduced from Mikoni [10] (under the Creative Commons Attribution License/public domain).

r_{ij}	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}	v_{16}
v_1	1.00	0.40	0.69	0.37	−0.60	−0.46	−0.47	0.33	−0.55	0.32	−0.72	0.80	0.83	0.77	0.36	0.60
v_1	0.40	1.00	0.86	0.97	0.12	0.03	0.21	−0.45	0.09	0.91	0.06	0.33	0.13	−0.03	−0.35	−0.30
v_1	0.69	0.86	1.00	0.90	−0.39	−0.47	−0.31	−0.34	−0.42	0.90	−0.44	0.40	0.34	0.21	−0.27	−0.05
v_1	0.37	0.97	0.90	1.00	−0.01	−0.11	0.08	−0.50	−0.05	0.96	−0.05	0.26	0.06	−0.08	−0.39	−0.33
v_1	−0.60	0.12	−0.39	−0.01	1.00	0.91	0.95	−0.25	0.99	−0.14	0.96	−0.23	−0.32	−0.35	−0.24	−0.55
v_1	−0.46	0.03	−0.47	−0.11	0.91	1.00	0.96	0.15	0.94	−0.29	0.85	0.07	−0.20	−0.21	0.18	−0.18
v_1	−0.47	0.21	−0.31	0.08	0.95	0.96	1.00	−0.08	0.97	−0.10	0.88	0.02	−0.22	−0.25	−0.02	−0.39
v_1	0.33	−0.45	−0.34	−0.50	−0.25	0.15	−0.08	1.00	−0.15	−0.59	−0.28	0.61	0.38	0.45	0.98	0.93
v_1	−0.55	0.09	−0.42	−0.05	0.99	0.94	0.97	−0.15	1.00	−0.20	0.94	−0.14	−0.25	−0.27	−0.13	−0.46
v_1	0.32	0.91	0.90	0.96	−0.14	−0.29	−0.10	−0.59	−0.20	1.00	−0.11	0.06	−0.10	−0.23	−0.52	−0.36
v_1	−0.72	0.06	−0.44	−0.05	0.96	0.85	0.88	−0.28	0.94	−0.11	1.00	−0.41	−0.55	−0.58	−0.30	−0.57
v_1	0.80	0.33	0.40	0.26	−0.23	0.07	0.02	0.61	−0.14	0.06	−0.41	1.00	0.79	0.76	0.70	0.68
v_1	0.83	0.13	0.34	0.06	−0.32	−0.20	−0.22	0.38	−0.25	−0.10	−0.55	0.79	1.00	0.99	0.43	0.52
v_1	0.77	−0.03	0.21	−0.08	−0.35	−0.21	−0.25	0.45	−0.27	−0.23	−0.58	0.76	0.99	1.00	0.50	0.56
v_1	0.36	−0.35	−0.27	−0.39	−0.24	0.18	−0.02	0.98	−0.13	−0.52	−0.30	0.70	0.43	0.50	1.00	0.90
v_1	0.60	−0.30	−0.05	−0.33	−0.55	−0.18	−0.39	0.93	−0.46	−0.36	−0.57	0.68	0.52	0.56	0.90	1.00

TABLE 3: The shortest distance between each pair of indexes.

d'_{ij}	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}	v_{16}
v_1	0.00	2.49	1.45	2.57	1.67	2.18	2.13	2.74	1.82	2.57	1.39	1.25	1.21	1.31	2.68	1.67
v_1	2.49	0.00	1.17	1.03	3.74	3.32	4.35	2.24	3.57	1.10	3.44	3.05	3.70	3.79	2.85	3.32
v_1	1.45	1.17	0.00	1.12	2.58	2.15	3.19	2.83	2.40	1.12	2.27	2.51	2.66	2.76	3.05	3.12
v_1	2.57	1.03	1.12	0.00	3.69	3.27	4.30	2.00	3.52	1.05	3.39	3.62	3.77	3.87	2.56	3.02
v_1	1.67	3.74	2.58	3.69	0.00	1.10	1.05	2.90	1.01	3.69	1.04	2.92	2.87	2.78	2.94	1.82
v_1	2.18	3.32	2.15	3.27	1.10	0.00	1.04	4.01	1.06	3.27	1.17	3.43	3.00	2.91	4.04	2.93
v_1	2.13	4.35	3.19	4.30	1.05	1.04	0.00	3.67	1.03	4.30	1.14	3.37	2.97	2.88	3.71	2.59
v_1	2.74	2.24	2.83	2.00	2.90	4.01	3.67	0.00	3.28	1.71	2.83	1.65	2.65	2.22	1.02	1.08
v_1	1.82	3.57	2.40	3.52	1.01	1.06	1.03	3.28	0.00	3.52	1.07	3.07	2.89	2.81	3.31	2.20
v_1	2.57	1.10	1.12	1.05	3.69	3.27	4.30	1.71	3.52	0.00	3.39	3.36	3.77	3.87	1.93	2.75
v_1	1.39	3.44	2.27	3.39	1.04	1.17	1.14	2.83	1.07	3.39	0.00	2.45	1.82	1.74	2.87	1.75
v_1	1.25	3.05	2.51	3.62	2.92	3.43	3.37	1.65	3.07	3.36	2.45	0.00	1.26	1.32	1.43	1.47
v_1	1.21	3.70	2.66	3.77	2.87	3.00	2.97	2.65	2.89	3.77	1.82	1.26	0.00	1.01	2.34	1.94
v_1	1.31	3.79	2.76	3.87	2.78	2.91	2.88	2.22	2.81	3.87	1.74	1.32	1.01	0.00	2.01	1.78
v_1	2.68	2.85	3.05	2.56	2.94	4.04	3.71	1.02	3.31	1.93	2.87	1.43	2.34	2.01	0.00	1.11
v_1	1.67	3.32	3.12	3.02	1.82	2.93	2.59	1.08	2.20	2.75	1.75	1.47	1.94	1.78	1.11	0.00

$$
\begin{aligned}
V_1 &= \{v_1\}, \\
V_2 &= \{v_2, v_3, v_4, v_{10}\}, \\
V_3 &= \{v_5, v_6, v_7, v_9, v_{11}\}, \\
V_4 &= \{v_8, v_{15}, v_{16}\}, \\
V_5 &= \{v_{12}\}, \\
V_6 &= \{v_{13}, v_{14}\},
\end{aligned}
\tag{8}
$$

which satisfies the conditions $V = V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5 \cup V_6$ and $V_1 \cap V_2 \cap V_3 \cap V_4 \cap V_5 \cap V_6 = \Phi$.

According to the construction method of graph G, six subgraphs ($G_1, G_2, \ldots, G_6$) were obtained, as shown in Figure 2. The result of the index classification is shown in Table 4.

5.3. Index Screening. Taking G_2 as an example, the relation graph of the shortest distance in graph G_2 is shown in Figure 3. The sum of the distances between U and the other indexes is 3.296, and the other indexes (Ra, $\sum \alpha$, and NH_4-N) have the sums 3.3977, 3.1888, and 3.2659, respectively. It can be seen that $\sum \alpha$ is minimized, which indicates that the relation is the closest. This finding is the reason why $\sum \alpha$ was chosen to be the representative element of graph G_2. The other representative element is shown in Table 4.

According to the standardized data of the representative element, principal component analysis can be realized by SPSS [11]. There are two principles for the selection of the principal component: an eigenvalue greater than 1 and over 85% of the cumulative contribution rate. In this paper, the first and second principal components were extracted, as shown in Table 5. In the first principal component, we select the indexes for which the absolute value of the factor load is over 0.9. In the second principal component, we select the index with the largest absolute value of the factor load. Table 6 shows that pH and F^- have a higher load in the first principal component and $\sum \alpha$ has a higher load in the second principal component.
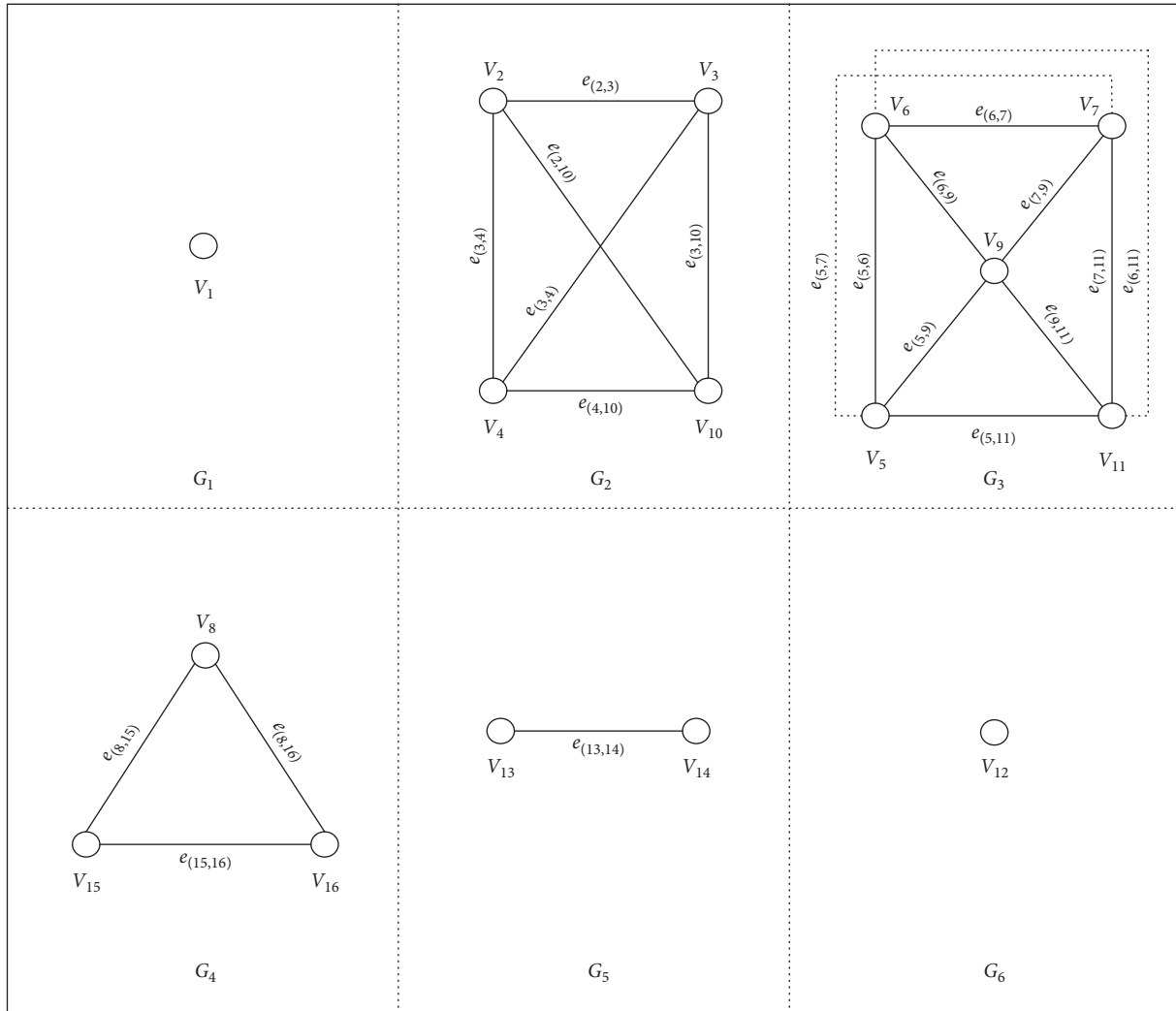
Figure 2: Index classification.

Table 4: The results of the index classification.

Category	Representative element	Index elements
G_1	pH	pH
G_2	$\sum \alpha$	U, Ra, $\sum \alpha$, NH_4-N
G_3	^{210}Pb	$\sum \beta$, ^{230}Th, $\sum Th$, ^{210}Pb, Mn
G_4	^{210}Po	^{210}Po, Zn, Cd
G_5	F^-	F^-
G_6	NO_3^-	NO_3^-, SO_4^{2-}

5.4. Determination of the Simplified Index System.
According to the screening results of the PCA, the selected indexes have a significant impact on the evaluation results and are applied as the important indexes to construct the index system. Because the same types of indexes exist relativity often in the index classification, the same types of indexes also have a significant impact and were also used as important indexes.

The seepage indexes were classified into 6 classes through the index classification, and the representative elements were selected. The screening results of the PCA are shown in
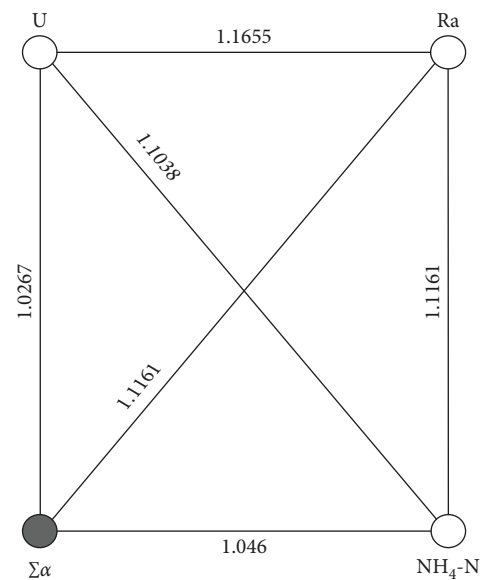

Figure 3: The relation graph of the shortest distance (G_2).

TABLE 5: The eigenvalue and variance contribution of the principal component.

Principal component	Composition 1	Composition 2
Characteristic value	3.053	1.529
Contribution rate	49.925%	26.43%
Cumulative contribution	49.925%	76.356%

TABLE 6: The factor loading matrix of the principal component.

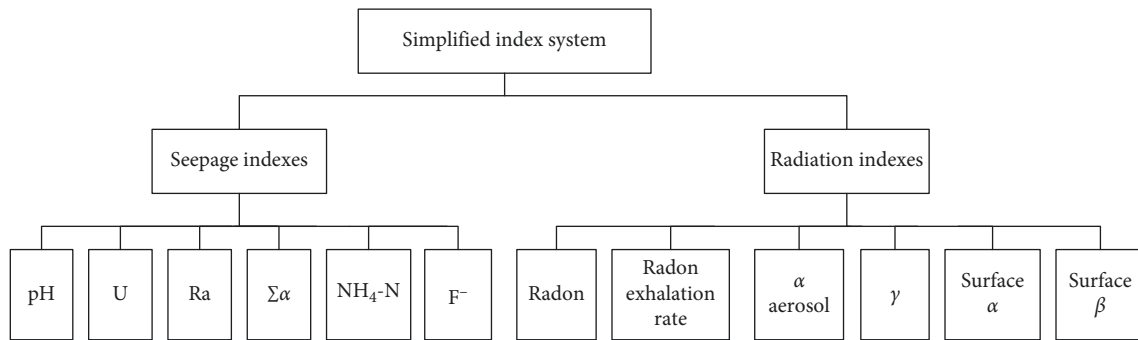| Index	Principal component		Screening results
	Composition 1	Composition 2	
pH	0.969	−0.138	Retain
$\sum \alpha$	0.293	−0.904	Retain
^{210}Pb	−0.495	0.088	Delete
^{210}Po	0.466	0.802	Delete
F$^-$	0.901	0.137	Retain
NO$_3$$^-$	0.834	0.282	Delete

FIGURE 4: The sifted index system for environmental quality assessment in decommissioning uranium tailing.

Table 6. The retained indexes were used as the important indexes of the evaluation index system, and the same class indexes as the retained indexes were also used as important indexes. The simplified index system of environmental quality assessment of decommissioning uranium tailing is shown in Figure 4.

6. Conclusions

(1) The seepage indexes were classified into 6 classes by using the graph distance classification: $G_1 = \{pH\}$, $G_2 = \{U, Ra, \sum \alpha, NH_4\text{-}N\}$, $G_3 = \{\sum \beta, ^{230}Th, \sum Th, ^{210}Pb, Mn\}$, $G_4 = \{^{210}Po, Zn, Cd\}$, $G_5 = \{F^-\}$, and $G_4 = \{NO_3^-, SO_4^{2-}\}$. The representative elements were selected, including pH, $\sum \alpha$, ^{210}Pb, ^{210}Po, F^-, and NO_3^-.

(2) On the basis of the index classification, the representative elements were analyzed by PCA, selecting the retained indexes, including pH, $\sum \alpha$, and F^-. Considering the classification, there were 6 indexes that remained in the final evaluation index system, including pH, U, Ra, $\sum \alpha$, NH_4-N, and F^-, which is consistent with the analysis results on the main pollution sources of seepage in the environmental monitoring report.

(3) In the graph distance classification method, the highly correlated elements were divided into one class. Then, the representative elements selected from the classification were analyzed by PCA, which can reduce the number of indexes when using PCA and the information processing workload and avoid repeated information analysis. According to the screening results of the PCA, the selected indexes were used as the important indexes with which to construct the index system, which indicates that the selected indexes had a significant impact on the evaluation results. At the same time, the same class indexes as the retained indexes also had a significant impact and were also used as important indexes, from which the loss of important indexes could be prevented. The method combined with graph distance classification and principal component analysis can make up for the deficiencies of PCA, which is suitable for multi-indexes and complex systems and favors the further application of PCA on the construction of index systems.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Zhang and Z. Qiu-hong, "An economic indicator screening method based on fundamental principle of principal components analysis," *Journal of Shandong University of Finance*, vol. 124, no. 2, pp. 52–61, 2013.

[2] J. Zhang, "Pipeline risk assessment method based on principle component-clustering analysis," *Oil-Gas Storage and Transportation*, vol. 33, no. 2, pp. 139–143, 2014.

[3] Y. Nazzal, F. K. Zaidi, I. Ahmed et al., "The combination of principal component analysis and geostatistics as a technique in assessment of groUndWater hydrochemistry in arid environment," *Research Communications*, vol. 108, no. 6, pp. 1138–1145, 2015.

[4] X.-l. Wang, Z.-b. Wei, P. Shi-tao et al., "Comprehensive assessment model of liquid pipeline leakage consequences based on principle component analysis," *Journal of Safety Science and Technology*, vol. 10, no. 5, pp. 84–89, 2014.

[5] R.-y. Zhou, A. Zhong, R. Jin-zhou et al., "An accident forecasting method of ANN based on PCA and its application," *China Safety Science Journal*, vol. 23, no. 7, pp. 55–60, 2013.

[6] C. Guo-tai, T.-t. Cao, and K. Zhang, "The establishment of human all-around development evaluation indicators system based on correlation-principle component analysis," *Systems Engineering-Theory and Practice*, vol. 32, no. 1, pp. 111–119, 2012.

[7] F. Fu-gui, "Study on the algorithm and applications in graph theory," *Computer and Digital Engineering*, vol. 40, no. 2, pp. 115–117, 2012.

[8] Y.-j. MAO, "Floyd algorithm and MATLAB program realization of shortest path problem," *Journal of Hebei North University (Natural Science Edition)*, vol. 29, no. 5, pp. 13-14, 2013.

[9] L. Jin-tao, X.-b. Li, F.-q. Gong et al., "Recognizing of mine water inrush sources based on principal components analysis and Fisher discrimination analysis method," *China Safety Science Journal*, vol. 22, no. 7, pp. 109–115, 2012.

[10] S. V. Mikoni, "Neural network approach to the formation models of multiattribute utility," *International Journal Information Models & Analyses*, vol. 3, no. 1, pp. 3–9, 2014.

[11] R.-l. Huang, *Data Statistical Analysis: SPSS Principles and Applications*, Higher Education Press, Vol. 7, Beijing, China, 2010.

The Scientific World Journal

Journal of Applied Chemistry

Journal of Nanomaterials

Scientifica

International Journal of Polymer Science

Advances in Chemistry

Advances in Physical Chemistry

International Journal of Analytical Chemistry

Hindawi

Submit your manuscripts at
www.hindawi.com

Advances in Condensed Matter Physics

Journal of Chemistry

International Journal of Biomaterials

Advances in High Energy Physics

Journal of Engineering

Journal of Nanotechnology

International Journal of Corrosion

Journal of Materials

BioMed Research International

Advances in Tribology

Advances in Materials Science and Engineering