

Adsorption Analysis user guide

This document is a guide on how to install and use the Adsorption-Analysis tool.

Table of contents

How the program works

Installation

Execution

Expected XYZ file format

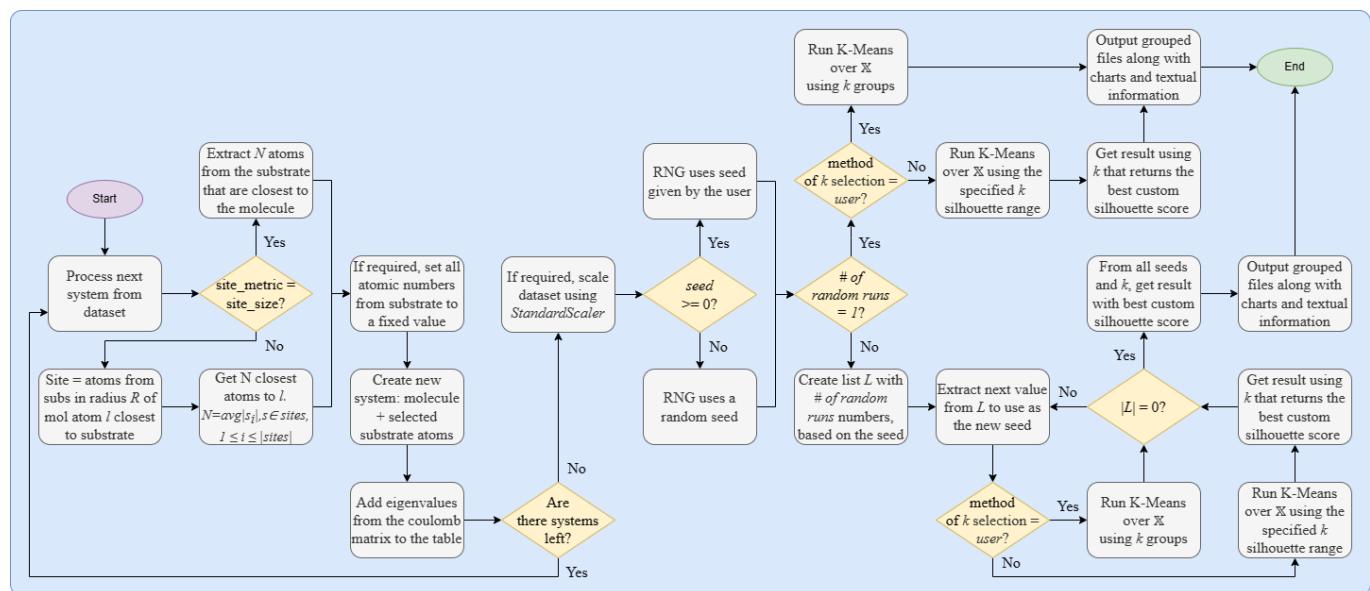
The input file

Output

Examples

Textual output

How the program works



First, the program reads all XYZ files provided by the user, then it applies a custom "local" version of the Coulomb Matrix (through `site_metric` and other settings) to each data entry. After some pre-processing, the data is sent to K-Means clustering algorithm, which clusters it according to the present adsorption modes. The amount of clusters can be specified by the user, or automatically picked using a custom silhouette metric. The program can also run many iterations of the clustering process to make sure a good local minimum is found by K-Means. Finally, Adsorption-Analysis outputs the clustered files, as well as textual information and charts.

Installation

This tool uses Anaconda to create a virtual environment in which all the dependencies are downloaded and contained.

First, clone or download this repository and install [Anaconda](#) for your operating system. Then, open your terminal emulator in the root folder of this repository and run:

```
$ conda env create -f environment.yml
```

This command will create a new Anaconda virtual environment with all the necessary dependencies to run Adsorption-Analysis.

Execution

Instructions

To run the software, the first step is to activate `adsp` Anaconda environment:

```
$ conda activate adsp
```

Then, run the program using Python:

```
$ python3 adsorption_analysis.py
```

Since the program is being ran without arguments, the following message should appear:

```
Usage: python3 adsorption_analysis.py input_file
```

This message is informing that an input file is needed to proceed with the execution.

Run example

An example dataset and input file have been provided. To run the program for these items, execute the following command (inside `adsp` conda environment):

```
$ python3 adsorption_analysis.py example_input.json
```

Expected XYZ file format

The format expected for the XYZ files is the following:

```

N_atoms [energy]

atomic_number_0      x_coord_0      y_coord_0      z_coord_0
atomic_number_1      x_coord_1      y_coord_1      z_coord_1
...
...                  ...
atomic_number_N      x_coord_N      y_coord_N      z_coord_N

```

where `[energy]` is only required if the parameter `use_energy` from the input file (presented below) is set as `true`.

The input file

The input file should be formatted as JavaScript Object Notation ([JSON](#)). This means that the user needs to follow certain rules to generate a valid file that the program understands.

Data types

The data types expected by Adsorption-Analysis in the input file are:

- `integer` is an integer number, e.g. `-302`;
- `float` is a floating point number, e.g. `3.141592`;
- `string` is a string of characters, e.g. `"abc def ghi"`;
- `boolean` is either `true` or `false`;
- `list` is a list of elements, e.g. `[2, 15, 3]`.

Parameters

The parameters necessary in the JSON are the following:

Parameter	Type	Description
<code>seed</code>	integer	Seed for the random number generator, where <code>-1</code> is random and any other value is a fixed/reproducible seed
<code>system_name</code>	string	Name of the system being clustered. This is used in charts and textual outputs
<code>input_folder_path</code>	string	Path to the XYZ files that will be clustered
<code>output_folder_path</code>	string	Path to save the clustered files, as well as charts and textual data about the clustering process
<code>method_of_k_selection</code>	string	How the value of K will be selected. Can be either user or silhouette
<code>number_k_of_clusters</code>	integer	Only used if <code>method_of_k_selection</code> is user . This defines the amount of clusters to be generated

Parameter	Type	Description
<code>silhouette_range</code>	list of 3 integers	Only used if <code>method_of_k_selection</code> is silhouette . This should be a list with 3 values: starting K , ending K and step. The program clusters with the informed range of K and automatically chooses a value based on the <i>Silhouette Criterion</i>
<code>number_of_random_runs</code>	integer	Execute the clustering procedure this amount of times and get the best outcome
<code>molecule_indices</code>	list of N integers	List containing the indices (from XYZ files) of the atoms that compose the molecule
<code>site_metric</code>	string	How the site atoms will be selected. Can be either site_size or site_radius
<code>site_size</code>	integer	Only used if <code>site_metric</code> is site_size . This informs how many of the closest atoms from the substrate to the molecule should be used to compute the <i>Coulomb Matrix</i>
<code>site_radius</code>	float	Only used if <code>site_metric</code> is site_radius . This informs the radius from the closest atom from the molecule to the substrate, where atoms within it are counted and the average amount of atoms is uatoms within it are counted and the average amount of atoms is used to pick this same amount from all the systems, to use in the <i>Coulomb Matrix</i> calculation
<code>fixed_substrate_atomic_number</code>	integer	Substitute every atom by the one represented by this atomic number
<code>z_exp</code>	float	Z exponent of the <i>Coulomb Matrix</i>
<code>d_exp</code>	float	D exponent of the <i>Coulomb Matrix</i>
<code>use_energy</code>	boolean	If true , use the minimum energy elements in each cluster as representatives. If false , use the closest elements to the centroids
<code>scale_dataset</code>	boolean	If true , scale the dataset in order to center the features in an average of 0 with a standard deviation of 1, using scikit-learn's <i>StandardScaler</i> . If false , do not scale the dataset
<code>projection_numbers</code>	boolean	If true , show the id of each element in the projection. If false , do not show the id of each element
<code>projection_repr</code>	boolean	If true , show the representatives in the projection. If false , do not show the representatives in the projection

Parameter	Type	Description
projection_tsne	boolean	If true , also generate <i>t-SNE</i> projection (which is more computationally expensive). If false , generate only PCA

Build the JSON

Once you understood the necessary parameters and their respective data types, build an input file as the example below:

```
{
  "seed": -1,
  "system_name": "SystemName",
  "input_folder_path": "path/to/input",
  "output_folder_path": "path/to/output",
  "method_of_k_selection": "silhouette",
  "number_k_of_clusters": 10,
  "silhouette_range": [2, 10, 2],
  "number_of_random_runs": 10,
  "molecule_indices": [0, 1, 2, 3],
  "site_metric": "site_size",
  "site_size": 4,
  "site_radius": 10.0,
  "fixed_substrate_atomic_number": 40,
  "z_exp": 2.4,
  "d_exp": 1.0,
  "use_energy": false,
  "scale_dataset": true,
  "projection_numbers": false,
  "projection_repr": false,
  "projection_tsne": true
}
```

Notice that, even though `number_k_of_clusters` is not being used (since `method_of_k_selection` is set as `silhouette`), it still must be present and can assume any valid value. The same applies for `silhouette_range` if `method_k_of_selection` is set as `user`, as well as `site_size` and `site_radius`, regardless of the chosen `site_metric`.

Output

After all the operations are done, the program outputs the following files:

File	Description
NMI	Normalized Mutual Information chart
Output	Textual output
PCA	Principal Component Analysis chart
TSNE	t-SNE chart (if enabled)

Examples

The following sections showcase numerous examples utilizing the provided data (found in the data folder). Some examples involve data from multiple runs, where only a single parameter in the input varies. These fields are marked with `<*>` and wouldn't function in the actual program, as users can only input one value at a time.

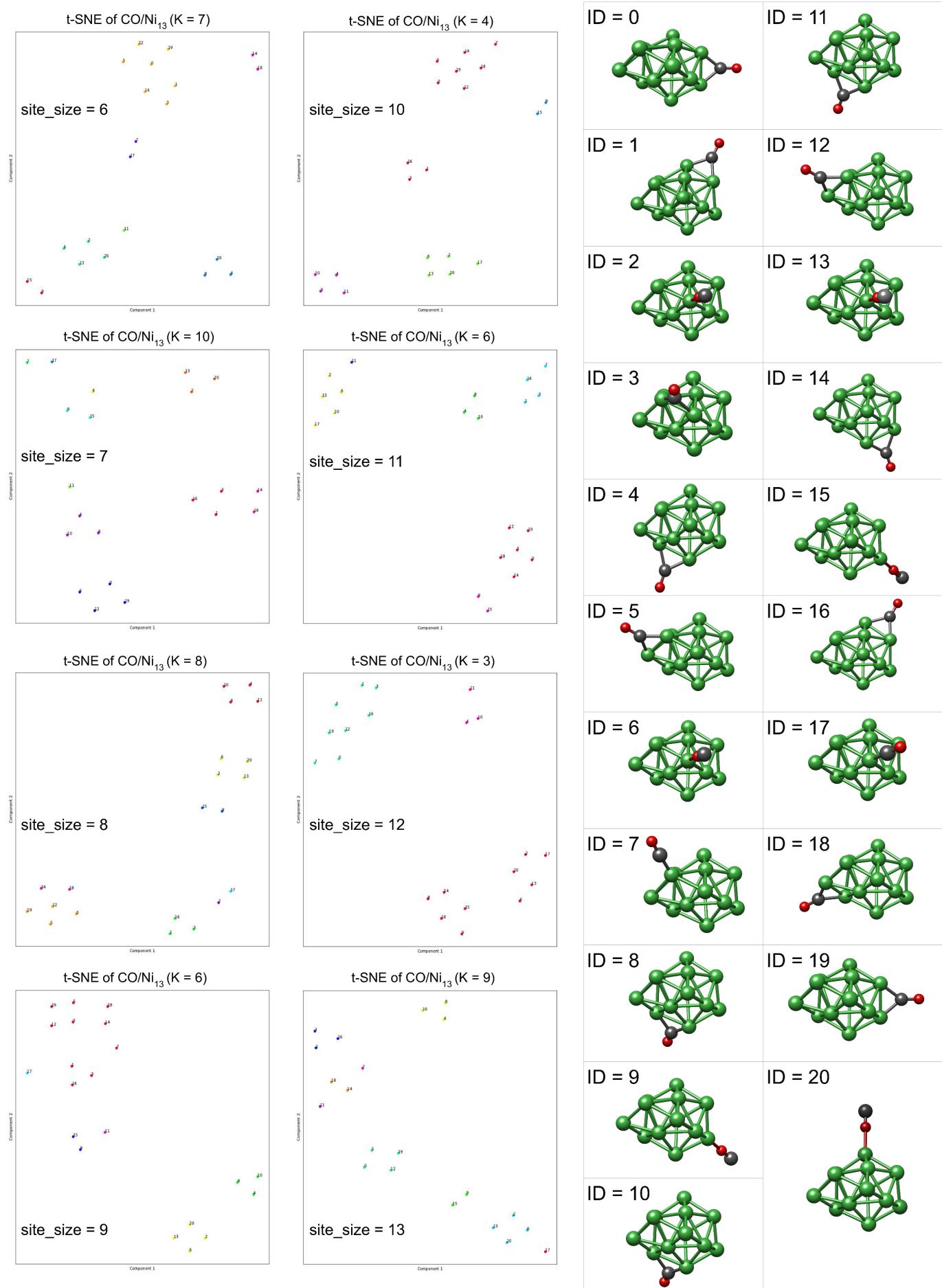
Example 1

This example showcases multiple runs, using different values for `site_size`, which is useful to analyze how smaller or bigger sites help or hinder the clustering quality.

JSON

```
{  
    "seed": -1,  
    "system_name": "CO-Ni13",  
    "input_folder_path": "data/Ni13-CO_structures_pbe/",  
    "output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
    "method_of_k_selection": "silhouette",  
    "number_k_of_clusters": 2,  
    "silhouette_range": [2, 20, 1],  
    "number_of_random_runs": 50,  
    "molecule_indices": [0, 1],  
    "site_metric": "site_size",  
    "<*> \"site_size\": 6, 7, 8, 9, 10, 11, 12, 13,  
    "site_radius": 0,  
    "fixed_substrate_atomic_number": 0,  
    "z_exp": 2.4,  
    "d_exp": 1.0,  
    "use_energy": false,  
    "scale_dataset": true,  
    "projection_numbers": true,  
    "projection_repr": true,  
    "projection_tsne": true  
}
```

Charts



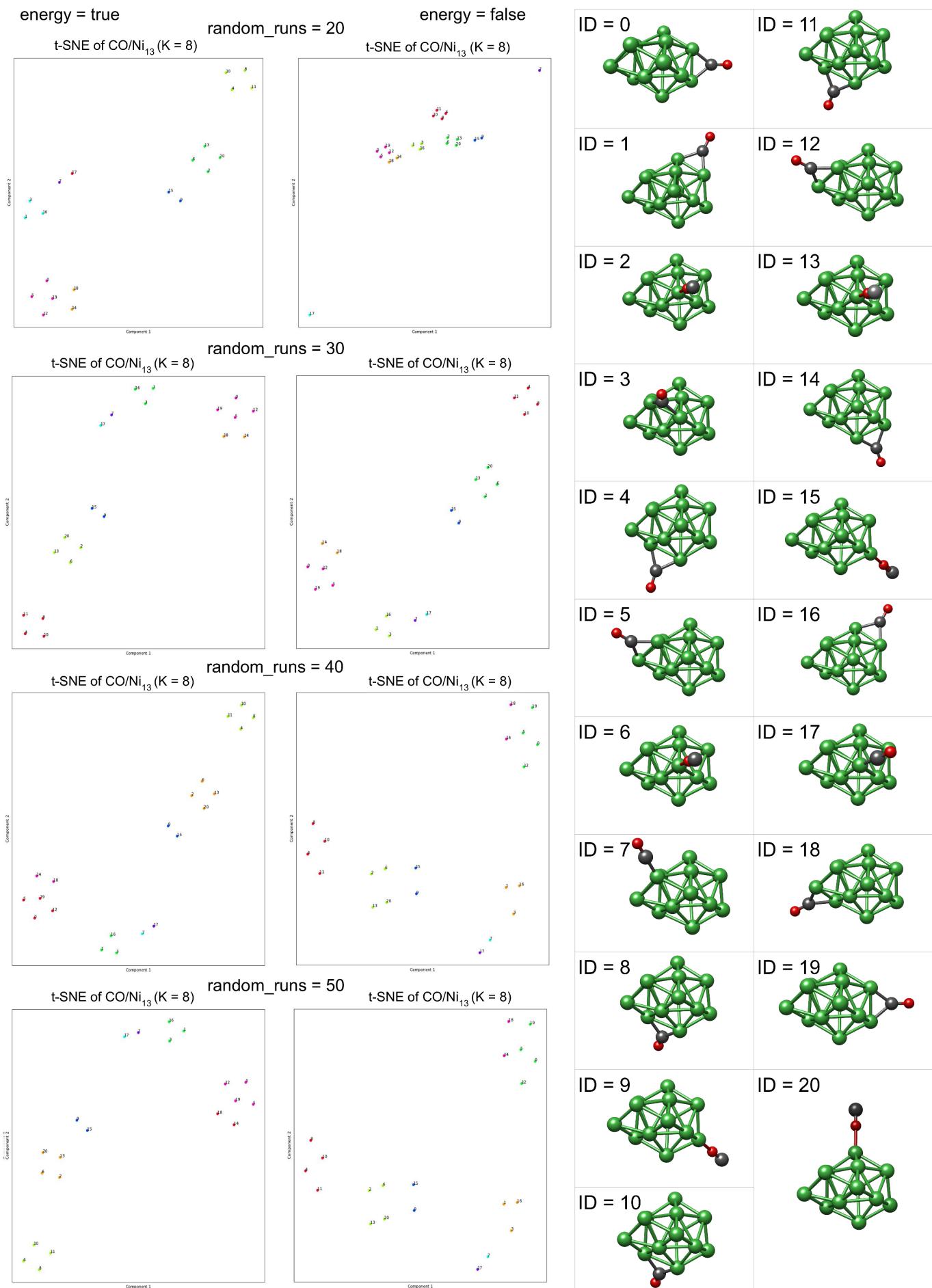
Example 2

This example varies the number of random executions, as well as if the representatives should be the elements with lowest energy, or those closest to the centroids. The largest the amount of random runs is, the most likely K-Means is to find a very good local optimum. The choice between using energy or centroids to obtain the representatives should be decided through specific domain knowledge, since some applications benefit from having configurations with lower energies.

JSON

```
{  
    "seed": -1,  
    "system_name": "CO-Ni13",  
    "input_folder_path": "data/Ni13-CO_structures_pbe/",  
    "output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
    "method_of_k_selection": "silhouette",  
    "number_k_of_clusters": 2,  
    "silhouette_range": [2, 20, 1],  
    <*> "number_of_random_runs": 20, 30, 40, 50,  
    "molecule_indices": [0, 1],  
    "site_metric": "site_size",  
    "site_size": 8,  
    "site_radius": 0,  
    "fixed_substrate_atomic_number": 0,  
    "z_exp": 2.4,  
    "d_exp": 1.0,  
    <*> "use_energy": false, true,  
    "scale_dataset": true,  
    "projection_numbers": true,  
    "projection_repr": true,  
    "projection_tsne": true  
}
```

Charts



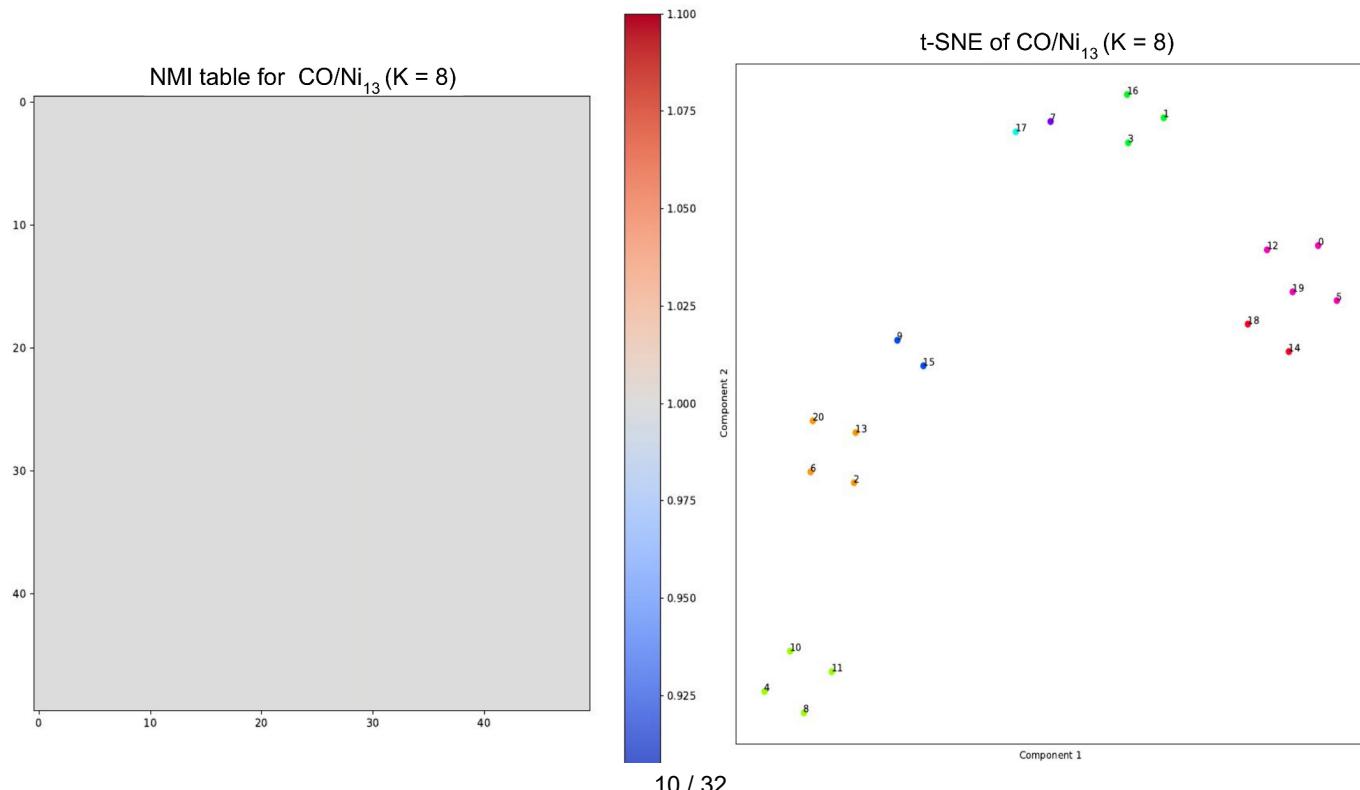
Example 3

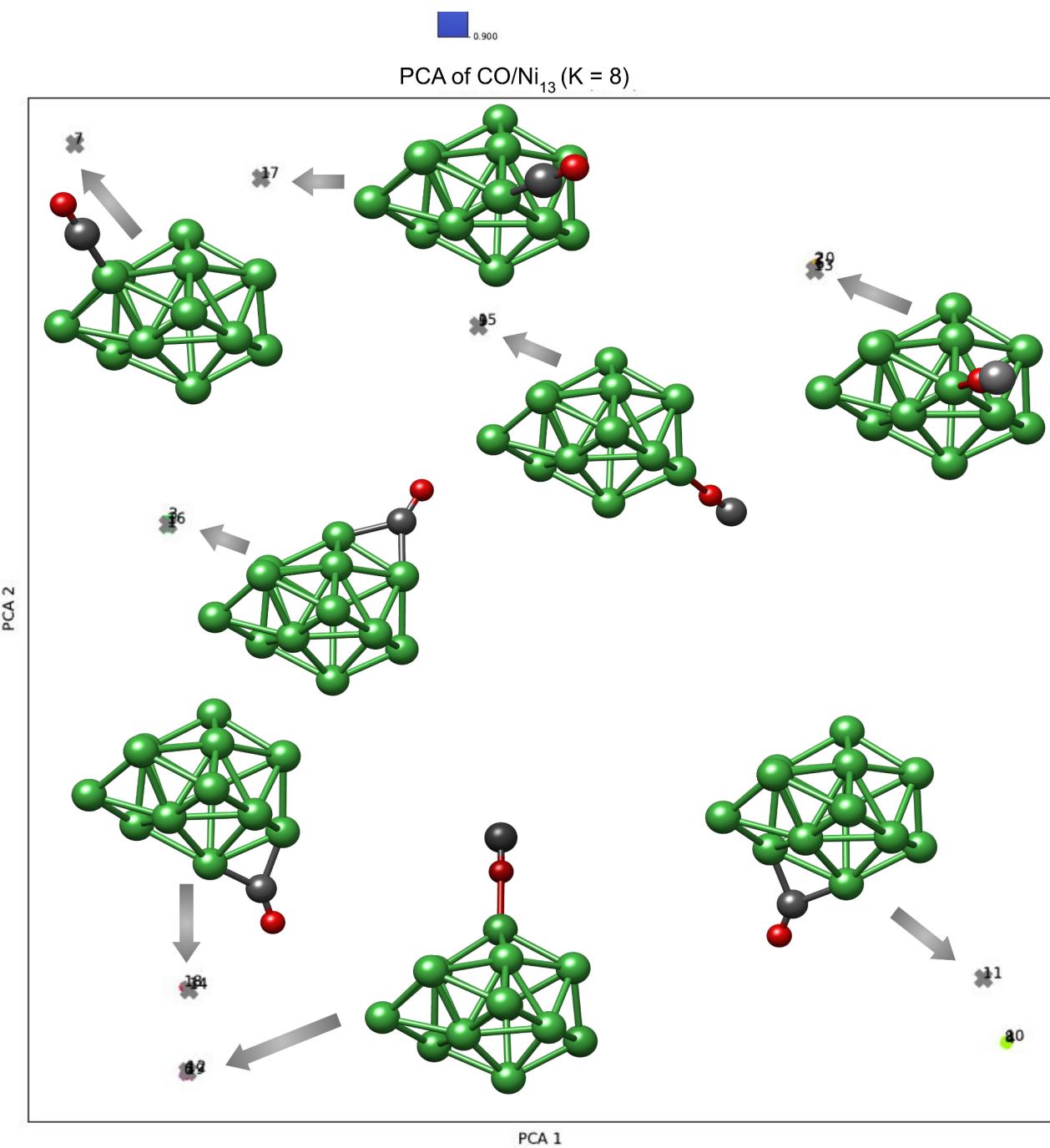
Here, you can observe the outcome of a highly reliable clustering process, as indicated by the *Normalized Mutual Information* (NMI) chart appearing completely gray. This signifies that across multiple random executions, K-Means consistently clustered the data in precisely the same manner.

JSON

```
{  
    "seed": -1,  
    "system_name": "CO-Ni13",  
    "input_folder_path": "data/Ni13-CO_structures_pbe/",  
    "output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
    "method_of_k_selection": "silhouette",  
    "number_k_of_clusters": 2,  
    "silhouette_range": [2, 20, 1],  
    "number_of_random_runs": 50,  
    "molecule_indices": [0, 1],  
    "site_metric": "site_size",  
    "site_size": 8,  
    "site_radius": 0,  
    "fixed_substrate_atomic_number": 0,  
    "z_exp": 2.4,  
    "d_exp": 1.0,  
    "use_energy": true,  
    "scale_dataset": true,  
    "projection_numbers": true,  
    "projection_repr": true,  
    "projection_tsne": true  
}
```

Charts





Example 4

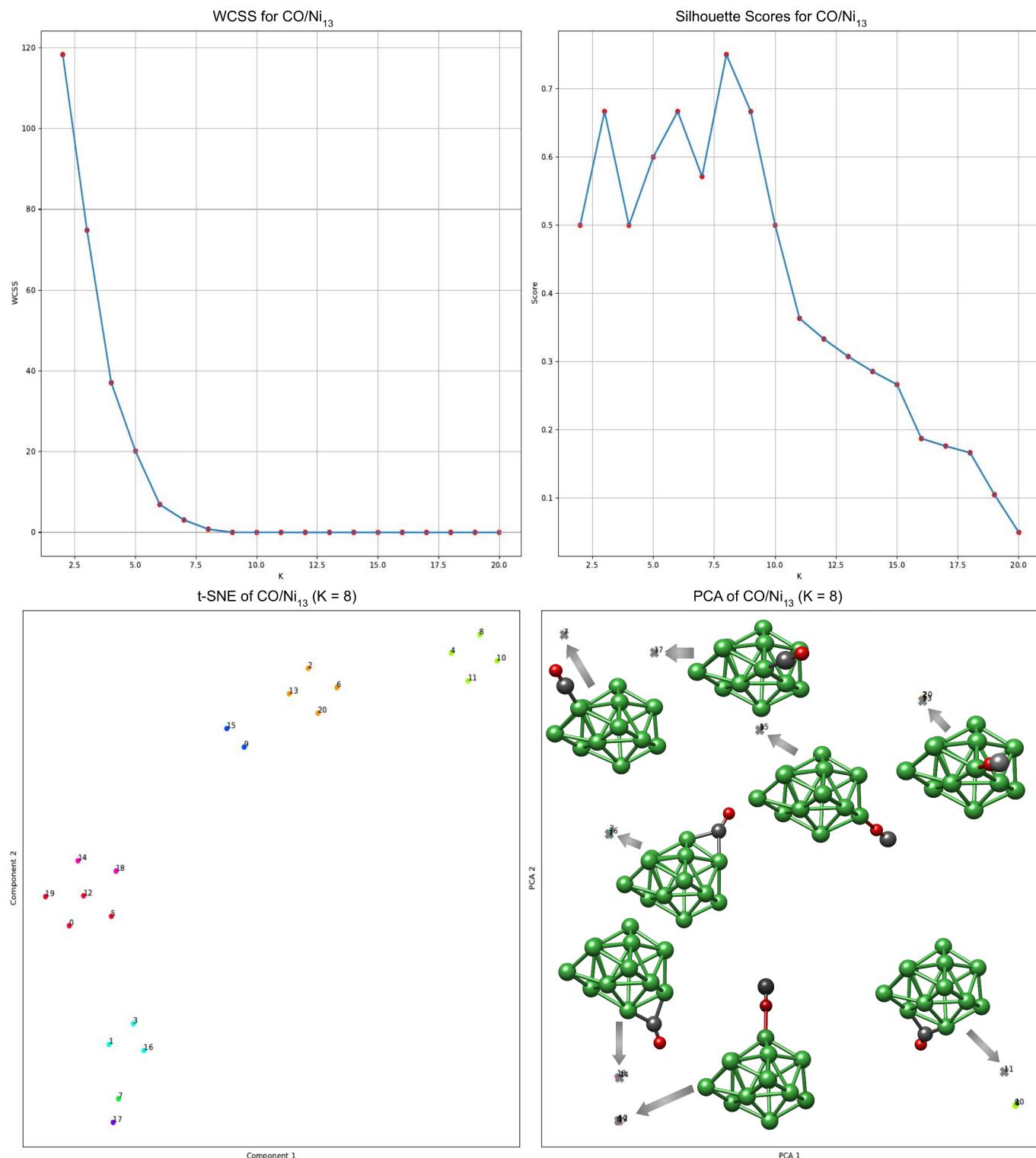
This example employs `user` rather than `silhouette` to determine the number `K` of clusters for K-Means. Here, the user specified that K-Means should generate 8 clusters. As evident from the silhouette score chart, this selection exhibits the highest custom silhouette score among all other alternatives. While the program would have defaulted to `K=8`, users have the flexibility to override automatic selection by manually specifying a value, recognizing that the provided silhouette criterion may not be the best choice for that specific situation.

JSON

```
{
  "seed": -1,
  "system_name": "CO-Ni13",
```

```
"input_folder_path": "data/Ni13-CO_structures_pbe/",
"output_folder_path": "data/Ni13-CO_structures_pbe/output/",
"method_of_k_selection": "user",
"number_k_of_clusters": 8,
"silhouette_range": [2, 20, 1],
"number_of_random_runs": 1,
"molecule_indices": [0, 1],
"site_metric": "site_size",
"site_size": 8,
"site_radius": 0,
"fixed_substrate_atomic_number": 0,
"z_exp": 2.4,
"d_exp": 1.0,
"use_energy": true,
"scale_dataset": true,
"projection_numbers": true,
"projection_repr": true,
"projection_tsne": true
}
```

Charts



Example 5

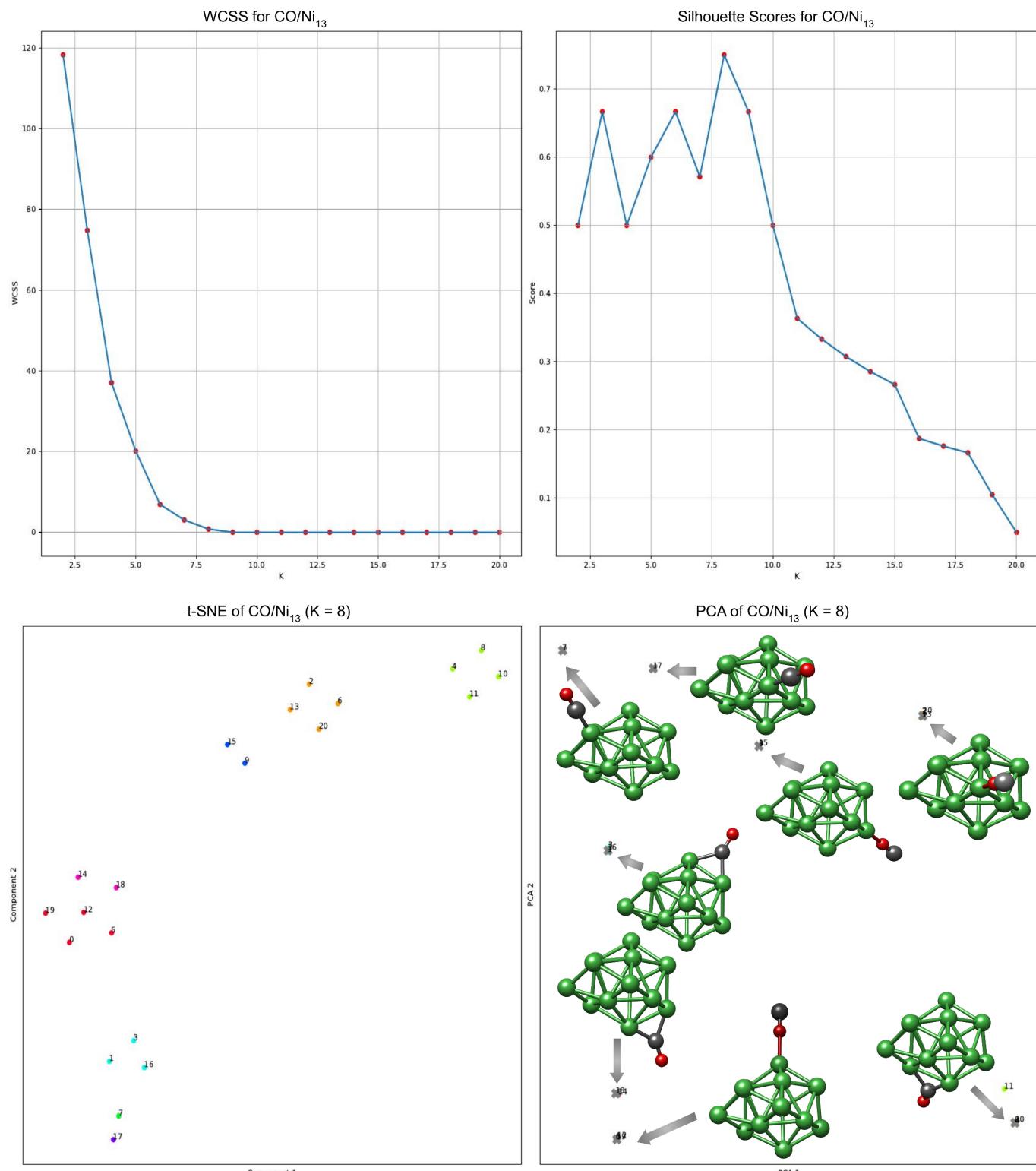
Here, it's employed a very similar approach to the example above, with a slight variation: instead of proximity to centroids, the lowest energy elements are elected representatives of their respective clusters. For instance, **ID 11** is a representative in **Example 4**, but not here.

JSON

```
{
  "seed": -1,
  "system_name": "CO-Ni13",
```

```
"input_folder_path": "data/Ni13-CO_structures_pbe/",
"output_folder_path": "data/Ni13-CO_structures_pbe/output/",
"method_of_k_selection": "user",
"number_k_of_clusters": 8,
"silhouette_range": [2, 20, 1],
"number_of_random_runs": 1,
"molecule_indices": [0, 1],
"site_metric": "site_size",
"site_size": 8,
"site_radius": 0,
"fixed_substrate_atomic_number": 0,
"z_exp": 2.4,
"d_exp": 1.0,
"use_energy": true,
"scale_dataset": true,
"projection_numbers": true,
"projection_repr": true,
"projection_tsne": true
}
```

Charts



Example 6

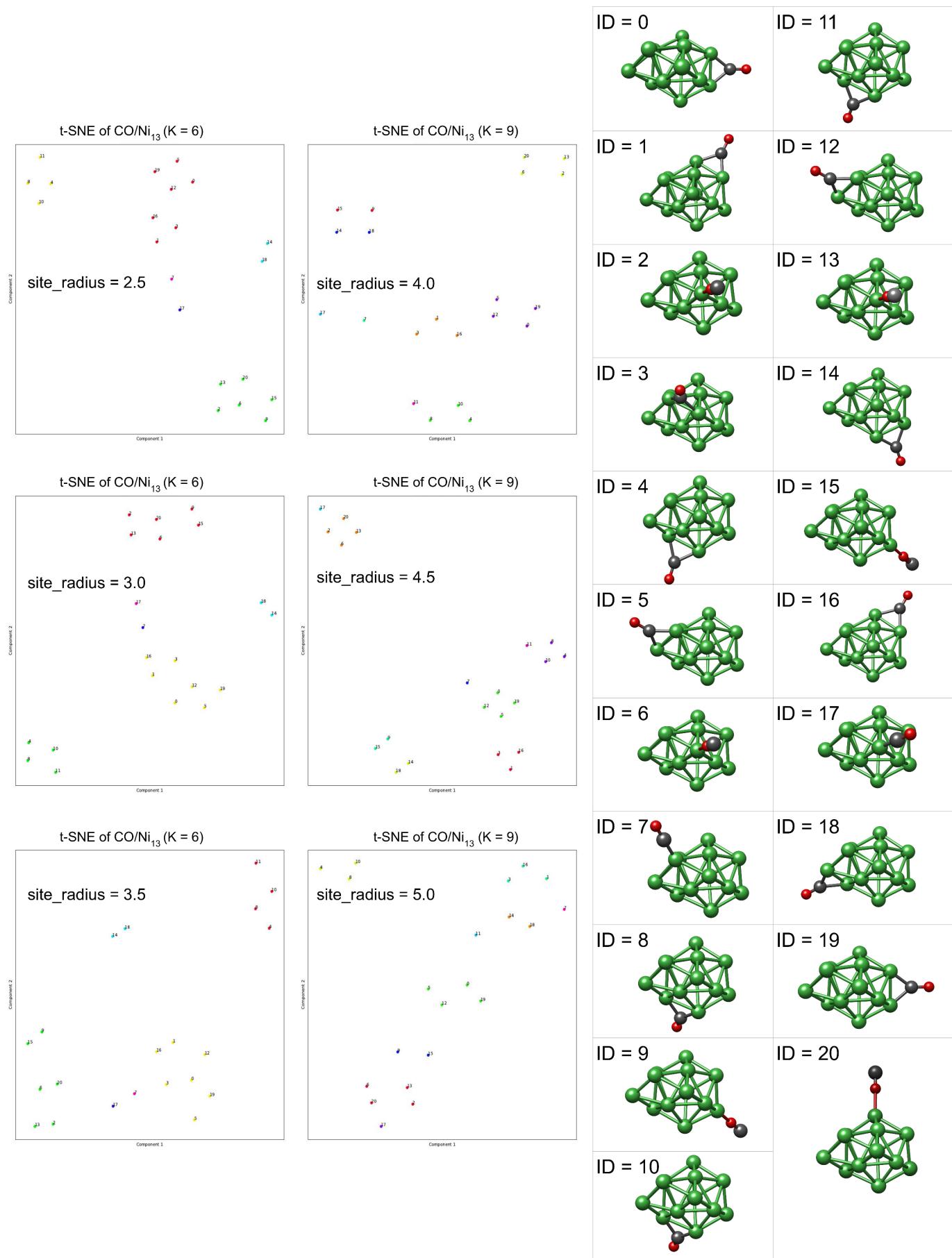
This example is the same as [Example 1](#), but using and varying `site_radius` instead of `site_size`.

JSON

```
{
  "seed": -1,
  "system_name": "CO-Ni13",
  "input_folder_path": "data/Ni13-CO_structures_pbe/",
```

```
"output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
"method_of_k_selection": "silhouette",  
"number_k_of_clusters": 2,  
"silhouette_range": [2, 20, 1],  
"number_of_random_runs": 50,  
"molecule_indices": [0, 1],  
"site_metric": "site_radius",  
"site_size": 0,  
<*> "site_radius": 2.5, 3.0, 3.5, 4.0, 4.5, 5.0,  
"fixed_substrate_atomic_number": 0,  
"z_exp": 2.4,  
"d_exp": 1.0,  
"use_energy": true,  
"scale_dataset": true,  
"projection_numbers": true,  
"projection_repr": true,  
"projection_tsne": true  
}
```

Charts



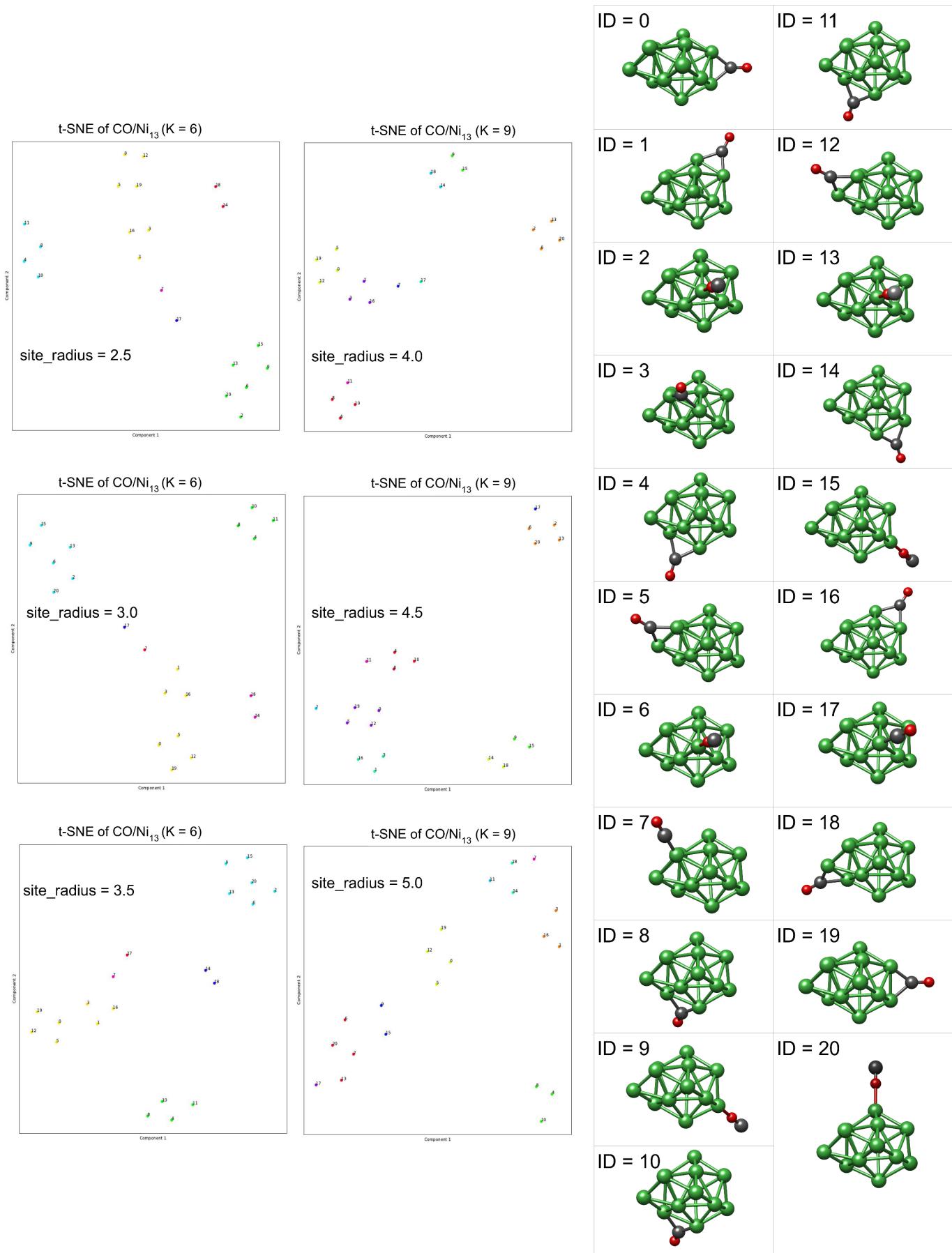
Example 7

Here we have the same parameters as **Example 6**, but using the energy to obtain the representative samples.

JSON

```
{  
    "seed": -1,  
    "system_name": "CO-Ni13",  
    "input_folder_path": "data/Ni13-CO_structures_pbe/",  
    "output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
    "method_of_k_selection": "silhouette",  
    "number_k_of_clusters": 2,  
    "silhouette_range": [2, 20, 1],  
    "number_of_random_runs": 50,  
    "molecule_indices": [0, 1],  
    "site_metric": "site_radius",  
    "site_size": 0,  
    "<*> site_radius": 2.5, 3.0, 3.5, 4.0, 4.5, 5.0,  
    "fixed_substrate_atomic_number": 0,  
    "z_exp": 2.4,  
    "d_exp": 1.0,  
    "use_energy": true,  
    "scale_dataset": true,  
    "projection_numbers": true,  
    "projection_repr": true,  
    "projection_tsne": true  
}
```

Charts



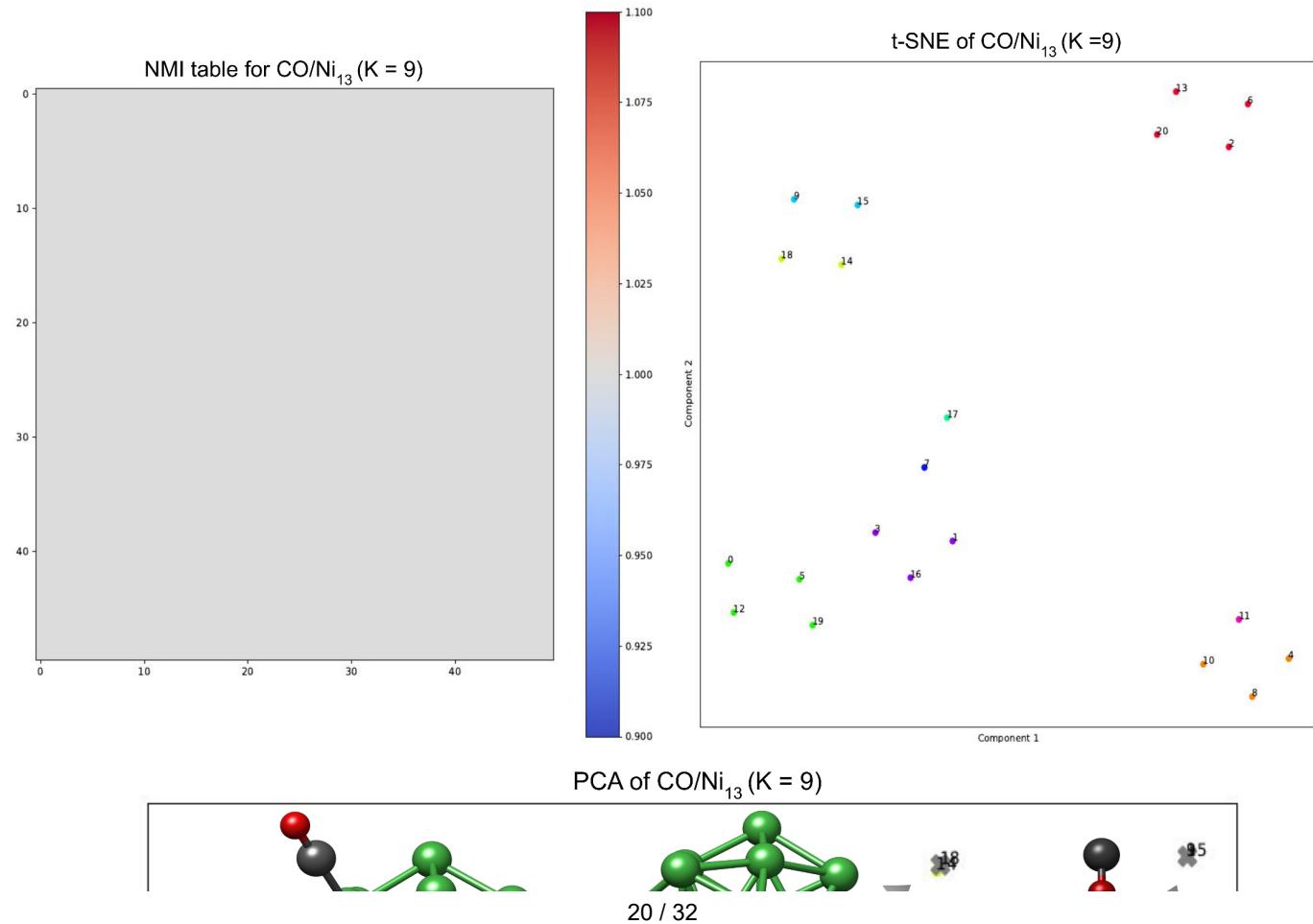
Example 8

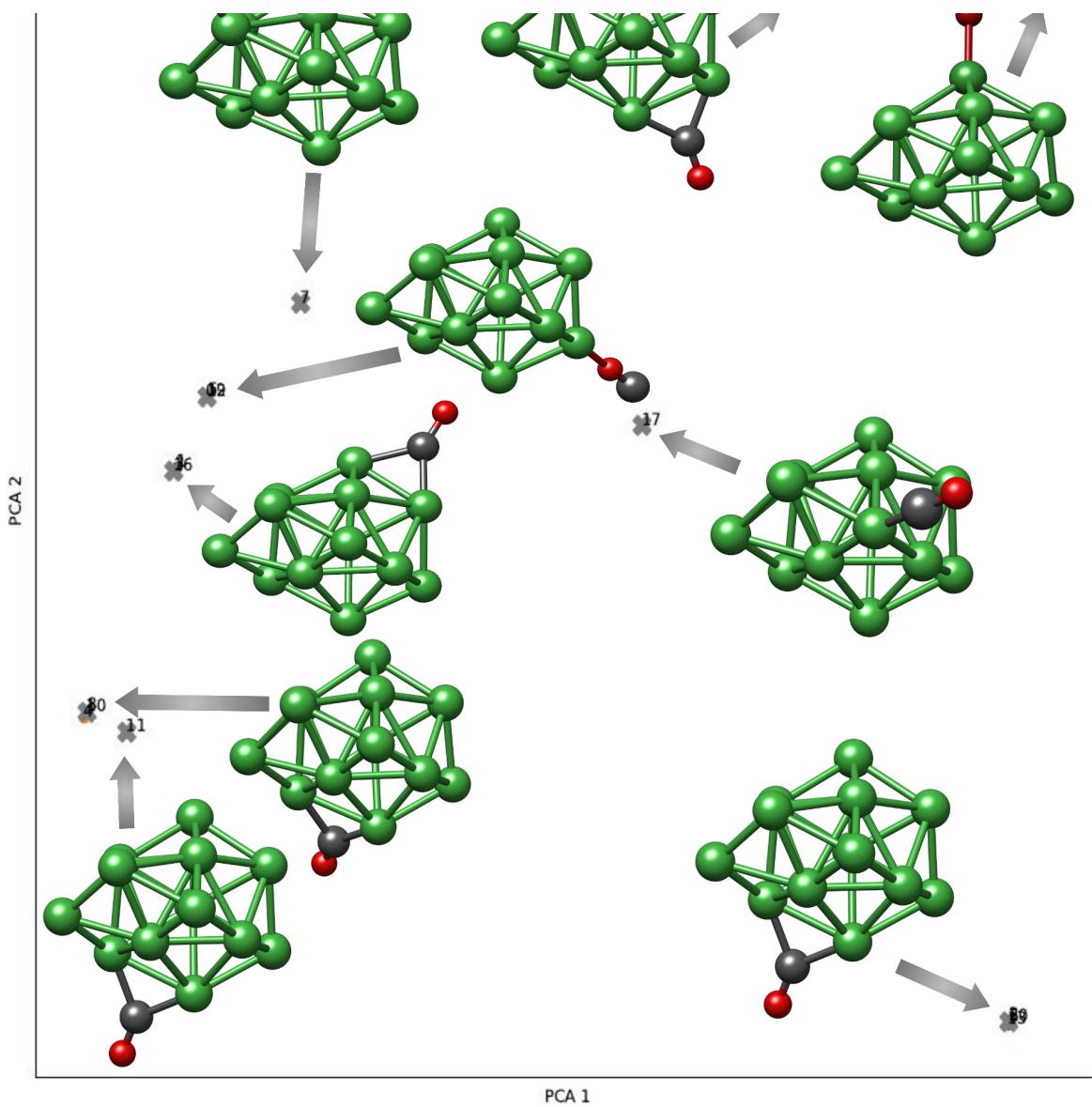
Same parameters as **Example 2**, but using `site_radius` instead of `site_size`, and all the tests were made using the energy.

JSON

```
{
  "seed": -1,
  "system_name": "CO-Ni13",
  "input_folder_path": "data/Ni13-CO_structures_pbe/",
  "output_folder_path": "data/Ni13-CO_structures_pbe/output/",
  "method_of_k_selection": "silhouette",
  "number_k_of_clusters": 2,
  "silhouette_range": [2, 20, 1],
  <*> "number_of_random_runs": 20, 30, 40, 50,
  "molecule_indices": [0, 1],
  "site_metric": "site_radius",
  "site_size": 0,
  "site_radius": 4.0,
  "fixed_substrate_atomic_number": 0,
  "z_exp": 2.4,
  "d_exp": 1.0,
  "use_energy": true,
  "scale_dataset": true,
  "projection_numbers": true,
  "projection_repr": true,
  "projection_tsne": true
}
```

Charts





Example 9

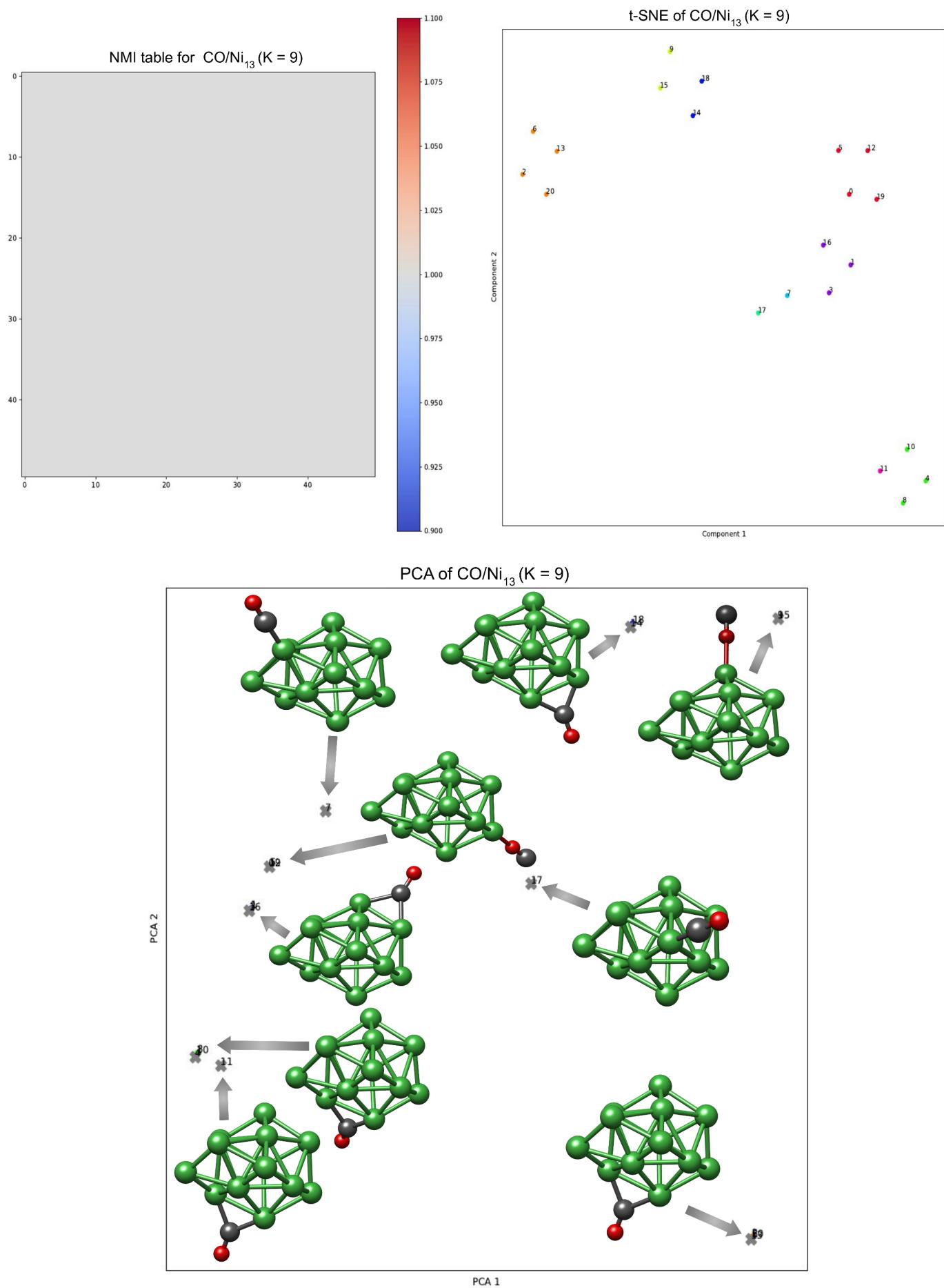
This example uses the same parameters as [Example 8](#), but fixes the number of runs as 50 and uses centroids, instead of energies to get representatives.

JSON

```
{
  "seed": -1,
  "system_name": "CO-Ni13",
  "input_folder_path": "data/Ni13-CO_structures_pbe/",
  "output_folder_path": "data/Ni13-CO_structures_pbe/output/",
  "method_of_k_selection": "silhouette",
  "number_k_of_clusters": 2,
  "silhouette_range": [2, 20, 1],
  "number_of_random_runs": 50,
  "molecule_indices": [0, 1],
  "site_metric": "site_radius",
  "site_size": 0,
  "site_radius": 4.0,
```

```
"fixed_substrate_atomic_number": 0,  
"z_exp": 2.4,  
"d_exp": 1.0,  
"use_energy": false,  
"scale_dataset": true,  
"projection_numbers": true,  
"projection_repr": true,  
"projection_tsne": true  
}
```

Charts



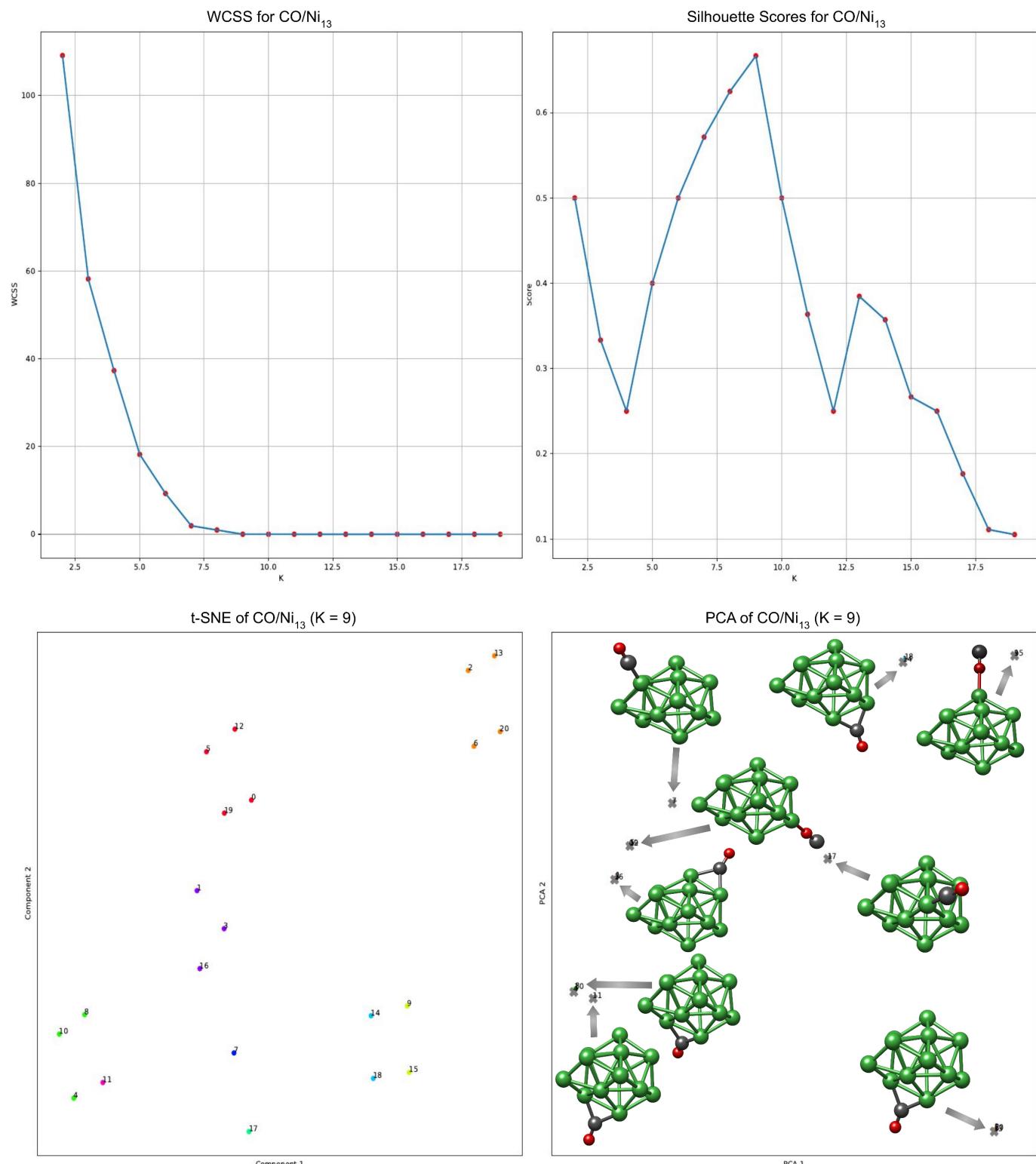
Example 10

This example configures the `method_of_k_selection` as `user` and sets `K=9`. Similar to **Example 4**, this choice of `K` yields the highest custom silhouette score, indicating it would have been automatically selected by the program. Again, this option is particularly useful when the specialist possesses prior knowledge of the correct number of clusters or prefers an alternative criterion over the default program's choice.

JSON

```
{  
    "seed": -1,  
    "system_name": "CO-Ni13",  
    "input_folder_path": "data/Ni13-CO_structures_pbe/",  
    "output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
    "method_of_k_selection": "user",  
    "number_k_of_clusters": 9,  
    "silhouette_range": [2, 20, 1],  
    "number_of_random_runs": 1,  
    "molecule_indices": [0, 1],  
    "site_metric": "site_radius",  
    "site_size": 0,  
    "site_radius": 4.0,  
    "fixed_substrate_atomic_number": 0,  
    "z_exp": 2.4,  
    "d_exp": 1.0,  
    "use_energy": true,  
    "scale_dataset": true,  
    "projection_numbers": true,  
    "projection_repr": true,  
    "projection_tsne": true  
}
```

Charts



Example 11

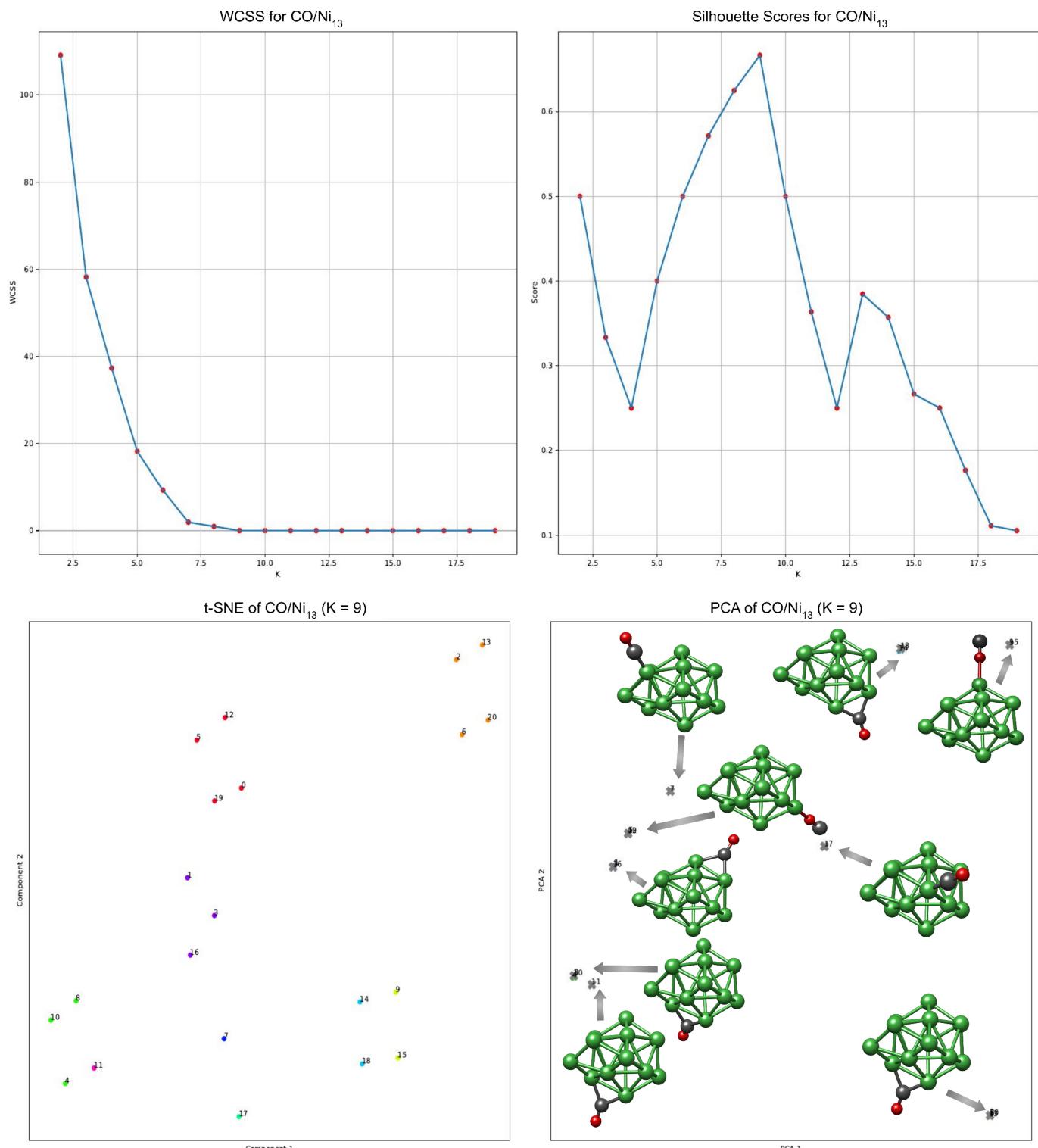
Same as [Example 10](#), but using proximity to centroids instead of lowest energies to select the representative samples.

JSON

```
{
  "seed": -1,
  "system_name": "CO-Ni13",
```

```
"input_folder_path": "data/Ni13-CO_structures_pbe/",  
"output_folder_path": "data/Ni13-CO_structures_pbe/output/",  
"method_of_k_selection": "user",  
"number_k_of_clusters": 9,  
"silhouette_range": [2, 20, 1],  
"number_of_random_runs": 1,  
"molecule_indices": [0, 1],  
"site_metric": "site_radius",  
"site_size": 0,  
"site_radius": 4.0,  
"fixed_substrate_atomic_number": 0,  
"z_exp": 2.4,  
"d_exp": 1.0,  
"use_energy": false,  
"scale_dataset": true,  
"projection_numbers": true,  
"projection_repr": true,  
"projection_tsne": true  
}
```

Charts



Textual output

This section displays examples of textual outputs generated by the program.

Using centroid closeness to pick representatives

```
Seed: 1
```

```
Number of random runs: 50
```

```
Number K of clusters: 8
```

Silhouette Score: 0.75

Within-Cluster Sum of Squares (WCSS): 0.8053732249808837

Number of atoms in site: 8

Fixed substrate atomic number: Disabled

Coulomb Matrix exponents: z=2.400000 d=1.000000

Scaled dataset: Yes

Labels and Substrate Z Averages:

ID: 000 - File: data/Ni13-CO_structures_pbe/Ni13-CO_001.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 001 - File: data/Ni13-CO_structures_pbe/Ni13-CO_002.xyz - Label: 3 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 002 - File: data/Ni13-CO_structures_pbe/Ni13-CO_003.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 003 - File: data/Ni13-CO_structures_pbe/Ni13-CO_005.xyz - Label: 3 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 004 - File: data/Ni13-CO_structures_pbe/Ni13-CO_006.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 005 - File: data/Ni13-CO_structures_pbe/Ni13-CO_007.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 006 - File: data/Ni13-CO_structures_pbe/Ni13-CO_008.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 007 - File: data/Ni13-CO_structures_pbe/Ni13-CO_009.xyz - Label: 4 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 008 - File: data/Ni13-CO_structures_pbe/Ni13-CO_010.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 009 - File: data/Ni13-CO_structures_pbe/Ni13-CO_012.xyz - Label: 5 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 010 - File: data/Ni13-CO_structures_pbe/Ni13-CO_013.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 011 - File: data/Ni13-CO_structures_pbe/Ni13-CO_014.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 012 - File: data/Ni13-CO_structures_pbe/Ni13-CO_015.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 013 - File: data/Ni13-CO_structures_pbe/Ni13-CO_017.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 014 - File: data/Ni13-CO_structures_pbe/Ni13-CO_018.xyz - Label: 7 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 015 - File: data/Ni13-CO_structures_pbe/Ni13-CO_020.xyz - Label: 5 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 016 - File: data/Ni13-CO_structures_pbe/Ni13-CO_021.xyz - Label: 3 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 017 - File: data/Ni13-CO_structures_pbe/Ni13-CO_022.xyz - Label: 6 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 018 - File: data/Ni13-CO_structures_pbe/Ni13-CO_023.xyz - Label: 7 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 019 - File: data/Ni13-CO_structures_pbe/Ni13-CO_024.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni

```
ID: 020 - File: data/Ni13-CO_structures_pbe/Ni13-CO_026.xyz - Label: 1 - Z Avg:  
28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
```

Representatives:

```
Cluster 000: data/Ni13-CO_structures_pbe/Ni13-CO_024.xyz  
Cluster 001: data/Ni13-CO_structures_pbe/Ni13-CO_008.xyz  
Cluster 002: data/Ni13-CO_structures_pbe/Ni13-CO_010.xyz  
Cluster 003: data/Ni13-CO_structures_pbe/Ni13-CO_021.xyz  
Cluster 004: data/Ni13-CO_structures_pbe/Ni13-CO_009.xyz  
Cluster 005: data/Ni13-CO_structures_pbe/Ni13-CO_012.xyz  
Cluster 006: data/Ni13-CO_structures_pbe/Ni13-CO_022.xyz  
Cluster 007: data/Ni13-CO_structures_pbe/Ni13-CO_023.xyz
```

Substrate information:

```
Substrates Average of averages: 28.000000  
Substrates Std.Dev of averages: 0.000000
```

PCA variance: 0.4783495699861799 0.28885479069482517

Information about automatic K selection

```
k: 002 - mean: 0.500000 - var: 0.000000  
k: 003 - mean: 0.573333 - var: 0.022400  
k: 004 - mean: 0.500000 - var: 0.000000  
k: 005 - mean: 0.600000 - var: 0.000000  
k: 006 - mean: 0.666667 - var: 0.000000  
k: 007 - mean: 0.571429 - var: 0.000000  
k: 008 - mean: 0.750000 - var: 0.000000  
k: 009 - mean: 0.666667 - var: 0.000000  
k: 010 - mean: 0.500000 - var: 0.000000  
k: 011 - mean: 0.363636 - var: 0.000000  
k: 012 - mean: 0.333333 - var: 0.000000  
k: 013 - mean: 0.304615 - var: 0.000227  
k: 014 - mean: 0.285714 - var: 0.000000  
k: 015 - mean: 0.266667 - var: 0.000000  
k: 016 - mean: 0.241250 - var: 0.000470  
k: 017 - mean: 0.176471 - var: 0.000000  
k: 018 - mean: 0.166667 - var: 0.000000  
k: 019 - mean: 0.105263 - var: 0.000000
```

Using lowest energy to pick representatives

Seed: 1

Number of random runs: 50

Number K of clusters: 8

Silhouette Score: 0.75

Within-Cluster Sum of Squares (WCSS): 0.8053732249808837

Number of atoms in site: 8

Fixed substrate atomic number: Disabled

Coulomb Matrix exponents: z=2.400000 d=1.000000

Scaled dataset: Yes

Labels and Substrate Z Averages:

ID: 000 - File: data/Ni13-CO_structures_pbe/Ni13-CO_001.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 001 - File: data/Ni13-CO_structures_pbe/Ni13-CO_002.xyz - Label: 3 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 002 - File: data/Ni13-CO_structures_pbe/Ni13-CO_003.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 003 - File: data/Ni13-CO_structures_pbe/Ni13-CO_005.xyz - Label: 3 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 004 - File: data/Ni13-CO_structures_pbe/Ni13-CO_006.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 005 - File: data/Ni13-CO_structures_pbe/Ni13-CO_007.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 006 - File: data/Ni13-CO_structures_pbe/Ni13-CO_008.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 007 - File: data/Ni13-CO_structures_pbe/Ni13-CO_009.xyz - Label: 4 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 008 - File: data/Ni13-CO_structures_pbe/Ni13-CO_010.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 009 - File: data/Ni13-CO_structures_pbe/Ni13-CO_012.xyz - Label: 5 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 010 - File: data/Ni13-CO_structures_pbe/Ni13-CO_013.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 011 - File: data/Ni13-CO_structures_pbe/Ni13-CO_014.xyz - Label: 2 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 012 - File: data/Ni13-CO_structures_pbe/Ni13-CO_015.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 013 - File: data/Ni13-CO_structures_pbe/Ni13-CO_017.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 014 - File: data/Ni13-CO_structures_pbe/Ni13-CO_018.xyz - Label: 7 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 015 - File: data/Ni13-CO_structures_pbe/Ni13-CO_020.xyz - Label: 5 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 016 - File: data/Ni13-CO_structures_pbe/Ni13-CO_021.xyz - Label: 3 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 017 - File: data/Ni13-CO_structures_pbe/Ni13-CO_022.xyz - Label: 6 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 018 - File: data/Ni13-CO_structures_pbe/Ni13-CO_023.xyz - Label: 7 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 019 - File: data/Ni13-CO_structures_pbe/Ni13-CO_024.xyz - Label: 0 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni
ID: 020 - File: data/Ni13-CO_structures_pbe/Ni13-CO_026.xyz - Label: 1 - Z Avg: 28.000000 - Site atoms: Ni Ni Ni Ni Ni Ni Ni Ni

Lowest energy elements:

Cluster 0: data/Ni13-CO_structures_pbe/Ni13-CO_015.xyz

```
Cluster 1: data/Ni13-CO_structures_pbe/Ni13-CO_017.xyz
Cluster 2: data/Ni13-CO_structures_pbe/Ni13-CO_014.xyz
Cluster 3: data/Ni13-CO_structures_pbe/Ni13-CO_021.xyz
Cluster 4: data/Ni13-CO_structures_pbe/Ni13-CO_009.xyz
Cluster 5: data/Ni13-CO_structures_pbe/Ni13-CO_012.xyz
Cluster 6: data/Ni13-CO_structures_pbe/Ni13-CO_022.xyz
Cluster 7: data/Ni13-CO_structures_pbe/Ni13-CO_018.xyz
```

Substrate information:

Substrates Average of averages: 28.000000
Substrates Std.Dev of averages: 0.000000

PCA variance: 0.4783495699861799 0.28885479069482517

Information about automatic K selection

```
k: 002 - mean: 0.500000 - var: 0.000000
k: 003 - mean: 0.573333 - var: 0.022400
k: 004 - mean: 0.500000 - var: 0.000000
k: 005 - mean: 0.600000 - var: 0.000000
k: 006 - mean: 0.666667 - var: 0.000000
k: 007 - mean: 0.571429 - var: 0.000000
k: 008 - mean: 0.750000 - var: 0.000000
k: 009 - mean: 0.666667 - var: 0.000000
k: 010 - mean: 0.500000 - var: 0.000000
k: 011 - mean: 0.363636 - var: 0.000000
k: 012 - mean: 0.333333 - var: 0.000000
k: 013 - mean: 0.304615 - var: 0.000227
k: 014 - mean: 0.285714 - var: 0.000000
k: 015 - mean: 0.266667 - var: 0.000000
k: 016 - mean: 0.241250 - var: 0.000470
k: 017 - mean: 0.176471 - var: 0.000000
k: 018 - mean: 0.166667 - var: 0.000000
k: 019 - mean: 0.105263 - var: 0.000000
```

Understanding the output

Seed, Number of random runs, Number K of clusters, Silhouette Score, Within-Cluster Sum of Squares (WCSS), Number of atoms in site, Fixed substrate atomic number, Coulomb Matrix exponents and Scaled dataset are self-explanatory and refer mostly to parameters the user input into the system and some general information.

The Labels and Substrate Z Averages table shows the ID number of the data entry (shown in charts if projection_numbers is true), followed by the corresponding XYZ file, number of the cluster to which this entry belongs, average atomic number and site atoms species.

Representatives and Lowest energy elements list the closest elements to their cluster's centroid and elements with lowest energy from each group, respectively.

Substrate information displays the mean atomic number averages across the data entries, along with their standard deviation.

[PCA variance](#) indicates the degree of variance accounted for in each dimension of the Principal Component Analysis (PCA) projection.

[Information about automatic K selection](#) is a table that demonstrates, for each amount [K](#) of clusters, the mean and variance of the custom silhouette score for all random runs.