Measuring uniqueness

Let $D = \{S_1, \dots, S_N\}$ be a dataset of N individuals, in the form of $S_i = (A_1 = [a_1^1, a_2^1, .., a_m^1], \dots, A_k = [a_1^k, a_2^k, .., a_m^k])$, containing m readings for k activities, a_i^j is the value of the i-th reading for the j-th activity.

Considerations:

1. Assume to conduct the uniqueness analysis on each activity separately. In other words, we compute a uniqueness score $u_i(\cdot)$, for each activity A_i. Here, we can start by considering the top-3 activity that you have already identified using the KL-divergence.
2. Consider a threshold value th to match the reading

Uniqueness score:
Consider a positive integer x, we generate x random readings $I_i{}^x = (i_1, i_2, \dots, i_x)$, for the activity A_i. An individual S_j is **compatible** with $I_i(x)$, if there exists x readings in S_j, $S_j^x = (A'_i = [a_{l1}^i, a_{l2}^i, .., a_{lx}^i])$ such that $|a_{l1}^i - i_1| \le th$,..., $|a_{lx}^i - i_x| \le th$ (i.e., all the reading are matched within a threshold).

Furthermore, let $|D(I_i{}^x)|$ denote the number of individuals in D that are compatible with $I_i{}^x$. Then an individual is **uniquely identified** by $I_i{}^x$ if $|D(I_i{}^x)| = 1$. Let $u_1(x)$ denote the fraction of time-series in D for which $|D(I_i{}^x)| \le 1$.

TODO: Compute the uniqueness for the top-3 activities with $x \in \{1, \dots, 5\}$ and

- $th \in \{1, 5\}$ for discrete activities (e.g., number of steps)
- $th \in \{0.1, 0.2\}$ for real-value activities (e.g., distance-based activities)

Example.--
$D = \{S_1, S_2, S_3\} = \{[10,13,5], [0,13,10], [10,5,6]\}$

$x = 1$
{(0),(5),(6),(10),(13)} universe of 1-reading from the data
I^1 list of matching individuals
(0) -> $\{S_2\}$
(5) -> $\{S_{1,}, S_3\}$
(6) -> $\{S_3\}$
(10) -> $\{S_{1,} S_2, S_3\}$
(13) -> $\{S_{1,} S_2\}$
$u_1(1) = 2/3$ -> because only S2 and S3 are uniquely identified

$x = 2$
{(0,10),(0,13),(5,6),(10,5),(10,6),(10,13),(13,5)} universe of 2-readings from the data
I^2 list of matching individuals
(0,10) -> $\{S_2\}$
(0,13) -> $\{S_2\}$
(5,6) -> $\{S_3\}$
(10,5) -> $\{S_{1,} S_2\}$
(10,6) -> $\{S_3\}$
(10,13) -> $\{S_1\}$
(13,5) -> $\{S_1\}$
$u_1(2) = 3/3$ all the individuals can be uniquely identified with 2 readings