Thanks for preparing a thorough document with many examples to help me understand the problem.

Here I will give you my viewpoints on your questions and I will also ask you some questions to declare the problem.

- First of all, you wrote that: "The neighbouring landuses are codded as unique existence of types in the set of landuse. It corresponds to transactions in item sets (like Market Basket Analysis). It means that repetition of the same landuse is considered a one-time appearance."

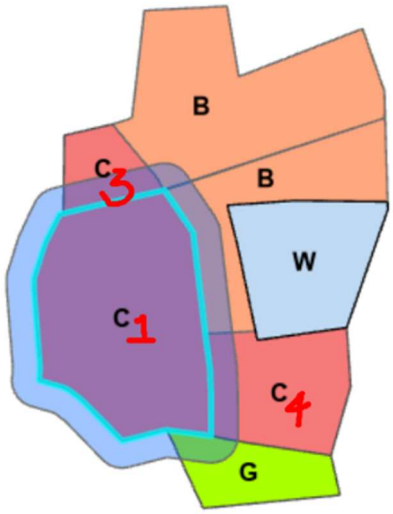
→ Is it the conventional way the Geoinformatics community considers only one-time appearance neighbor of a landuse? I mean that the repetition of the same landuse is redundant and you consider it as a one-time appearance?

- When I look at Figure 2, I think that the repetition of the same landuse is important, and we should keep such repetition in the dataset. Such repetition is similar to the weight of a vertice in a weighted graph. If we ignore it, we may consider all of the polygons as equal weight-> As you say, it may hide some non-trivial info.

Let me take an example from Fig. 2. If I use one time appearance, I will get three polygon C as you show in Table 1

Table 1

Polygon No.	Input polygon	Neighbours
1	C	C B G
2	W	C B
3	C	C B
4	C	C G B W



It seems that C_1 and C_4 have similar weight (even C_4 has more neighbor landuse than C_1). However, when we look at to the map in figure 2, If I have to choose a landuse to buy, I will buy C_1 instead of C_4.

Now, If I keep the repetition of landuses, your table will be as follows, I just show an example for C

Polygon No.	Input polygon	Neighbours
1	C	BBCCG
3	C	BBC
4	C	BCGW

We can list some patterns as neighbor landuses of the polygon C (no we donot care the Polygon No.) having a high frequency (assume that the minSup=2)

- B

- B B

- B B C

- B C

- B C G

- C

- CG

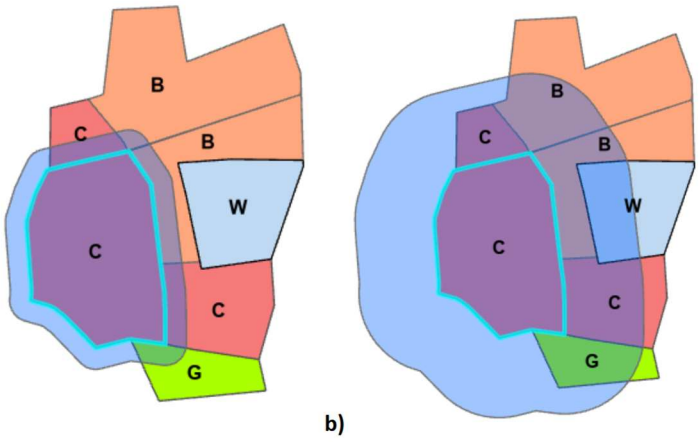
So I think it depends on the way of interpretation, you may decide to keep the repetition or not. Technically, detecting frequent patterns having repetition can be solved.

For your questions:

QUESTION:
It is correct description of the situation of neighbours? The data maybe contain some **hidden** "repetition"!
2 . row - W (water) contains information about water´s neighbour's – **C** and **B**.
4. row - C (commercial) has neighbour W (beside C, G, B) .
Is could be considered something like symmetry?
→ Yes, if you do not consider the repetition, then the symmetry exists since if we know C is a neighbor of W then W is a neighbor of C.
Are the input data correct for discovering frequent item sets? Will be there any bias by the "repettion"? Can we do not mind this "repetition"?

→Yes, the data in Table 1 can be used as input in frequent itemsets. For the repetition, I have comments before, it depends on the interpretation. I personally want to keep the repetition.

I think if we use frequent patterns mining, we can find all patterns that support for your hypothesis.



b)

For this Figure, I agree that the radius has a high impact on the converted data. This parameter can be practically tuned so that the generated dataset contains transactions with different lengths. You can set a parameter called average_length to adjust for the number of neighbors on each landuse.

- When I see the map of landuses, I have also several wrap ideas for the map:

+ Convert the map to a (weighted) graph and do graph mining.

+ Since it contains the neighbor information, we can build a distance matrix, then use hierarchical clustering to build the dendrogram to find some interesting groups. We also can use HDBScan (an improved version of DBScan) to do clustering.

+ In pattern mining, we have a subtopic called frequent sequential pattern mining (FSPM) which is an extension of frequent itemset mining (FIM). FSPM finds the patterns that contain items bought in sequential order. The following input can be used:

ID	Sequences
S1	(1), (1 2 3), (1 3), (4), (3 6)
S2	(1 4), (3), (2 3), (1 5)
S3	(5 6), (1 2), (4 6), (3), (2)
S4	(5), (7), (1 6), (3), (2), (3)

1 : bread
2 : hotdog
3 : beer
4 : banana
5 : apple

Take S1 as an example: there are five transactions within the sequence S1. Assume that these transactions were made by a customer X in the time period. For example, yesterday morning X bought (bread) and yesterday evening X bought (bread and hotdog beer), then today morning X bought (bread and beer) and today noon X bought (banana) and then today evening he bought (beer apple). So you see that even sequence S1 may contain repitition item 1 (banana), but it different in terms of time purchasing (you can

see from this post: https://www.philippe-fournier-viger.com/spmf/FAST.php). I try to map the FSPM but maybe the nature is different.

I also think the way that you consider the neighbors of each land use according to its directions: neighbor in the north, northeast, east, south east, south, south west, west, and northwest (direction by the clockwise). We can build a dataset that has fixed columns (directions) and then neighbors will be filled in each column (may be null if no neighbor). If so, we have an input like S1 that contains multiple transactions (columns) and can apply FSPM for discovering the patterns.

Generally, I think frequent pattern mining can be used for finding evidence support for your hypothesis.