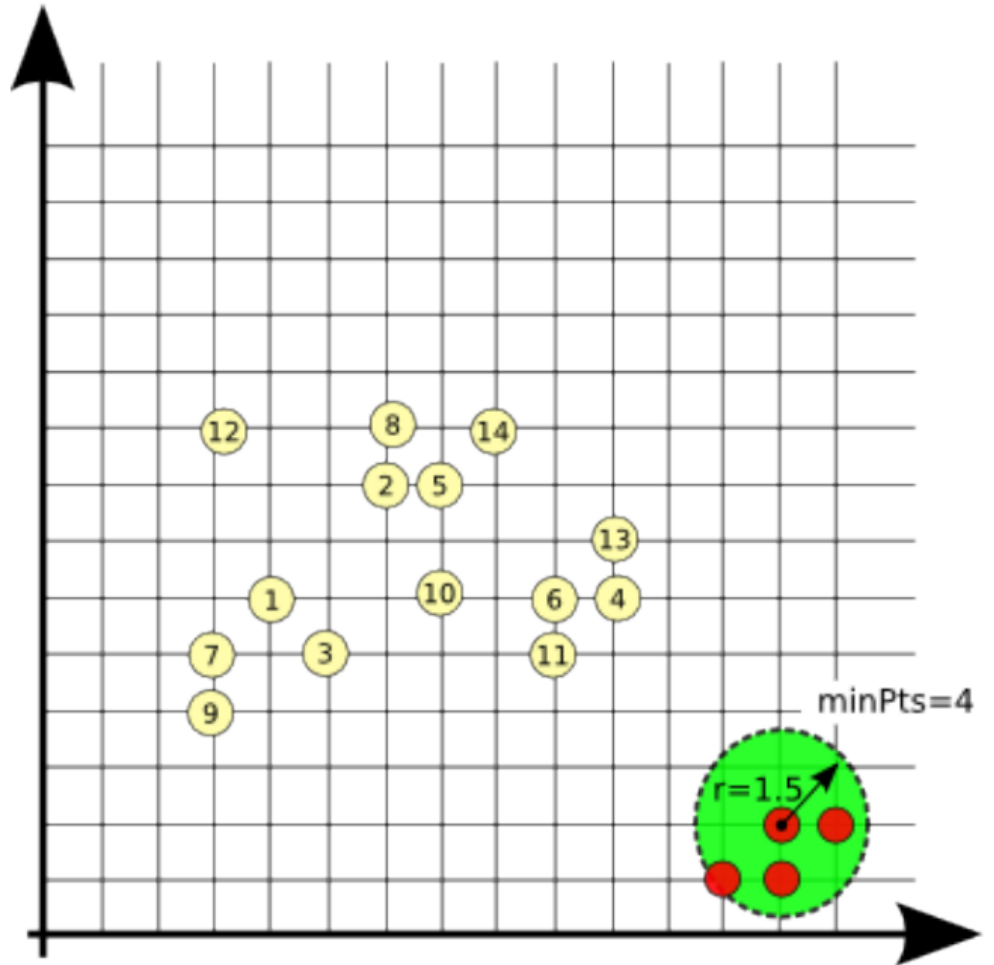
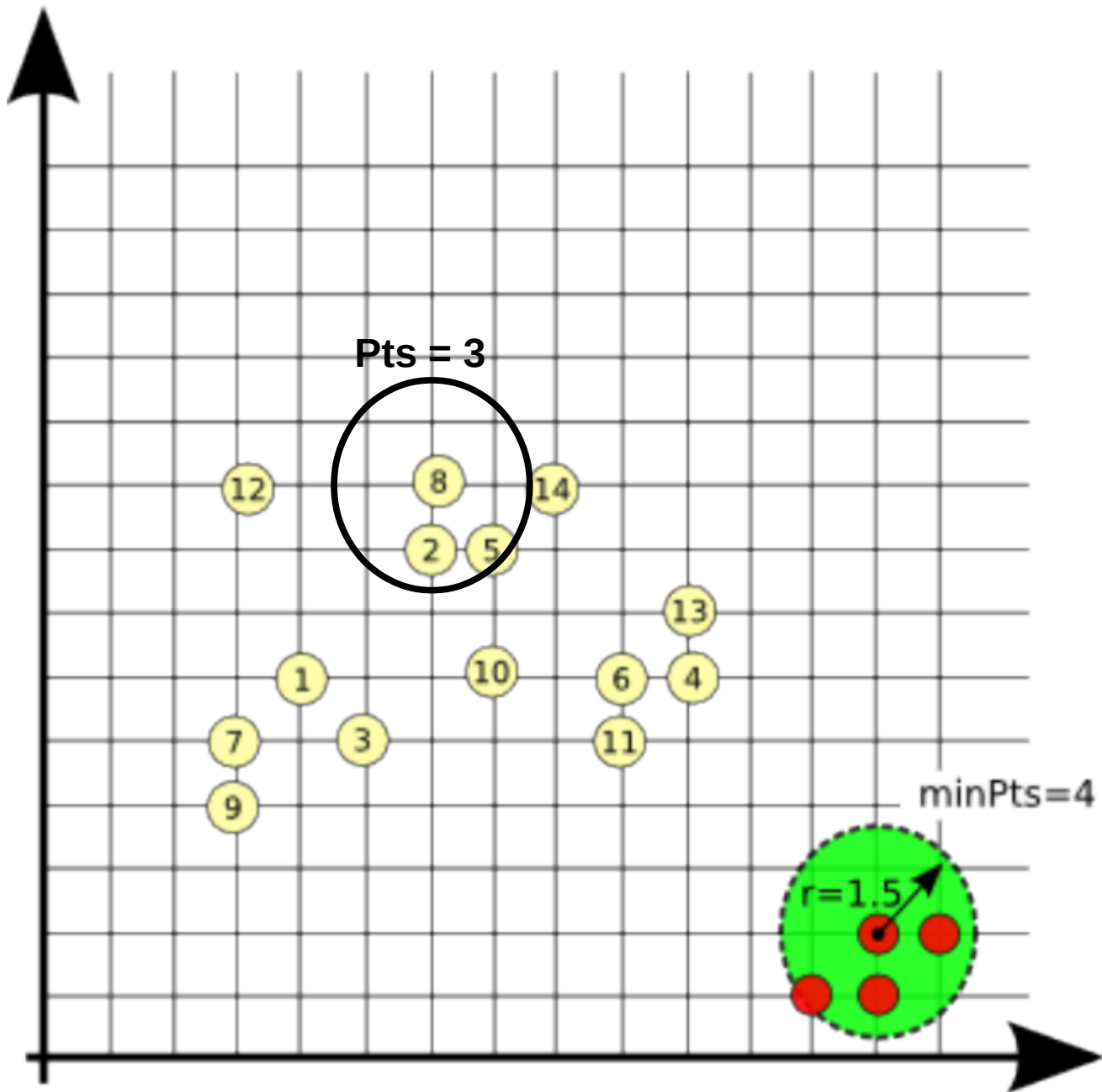
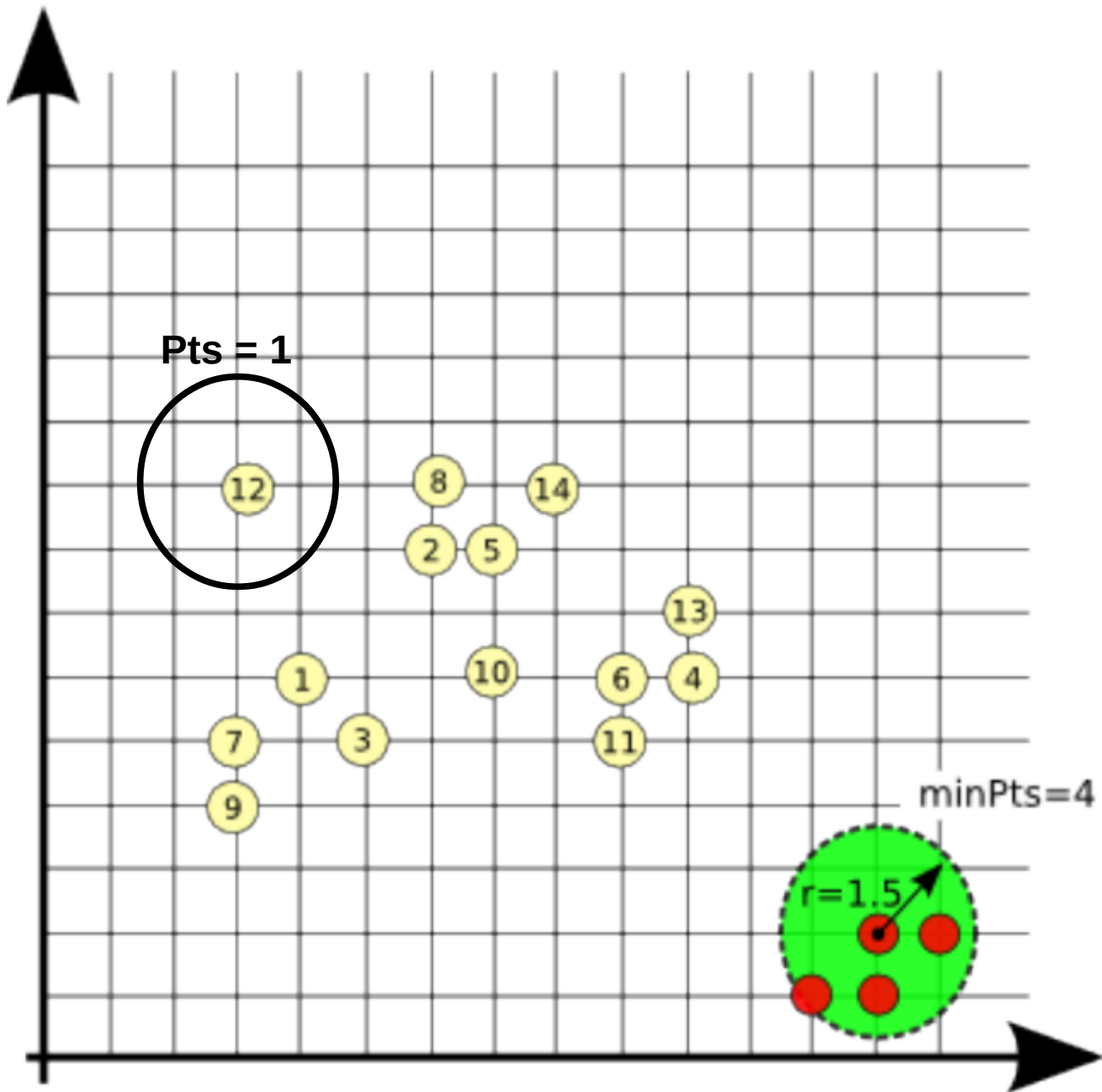


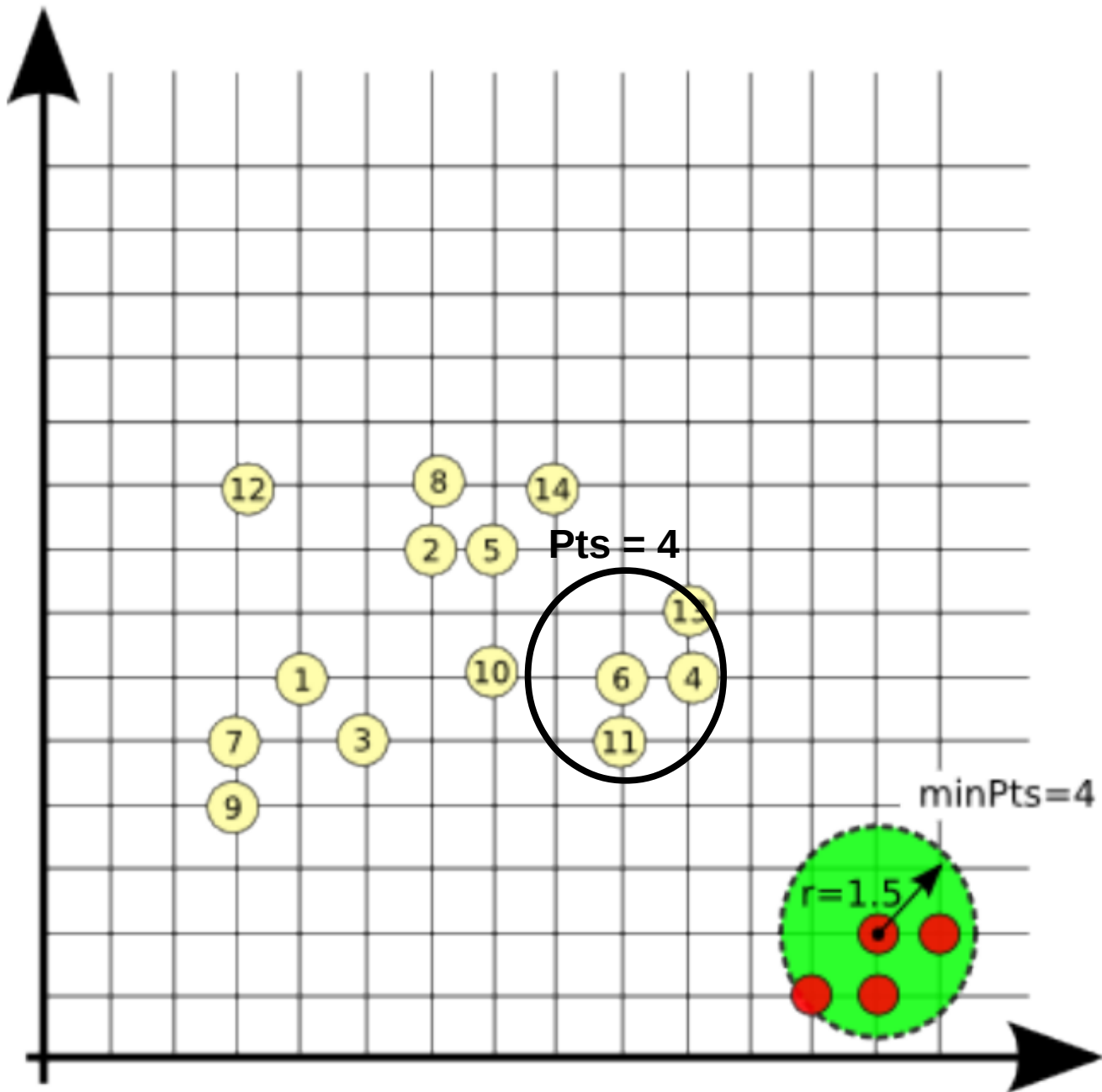
# Distance-based outlier detection

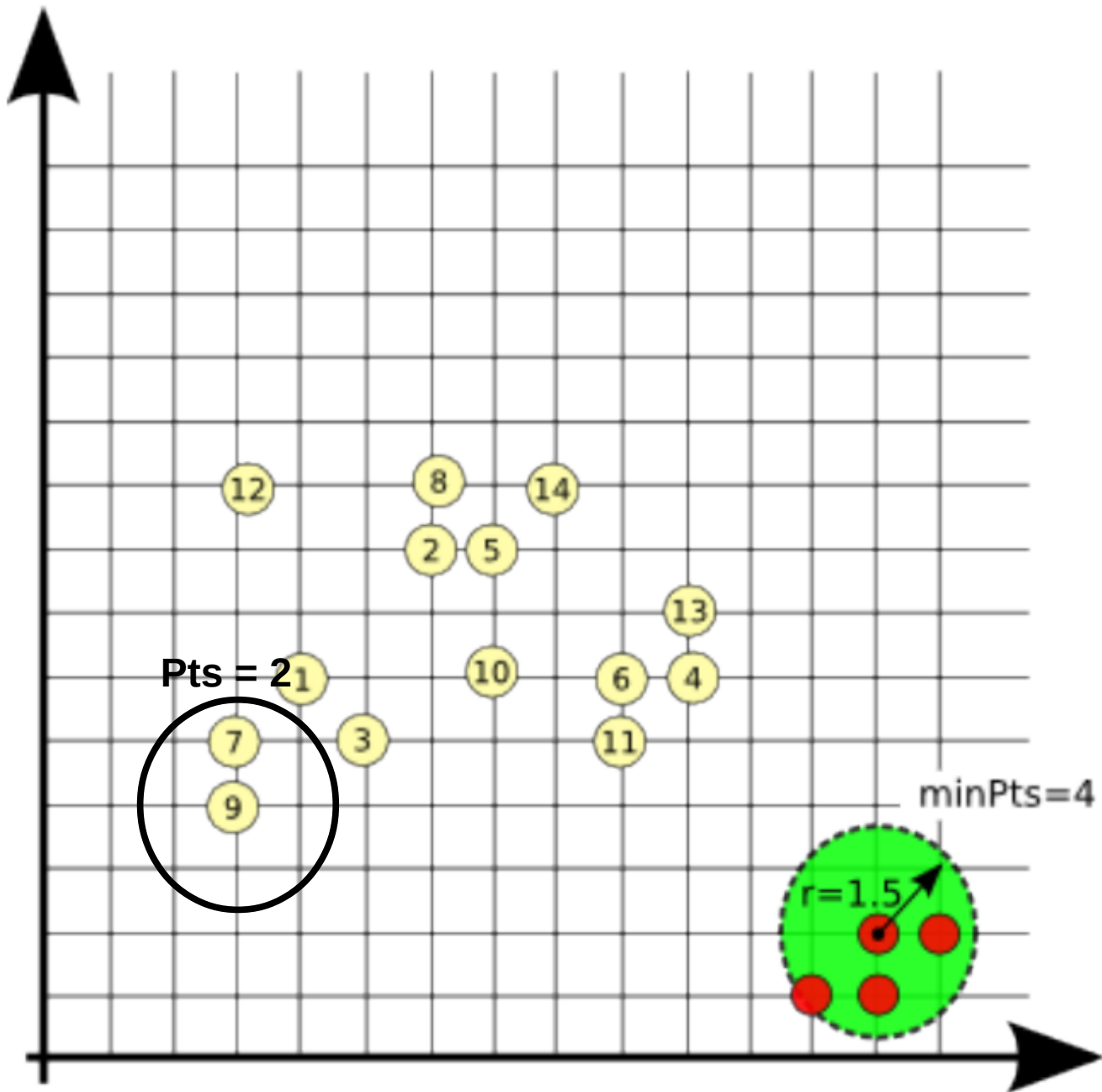
- Given the dataset of the right, find the outliers according to the basic  $DB(\epsilon, \pi)$ -Outliers method with:
  - $\epsilon = 1.5$
  - $\pi = 0.28 = 4$  points (including the reference point we are analyzing)

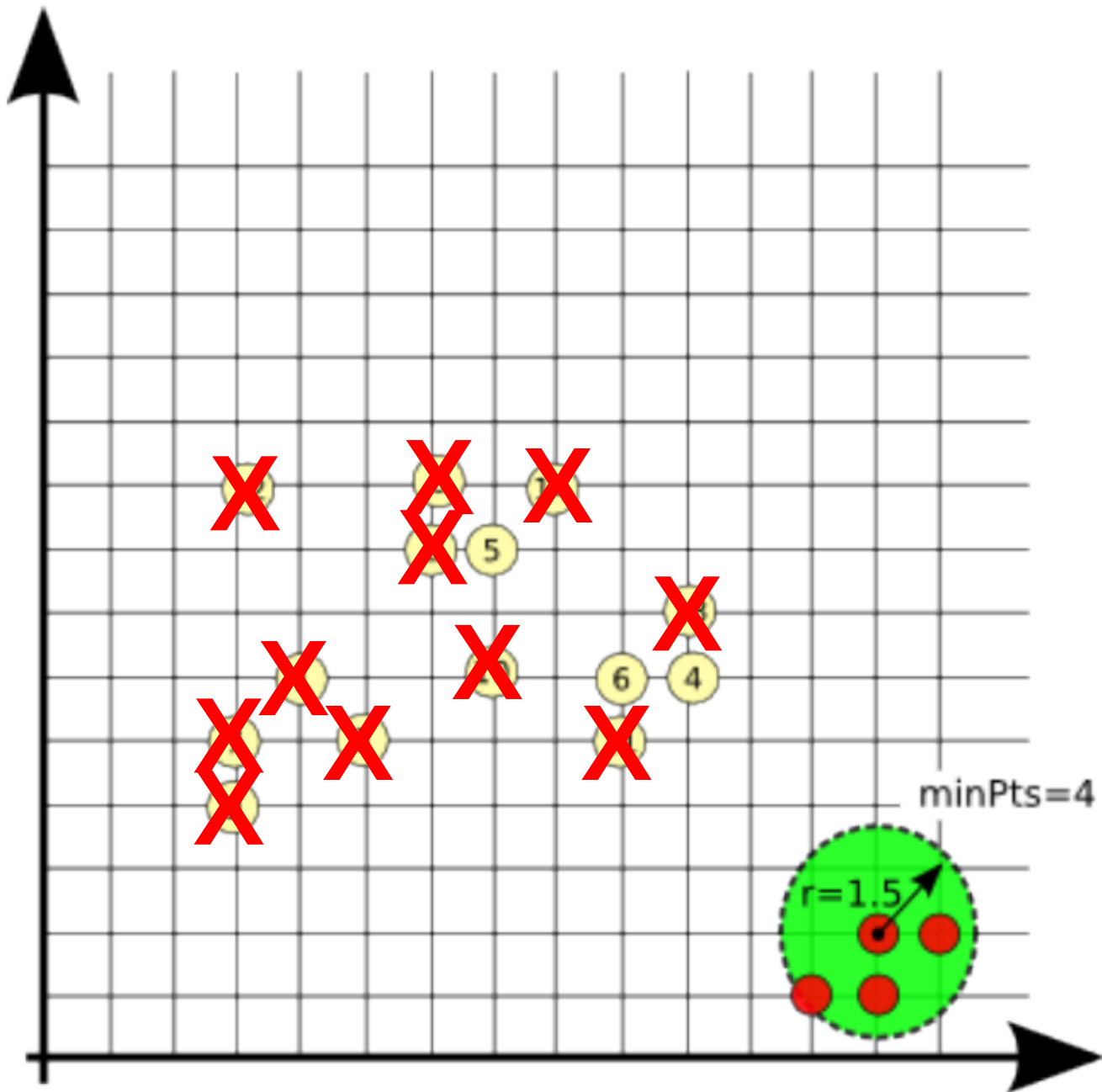






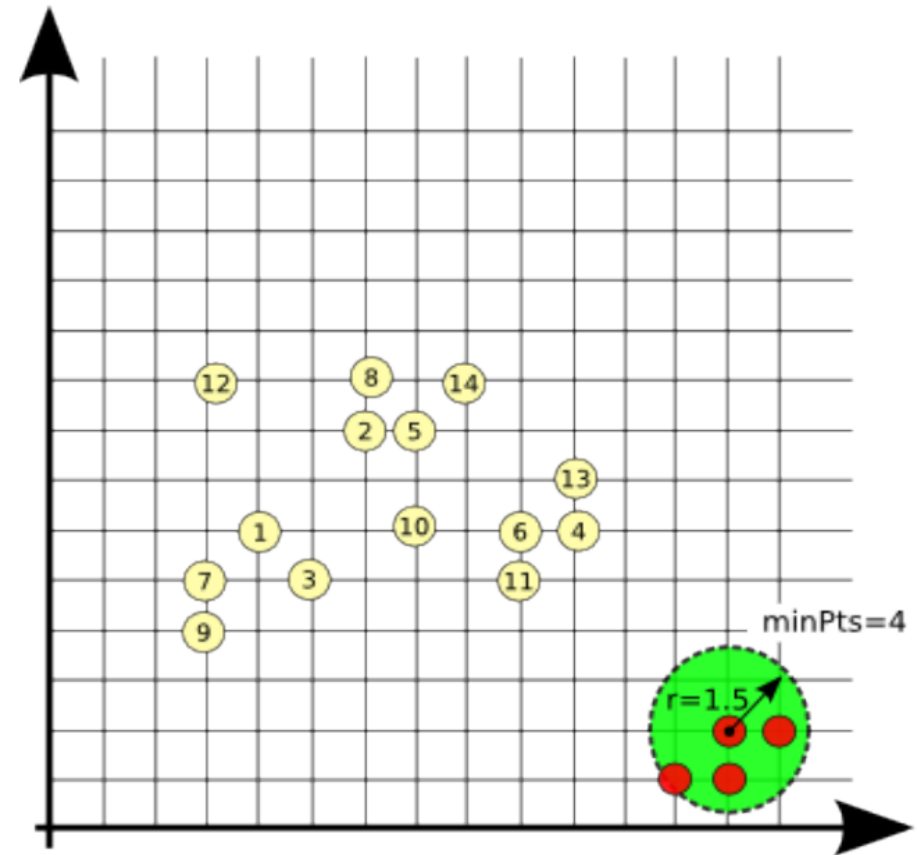






# Density-based outlier detection

- Compute LOF score for points 5, 1, 10 of the dataset:
- $K = 3$ 
  - The reference point is not considered
- Simplify: use distance in place of reachability distance ( $\text{reach\_dist}(p,o)$ )



$$lrd_k(p) = 1 / \left( \frac{\sum_{o \in kNN(p)} \text{reach-dist}_k(p,o)}{\text{Card}(kNN(p))} \right)$$

$$LOF_k(p) = \frac{\sum_{o \in kNN(p)} \frac{lrd_k(o)}{lrd_k(p)}}{\text{Card}(kNN(p))}$$

$$NN_3(5) = (2), (8), (14)$$

$$Lrd(5) = 1 / \left( \frac{1 + 1.41 + 1.41}{3} \right) = 0.78$$

$$\rightarrow Lrd(2) = 1 / \left( \frac{1 + 1 + 2.23 + 2.23}{4} \right) = 0.62$$

$$\rightarrow Lrd(8) = 1 / \left( \frac{1 + 1.41 + 2}{3} \right) = 0.68$$

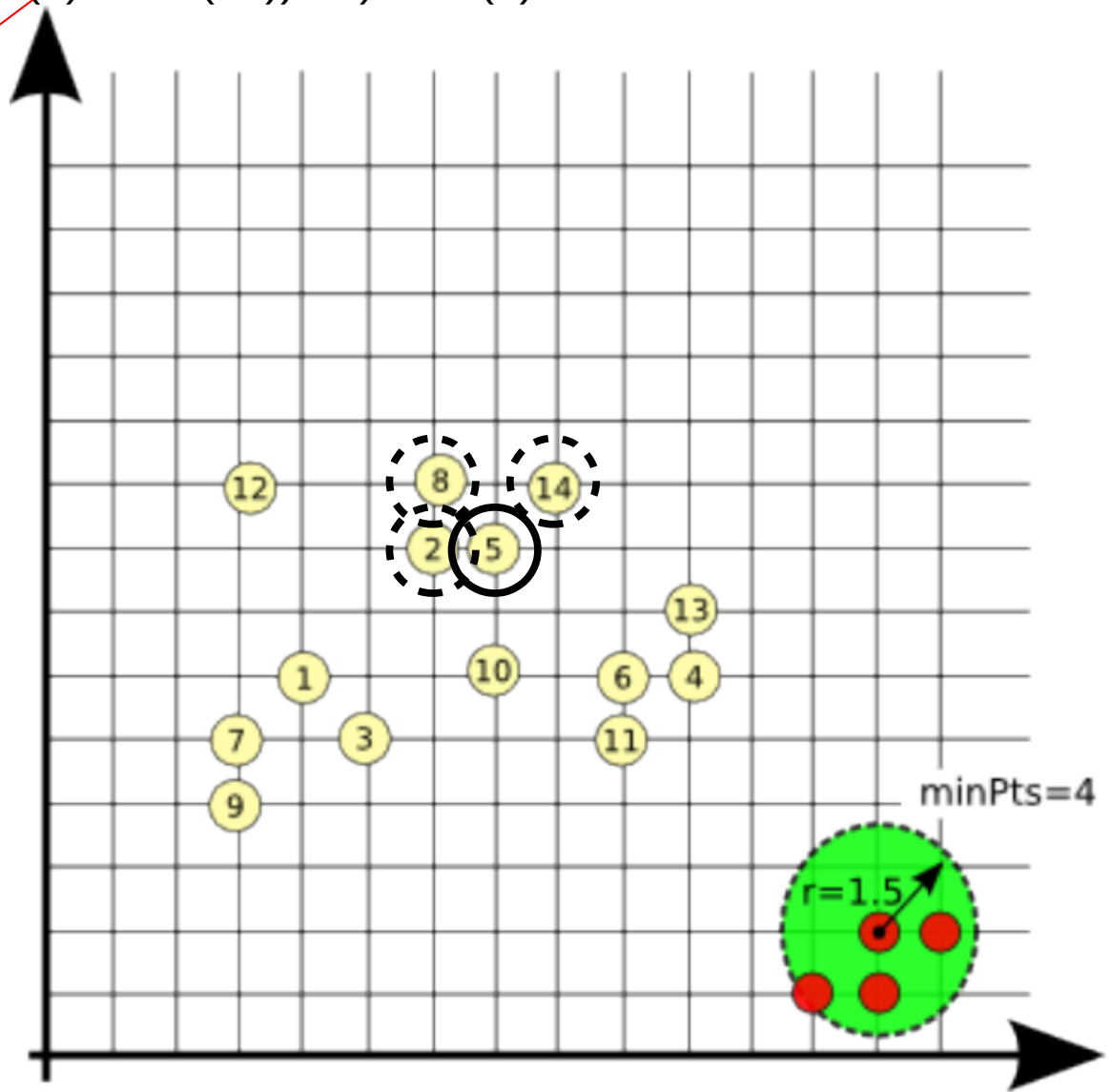
$$\rightarrow Lrd(14) = 1 / \left( \frac{2 + 1.41 + 2.23}{3} \right) = 0.53$$

$$LOF(5) = ((Lrd(2) + Lrd(8) + Lrd(14)) / 3) / Lrd(5) = 0.61 / 0.78 = \mathbf{0.78}$$

$$lrd_k(p) = 1 / \left( \frac{\sum_{o \in kNN(p)} reach-dist_k(p, o)}{Card(kNN(p))} \right)$$

$$LOF_k(p) = \frac{\sum_{o \in kNN(p)} \frac{lrd_k(o)}{lrd_k(p)}}{Card(kNN(p))}$$

Notice that (2) has 4 points in its 3-NN set, because (14) and (10) have the same distance. In case of ties we take all the points having that same distance.





$NN_3(1) = (3), (7), (9)$

$Lrd(1) = 1 / ( ( 1.41 + 1.41 + 2.23 ) / 3 ) = 0.59$

→  $Lrd(3) = 1 / ( ( 1.41 + 2 + 2.23 + 2.23 ) / 4 ) = 0.50$

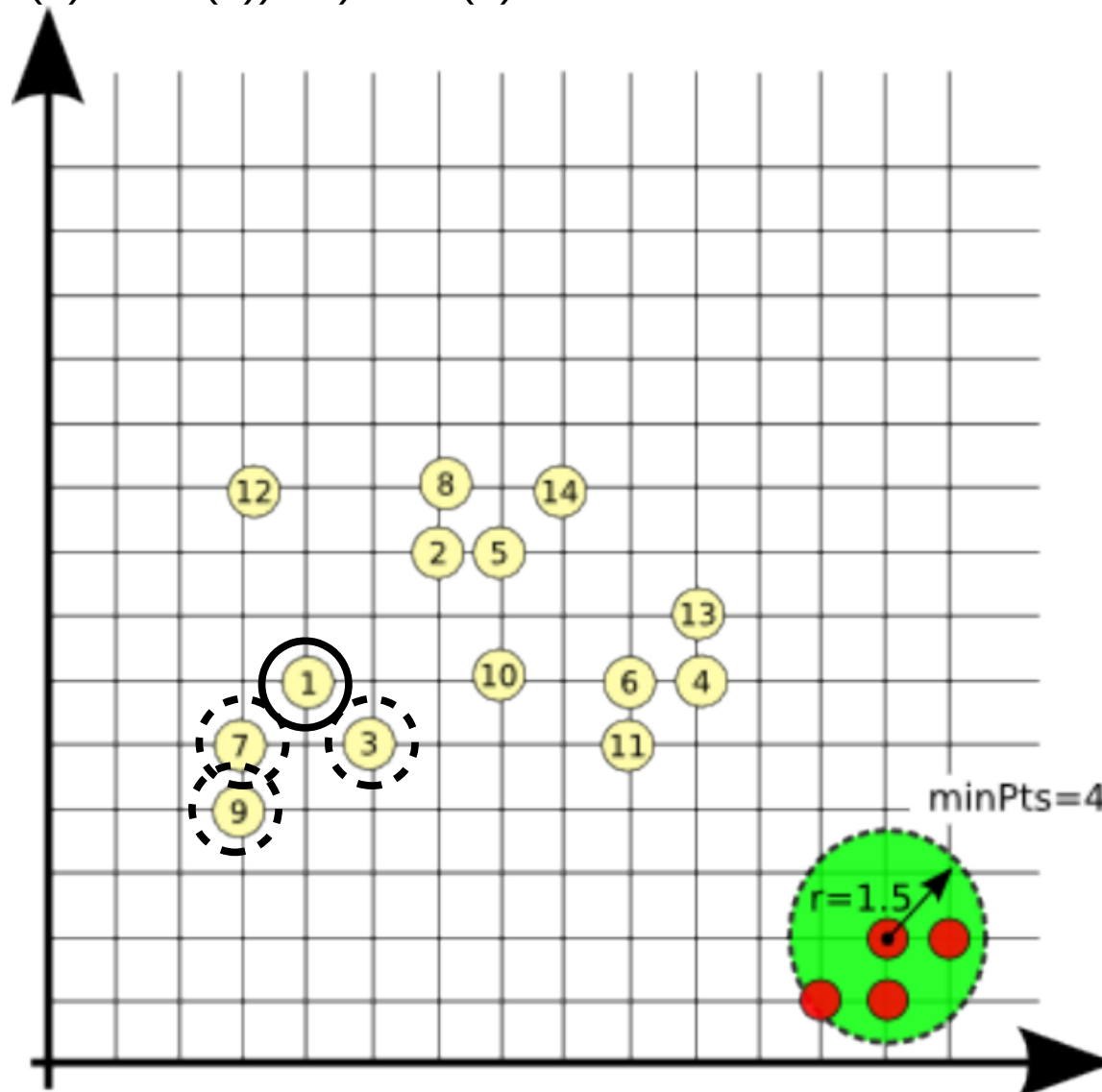
→  $Lrd(7) = 1 / ( ( 1 + 1.41 + 2 ) / 3 ) = 0.68$

→  $Lrd(9) = 1 / ( ( 1 + 2.23 + 2.23 ) / 3 ) = 0.55$

**LOF(5) = ((Lrd(3) + Lrd(7) + Lrd(9)) / 3 ) / Lrd(1) = 0.57 / 0.59 = 0.96**

$$lrd_k(p) = 1 / \left( \frac{\sum_{o \in kNN(p)} reach-dist_k(p, o)}{Card(kNN(p))} \right)$$

$$LOF_k(p) = \frac{\sum_{o \in kNN(p)} \frac{lrd_k(o)}{lrd_k(p)}}{Card(kNN(p))}$$



$NN_3(10) = (2), (5), (6), (11), (3)$

$Lrd(10) = 1 / ( ( 2 + 2.23 + 2 + 2.23 + 2.23 ) / 5 ) = 0.46$

→  $Lrd(2) = 1 / ( ( 1 + 1 + 2.23 + 2.23 ) / 4 ) = 0.62$

→  $Lrd(5) = 1 / ( ( 1 + 1.41 + 1.41 ) / 3 ) = 0.78$

→  $Lrd(6) = 1 / ( ( 1 + 1 + 1.41 ) / 3 ) = 0.87$

→  $Lrd(11) = 1 / ( ( 1 + 1.41 + 2.23 + 2.23 ) / 4 ) = 0.58$

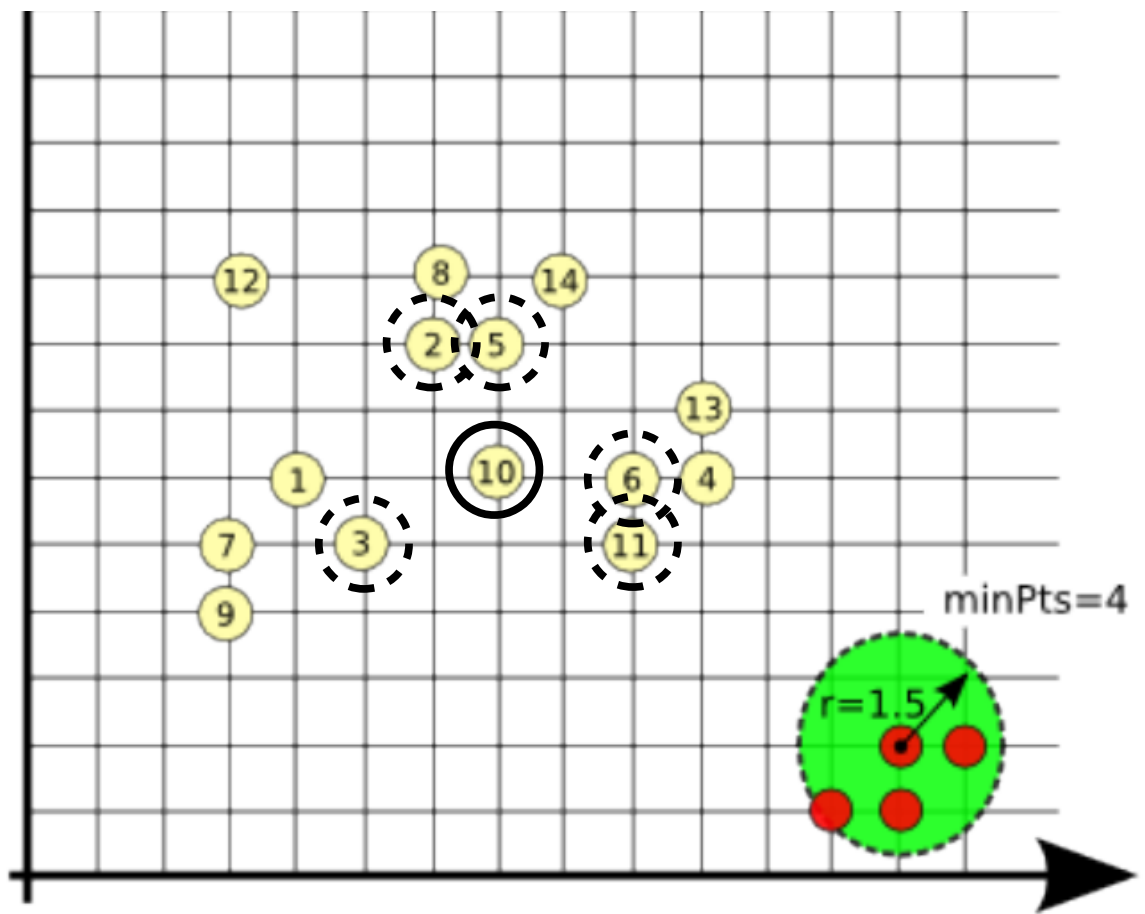
→  $Lrd(3) = 1 / ( ( 1.41 + 2 + 2.23 + 2.23 ) / 4 ) = 0.50$

$$lrd_k(p) = 1 / \left( \frac{\sum_{o \in kNN(p)} reach-dist_k(p, o)}{Card(kNN(p))} \right)$$

$$LOF_k(p) = \frac{\sum_{o \in kNN(p)} \frac{lrd_k(o)}{lrd_k(p)}}{Card(kNN(p))}$$

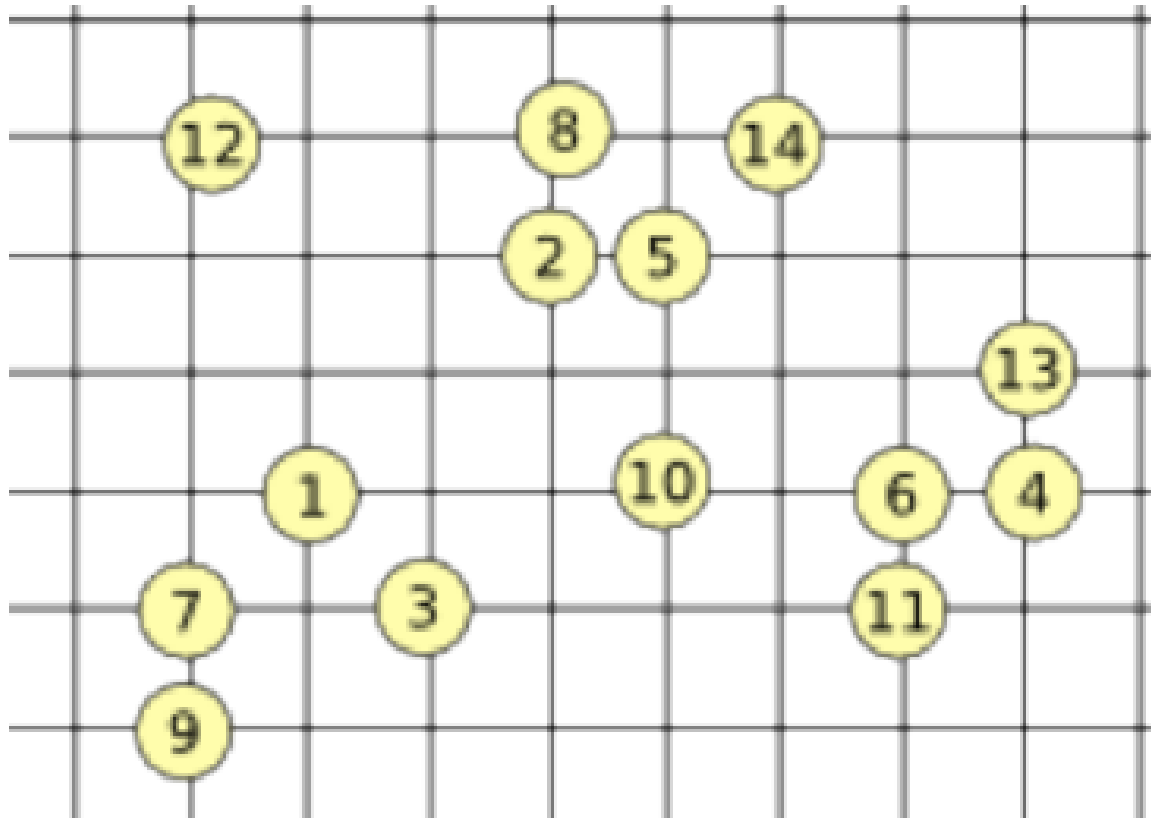
**LOF(10) = ((Lrd(2) + Lrd(5) + Lrd(6) + Lrd(11) + Ldr(3)) / 5 ) / Lrd(10) = 0.67 / 0.46 = 1.45**

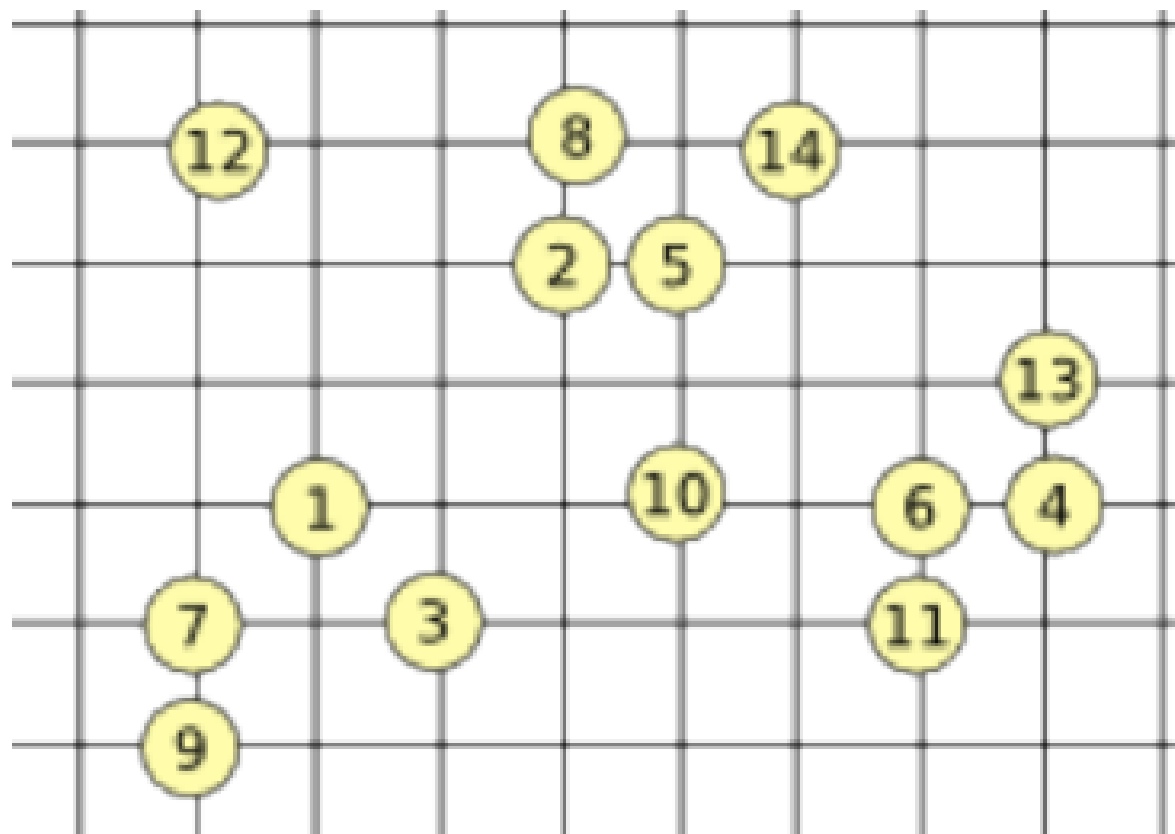
Here (10) has 3 points in a tie for the 3<sup>rd</sup> place in the 3-NN set (2, 3, 11), resulting in 5 NNs.

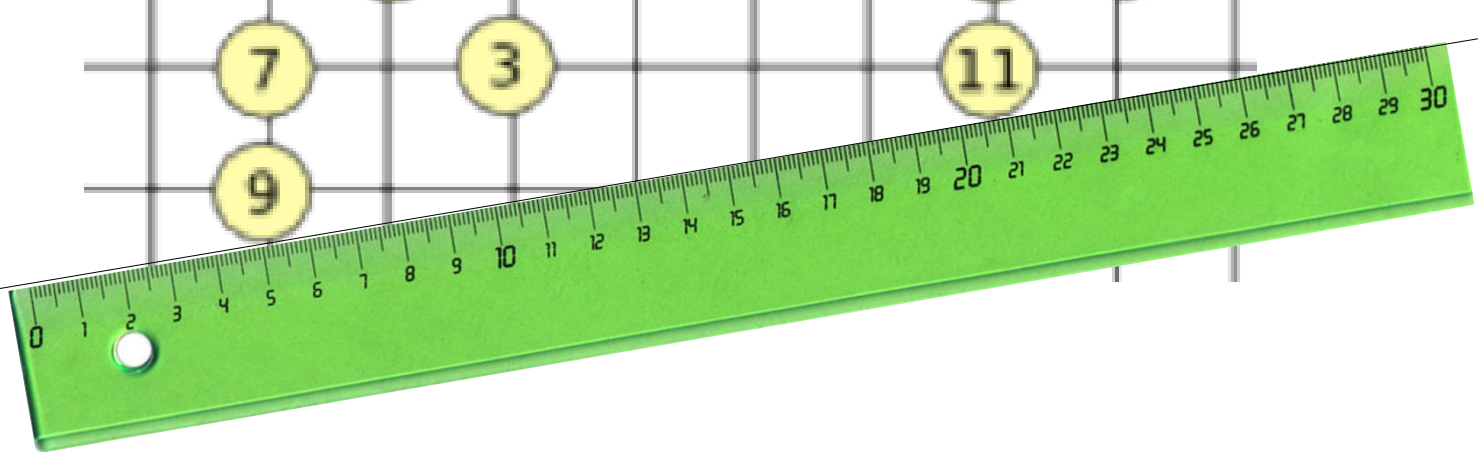
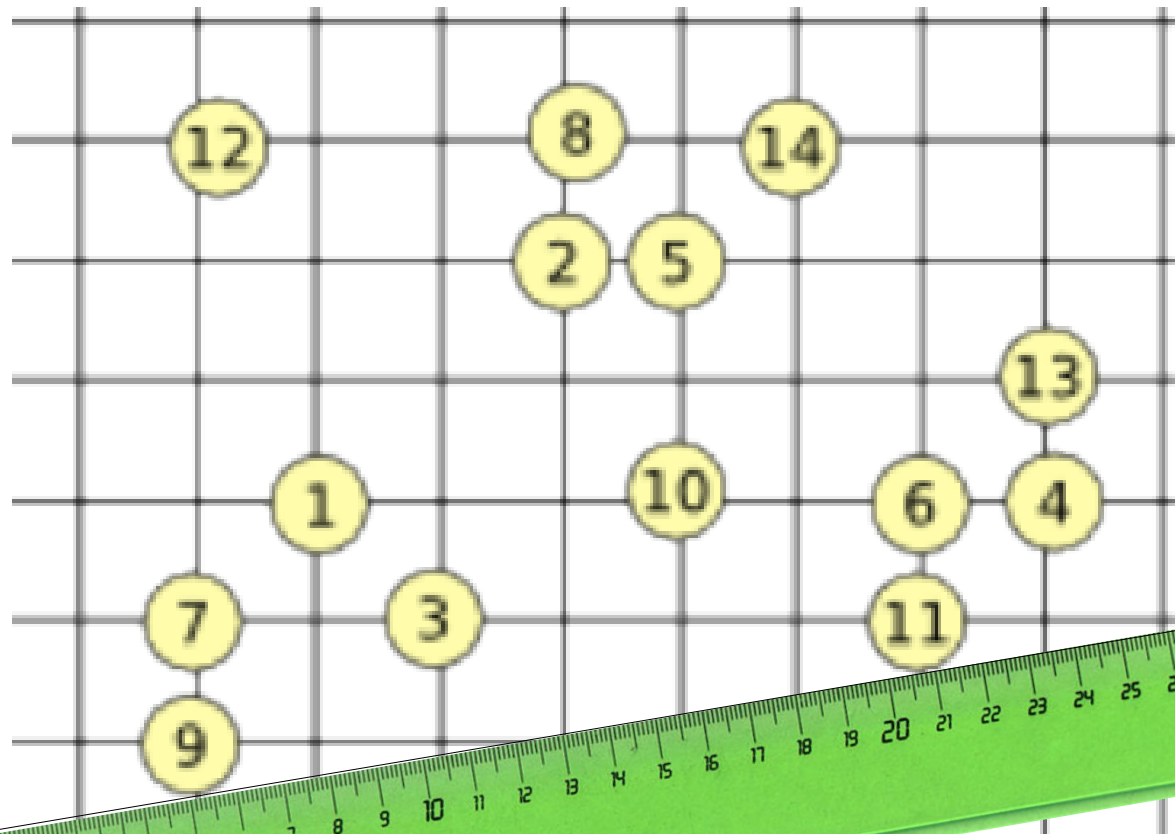


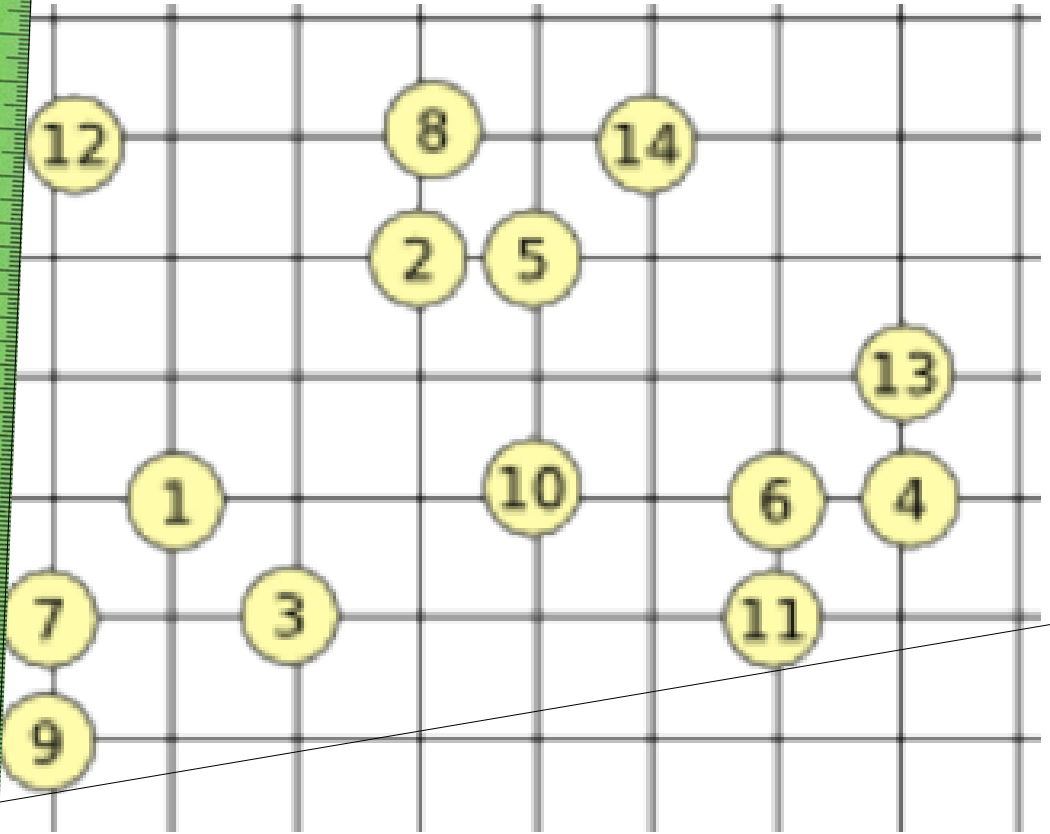
# Depth-based outliers

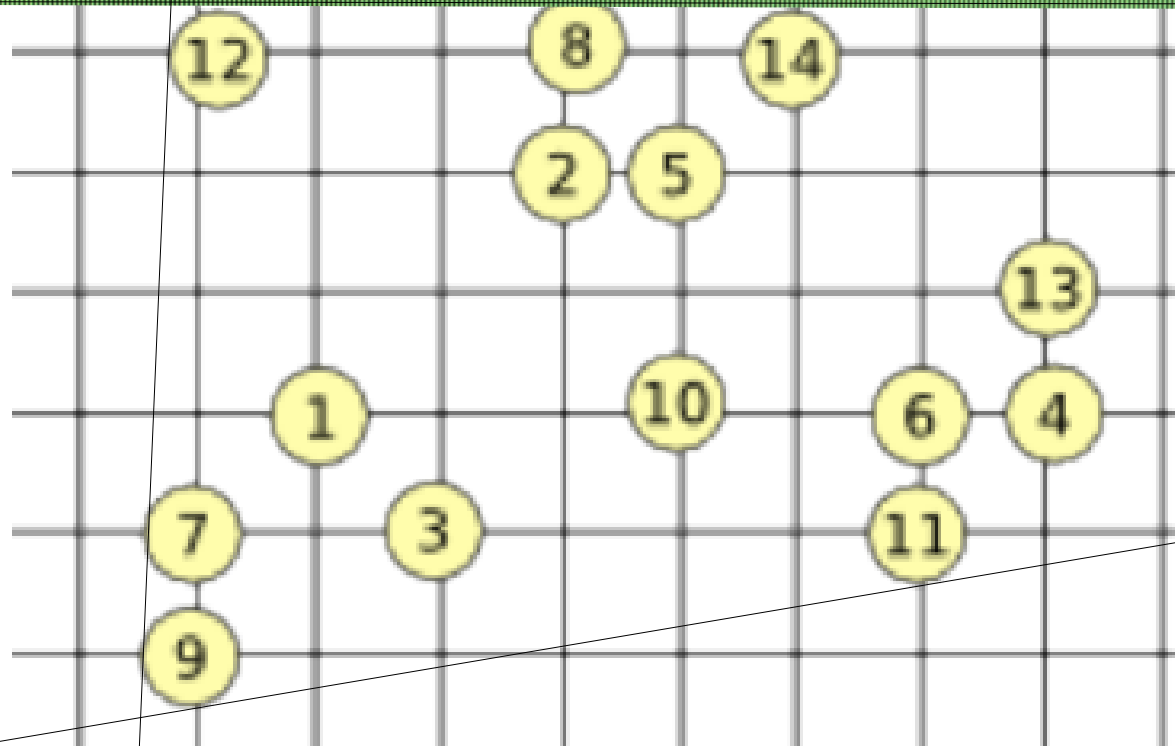
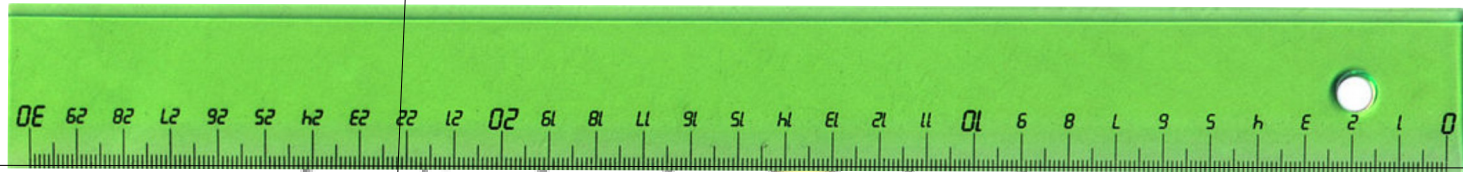
- Compute the depth score of each point

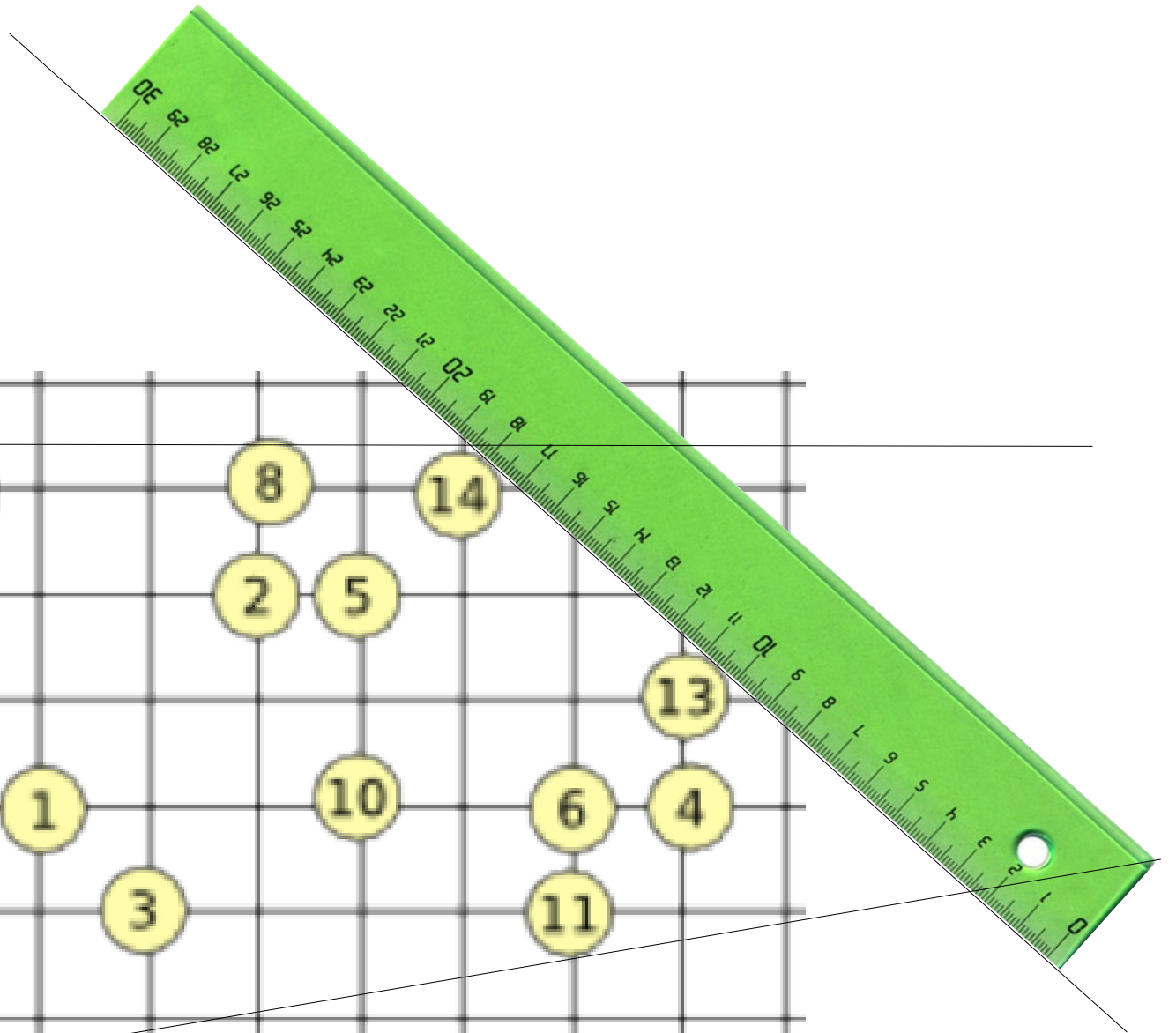
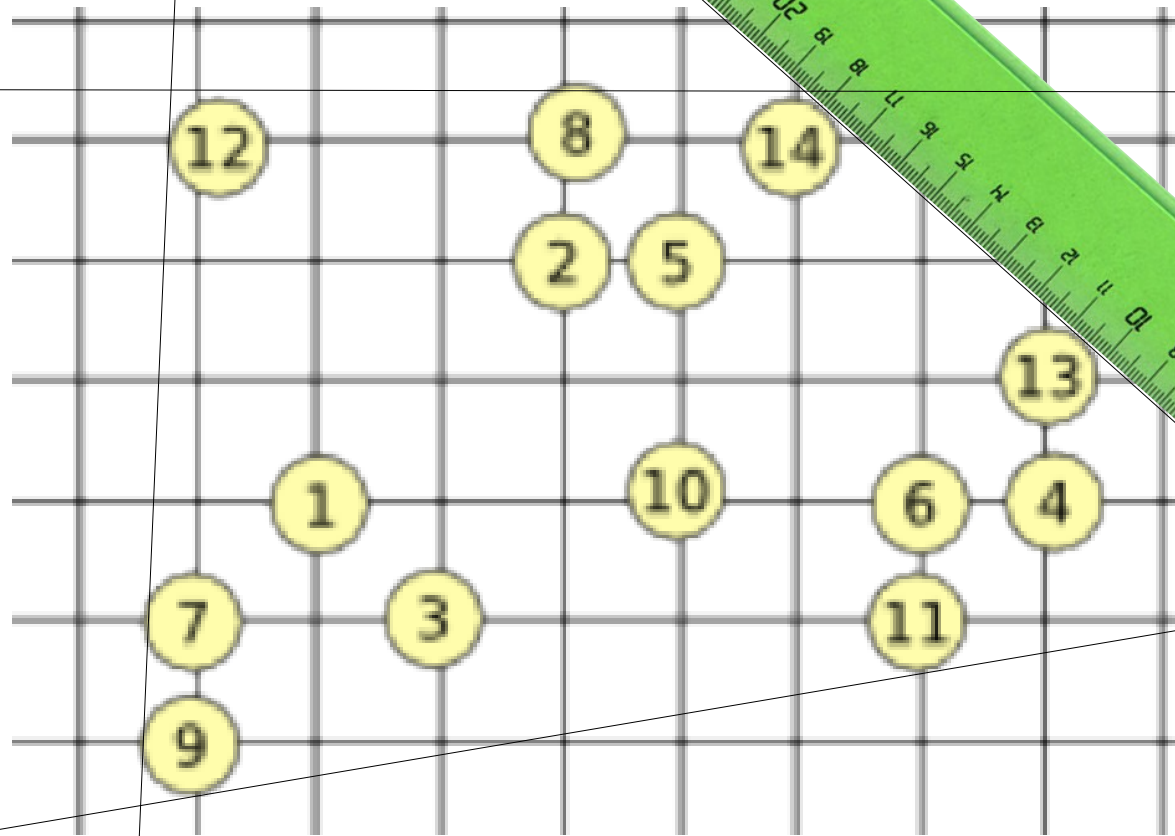




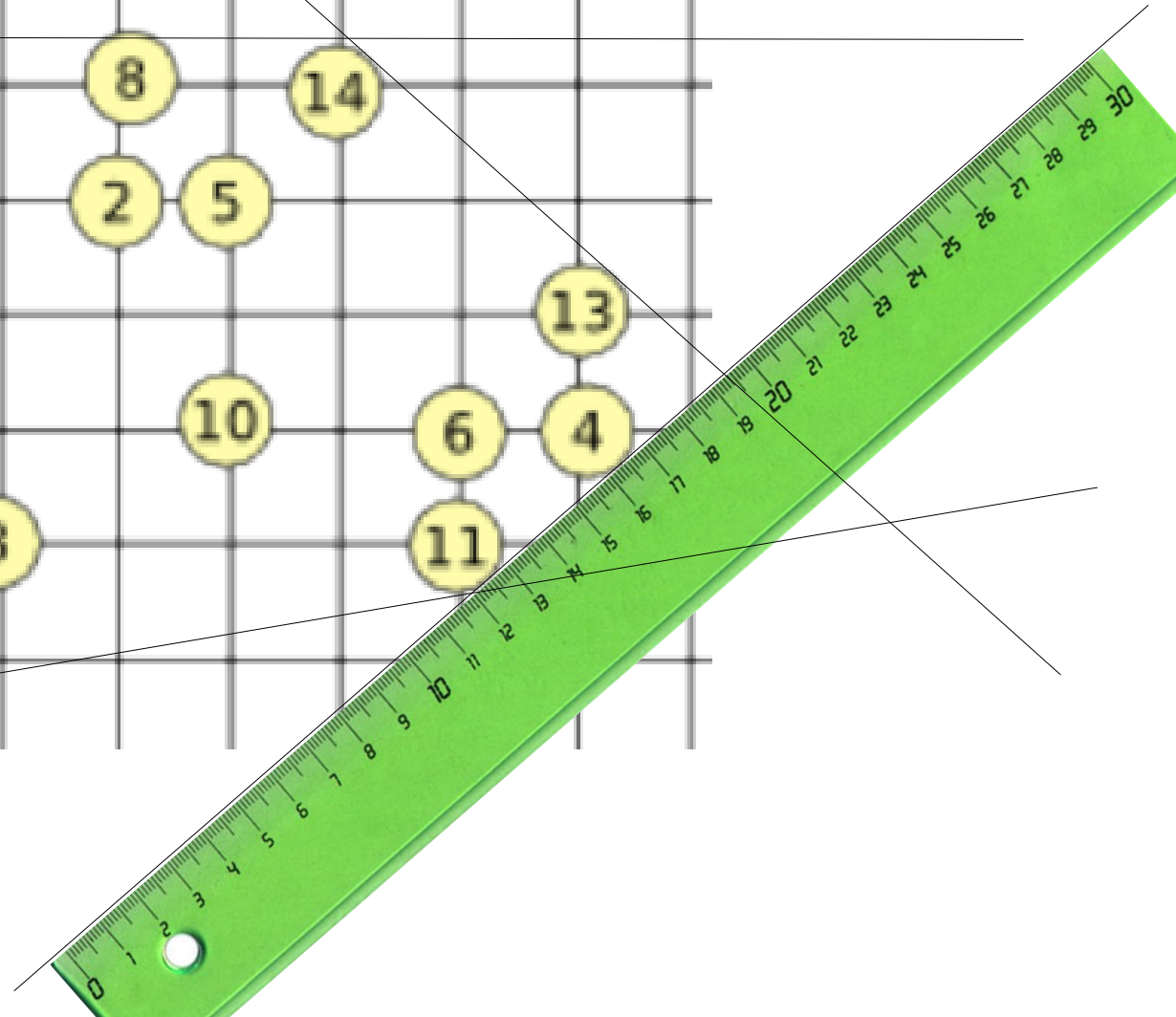
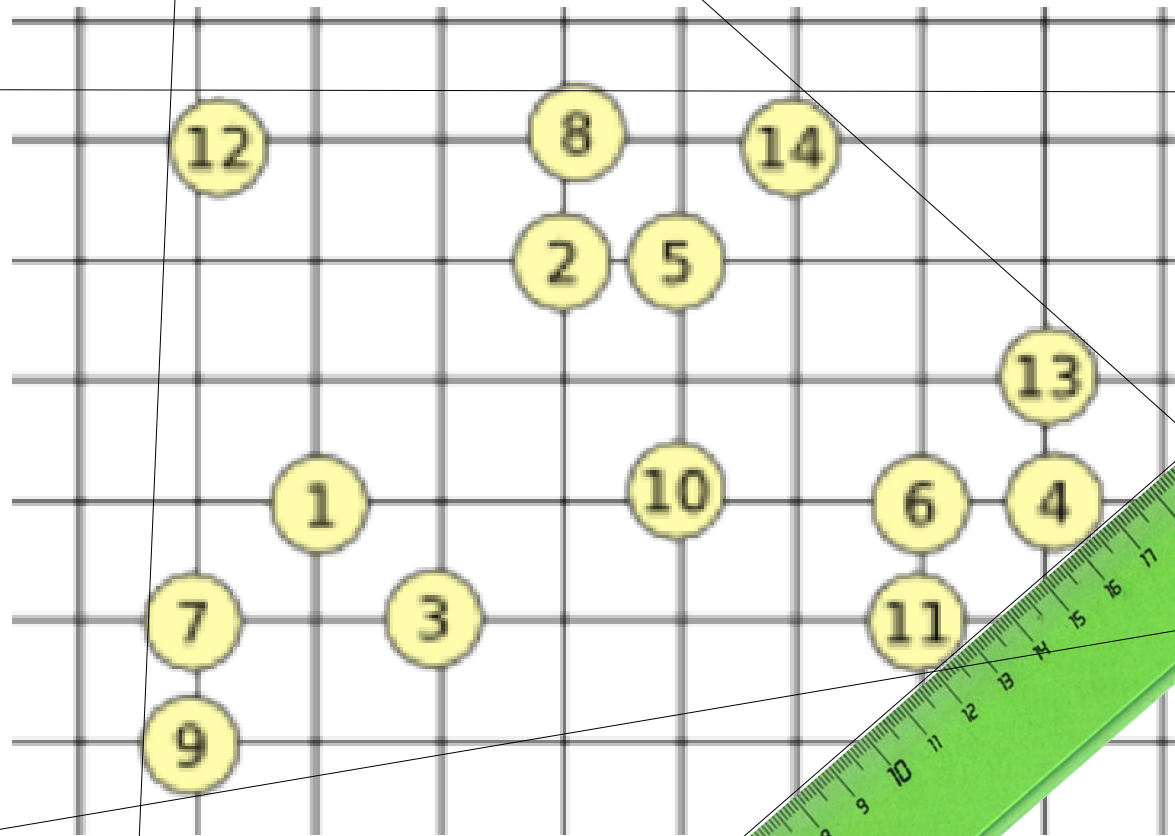


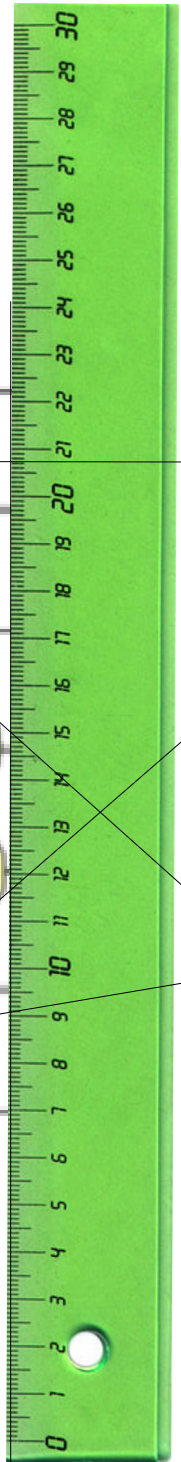
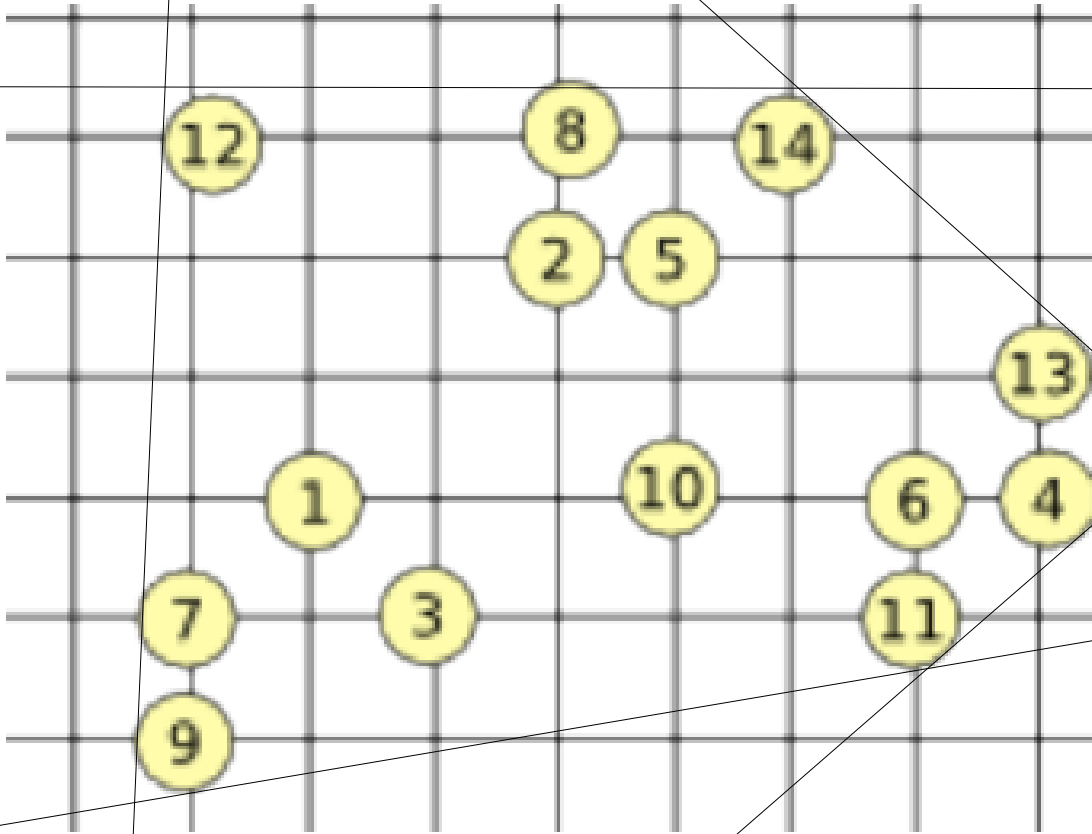


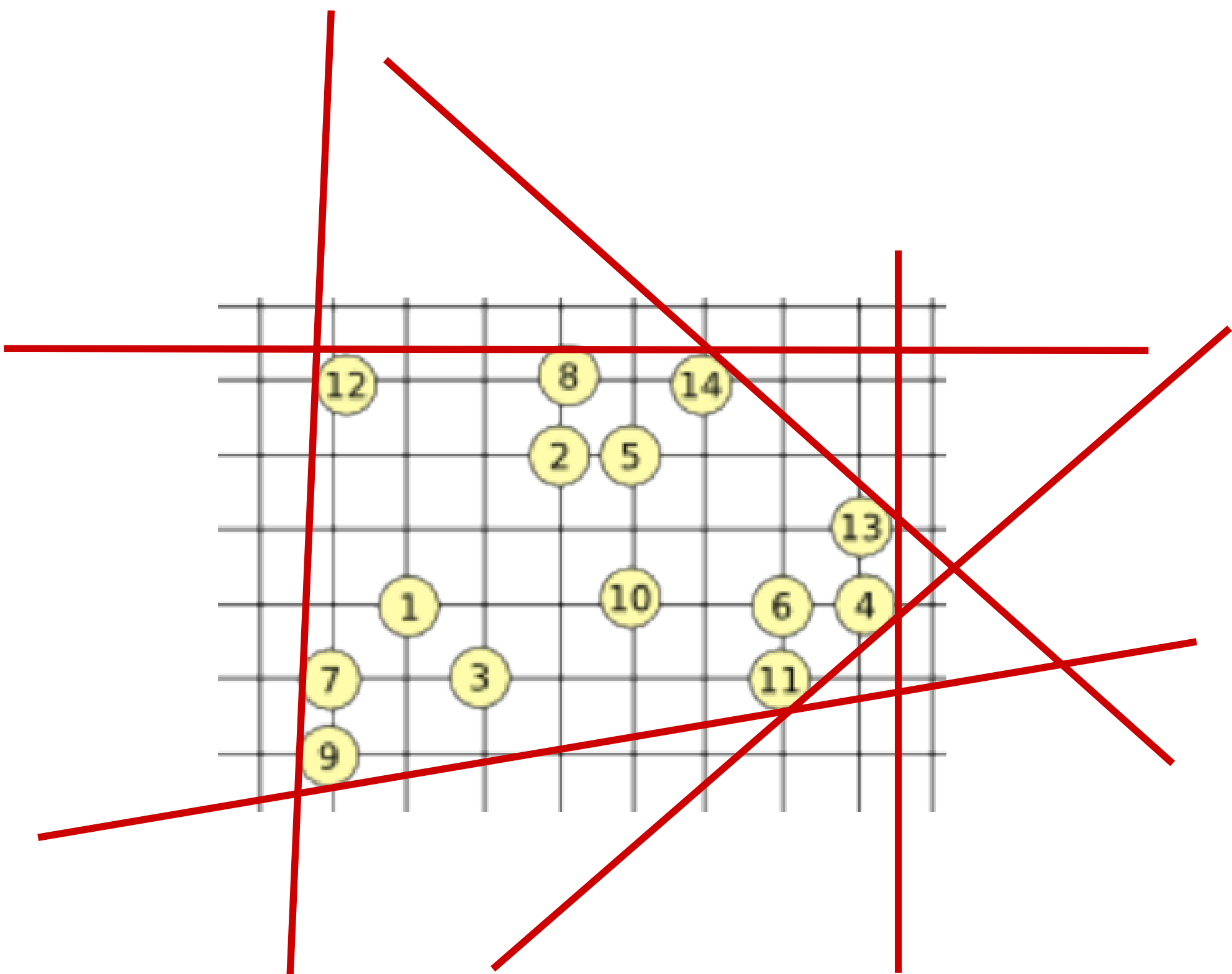




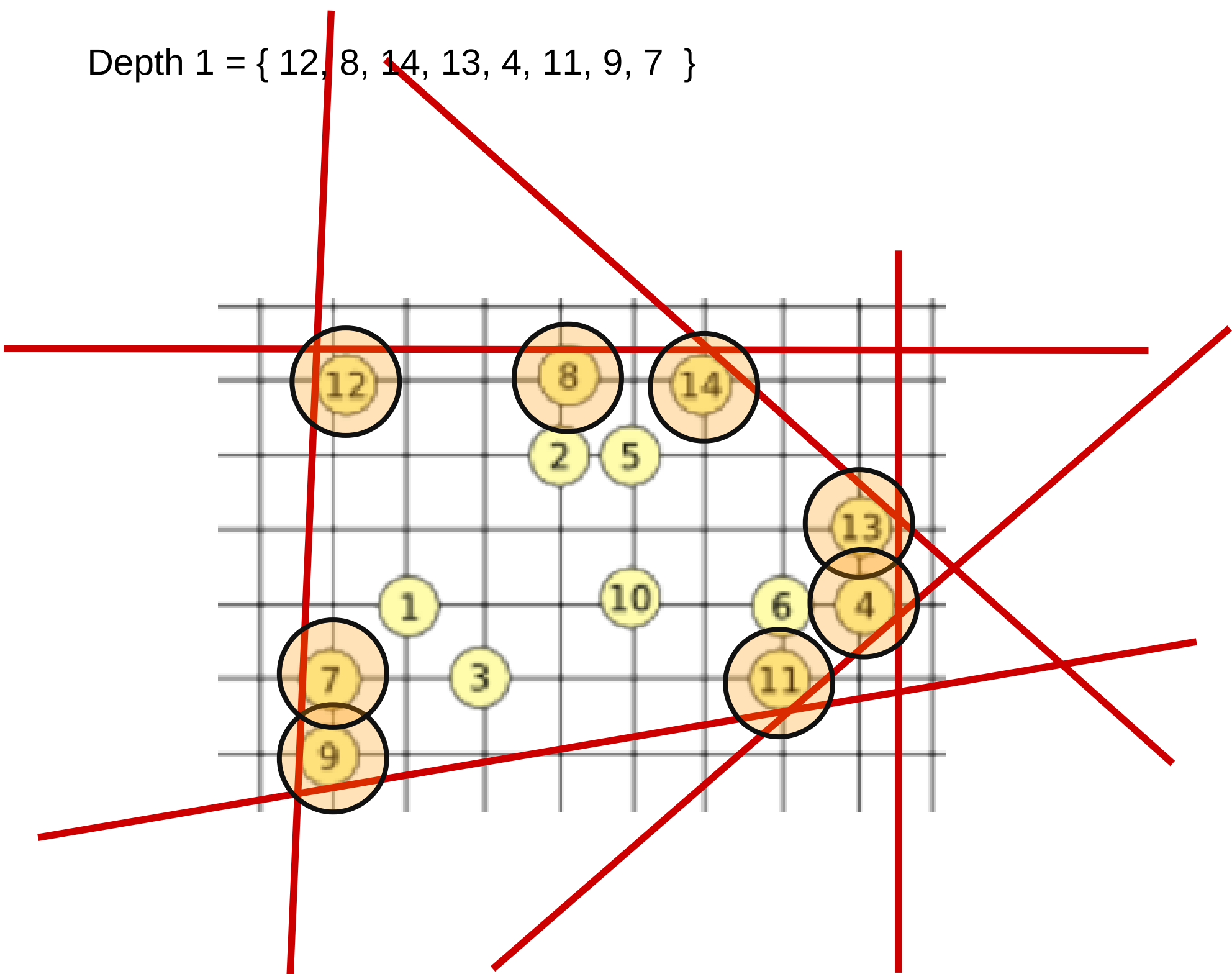




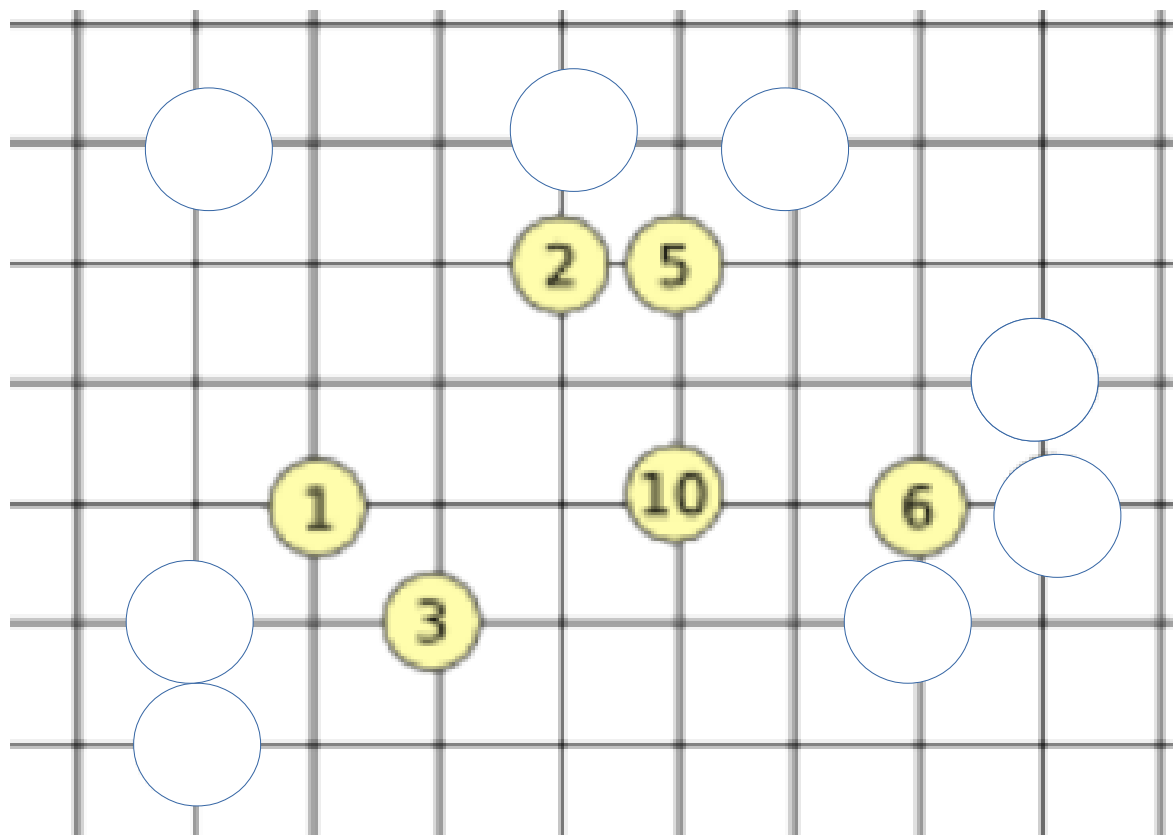




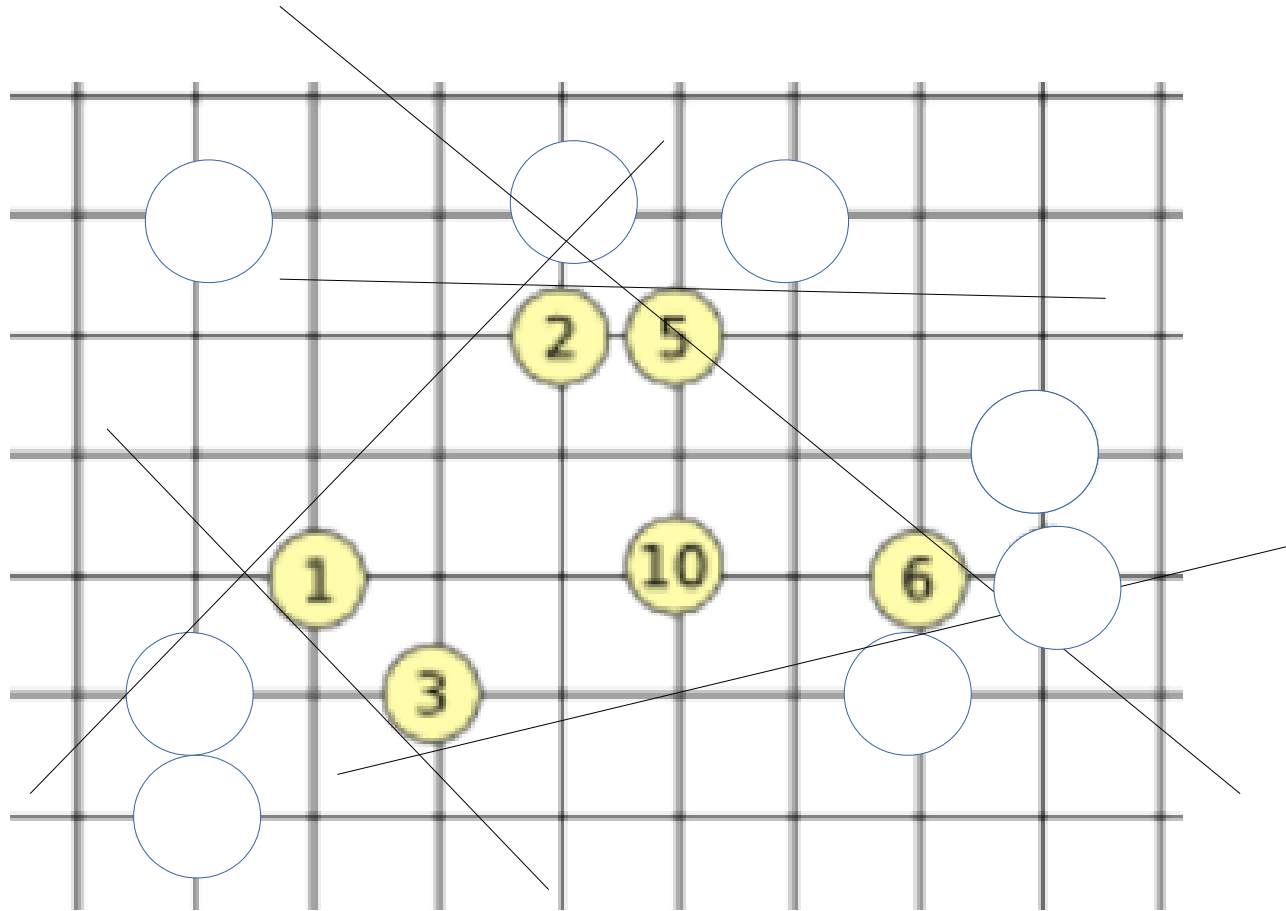
Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }



Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }

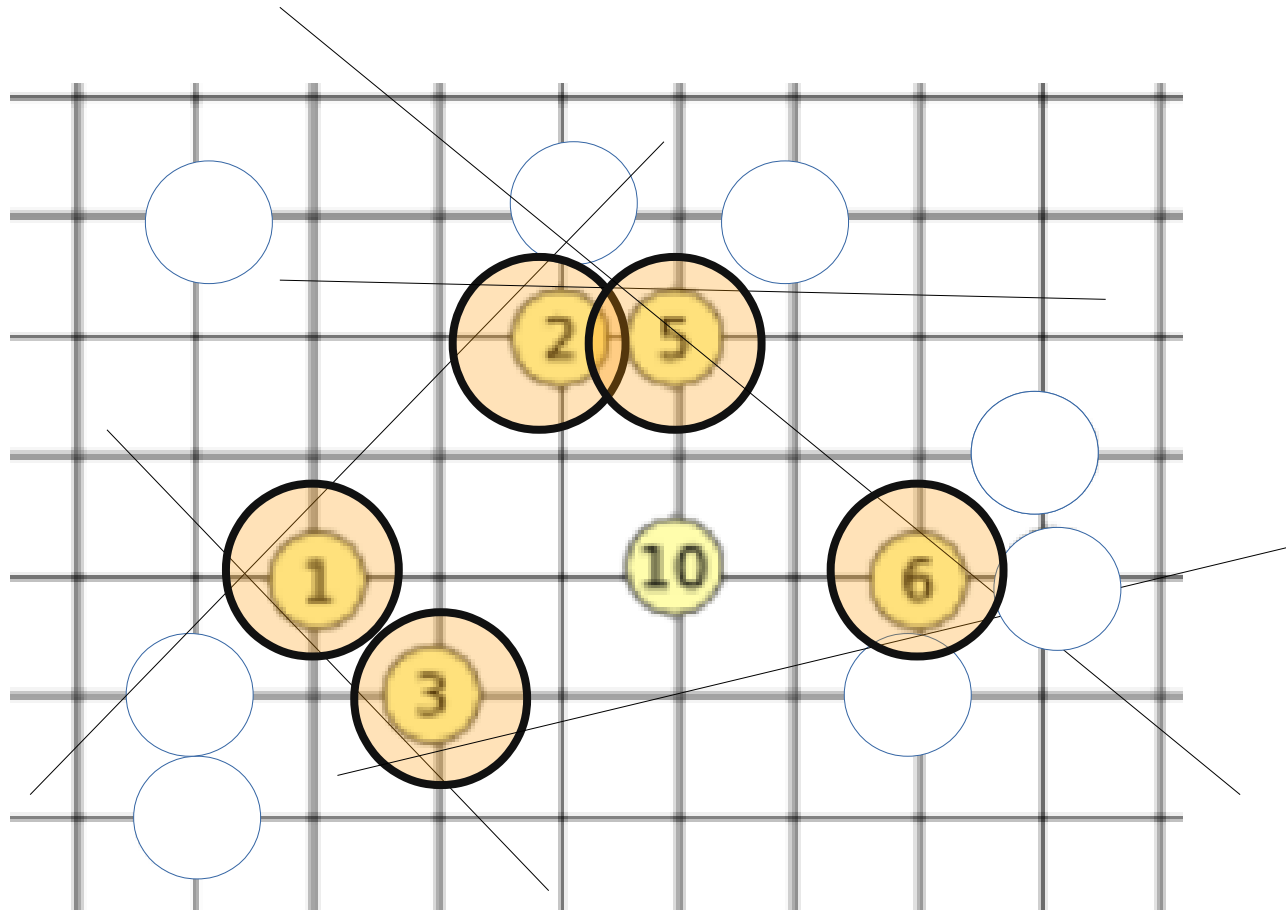


Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }



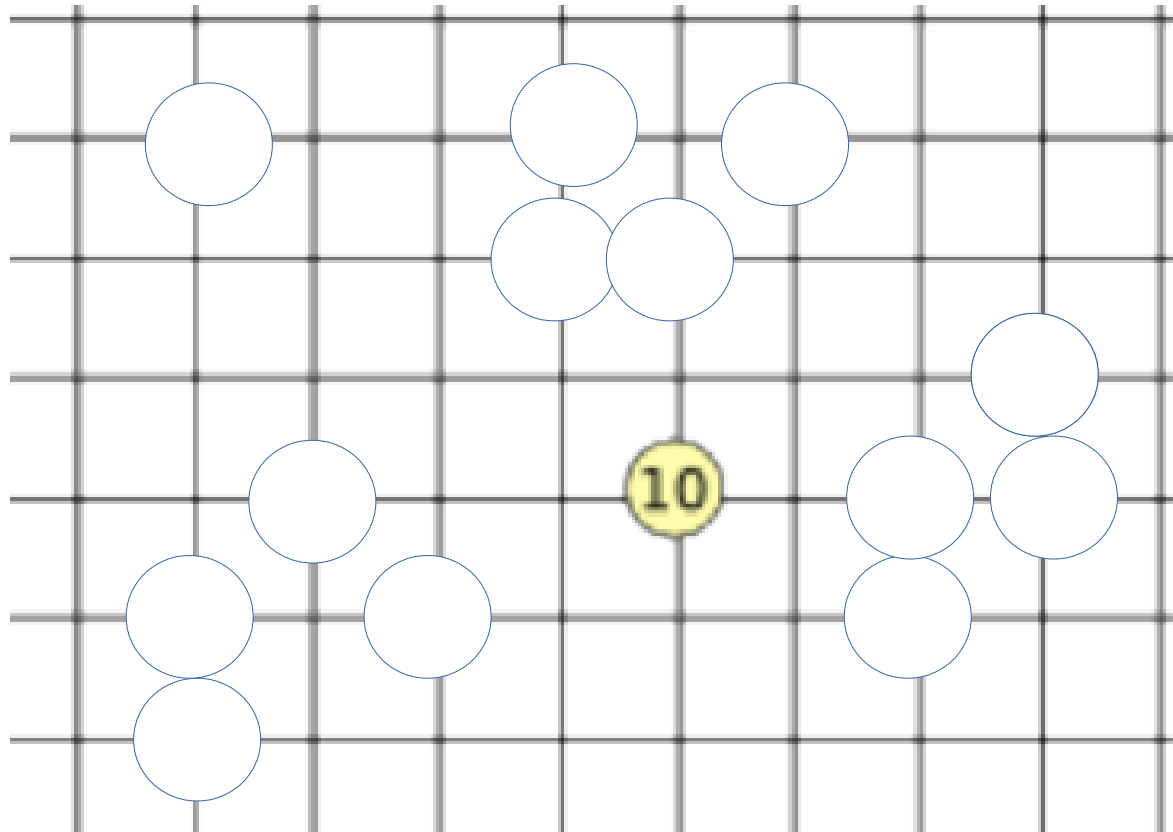
Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }

Depth 2 = { 2, 5, 6, 3, 1 }



Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }

Depth 2 = { 2, 5, 6, 3, 1 }

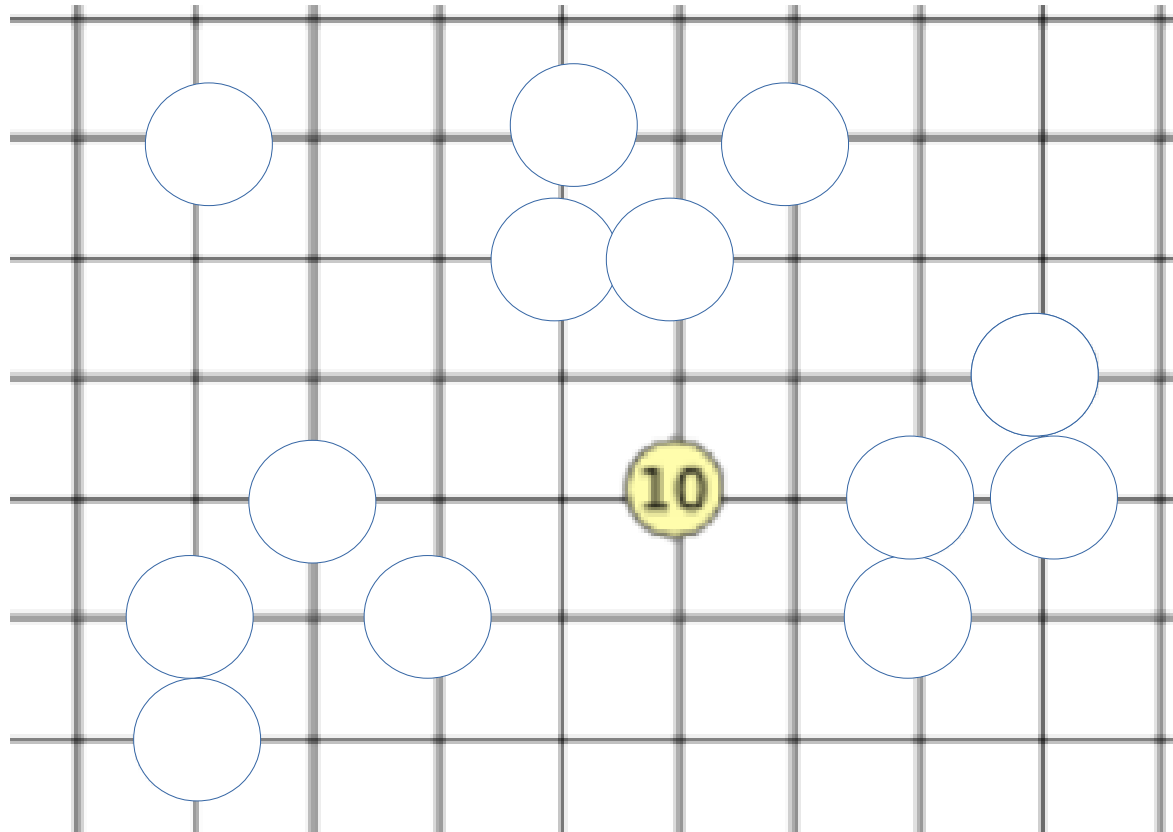




Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }

Depth 2 = { 2, 5, 6, 3, 1 }

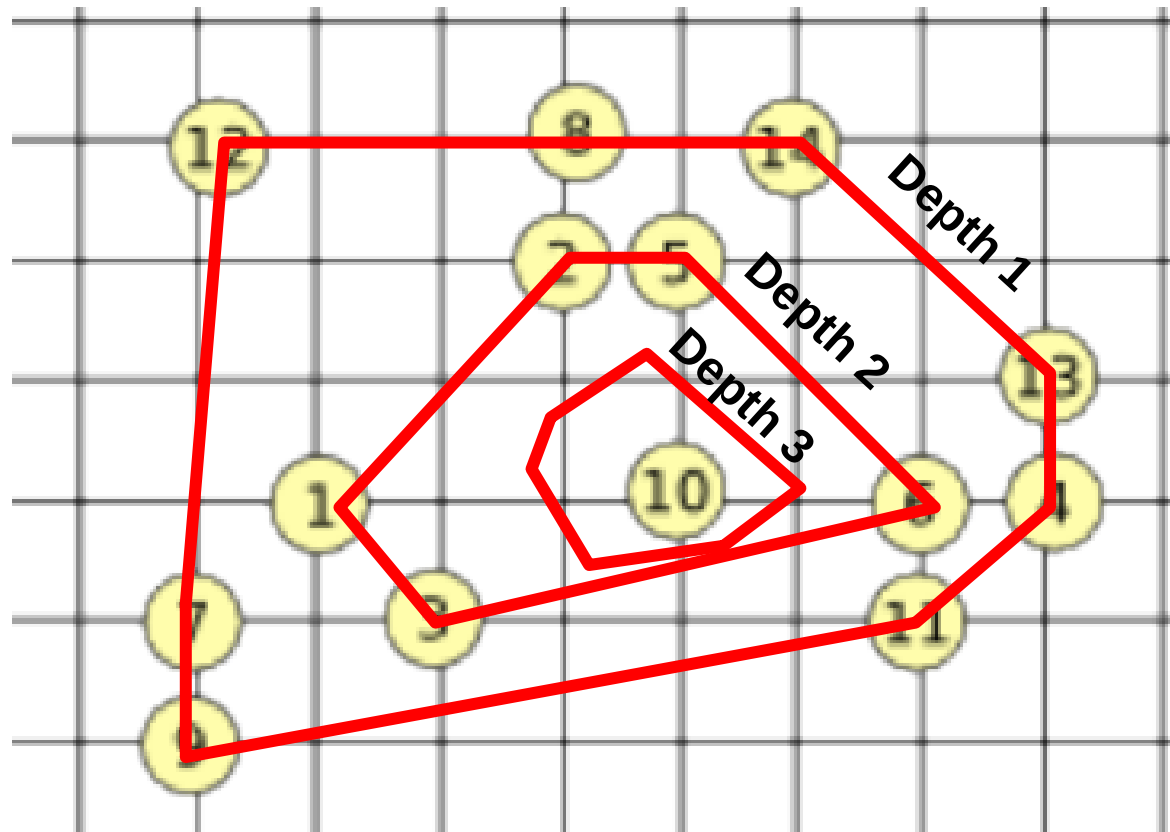
Depth 3 = { 10 }



Depth 1 = { 12, 8, 14, 13, 4, 11, 9, 7 }

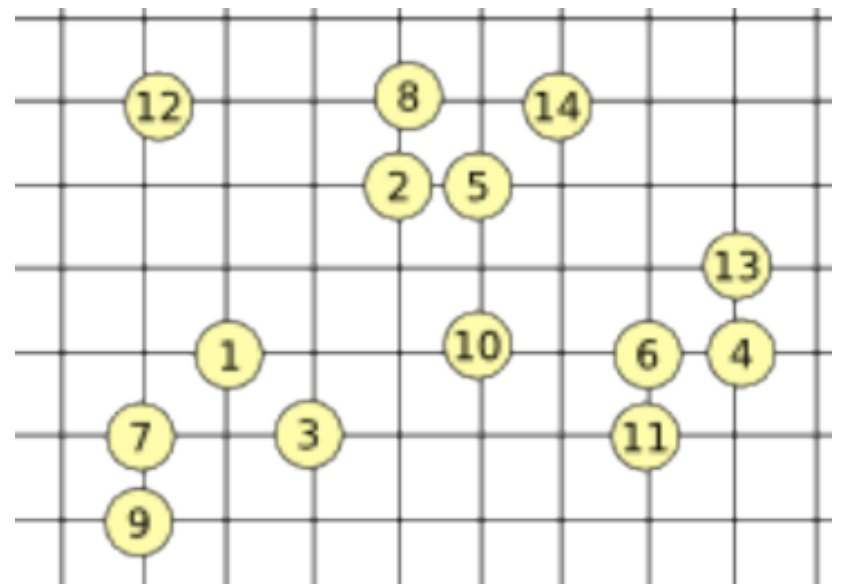
Depth 2 = { 2, 5, 6, 3, 1 }

Depth 3 = { 10 }



# Angle-based outliers

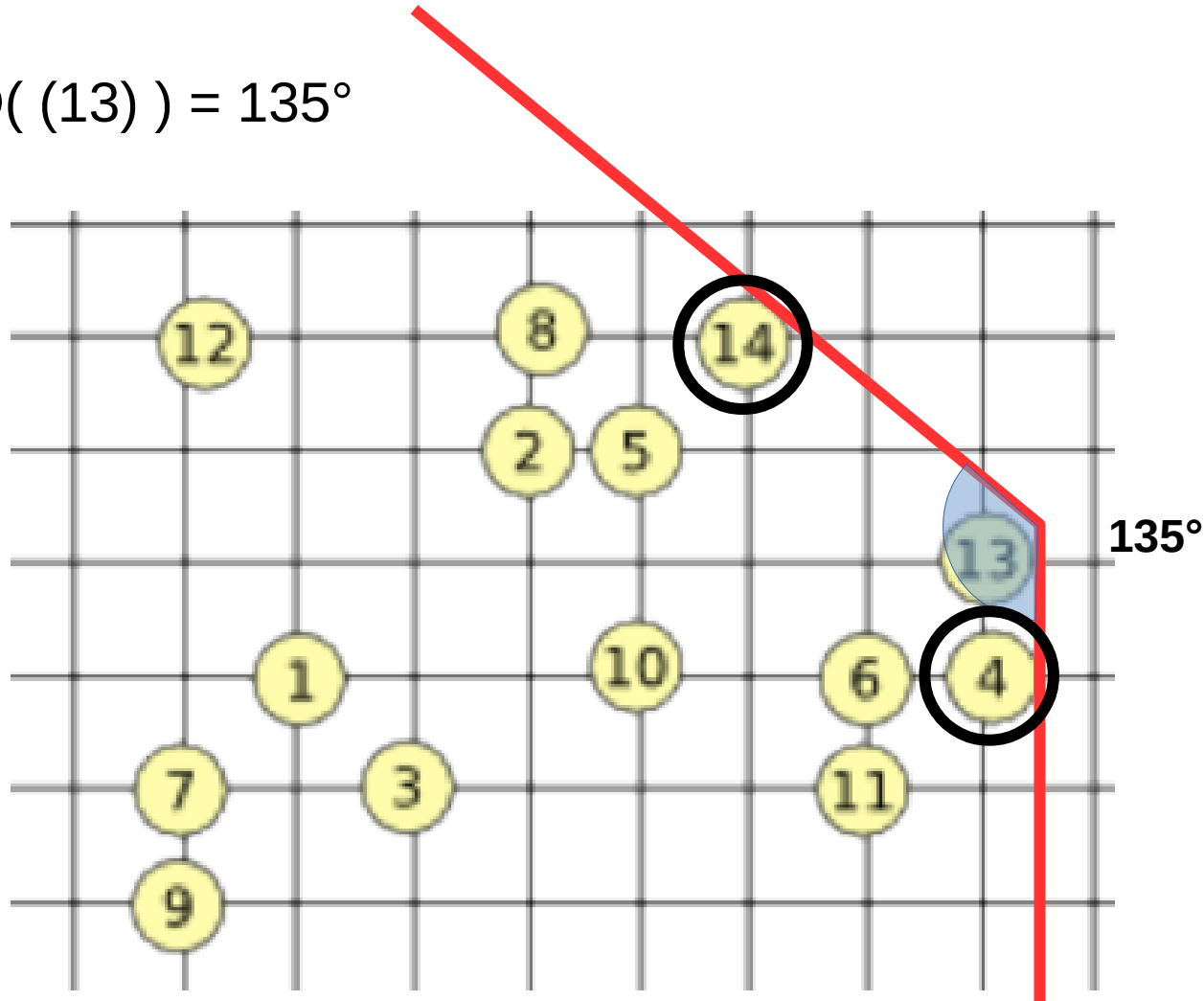
- Compute the simplified ABOD for each point
  - $\text{Simple-ABOD}(p) = \text{MAX}_{x,y \text{ in DB}} \{ \text{angle}(x\bar{p}y) \}$
- Equivalent to find the smallest angle in  $p$  that contains the whole dataset



# Angle-based outliers

$$\text{Simple-ABOD}(p) = \text{MAX}_{x,y \text{ in DB}} \{ \text{angle}(x\bar{p}y) \}$$

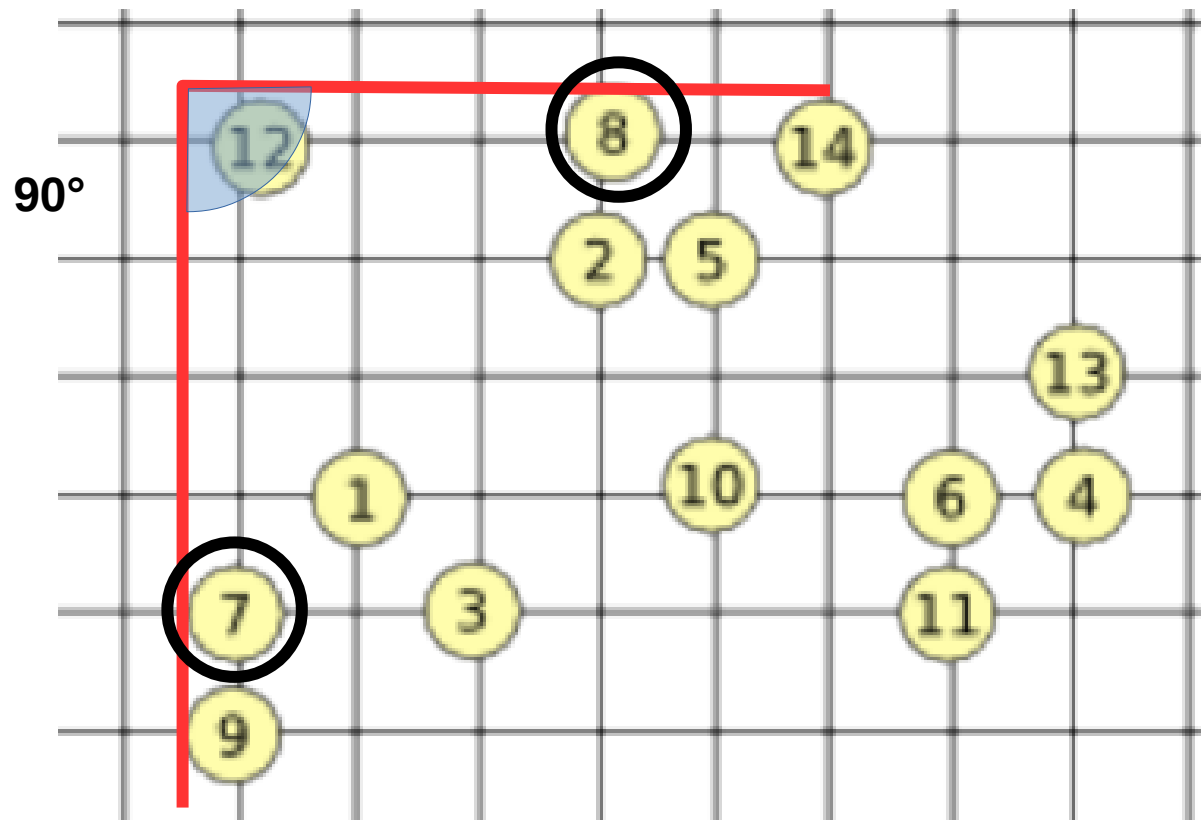
Simple-ABOD( (13) ) = 135°



# Angle-based outliers

$$\text{Simple-ABOD}(p) = \text{MAX}_{x,y \text{ in DB}} \{ \text{angle}(x\bar{p}y) \}$$

$$\text{Simple-ABOD}(12) = 90^\circ$$



# Angle-based outliers

$$\text{Simple-ABOD}(p) = \text{MAX}_{x,y \text{ in DB}} \{ \text{angle}(x\bar{p}y) \}$$

$$\text{Simple-ABOD}( (10) ) = 180^\circ$$

