

Pseudo-code

Algorithm 1 Finding maximal disks following a parallel approach.

Input: Set of points T ,
Maximum distance ϵ ,
Minimum number of points per disk μ
Output: Set of maximal disks M

```
pointsRT  $\leftarrow$  build.RTree( $T$ )  
 $P \leftarrow$  find.Pairs(pointsRT,  $\epsilon$ )  
 $C \leftarrow \emptyset$   
for each  $p_i$  in  $P$  do  
   $c_i^1, c_i^2 \leftarrow$  compute.Disks( $p_i, \epsilon$ )  
   $C.Add(c_i^1, c_i^2)$   
end for  
 $D \leftarrow \emptyset$   
for each  $c_i$  in  $C$  do  
   $d_i \leftarrow$  find.PointsAround( $c_i, \epsilon$ )  
  if  $|d_i| \geq \mu$  then  
     $D.Add(d_i)$   
  end if  
end for
```

Proof Sketch

It is assumed that the stages of pairs finding, disks generation and ‘less-than- μ ’ disk filtering have already been done. We start with a set of candidate disks which are represented by their centers and the set of original points which are enclosed by them. It is expected that they are organized in an R-Tree index as it is shown in figure 1. The following procedure aims to filter out those disks which set of points are a subset of another disk’s set of points. Just disks with a superset of points are kept and they are known as maximal disks.

There are two problems to report maximal disks following a parallel approach. First, it can be disks with the same set of points which are reported as maximals in different partitions. We should report only one of them. Second, it can be a disk with a subset of points which appears as maximal in one partition but there is another disk in another partition with a superset that contains that set. We should just report the disk with the superset of points. We will refer the first problem as *duplicate reporting* and the second one as *subset reporting*.

Duplicate reporting can be easily avoided by using a deterministic way to represent the center of each disk previous to the partitioning. It is, using an unique center for



Figure 1: An initial R-Tree over a set of candidate disks represented by their centers.

those disks with the same set of points. For example, changing the center of the disks to the centroid of the MBR of its points will lead to a unique representation of the disks. The change of the center does not affect the logic of the procedure due to the set of points of the disks remains the same.

Subset reporting demands more attention. In order to avoid this problem, we have to evaluate disks which potentially can contain the set of points of other disks. Keep in mind that just disks which intersect each other should be evaluated as that means that they have points in common. From figure 1, it is clear than disks close to the edge of each partition¹ could intersect other disks in the edge of contiguous partitions. Specifically, disks lying ϵ distance to the edge of their MBRs must be further analyzed but note that those remaining in the internal area are certainly safe.

As we have access to the MBRs, we perform an expansion applying a ϵ buffer around each of them (see figure 2). As the radius of each disk is $\frac{\epsilon}{2}$, it is clear that a disk in any specific partition only can intersect other disks if they lie in the extension area of that partition.

We re-index the original set of candidate disks using the new expanded MBRs. For those disks which intersect multiple expanded MBRs, a copy of the disk is sent to each of them. If a disk lie in the expansion area of one of the new MBR, it is marked accordingly. A disk can lie in the expansion area of multiple MBRs but it will be outside of the expansion area of its original MBR.

Now we can run a maximal pattern algorithm to find supersets of points and their correspond disks in each partition. Any disk with a subset of points will be obfuscated

¹In this context, we refer to partition or MBR indistinctly.

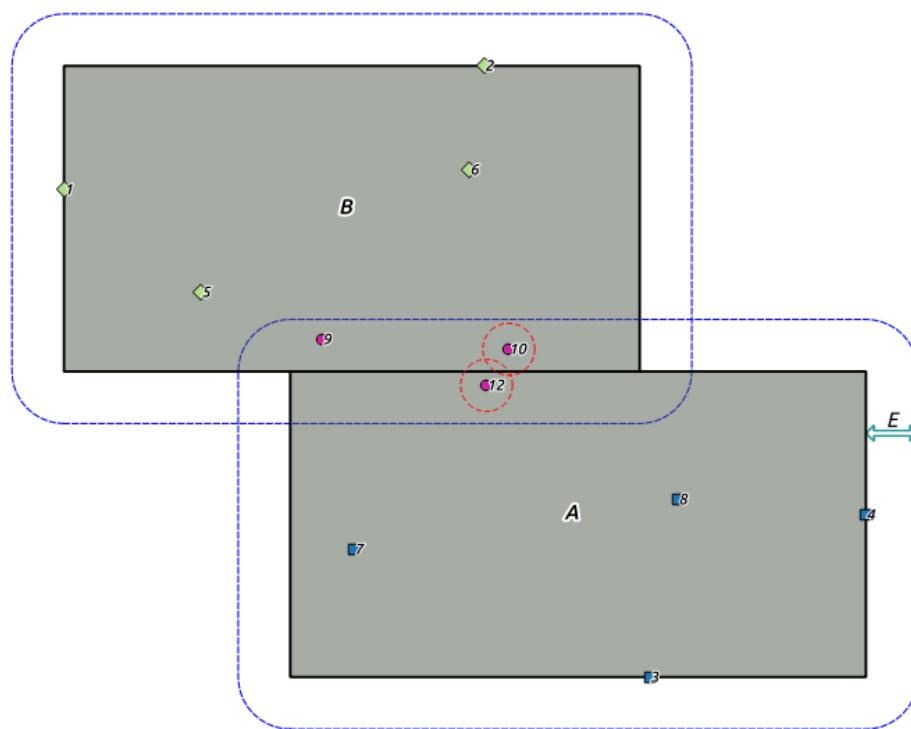


Figure 2: An ϵ expansion over the partitions of the R-Tree.

by its superset (if any) lying in the expansion area avoiding subset reporting. In order to avoid duplicate reporting we only report a maximal disk if it does not lie in the expansion area. As a disk will only be one time outside of an expansion area, we let this disk to be reported by its original partition.