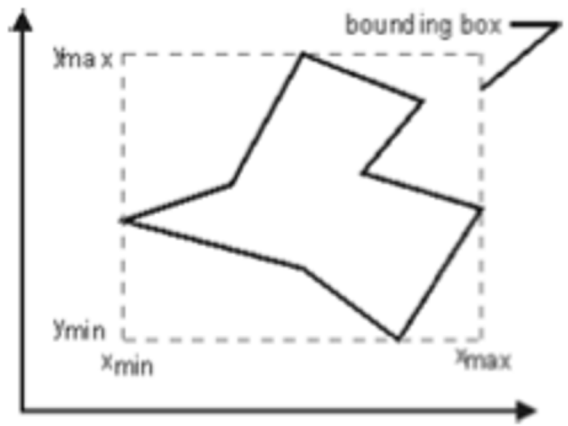
Filter and Refine Strategy

Geometry and its bounding rectangle



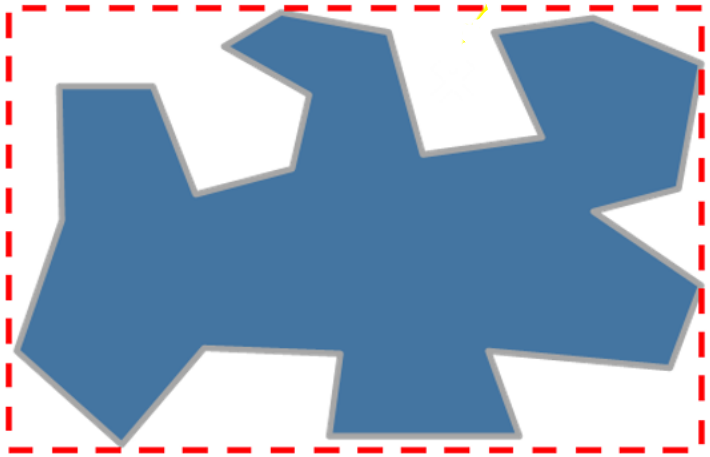
MBR = Rectangle
4 points in 2D

Polygon(20 30, 25 40, ..,)

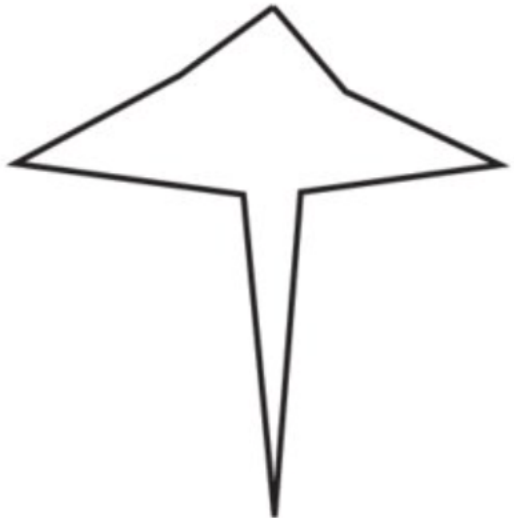
Lake with more than 100K pts

A polygon

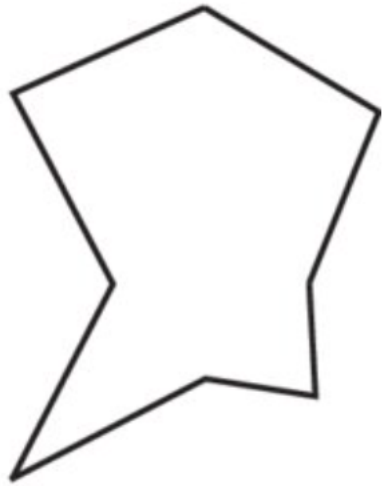
- Minimum bounding rectangle in red.



Minimum bounding rectangle



MBR Approximation



Compute a minimum bounding rectangle for

- POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))
- POLYGON ((x1 y1, x2 y2, , , x1 y1))

Approximating a geometry

- Less memory: Only two vertices to represent a rectangle

Approximating a geometry

- Positive

  Less memory: Only two vertices to represent a rectangle

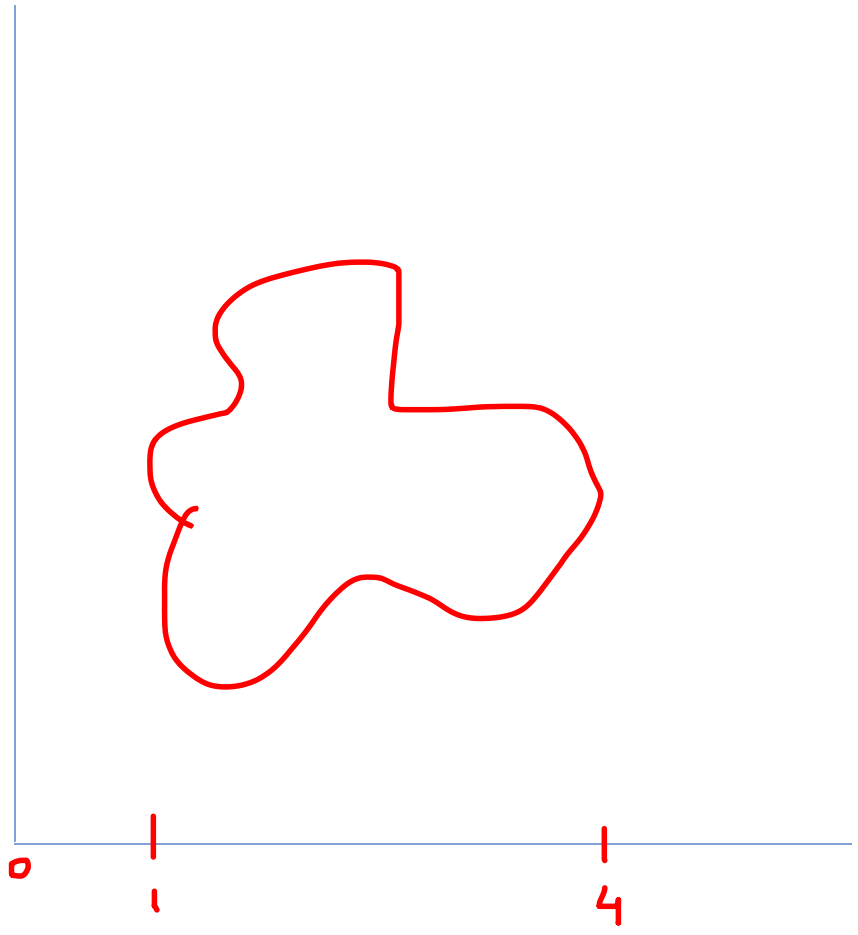
- Negative

  Dead space (empty area)

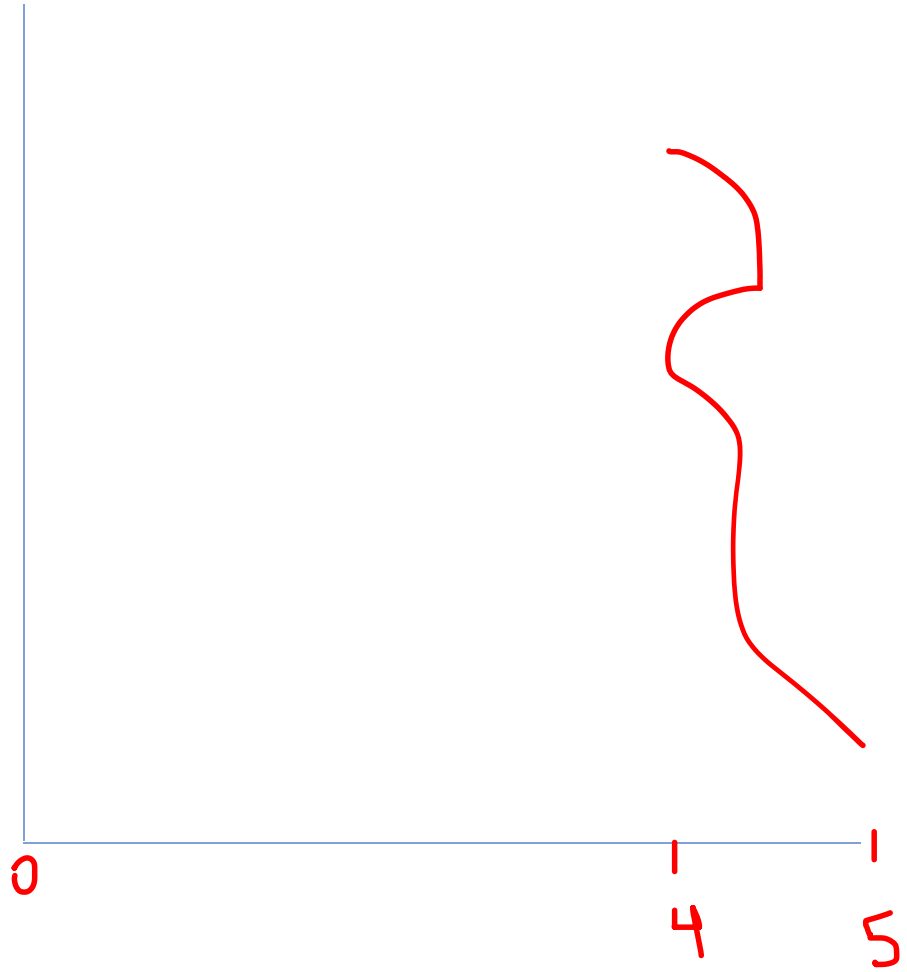
Any other approximation possible?

- For polygons

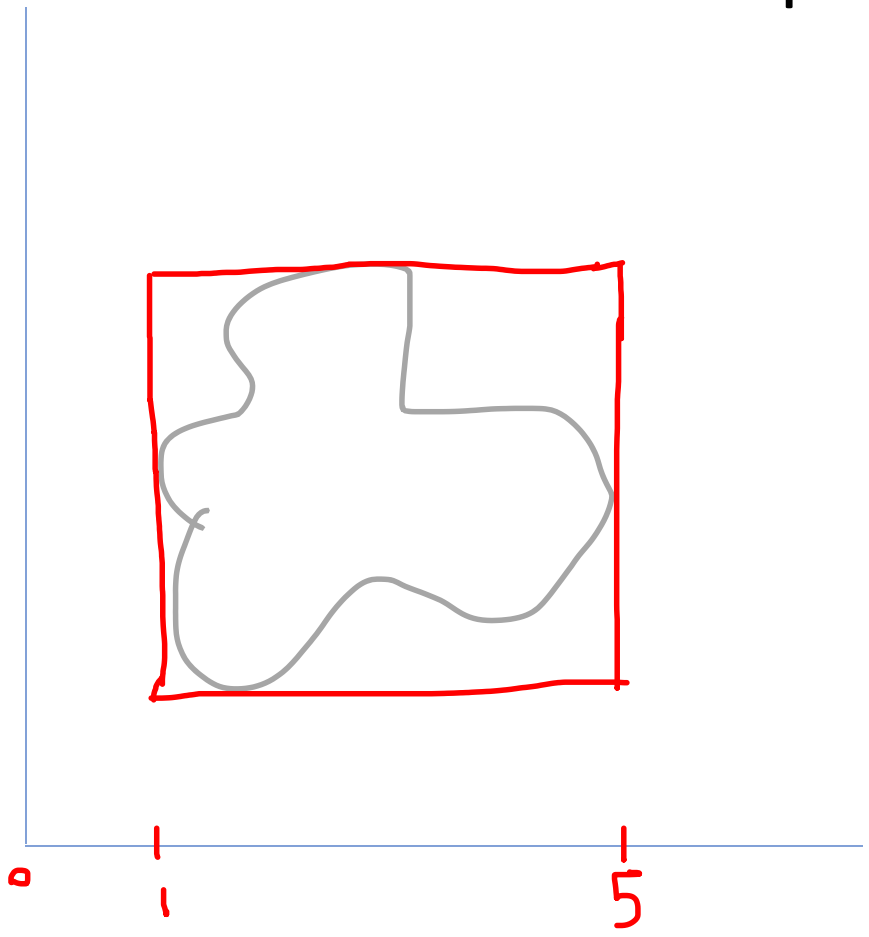
Is there an overlap between two geometries

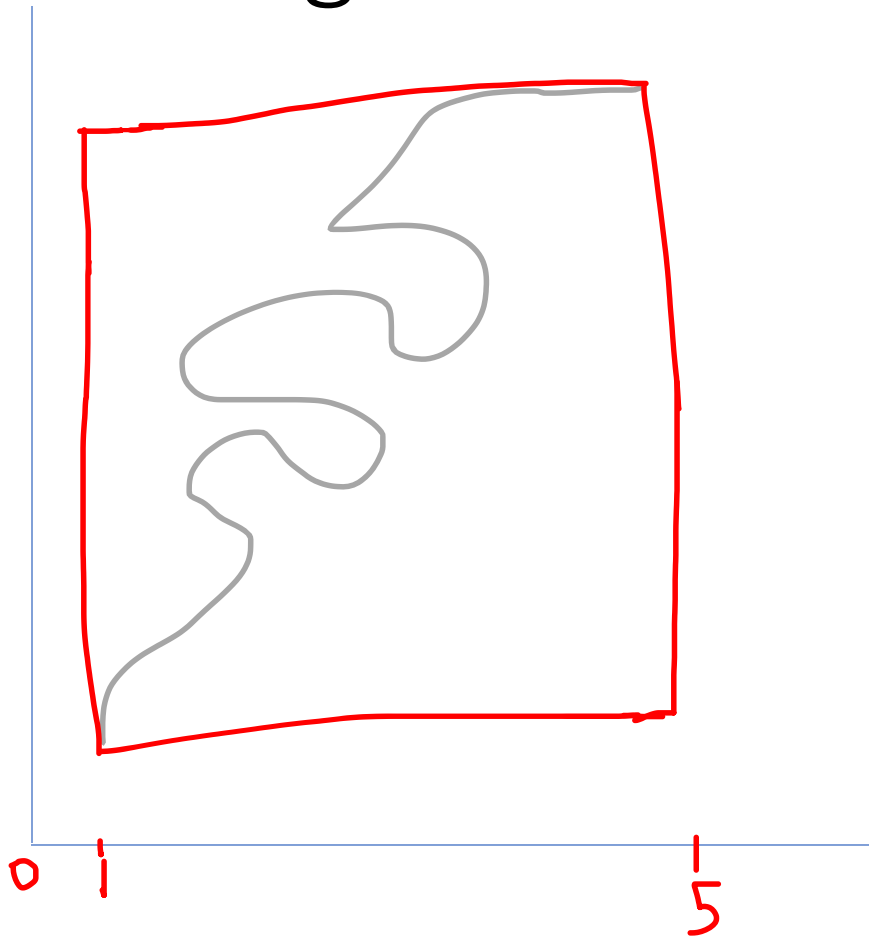


Is there an overlap between two geometries

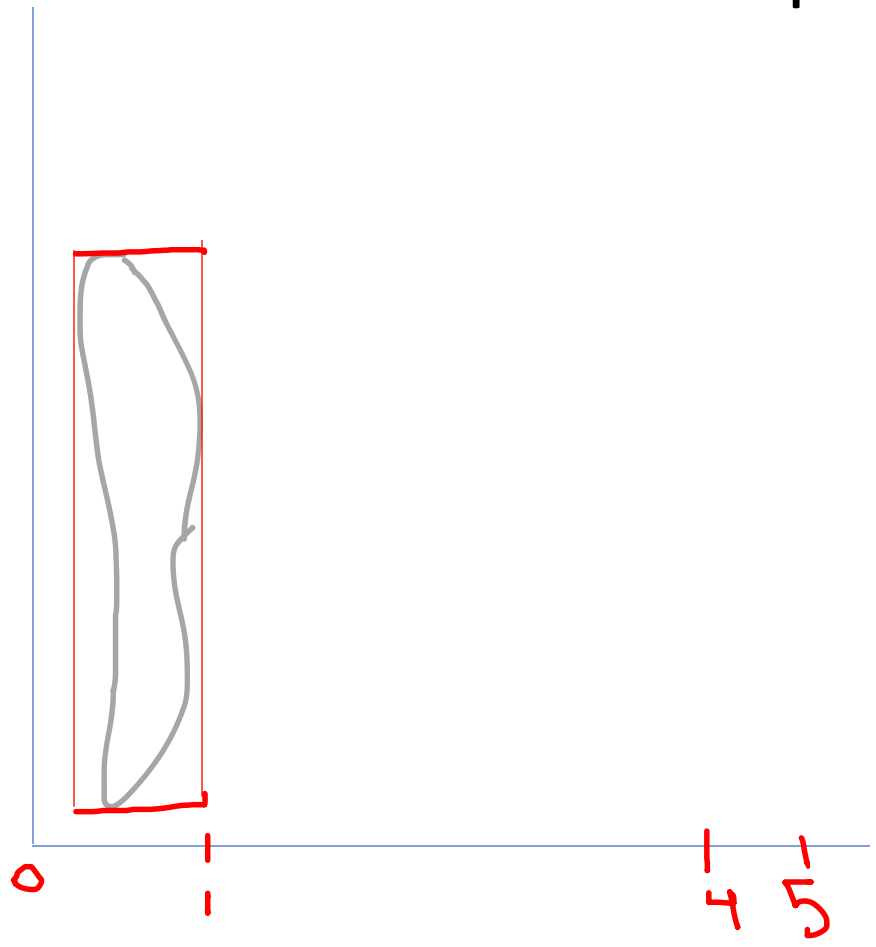


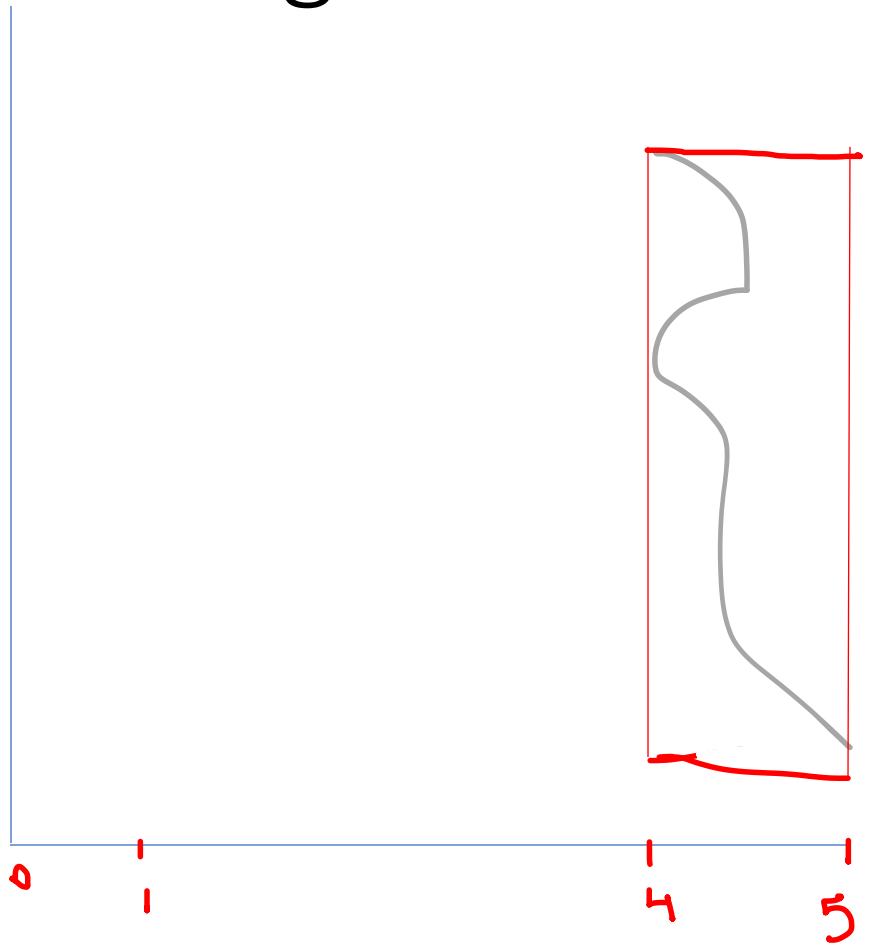
Using Red rectangle,
is there an overlap between two geometries





Using Min Bounding Rectangle, is there an overlap between two geometries





Filter and Refine Strategy

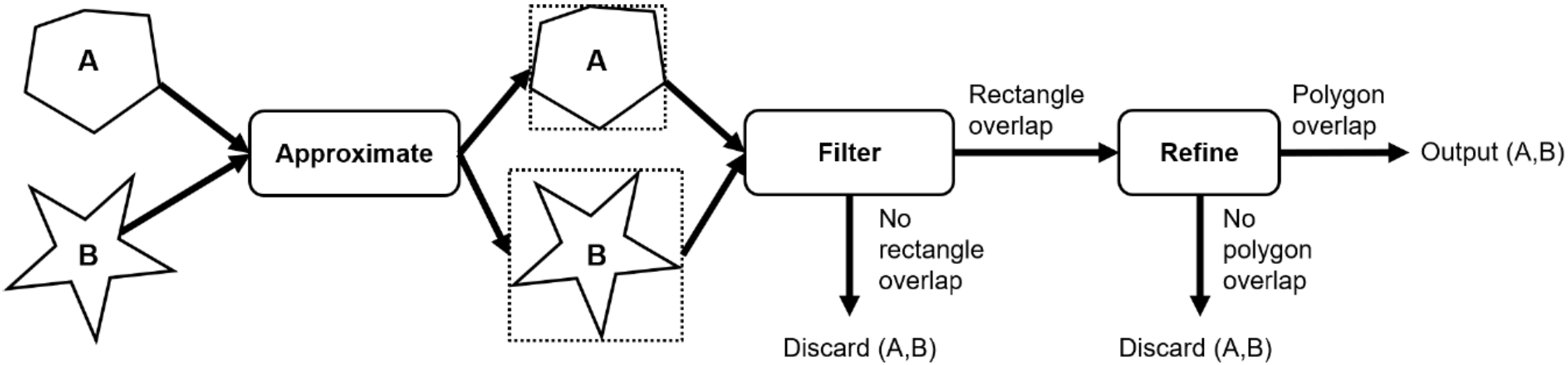


Fig. 1. The filter-and-refine strategy for one polygon pair.

Filter and Refine Strategy

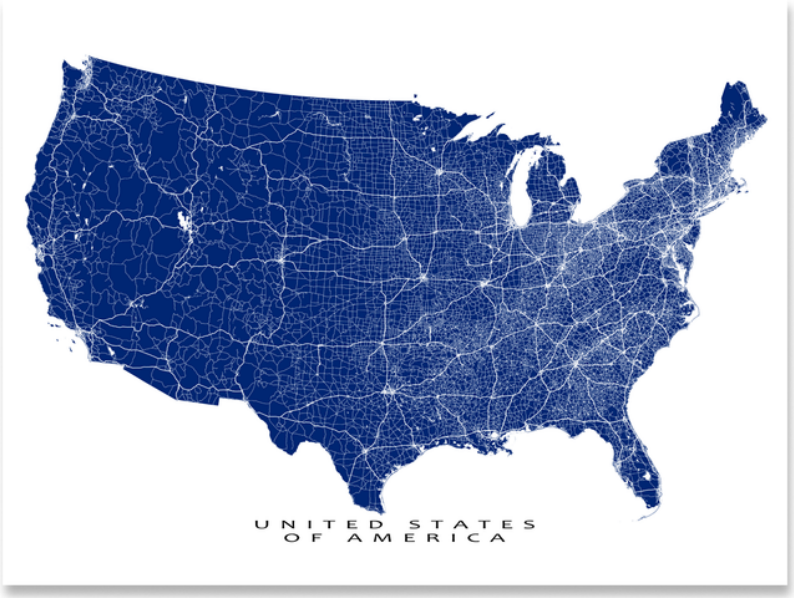
1) Filter = Work with approximations
   Weed out some input based on approximation

2) Refine = Work with reduced size of input

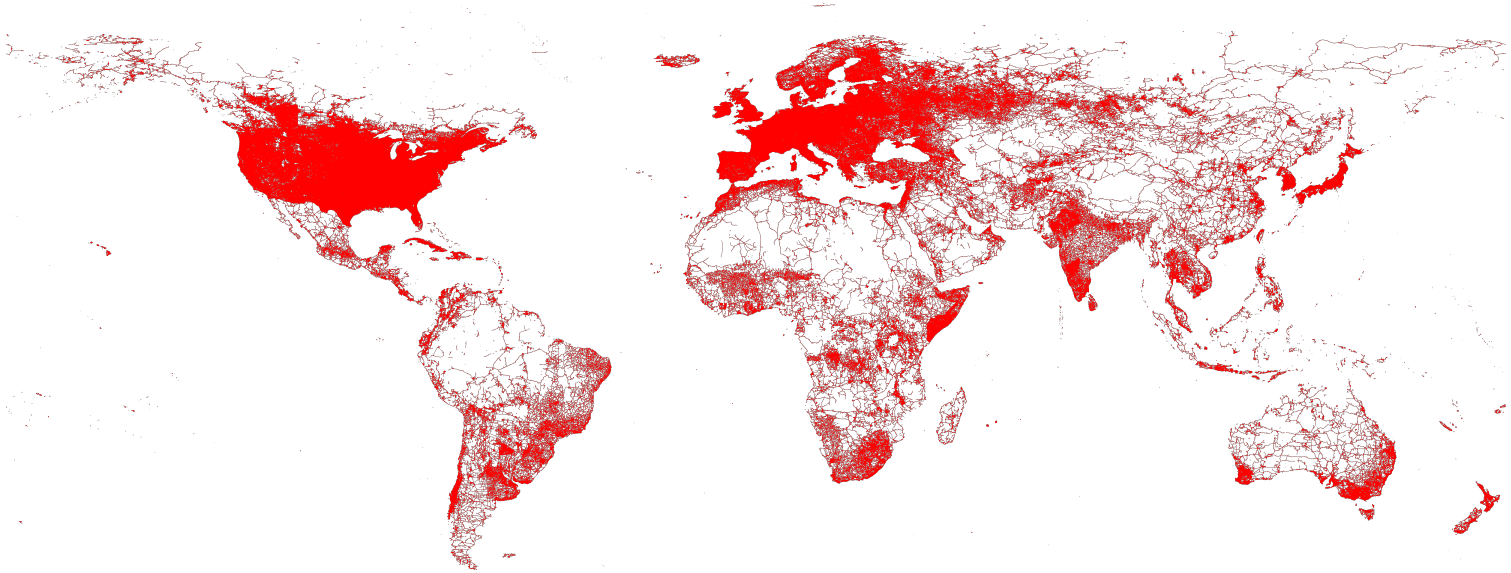
Rationale: Step 1 + Step 2 processing time should be less than working with actual input directly

Find which counties a river goes through?

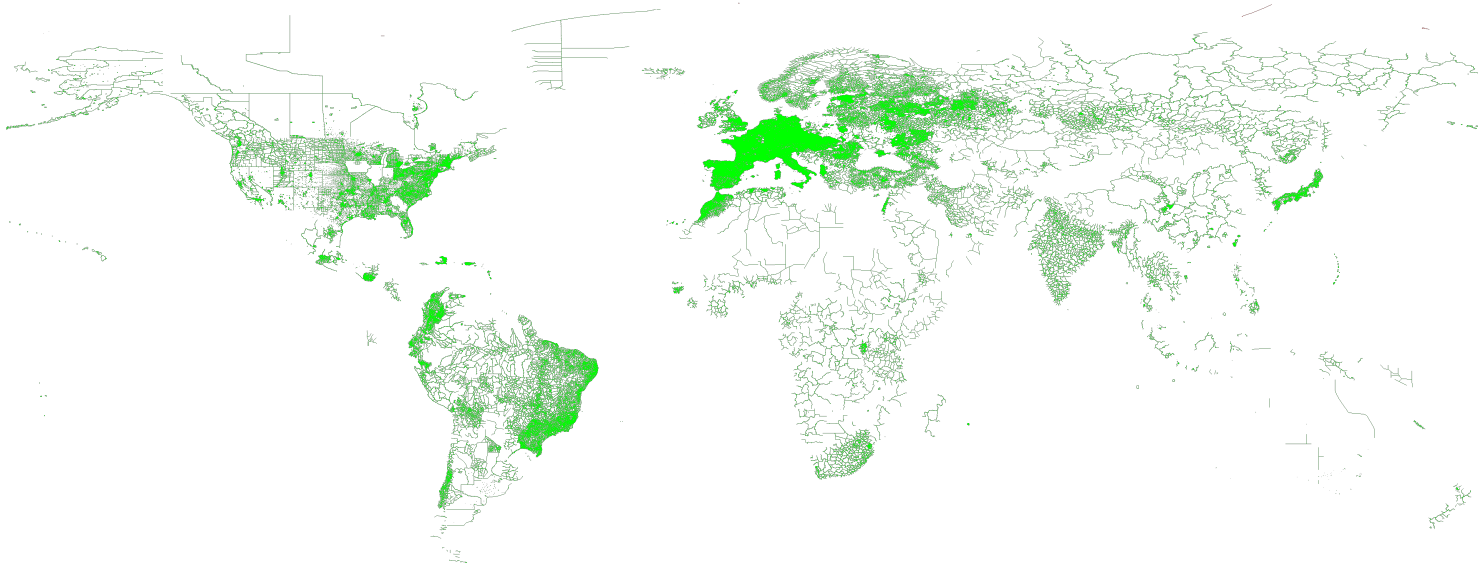
Find the roads intersecting rivers?



Spatial Join



Roads



Cities

- Given two data sets of spatial objects, *R* and *S*,

  find all pairs of spatial

  objects between *R* and *S*

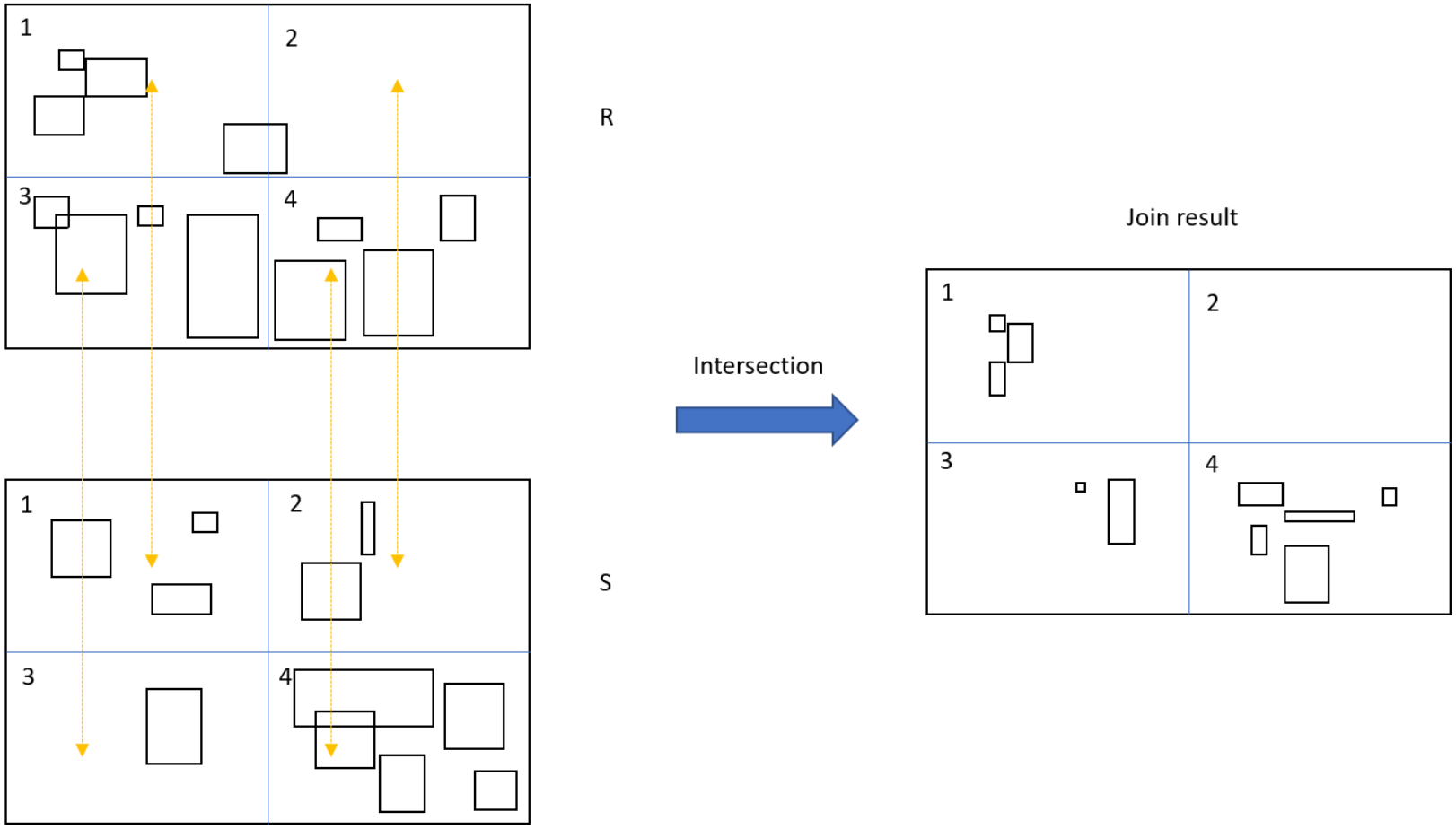
  satisfying a join condition.

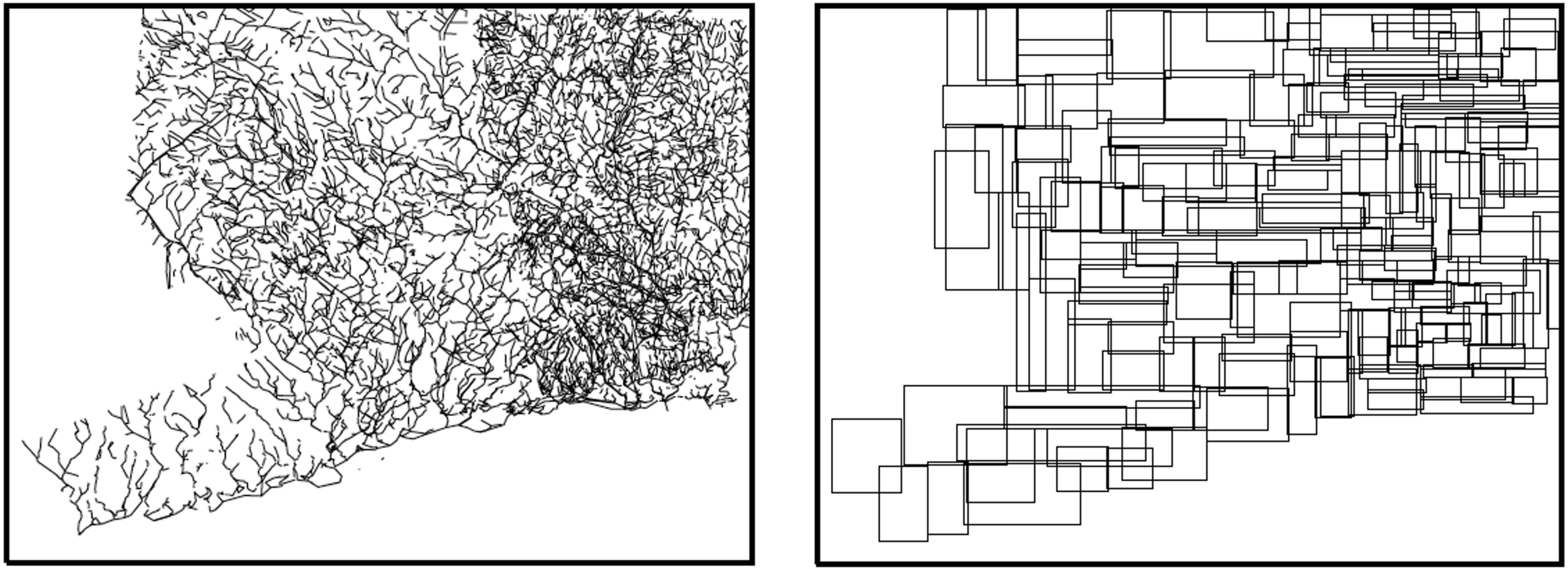
- Output:

  Roads crossing through cities

Pictures were taken from: http://spatialhadoop.cs.umn.edu/datasets.html

Partition based Spatial Join

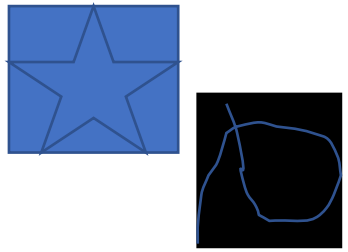




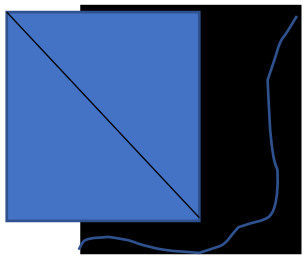
(a) (b)

Figure 6.23 Indexing the hydrography network of Connecticut: TIGER data (simplified) (a) and R-tree (leaf level) (b).

Filter and Refine example



Ignore



More processing required

Do Refinement

Conclusion: Do not intersect

Does two polygon intersect?

- Polygon P and Q
  P has 1000 vertices.
  Q has 500 vertices.

Does P and Q have overlap/intersect?

- Potentially (1000 * 500) operations.

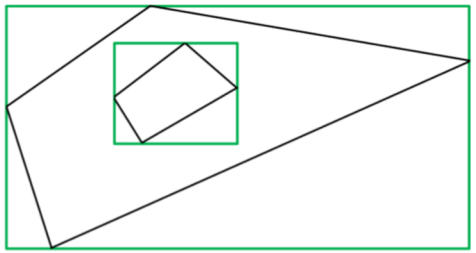
Alternative strategy : Does P's MBR overlap/intersect with Q's MBR?
P MBR has 4 vertices.
Q MBR has 4 vertices.

Benefits?

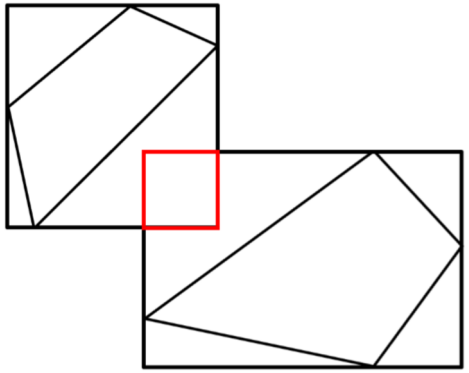
Pop Quiz

(A) Apply filter and refine technique

Green rectangle is the MBR



(B) Apply filter and refine technique



False hit

(C) Apply filter and refine technique