A Scalable DCEL implementation

Andres Calderon

University of California, Riverside

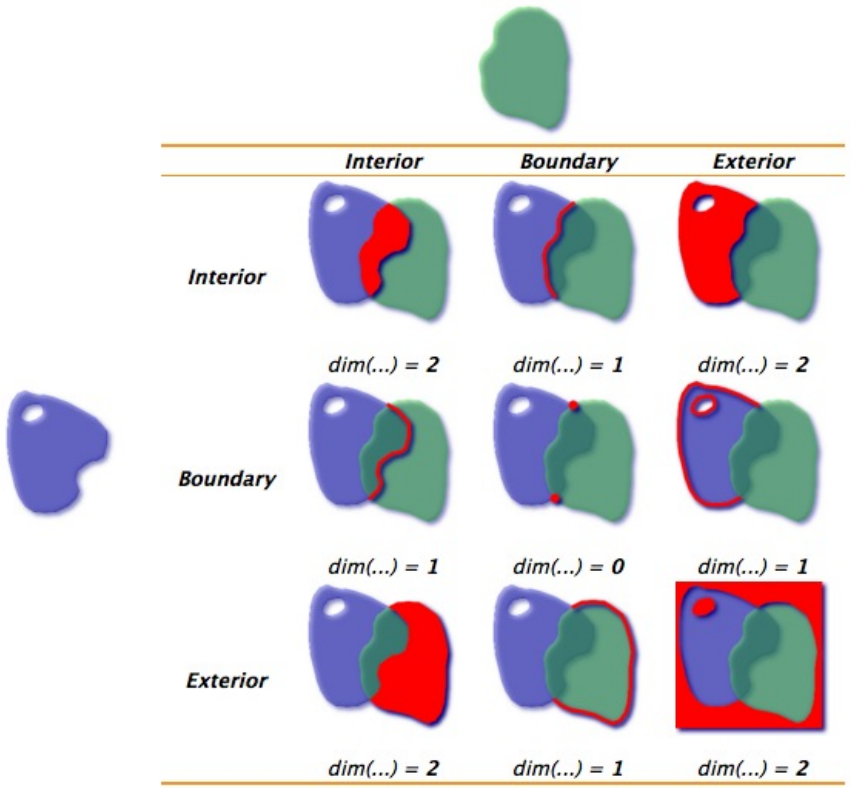
December 14, 2020

Dealing with problems about precision...

- The current dataset has overlapping polygons.
- Even after fixing the overlapped polygons did not match appropriately.
- The current dataset is the result of some transformations from the original dataset (re-project and truncate decimals).
- I decided to have a look at the original dataset.

Testing quality in original dataset...

- level 0 and level 1 from original dataset do not have any overlapping polygon.
- Using Dimensionally Extended 9-Intersection Model to test data quality.

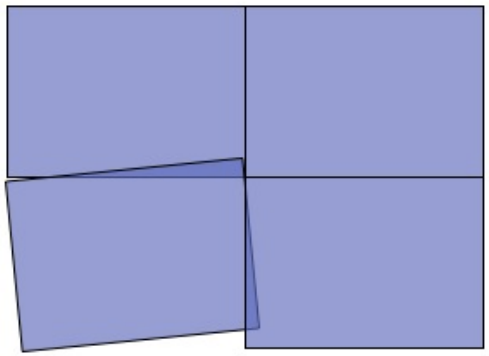


Testing quality in original dataset...

[1]

24.1. Data Quality Testing

The TIGER data is carefully quality controlled when it is prepared, so we expect our data to meet strict standards. For example: no census block should overlap any other census block. Can we test for that?



Tracts with an overlap?

Sure!

```
SELECT a.gid, b.gid
FROM nyc_census_blocks a, nyc_census_blocks b
WHERE ST_Intersects(a.geom, b.geom)
  AND ST_Relate(a.geom, b.geom, '2********')
  AND a.gid != b.gid
LIMIT 10;

-- Answer: 10, there's some funny business
```

[1] https://postgis.net/workshops/postgis-intro/de9im.html

What is next...

- Currently, I have problems during the execution in the cluster. Some tasks do not finish and just keep working.
- However, I have been able to finish on local mode with some sections of the data.
- I plan to test the full dataset by chunks of data until I can detect and fix the problem.