Cognitive Adequacy of Topological Consistency Measures

Nieves R. Brisaboa
Miguel R. Luaces

Database Laboratory
University of A Coruña
A Coruña, Spain

M. Andrea Rodríguez

University of Concepción
Concepción, Chile

Outline

- Motivation
- Consistency measures
- Evaluation framework
- Evaluation results
- Conclusions and Future Work

Outline

- Motivation
- Consistency measures
- Evaluation framework
- Evaluation results
- Conclusions and Future Work

Motivation

- GIS and spatial databases are widely used
  - In 2011, we all carry a GIS in our mobile phone
- Many tools exist to create, store, analyse and visualize geographic information
  - In 2011, It is fairly easy to create a GIS
- Few tools to check the quality of the information
- Data quality is a complex problem in GIS
  - Creation, and manipulation is difficult and specialized
  - Definition and evaluation of quality rules is hard

Motivation

- Dataset consistency in traditional DBMS
  - It is usually a binary property
  - In GIS, the degree of the error matters
- In a previous work
  - Integrity constraints and consistency measures
  - Definition of measures to evaluate the degree of violation of a dataset w.r.t integrity constraints
- The goal of this work:
  - Do users perceive as errors what we measure?

Outline

- Motivation
- Consistency measures
- Evaluation framework
- Evaluation results
- Conclusions and Future Work

Consistency measures

- Topological integrity constraint

$$\forall \bar{x}_1, \bar{x}_2, g_1, g_2 (P(\bar{x}_1, g_1) \wedge R(\bar{x}_2, g_2) \wedge \psi \rightarrow T(g_1', g_2'))$$

- Examples
  - A county must be within the state to which it belongs

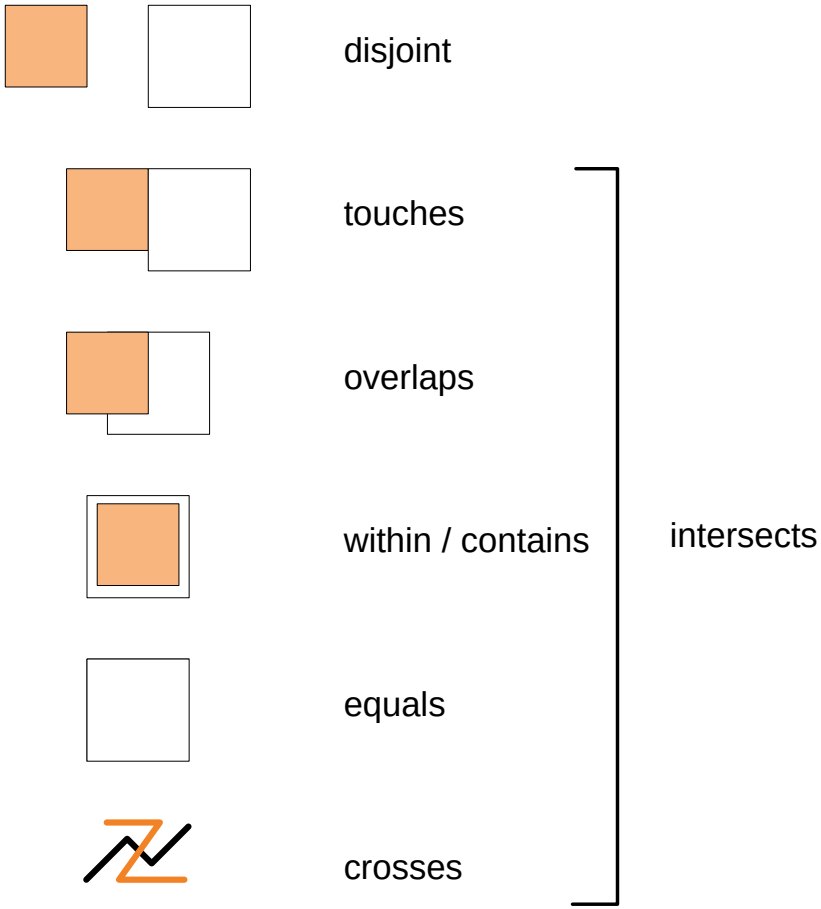
$$\forall idc, ids, g_{1,} g_2 (\text{state}(ids, g_1) \wedge \text{county}(idc, ids, g_2) \rightarrow \text{within}(g_1, g_2))$$

  - Land parcels that intersect must touch

$$\forall id_1, id_2, g_{1,} g_2 (\text{parcel}(id_1, g_1) \wedge \text{parcel}(id_2, g_2) \wedge$$
$$(id_1 \neq id_2) \wedge \text{intersects}(g_1, g_2) \rightarrow \text{touches}(g_1, g_2))$$

Consistency measures

- Topological relationships considered

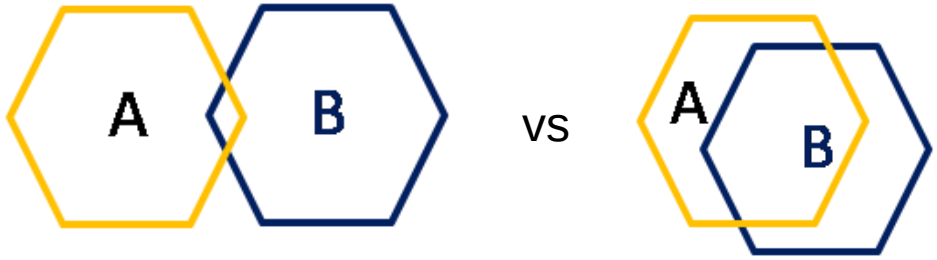


Consistency measures

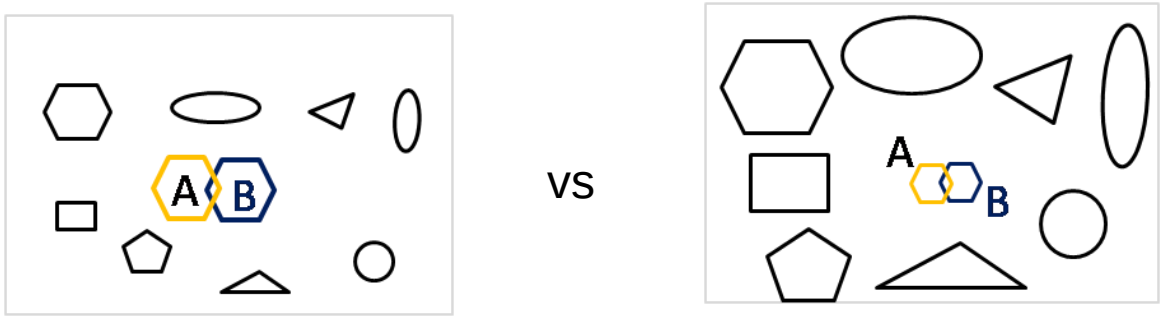
- A consistency measure evaluates the degree of violation of a topological integrity constraint
  - We published measures for
    - surface × surface
    - curve × curve
  - We defined measures the other combinations
    - surface × point
    - surface × point
    - curve × point

Consistency measures

- Consistency measures evaluate two aspects
  - The magnitude of the conflict

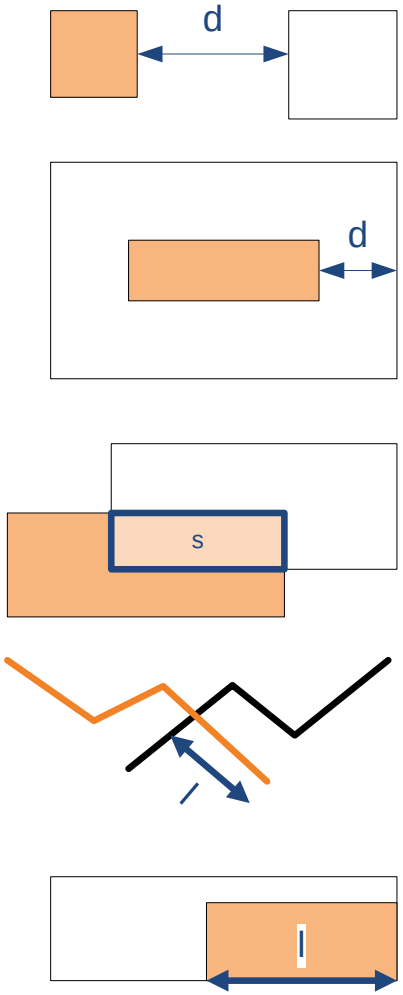


  - The relevance of the conflict



Consistency measures

- We defined five parameters to compute the magnitude of the conflict
- In our consistency measures we used the first four but we did not use the touching length



Outline

- Motivation
- Consistency measures
- Evaluation framework
- Evaluation results
- Conclusions and Future Work

Evaluation framework

- Three hypothesis were formulated:
  - H1: *External distance*, *internal distance*, *crossing length*, and *overlapping size* are perceived and used by subjects to evaluate the degree of violation of topological integrity constraints.
  - H2: *Touching length* is not considered by subjects to evaluate the degree of violation of topological integrity constraints.
  - H3: The *relative size of the geometries* that participate in the violation of topological integrity constraints with respect to other objects in the dataset affects the perceived violation degree.

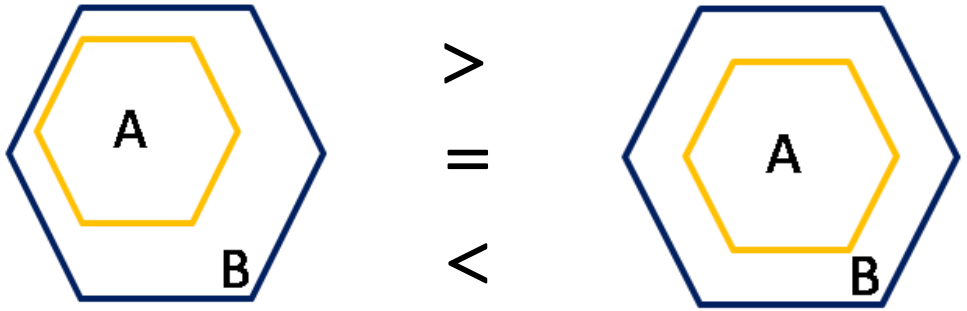
Evaluation framework

- Structure of the test:
  - Brief description of the topological relationships
  - Description of the objective of the test
  - Three sections each with a different task
- Background of the subjects
  - Second-year computer science students
  - No previous knowledge in GIS
  - No explanation of topological relationships beyond the description on the test
  - No rewards for answering the test

Evaluation framework

- Contents of the test
  - Section I: the parameters used are perceived by the subjects
    - Comparison of two figures with two geometries each

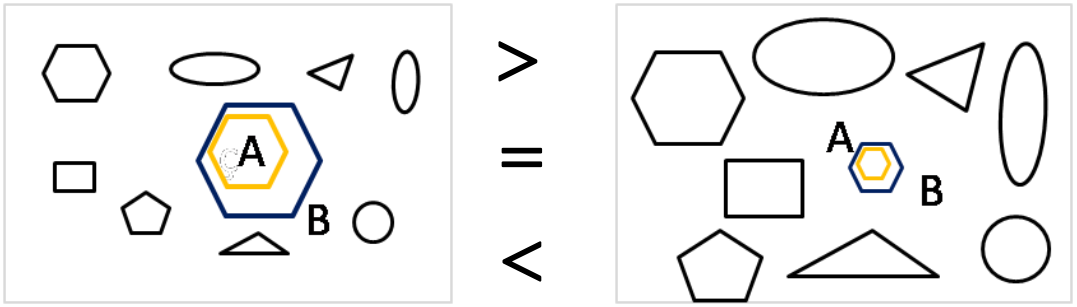
1) A and B should be disjoint



Evaluation framework

- Contents of the test
  - Section II: influence of the context
    - Comparison of two figures with the same geometries in a different context

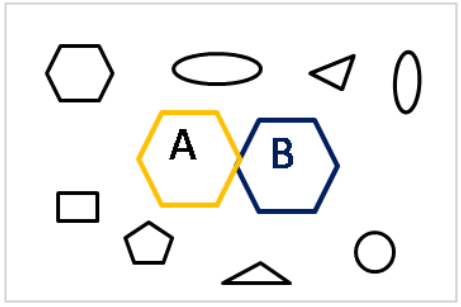
1) A and B should be disjoint

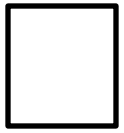


Evaluation framework

- Contents of the test
  - Section III: evaluation of the violation degree measures
    - Numeric evaluation of the violation degree of two geometries

1) A and B should overlap





rate from 0 to 100

Outline

- Motivation
- Consistency measures
- Evaluation framework
- Evaluation results
- Conclusions and Future Work

Evaluation results

- Raw data from the tests in section I

Exercise	Expected	Actual	Geometries	Parameter	Impact	Neutral	No impact
5	Overlaps	Disjoint	surface × surface	External distance	87 %	3 %	10 %
12	Disjoint	Overlaps	curve × curve	External distance	83 %	5 %	12 %
13	Disjoint	Overlaps	curve × curve	External distance	82 %	10 %	8 %
14	Disjoint	Overlaps	curve × curve	External distance	80 %	12 %	8 %
19	Disjoint	Overlaps	surface × curve	External distance	87 %	12 %	7 %
21	Disjoint	Overlaps	surface × curve	External distance	72 %	18 %	10 %
24	Disjoint	Overlaps	curve × point	External distance	67 %	27 %	7 %
1	Disjoint	Overlaps	surface × surface	Overlapping size	83 %	8 %	8 %
3	Touches	Overlaps	surface × surface	Overlapping size	68 %	20 %	12 %
8	Within	Overlaps	surface × surface	Overlapping size	68 %	17 %	15 %
10	Disjoint	Overlaps	curve × curve	Crossing length	68 %	23 %	8 %
16	Disjoint	Overlaps	surface × curve	Crossing length	83 %	8 %	8 %
20	Disjoint	Overlaps	surface × curve	Crossing length	68 %	23 %	8 %

Evaluation results

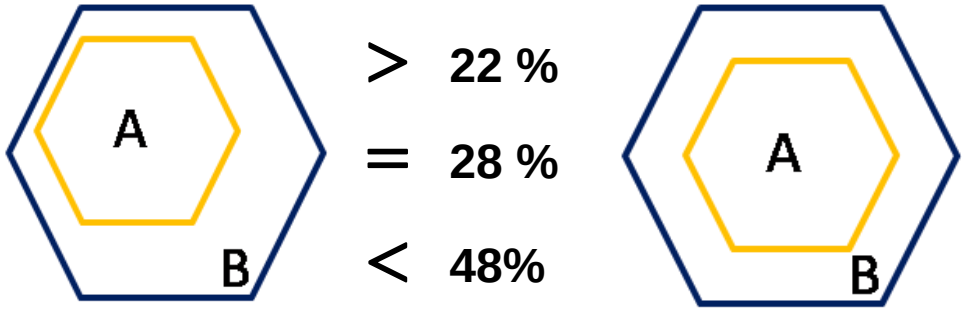
- Raw data from the tests in section I

Exercise	Expected	Actual	Geometries	Parameter	Impact	Neutral	No impact
2	Disjoint	Within	Surface × surface	Internal distance	48 %	28 %	23 %
4	Touches	Within	surface × surface	Internal distance	62 %	22 %	17 %
7	Overlaps	Within	surface × surface	Internal distance	53 %	40 %	7 %
17	Disjoint	Overlaps	surface × curve	Internal distance	45 %	45 %	10 %
18	Disjoint	Overlaps	surface × curve	Internal distance	50 %	28 %	20 %
22	Disjoint	Overlaps	surface × curve	Internal distance	47 %	43 %	10 %
23	Disjoint	Overlaps	surface × point	Internal distance	52 %	28 %	20 %
6	Overlaps	Touches	surface × surface	Touching length	32 %	48 %	20 %
9	Within	Touches	surface × surface	Touching length	20 %	65 %	15 %
11	Disjoint	Overlaps	curve × curve	Touching length	58 %	35 %	7 %
15	Disjoint	Overlaps	surface × curve	Touching length	40 %	52 %	8 %

Evaluation results

- Summary of the results from section I
  - Around 10% of subjects answered incorrectly
  - External distance, overlapping size and crossing length are used by subjects
  - Some tests for internal distance showed misunderstanding of *disjoint*

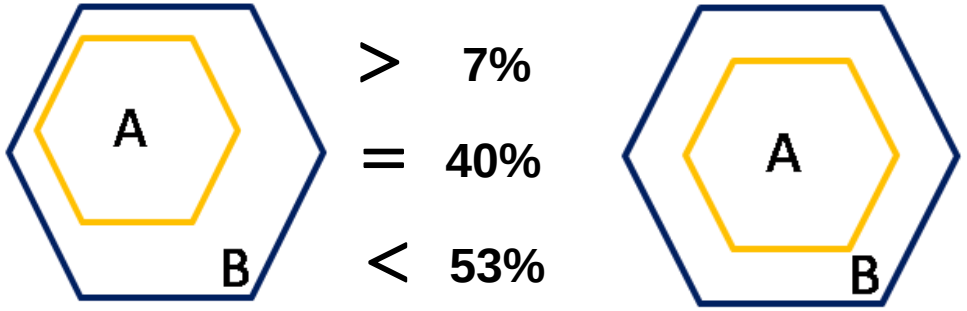
2) A and B should be disjoint



Evaluation results

- Summary of the results from section I
  - Some tests for internal distance showed misunderstanding of *overlaps*

7) A and B should overlap



  - Touching length is not used by subjects

Evaluation results

- Results from section II
  - 35%,35% and 30% answered that the size of geometries has a positive, neutral or negative impact on the violation degree (respectively)
- Results from section III
  - Violation degrees answered by the subjects vs violation degrees computed with our measures
    - Using the original measures results in a Pearson correlation coefficient of 0.54
    - Removing the relevance of the conflict from the measures results in a coefficient of 0.84

Outline

- Motivation
- Consistency measures
- Evaluation framework
- Evaluation results
- Conclusions and Future Work

Conclusions and future work

- Regarding the definition of the measures
  - H1 is partially confirmed and H2 is confirmed
    - External distance, crossing length, and overlapping size are perceived and used by subjects as a violation degree measure
    - Internal distance is not confirmed to be used
    - Touching length is not used by subjects
  - H3 is rejected
    - The relative size of geometries in conflict compared to other ones in the dataset does not impact the perceived violation degree

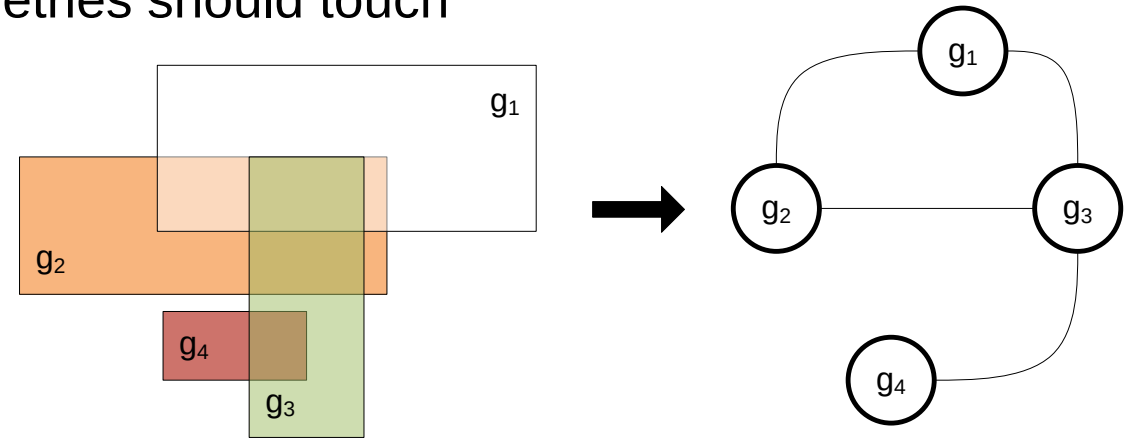
Conclusions and future work

- Regarding the evaluation framework
  - The task that has to be performed by the subject is very difficult
  - The knowledge of topological relationships is very important, and explaining the meanings may not be the solution
    - Subjects may use their intuition instead of the formal definition
    - Training subjects without imposing our view of the measures may be difficult

Conclusions and future work

- Define and perform a new study
  - Evaluate precisely the internal distance parameters
  - Differentiate subjects in trained and not trained
- Define alternative measures that consider all the geometries that participate in a conflict

Geometries should touch



Cognitive Adequacy of Topological Consistency Measures

Nieves R. Brisaboa
Miguel R. Luaces

Database Laboratory
University of A Coruña
A Coruña, Spain

M. Andrea Rodríguez

University of Concepción
Concepción, Chile

Contact:

Miguel R. Luaces
luaces@udc.es