

图数据库系统随机差分测试

郑莹莹, 窦文生, 汪钊丞, 秦政, 汤磊, 高钰, 王栋, 王伟, 魏峻

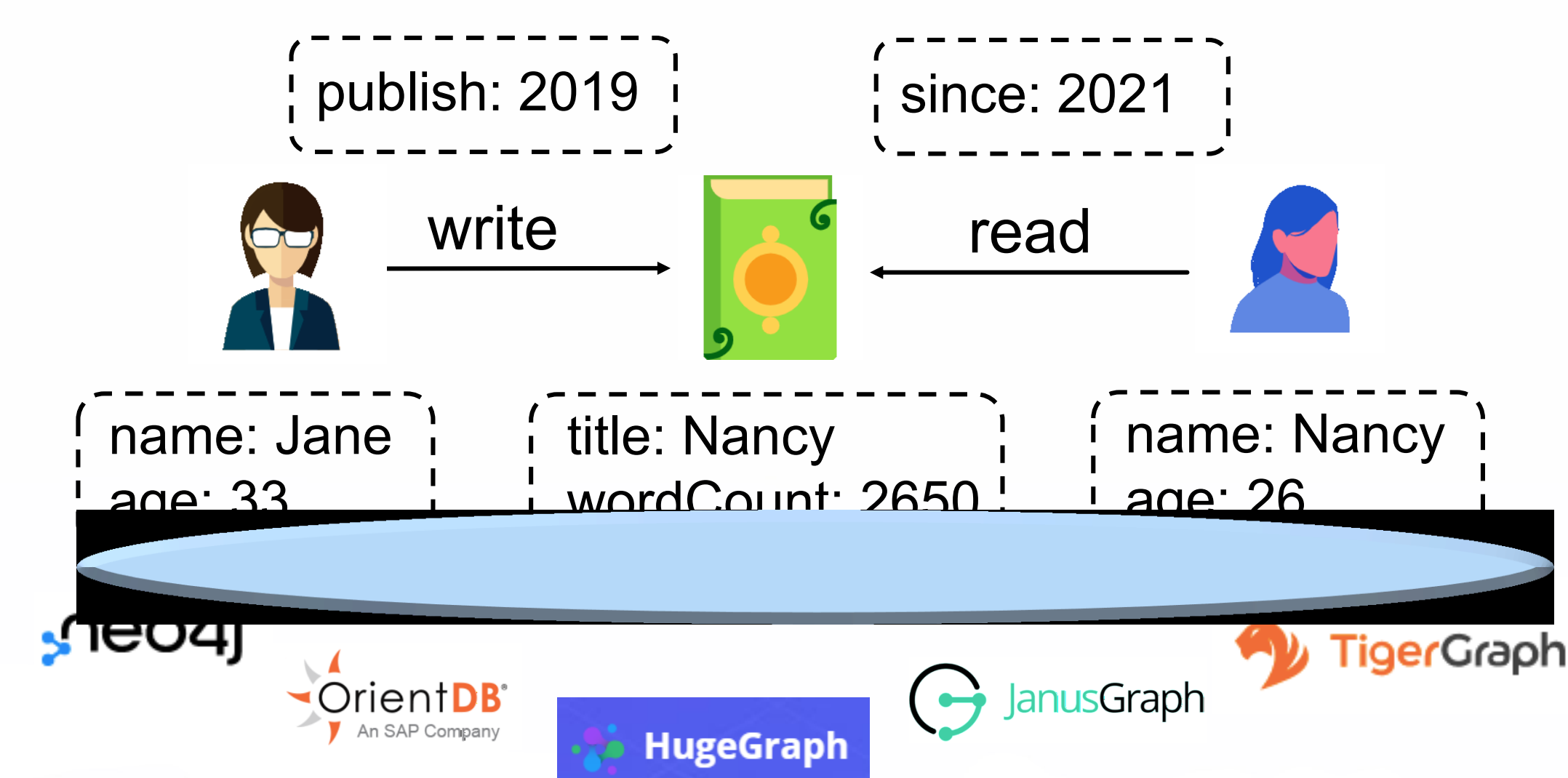
Finding Bugs in Gremlin-Based Graph Database Systems via
Randomized Differential Testing

The 31th of ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA'22)

联系方式: 郑莹莹, zhengyingying14@otcaix.iscas.ac.cn

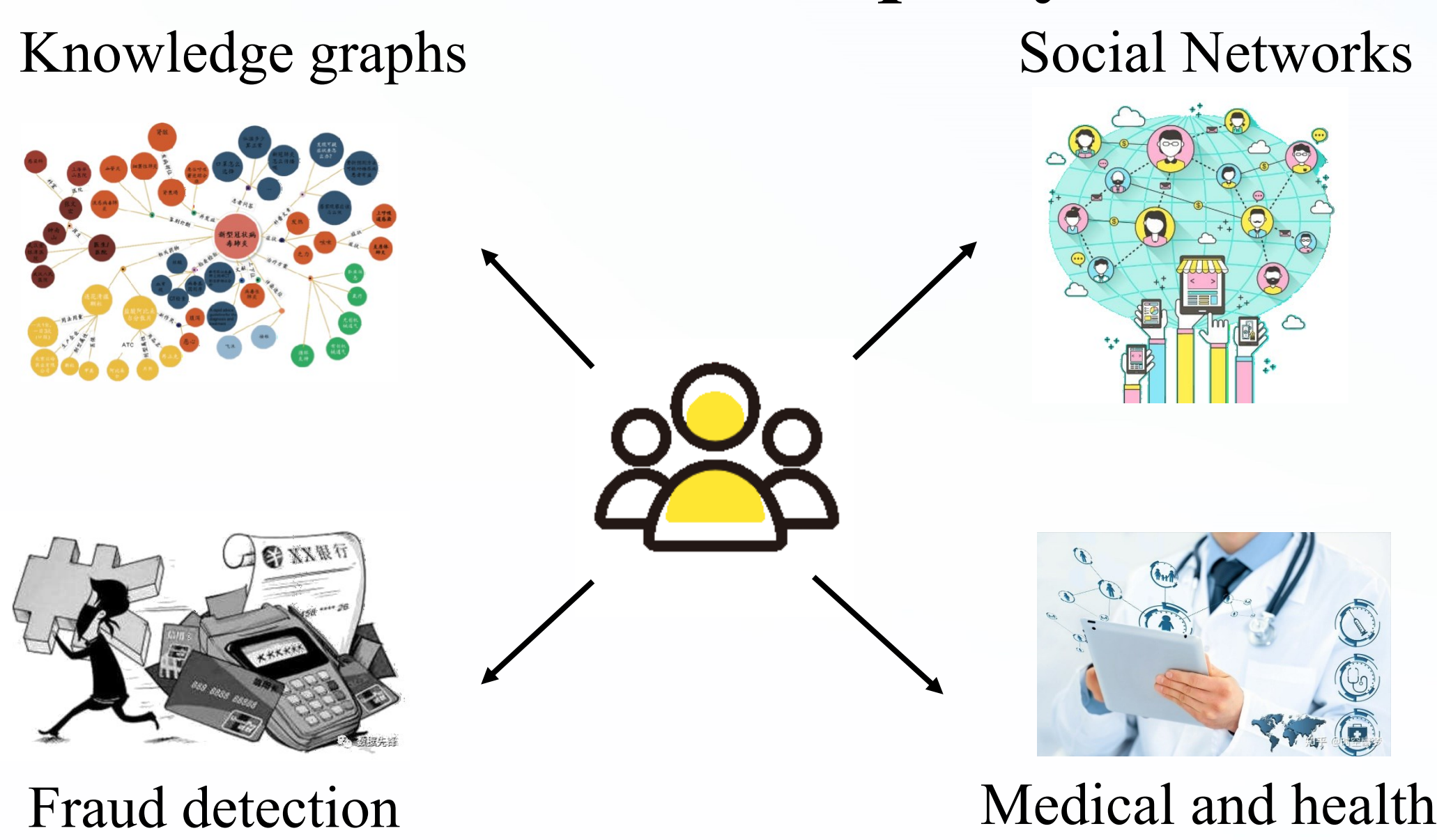
Graph Database Systems

GDBs support efficient storage and queries for graph data.



Reliable Data Access

Numerous application areas require GDBs return correct query results.



Bugs in GDBs

GDBs suffer from logic bugs, in which a query returns an unexpected result.

```

1 hugegraph.schema().vertexLabel("vLabel").properties("prop")
  .nullableKeys("prop").create();
2 hugegraph.schema().indexLabel("index").onV("vLabel").by("
  prop").shard().ifNotExist().create();
3
4 Vertex v1 = new Vertex("vLabel").property("prop", 5);
5 Vertex v2 = new Vertex("vLabel").property("prop", 1);
6 Vertex v3 = new Vertex("vLabel").property("prop", 3);
7 addVertices(Arrays.asList(v1, v2, v3));
8
9 g.V().has('prop', between(0,4).or(lt(2))).count();
10 -- {3} x {2} ✓
    
```



A real logic bug in HugeGraph.

No Testing Tools or Methods

Motivation

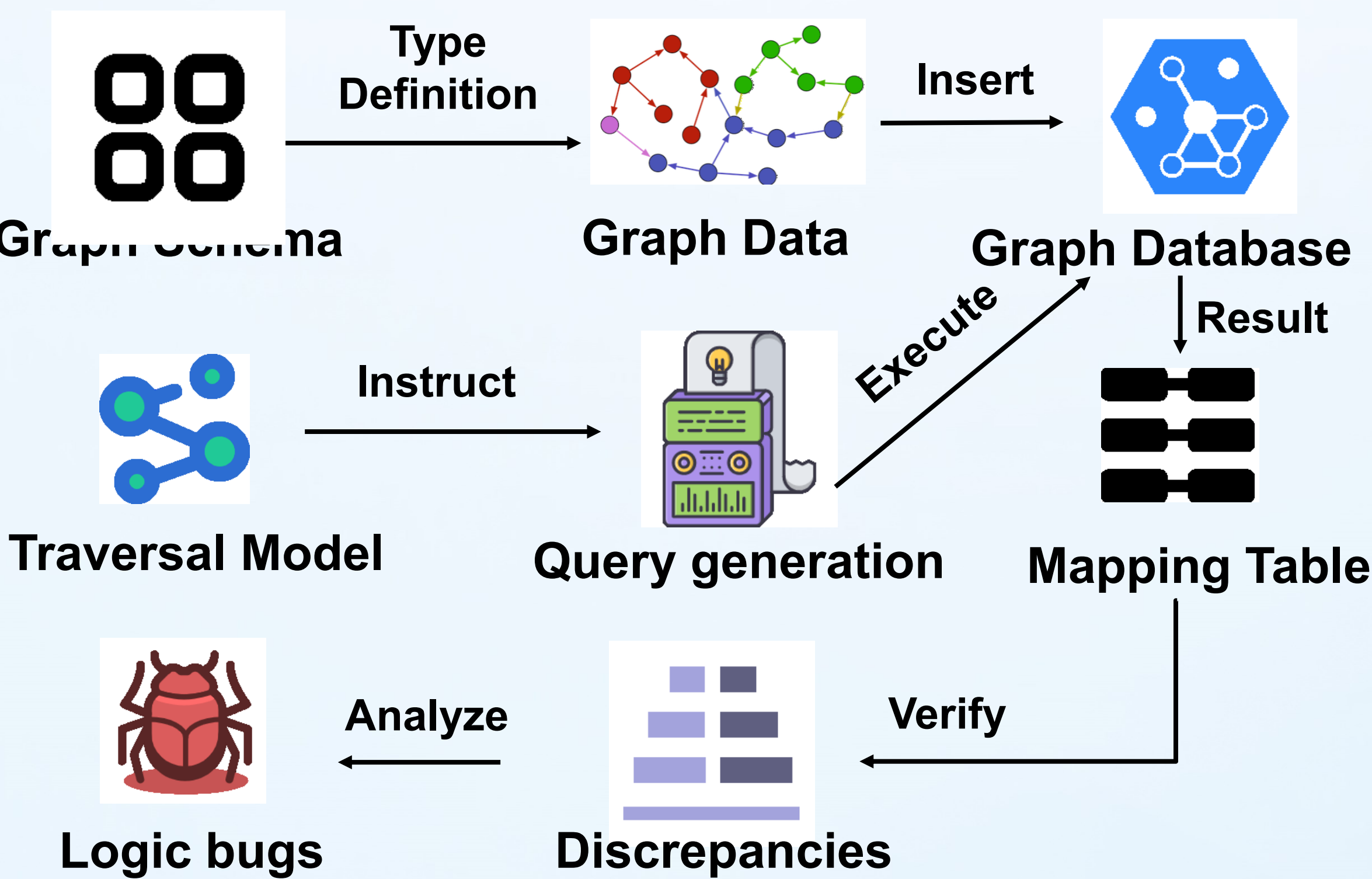
- DBMS testing tools and methods cannot be used to test GDB.
- GDB testing works cannot detect bugs.

Challenges

- How to generate syntactically correct and valid query?
- How to solve the test oracle problem?

Grand

Find bugs via differential testing



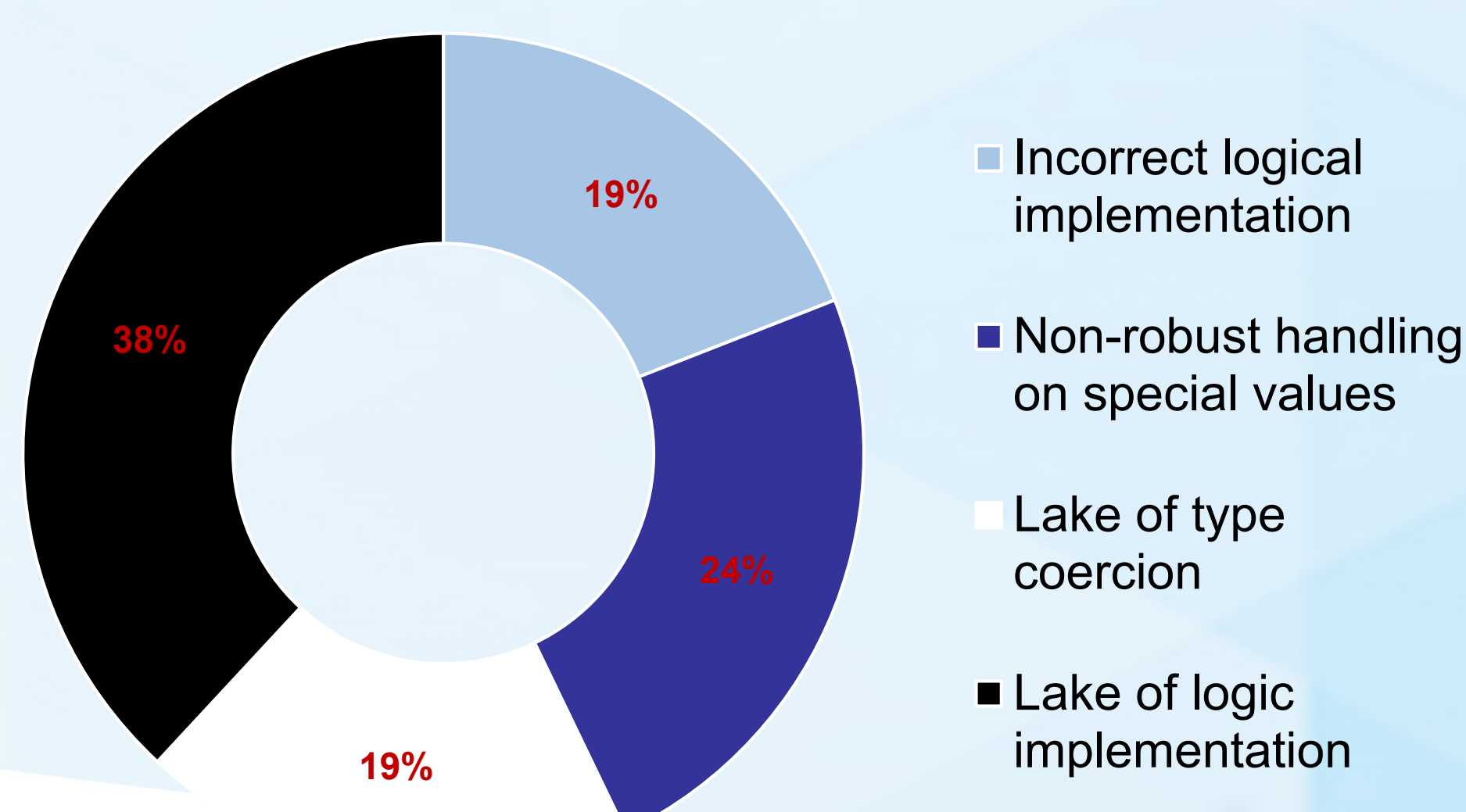
Found Bugs

21 found bugs in six widely-used GDBs

GDB	Detected	Confirmed	Fixed
Neo4j	3	2	1
OrientDB	1	0	0
JanusGraph	3	3	2
HugeGraph	9	9	3
TinkerGraph	3	3	1
ArcadeDB	2	1	0
Total	21	18	7

Bug Analysis

Summarize the found bugs into four categories according to root causes



Instruction Coverage

Achieve coverage from 32% to 61% for query engines

