



SIREN

INTERACTIVE REDESCRIPTION MINING

Esther Galbrun & Pauli Miettinen

galbrun@cs.bu.edu, Department of Computer Science,
Boston University, US-MA

pmiettinen@mpi-inf.mpg.de, Max Planck Institute for Informatics
Saarbrücken, Germany

► REDESCRIPTION MINING

Given a **set of entities** with **two sets of characterizing variables** find a **pair of queries** that describe approximately the **same entities** [1].

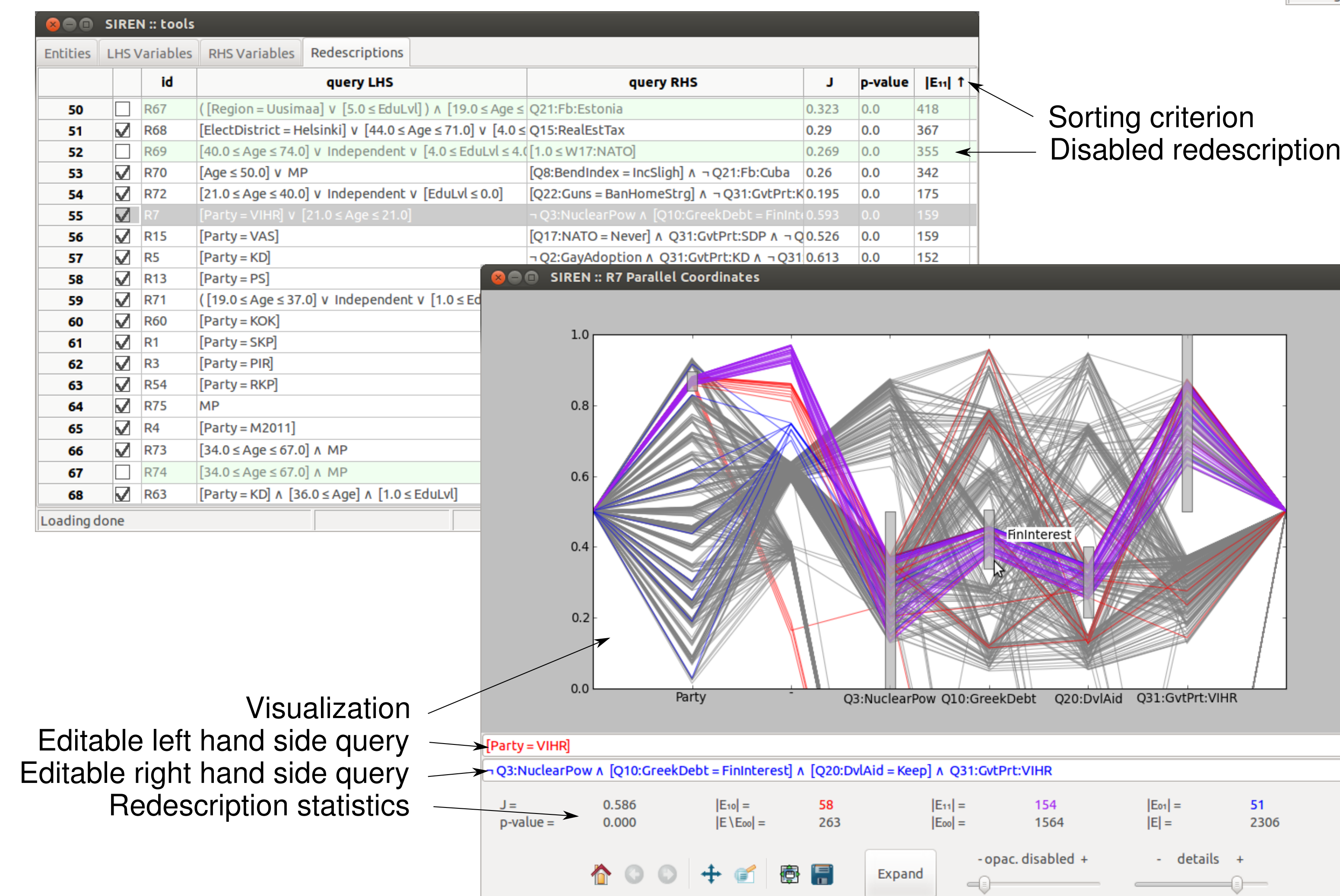
Finding **alternative descriptions of the same entities** is a problem that appears in many areas of science, for example biology.

Redescription mining is a **powerful data analysis tool**.

Two points of view:

- Coherent subsets of **entities**, which can be described in two ways.
- **Variables and conditions** appearing together in the queries.

Our **ReReMi** algorithm [2] extends this problem to **real-valued data**.



REFERENCES

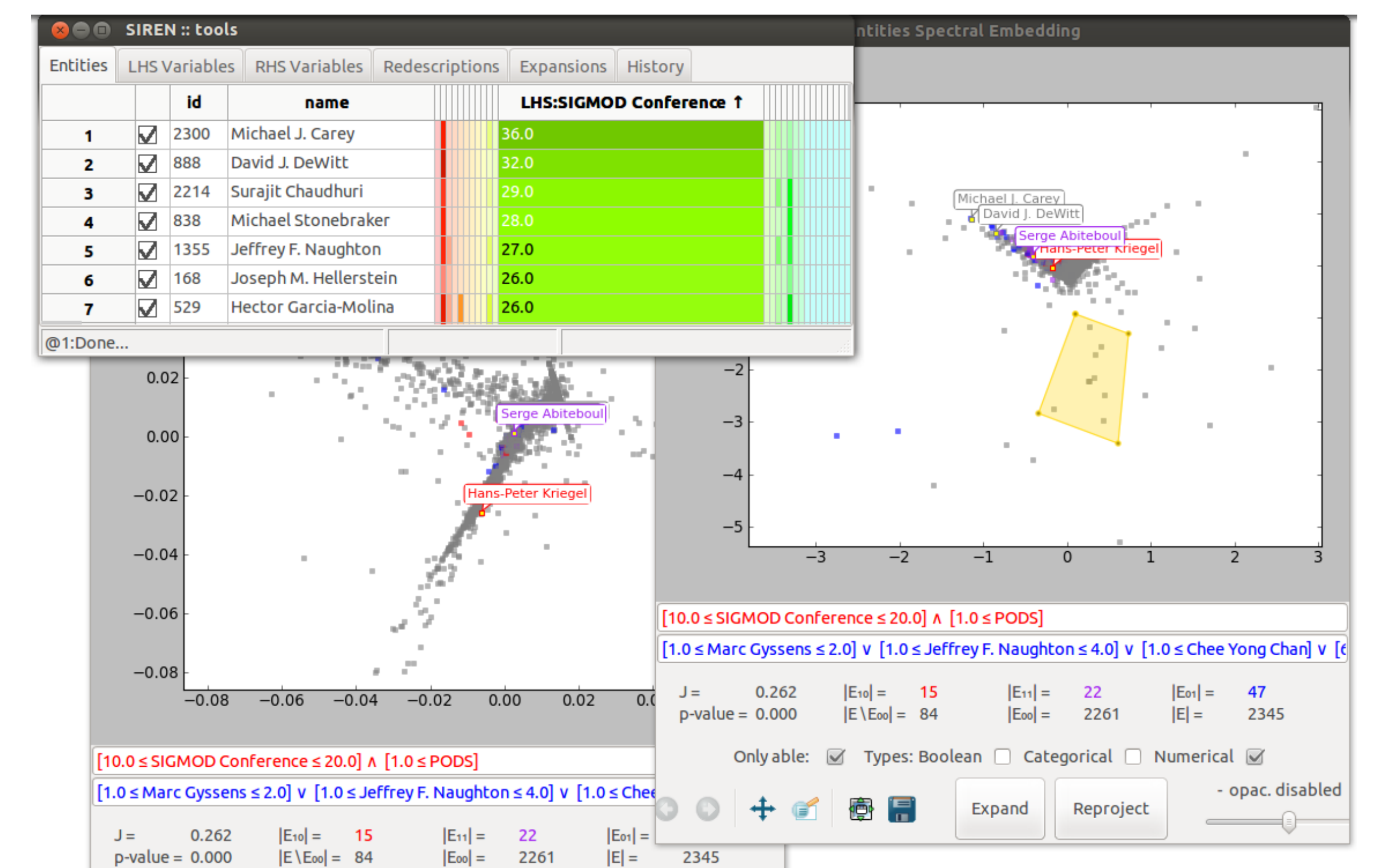
- [1] N. Ramakrishnan et al. Turning CARTwheels: An Alternating Algorithm for Mining Redescriptions. In KDD, 266–275. ACM, 2004.
- [2] E. Galbrun & P. Miettinen. From black and white to full color: extending redescription mining outside the Boolean world. Statistical Analysis and Data Mining, 5(4):284–303, 2012.
- [3] E. Galbrun & P. Miettinen. A Case of Visual and Interactive Data Analysis: Geospatial Redescription Mining. In Instant Interactive Data Mining Workshop @ ECML-PKDD, 2012.
- [4] E. Galbrun & P. Miettinen. Siren: an interactive tool for mining and visualizing geospatial redescriptions. In KDD, 1544–1547. ACM, 2012.
- [5] E. Galbrun. Methods for Redescription Mining. PhD thesis, University of Helsinki, 2014.

► INTERACTIVE DATA MINING

Mining data is generally an **iterative process**.

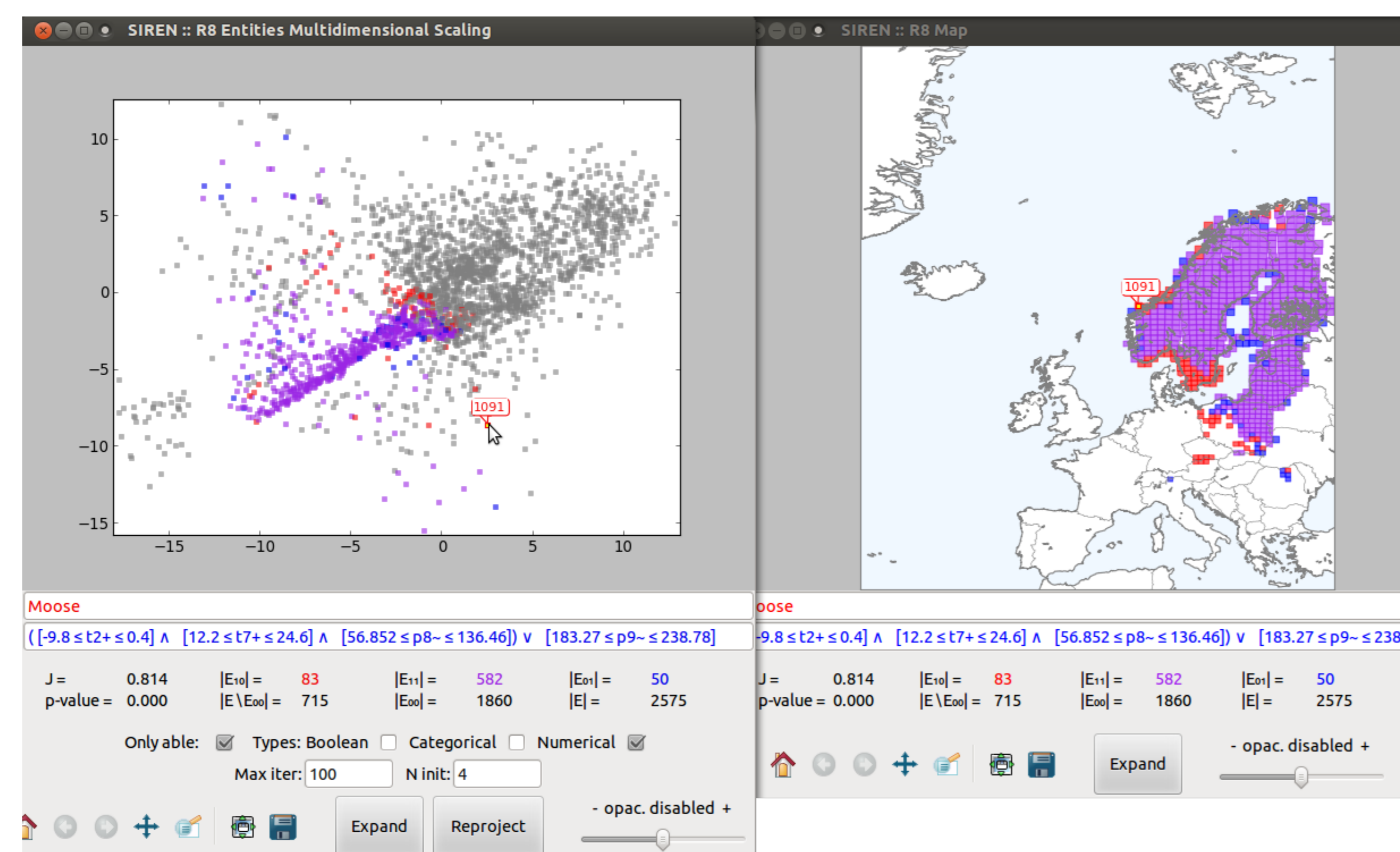
Results obtained at one step **give rise to hypotheses** which will be tested at a further step.

Siren improves the analysis by providing means to easily **interact with the redescription mining process**.



► FEATURES

Siren allows to mine, visualize, and edit redescriptions **interactively**. It provides multiple **visualizations** including parallel coordinates, maps and various projections. Its **brush-and-link workflow** and **tight integration** to the mining algorithm make it a **very effective exploratory data analysis tool**.



► USE CASES

We demonstrate the use of Siren on various tasks from different domains:

- Learn about the candidates to the Finnish 2011 Parliamentary elections
- Find biological niches, i.e. bioclimatic constraints on species habitat
- Explore bibliographic data in search of publication patterns



► DOWNLOAD

Siren is currently available for Linux, OS X and Windows at
<http://www.cs.helsinki.fi/u/galbrun/redescriptors/siren/>

max planck institut
informatik

BOSTON
UNIVERSITY