Data Polygamy: The Many-Many Relationships among Urban Spatio-Temporal Data Sets

Fernando Chirigati¹, Harish Doraiswamy¹, Theodoros Damoulas²,³, Juliana Freire¹

NYU ¹New York University  ²The University of Warwick  ³Alan Turing Institute

Urban Data Sets are Polygamous!
There are multiple interactions between entities of a city. These are captured by the relationships between urban data sets.

Relationship Queries
Find all data sets related to a given data set D
Enable hypothesis generation and hypothesis testing!

Hypothesis Testing
NYC residents often struggle to get a taxi when it is raining.
Long-standing hypothesis:
- Taxi drivers set an income goal
- They reach goal faster on rainy days
Can we test such hypothesis? Yes!

Challenge 1: How to define a data set relationship?

Our Approach: Computational Topology
1) Modeling the Data as a Terrain
\[ f : [S \times T] \rightarrow \mathbb{R} \]
- Critical points
- Peaks
- Valleys

2) Identifying and Computing Topological Features
Neighborhoods of critical points

3) Identifying Topology-based Relationships
Relationship between features
- Positive Features
- Negative Features

Challenge 2: Data Complexity
- Multiple spatio-temporal resolutions
- Large data sets
- Relationships can be between any of the attributes

Our Approach:
- Monte Carlo tests filter potentially coincidental relationships
- Further filtering using \( \tau \) and \( \rho \)

Reduces the number of output relationships in around 99%

Interesting Relationships
- Taxi and Wind Speed
  - No. taxis \( \times \) Wind speed
- Taxi and Rainfall
  - No. taxis \( \times \) Precipitation
  - Taxi fare \( \times \) Precipitation
- Weather and Citi Bike
  - Snow precipitation \( \times \) Trip duration
  - Snow precipitation \( \times \) Active stations

Weather is the most polygamous data set!