University of BRISTOL

Making Sense of (Multi-)Relational Data

Part I: Mining Relational Data – An Overview

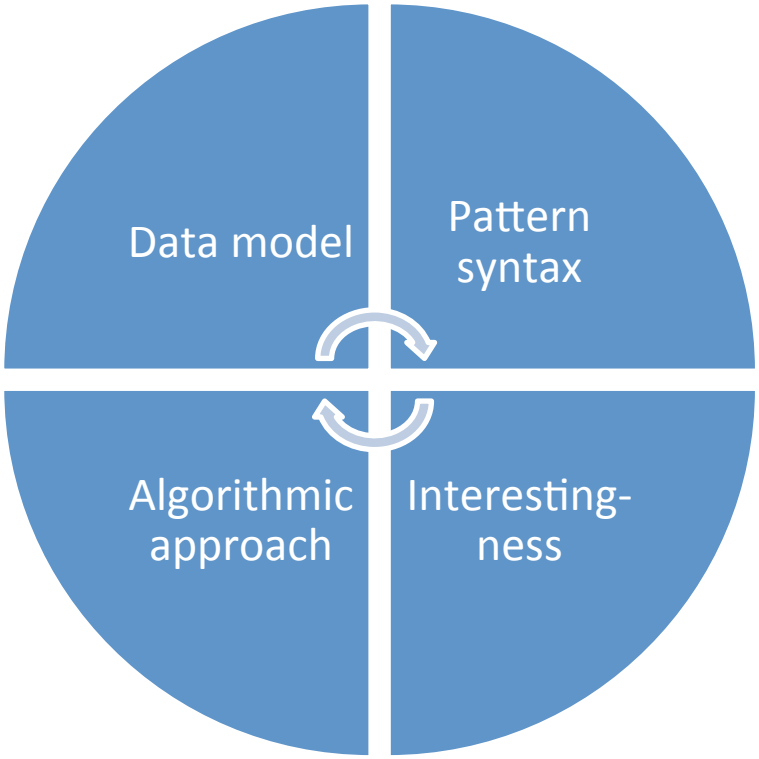
Jefrey Lijffijt

Eirini Spyropoulou

Tijl De Bie

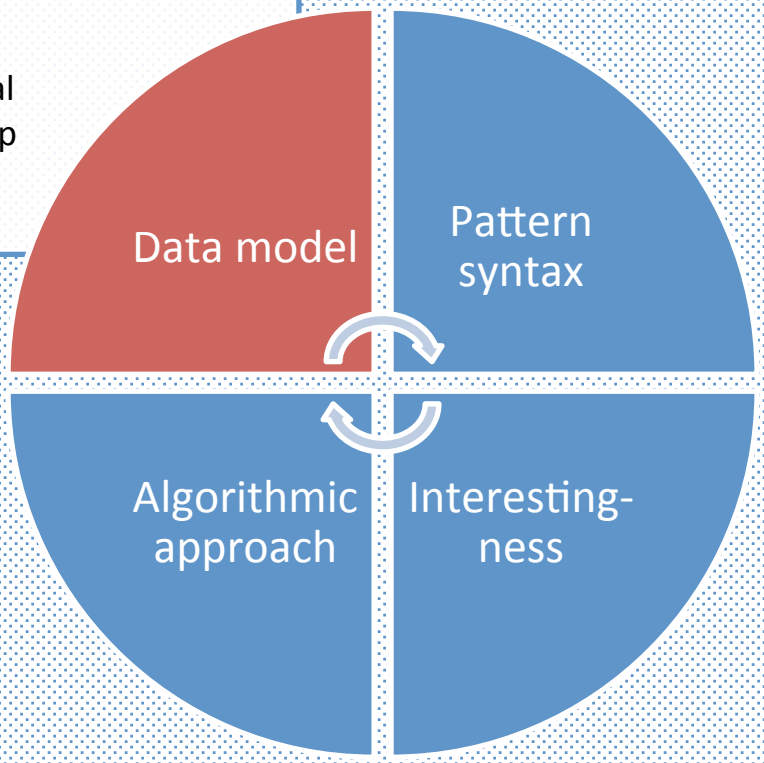
University of Bristol
Ghent University

Overview


Data model
Pattern syntax
Algorithmic approach
Interesting-ness

Relational Data Models

- Networks
- RDF
- Multi-dimensional
- Entity-relationship

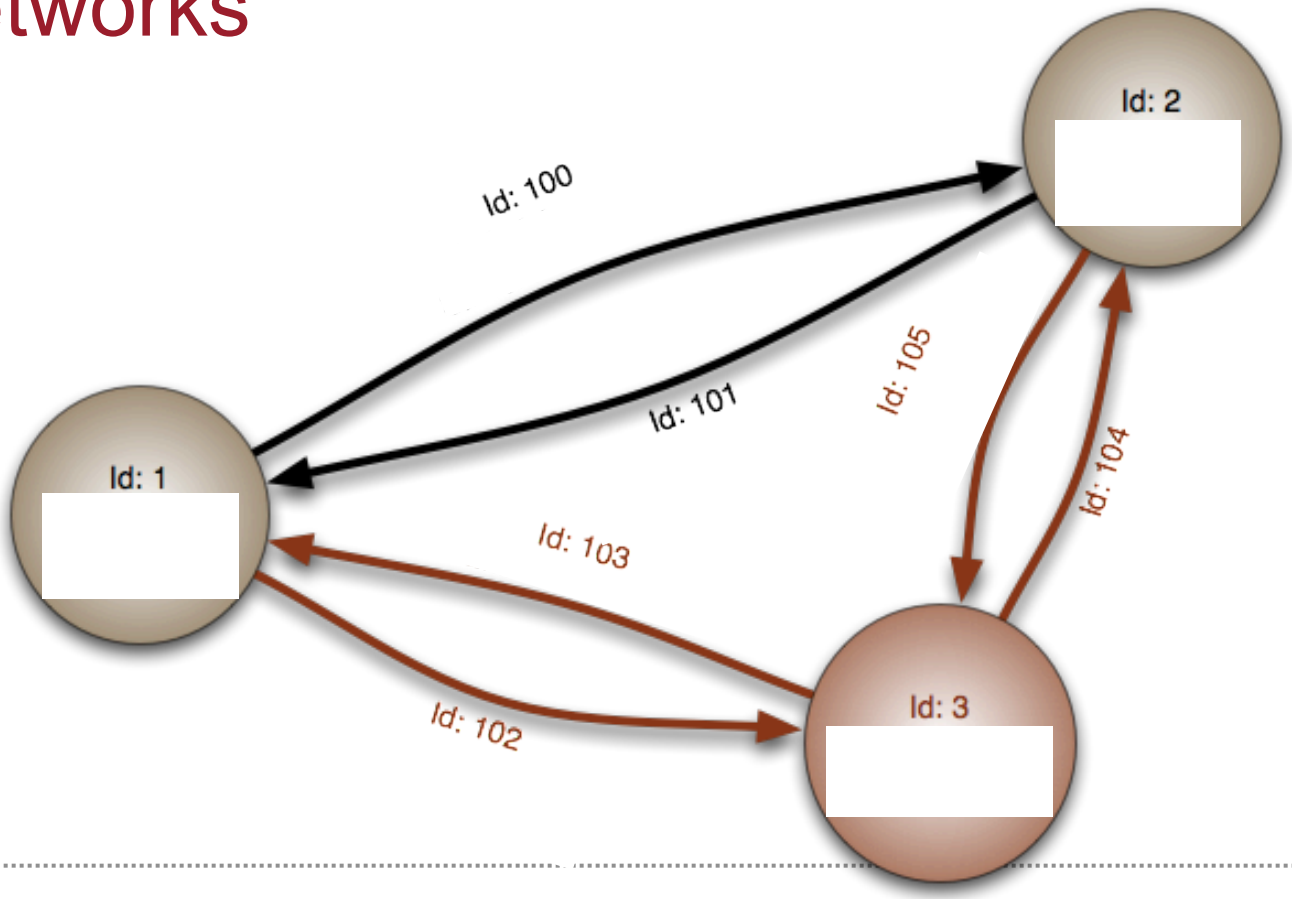


Relational data

- Data that is *not merely a set of unrelated points*
- More constructively, various frameworks:

Data model
Networks
RDF data model
Multidimensional data model
Entity-relationship data model

Networks



Source: wikipedia

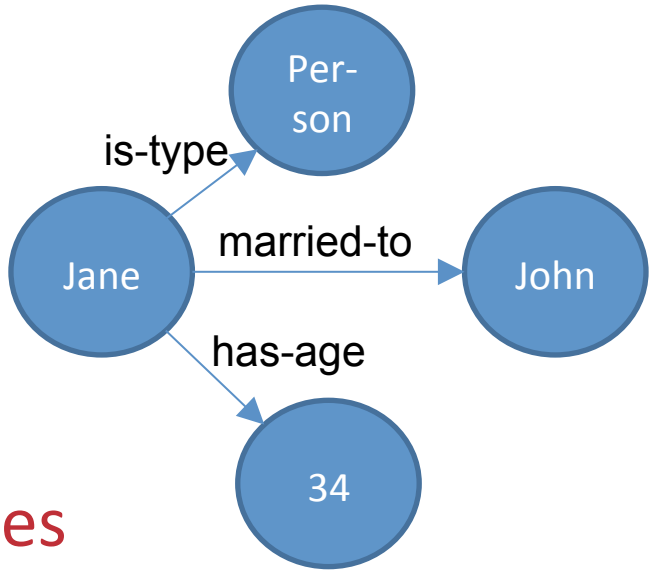
Networks

- Used for:
  - Social networks
  - …
- Technology:
  - Graph Database Management Systems (GDMS)

Titan

RDF data model

(Jane, is-type, person)
(Jane, married-to, John)
(Jane, has-age, 34)



- The data is a set of triples
  - Subject-predicate-object statements
- Graph-like structure
  - Focus on edges, less on nodes
- Also predicates can take the role of subject or object in other triples
  - This leads to ontologies

(has-age, type, object-property)
(has-age, domain, person)
(has-age, range, pos-integer)

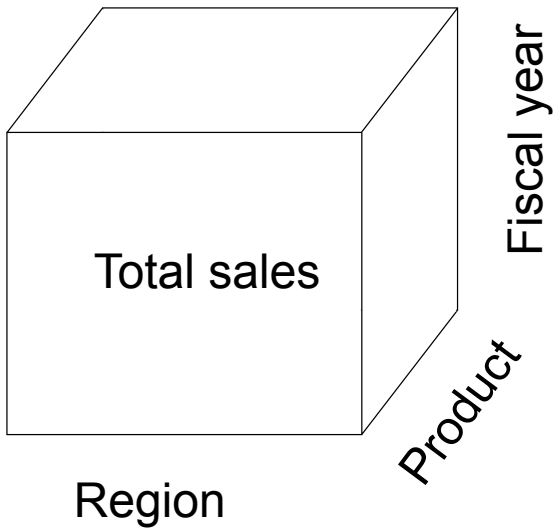
RDF data model

- Used for:
  - Semantic web, Linked Data
- Technology:
  - Triple stores
  - Often using relational database

GraphDB

Multidimensional data model

- Generalisation of spreadsheet to more than rows/columns
- Two concepts:
  - Dimensions
    - E.g. region, product, fiscal year
  - Facts
    - E.g. total sales



Multidimensional data model

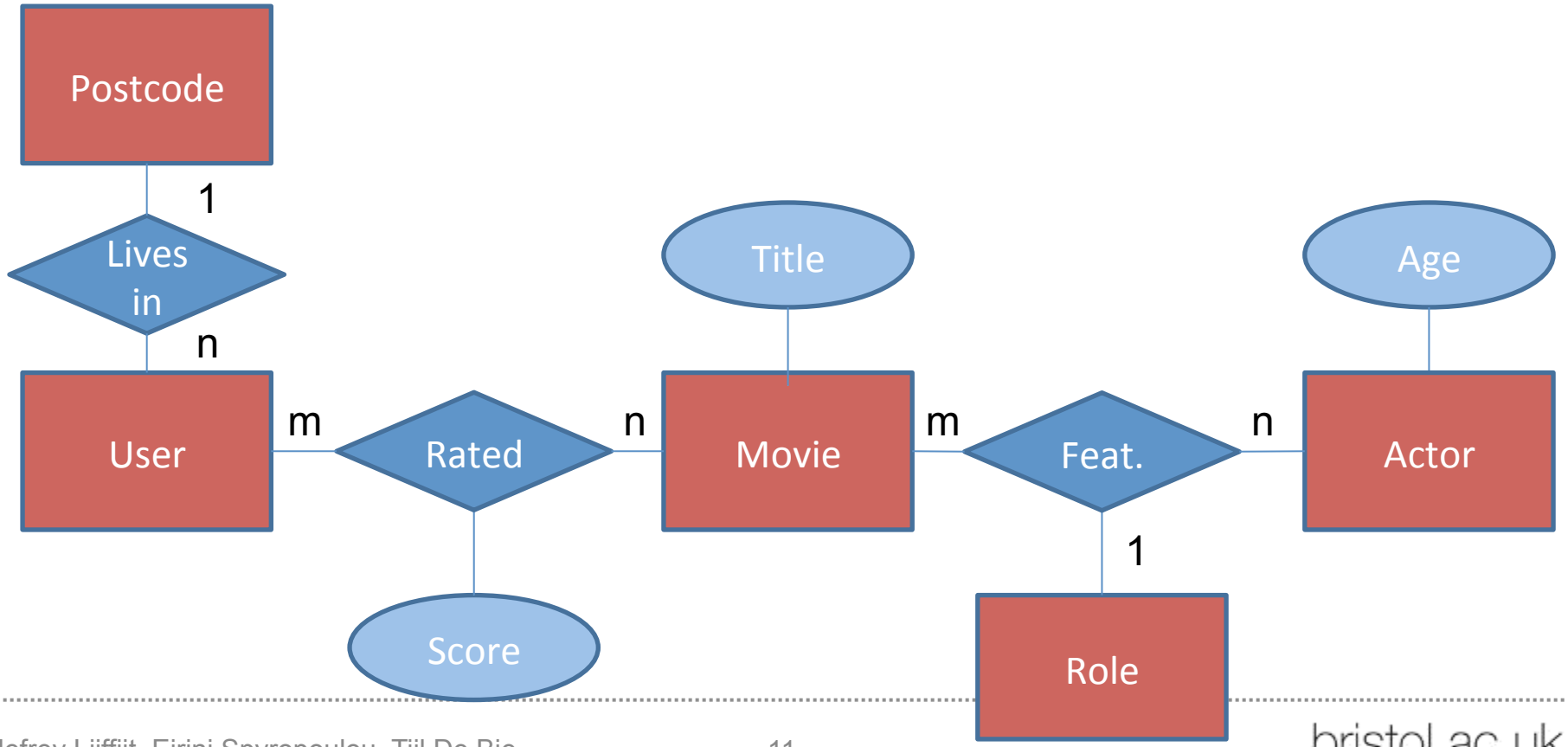
- Used for:
  - Online Analytic Processing (OLAP)
- Technology:
  - Multi-Dimensional Database Management System (MDDMS)

ORACLE
ENTERPRISE PERFORMANCE MANAGEMENT

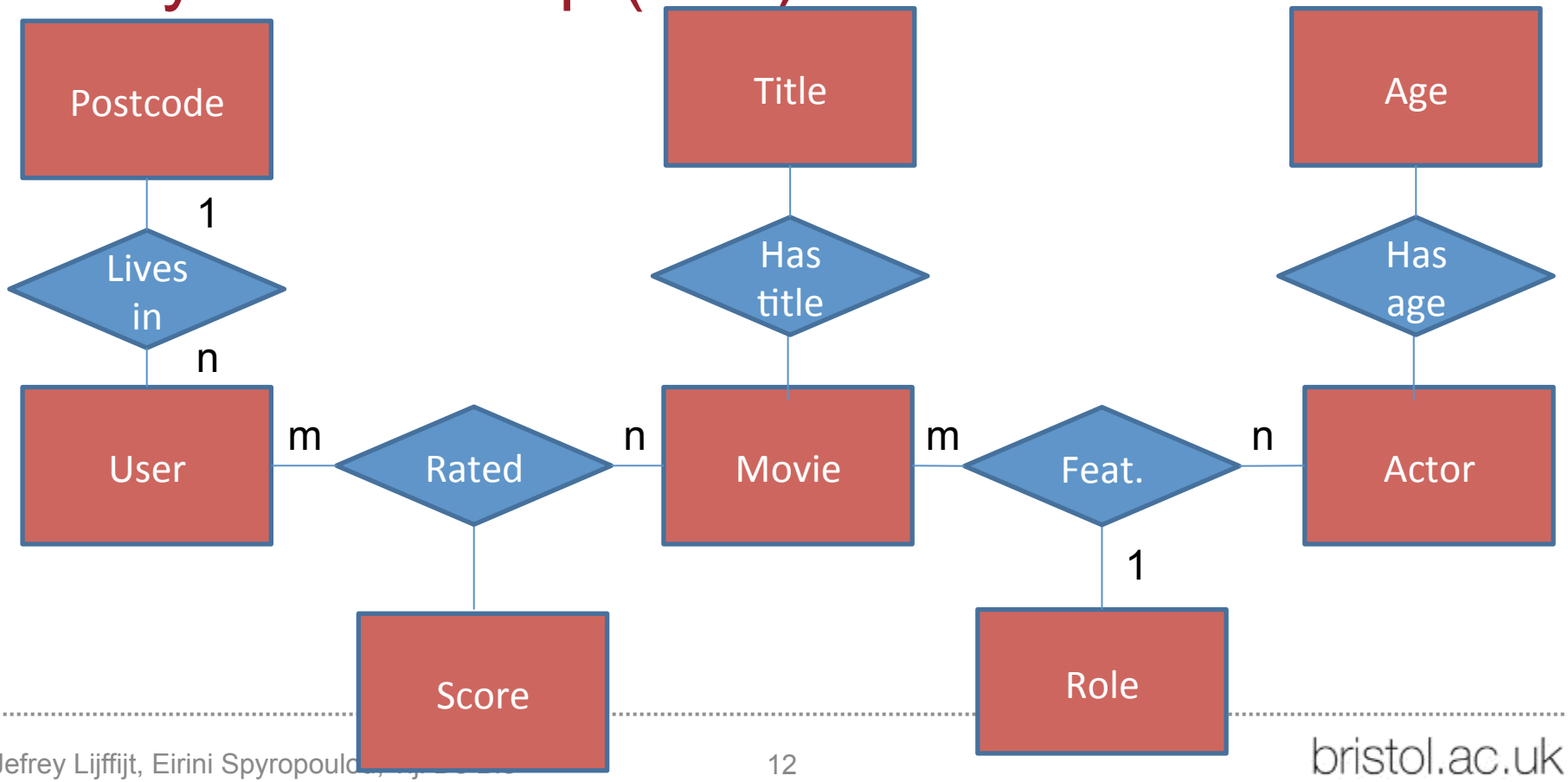
Oracle Essbase

Microsoft SQL Server
ANALYSIS SERVICES

Entity-relationship (E-R) data model

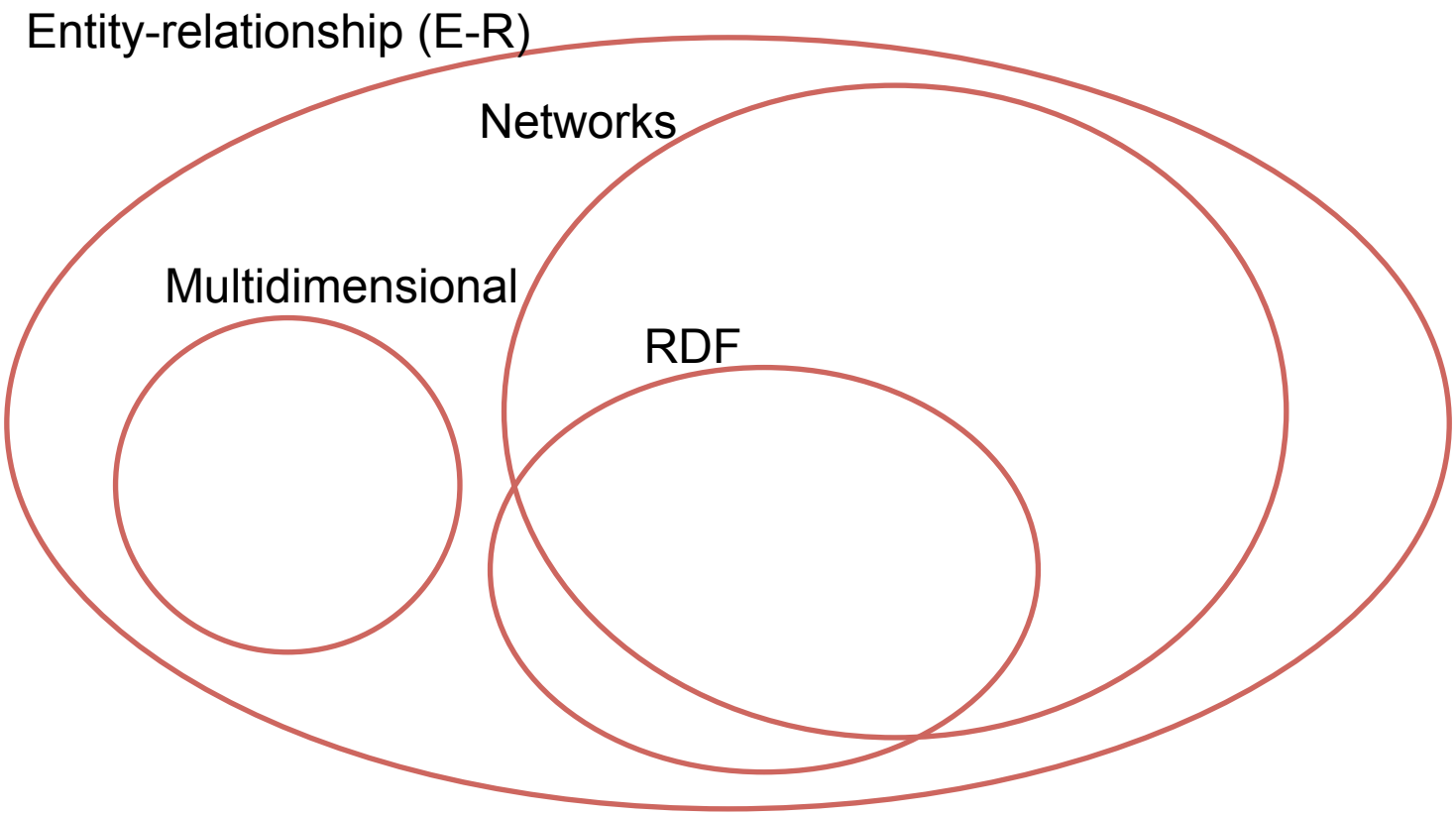


Entity-relationship (E-R) data model



Entity-relationship (E-R) data model

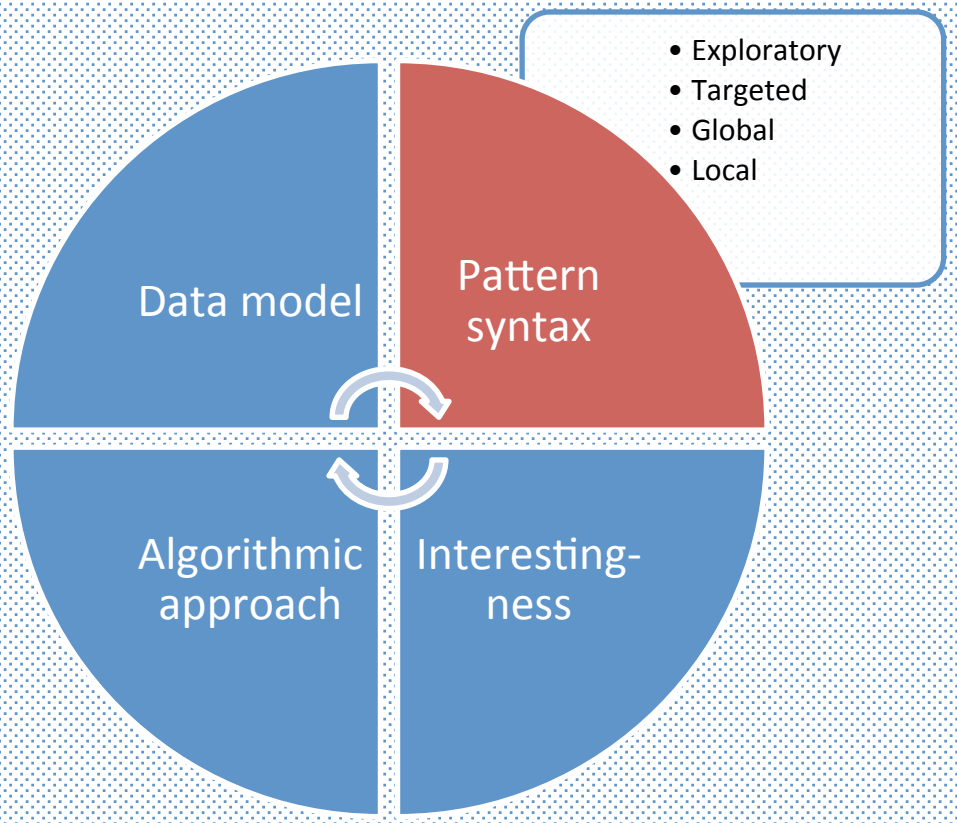
- Used for:
  - Wide diversity of applications
  - Capable of representing other data models
- Technology:
  - Relational Database Management System (RDMS)


Entity-relationship (E-R)
Networks
Multidimensional
RDF

Data models summary

Data model	Technology	Product
Networks	GDMS	Neo4J, Titan
RDF data model	Triple stores	Jena, Ontotext's GraphDB
Multidimensional data model	MDDMS	Oracle Essbase, MS SQL Server Analysis Services
Entity-relationship (E-R) data model	RDMS	MySQL, PostgreSQL

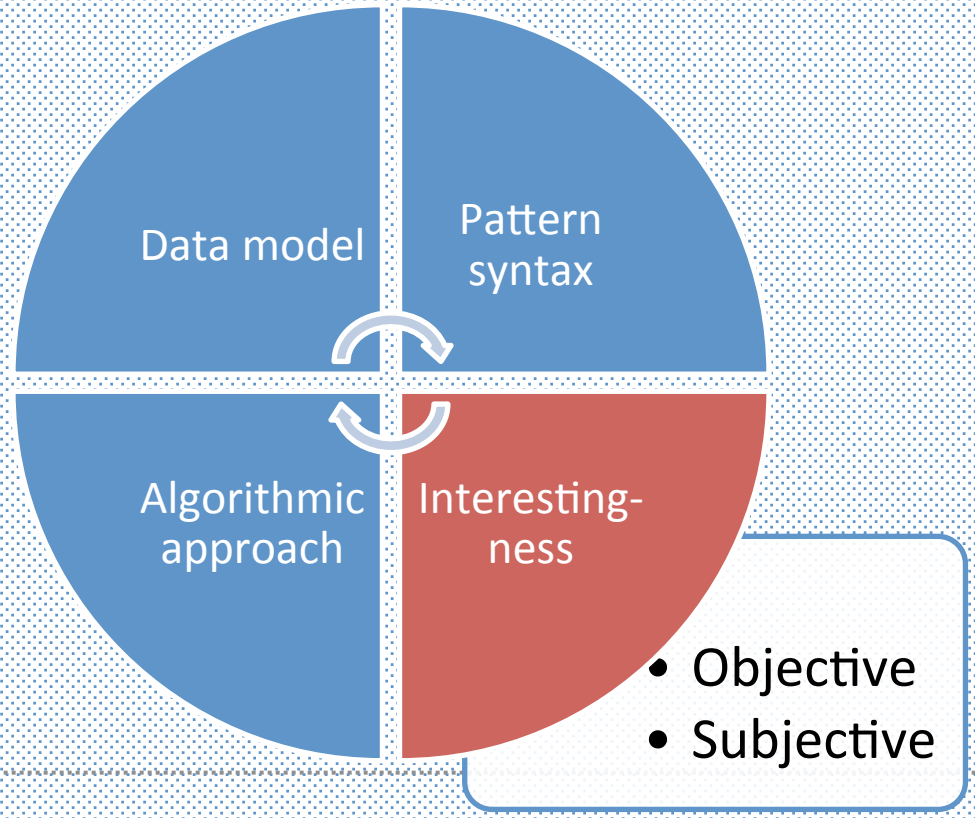
Pattern syntax



Pattern syntax

Data model	Typical pattern syntaxes	Targeted / Exploratory	Local / global
Network	Cliques Dense subgraphs	Exploratory	Local
RDF	[early days]	[Targeted?]	
Multi-dimensional	n-sets Tensor factorisations	Exploratory	
Entity-relationship (E-R)	Safarii / 'Multi-relational data mining' Relational Krimp Inductive Logic Programming	Targeted	
	MCCSs (RMiner) Smurfig Constraint Programming for closed relational sets Uncovering the plot	Exploratory	
	Coupled matrix-tensor factorisations	Exploratory	Global

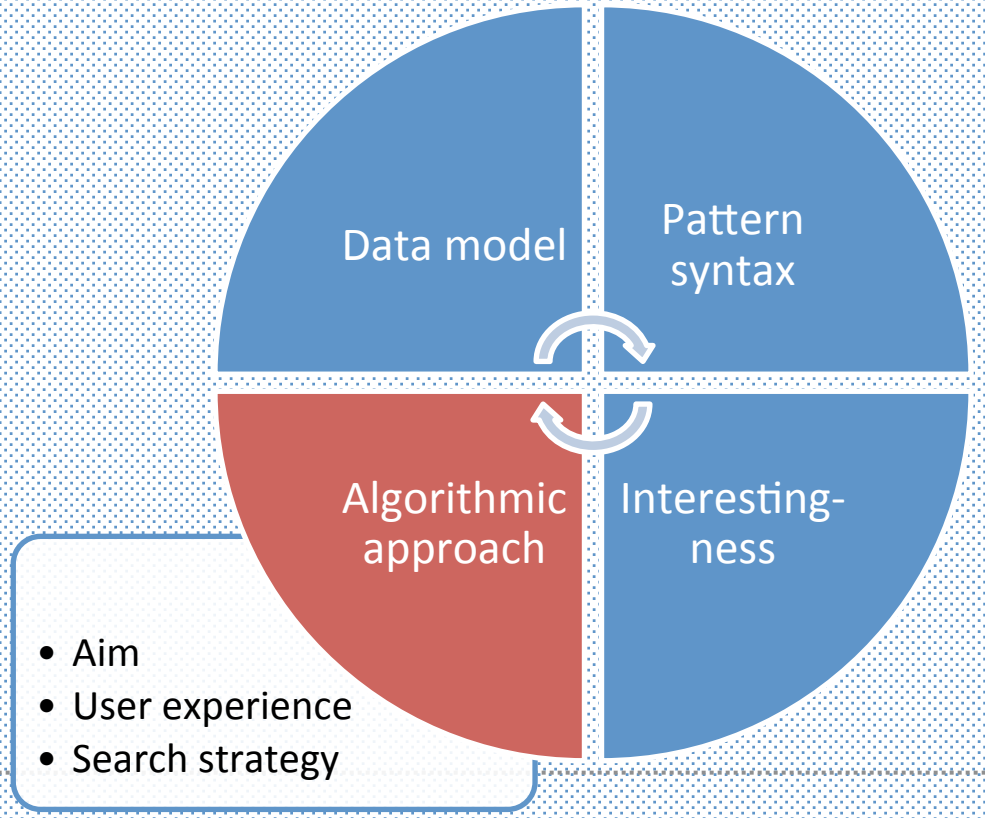
Interestingness



Interestingness

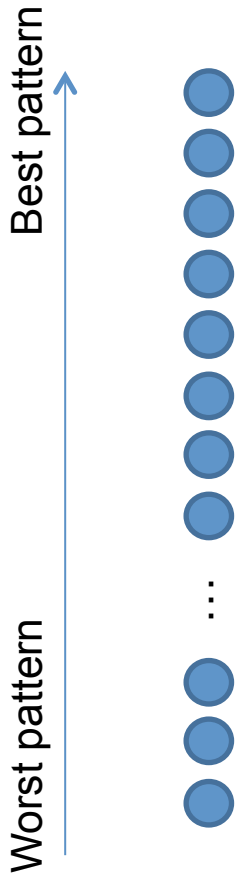
- Objective:
  - Independent of the user
    - Statistically inspired
    - Compression-based
    - Based on a physical (noise) model
    - Often pragmatic choice…
- Subjective:
  - Dependent on the user
    - Their prior knowledge or beliefs

Algorithmic approach



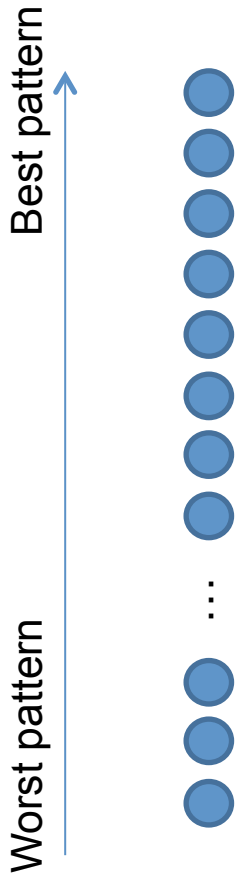
Algorithmic approach

- Aim
  - Exhaustive enumeration
  - Single best
  - Top-K best
  - Best set of patterns ("pattern set mining")
- Approximate / exact



Algorithmic approach

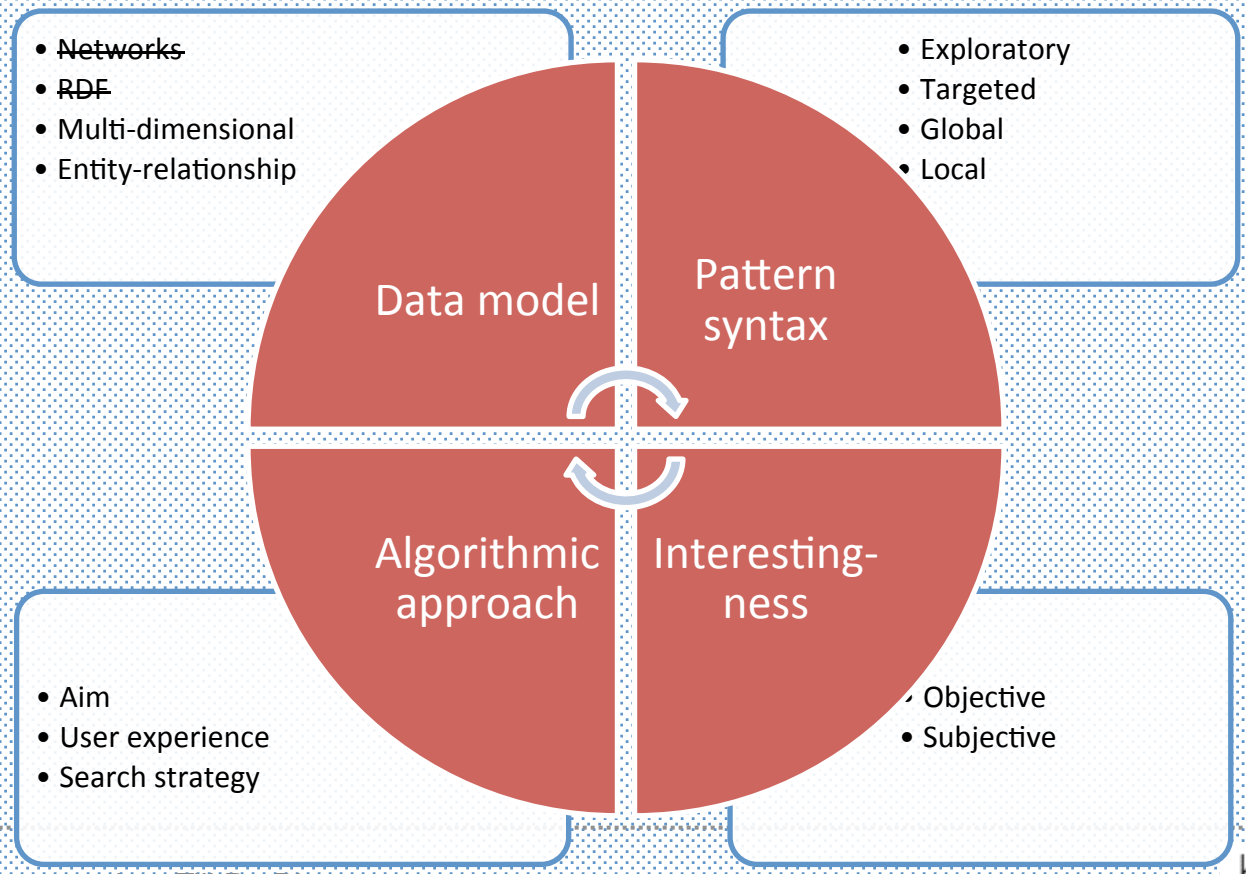
- User experience
  (in pattern set mining)
  - One-shot
  - Iterative
    - Greedy approximation for pattern set mining
  - Interactive
    - User affects ranking



Algorithmic approach

- Search strategy
  - Discrete search
    - Generic approaches (‘levelwise search’, branch-and-bound, divide-and-conquer…)
    - Generic tools, e.g. Constraint Programming solvers
    - Dedicated algorithms
  - Numerical optimisation
    - Convex
    - Eigenvalue problem
    - Nonlinear optimisation

Tutorial outline



Tutorial outline

- Part II: Targeted fully relational approaches
  - Safarii / 'Multi-relational data mining'
  - Relational Krimp
  - Inductive Logic Programming (ILP)
- Part III: Local semi-relational approaches
  - Frequent itemsets on the join
  - SMuRFIG
- Part IV: Local fully relational approaches
  - n-sets
  - RMiner
  - Constraint programming for closed relational sets
  - Uncovering the plot
- Part V: Global fully relational approaches
  - Coupled matrix-tensor factorisations
- Part VI: Perspectives

European Research Council

Established by the European Commission