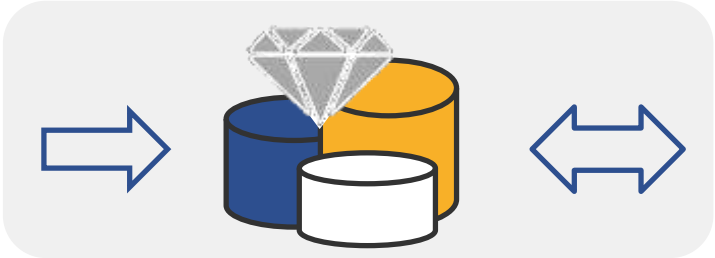
technische universität
dortmund

Chair of Data Science and
Data Engineering
Prof. Dr. Emmanuel Müller

01110010 011
00001 011
10111 001000
00 01100100 01
100001 011101
00 01100001



Trustworthy Machine Learning
- Methods, Systems, and Education -

Prof. Dr. Emmanuel Müller
Chair of Data Science and Data Engineering

Machine Learning – Applications

Digital Health

- Gene Expression Analysis
- Patient Health Records
- Drug Side Effect Analysis
- Personalized Medicine
- …

CORRELATION IS NOT CAUSATION

Data Science – Enabling Technologies

Ubiquitous Sensing
Open Data Sources

Cheap Storage
Efficient Processors
Cloud Computing

Machine Learning Methods
Artificial Intelligence
Big Data Analytics

Data Science – Missing Methods

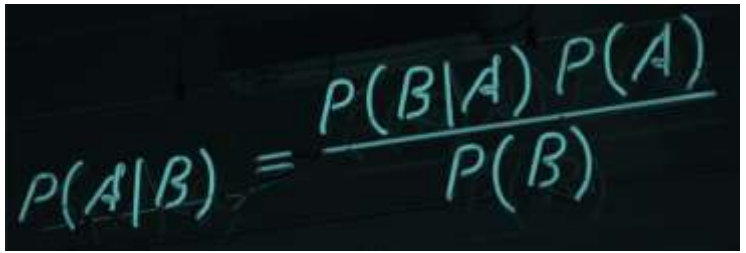
Ubiquitous Sensing
Open Data Sources

Cheap Storage
Efficient Processors
Cloud Computing

Machine Learning Methods
Artificial Intelligence
Big Data Analytics

(1) Data Science Methods

Bayes Theorem, Statistical Tests,
Correlation Measures, …



(+) Explainability of Results

Data Exploration, Descriptive Data Mining,
Human Computer Interaction, …

Data Science – Missing Infrastructures

Ubiquitous Sensing
Open Data Sources

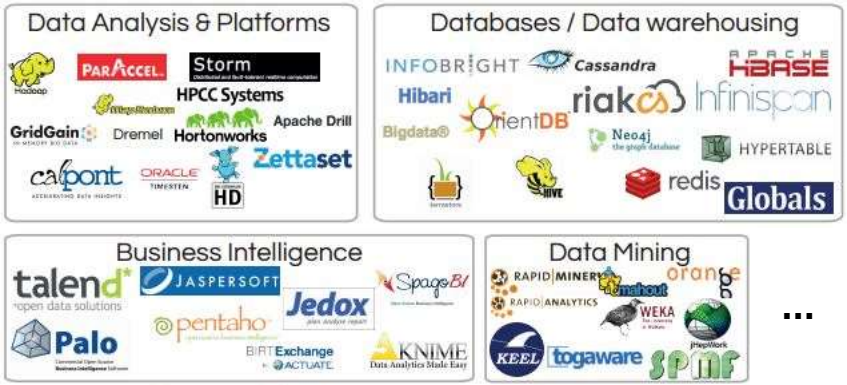
Cheap Storage
Efficient Processors
Cloud Computing

Machine Learning Methods
Artificial Intelligence
Big Data Analytics

(1) Data Science Methods

(2) Big Data Systems

Data Management, Data Analytics,
Data Visualization, …



Data Science – Missing Human Resources

Ubiquitous Sensing
Open Data Sources

Cheap Storage
Efficient Processors
Cloud Computing

Machine Learning Methods
Artificial Intelligence
Big Data Analytics

(1) Data Science Methods

(2) Big Data Systems

(3) Basic Education

Technical Experts, Domain Experts,
General Understanding, …

Methods and Systems – Predictive Maintenance

- Unsupervised **anomaly detection** and **description** in various domains…




We develop novel machine learning methods and systems

- Artificial neural networks for deep one-class classification by Lukas Ruff et al. @ ICML 2018 and ICLR 2020
- Representation learning for unsupervised stream anomaly detection by Bin Li et al. submitted to ICDE 2021

Bundesministerium für Verkehr und digitale Infrastruktur

Our Methods – Stream Data Analytics

multivariate data streams

…01101010101
…01001010100
…10100001001
…00110010011

Big Data Analytics

- predictive analytics
- cluster extraction
- anomaly detection
- **correlation / causality**

… many more …

We develop novel statistical methods

- Non-redundant feature selection methods
  by Arvind Shekar et al. @ ECML PKDD 2017 and DS 2018
- Two-Sample Testing for Event Impacts in Time Series
  by Erik Scharwächter et al. @ SDM 2020

Interactive Systems – Industrial Projects

Automobile Industry:

- driver assistance systems
- production optimization
- **predictive maintenance**
- autonomous driving
- …

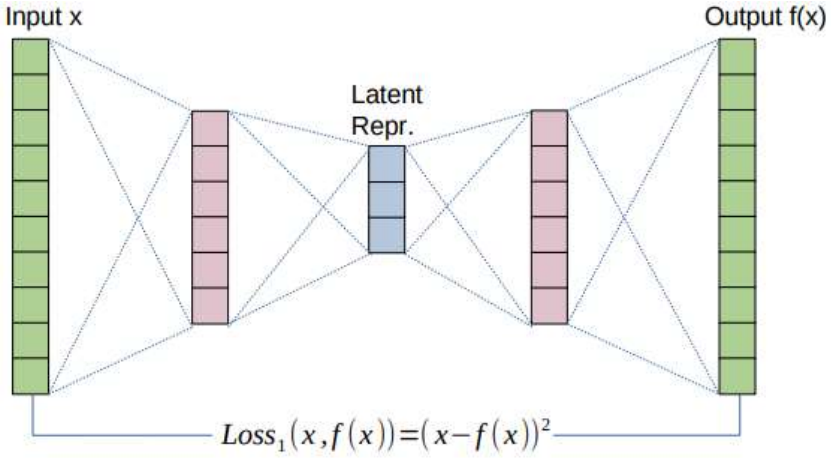
We develop novel data exploration systems

- We enable the understanding of statistical feature selection by Louis Kirsch et al. @ ECML PKDD 2017

Formal Methods – Guarantees and Explanations

Unsupervised Machine Learning
beyond given data sets

- **formal (provable) guarantees**
- uncertainty quantification
- feature importance
- …

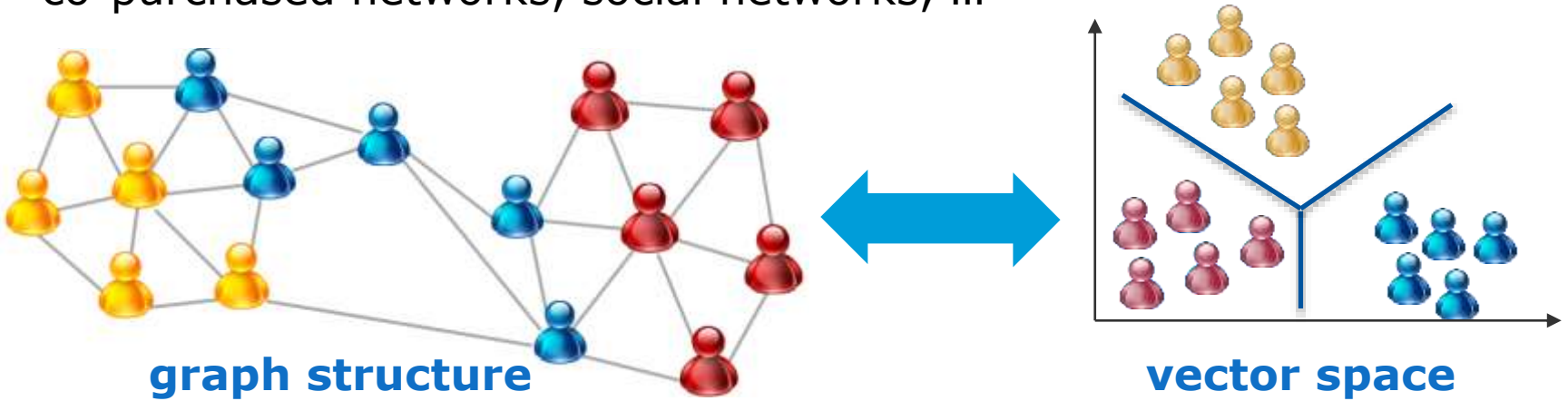


We study guarantees and explanation on unsupervised models

- We enable the formal quality guarantees via unsupervised adversarial attacks by Benedikt Böing et al. @ ECML PKDD 2020
- We study intuitive feature importance for unsupervised learning by Chiara Balestra et al. submitted to SDM 2021

Formal Methods – Graph Mining and Exploration

- Several application domains: communication networks, co-purchased networks, social networks, …

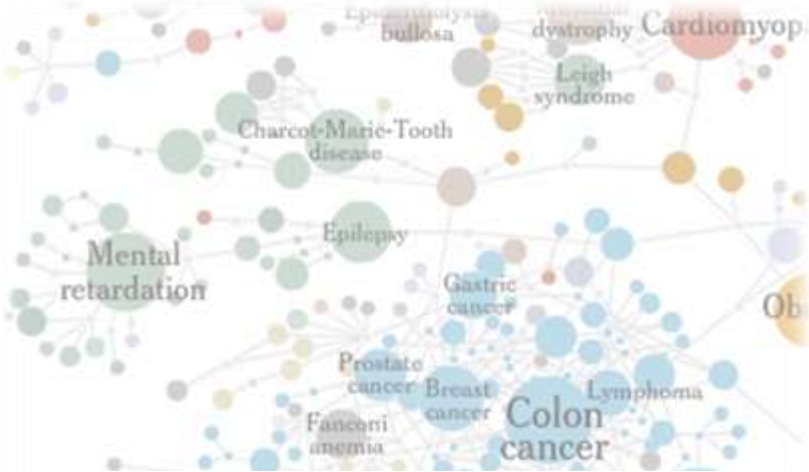


graph structure **vector space**

We develop novel graph mining methods

- Versatile graph embeddings and graph similarity measures
  by Anton Tsitsulin et al. @ WWW 2018, KDD 2018, WWW 2019, ICLR 2020

Exploration Systems – Interdisciplinary Projects



Bio-Medical Science:

- data-driven computational biology
- **knowledge representation**
- automated pattern discovery
- personalized health
- …

neo4j KNOWING HELMHOLTZ ASSOCIATION

We develop novel graph exploration systems

- Enabling interactive explanation and exploration of large knowledge graphs by Freya Behrens et al. @ WWW 2018

Trustworthy Machine Learning
- open challenges in research, industry, education -

Innovations for Large Scale Data Challenges

(1) Data Science Methods

(2) Big Data Systems

(3) Basic Education

Data Science and Data Engineering

Prof. Dr. Emmanuel Müller
Chair of Computer Science 9
Data Science and Engineering