technische universität dortmund

SFB 876 Providing Information by Resource-Constrained Data Analysis

Project A6

Resource-Efficient Graph Mining

Dr. Nils M. Kriege [from 2017], **Prof. Dr. Petra Mutzel, Prof. Dr. Christian Sohler, Prof. Dr. Kristian Kersting** [until 2017]

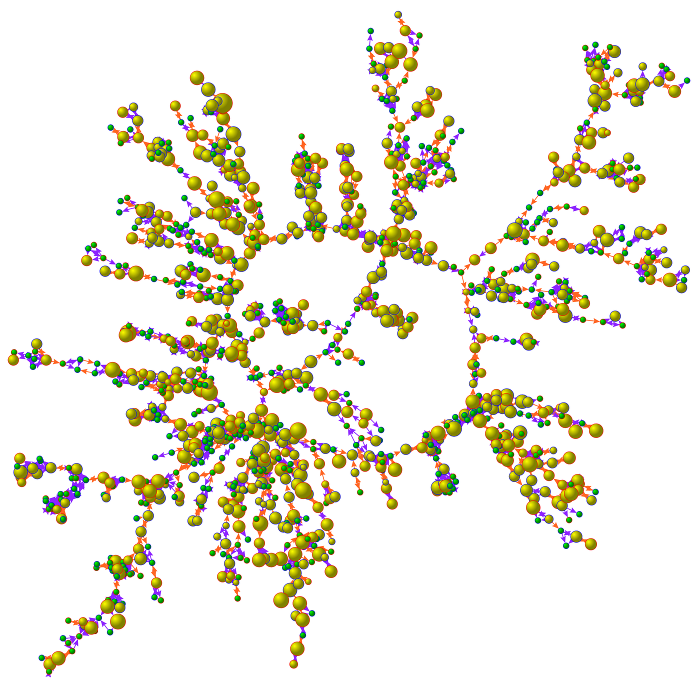
Problem

Graphs are ubiquitous and complex

Graphs occur, *e.g.*, in **chem-**, **bioinformatics**, and **social network analysis**

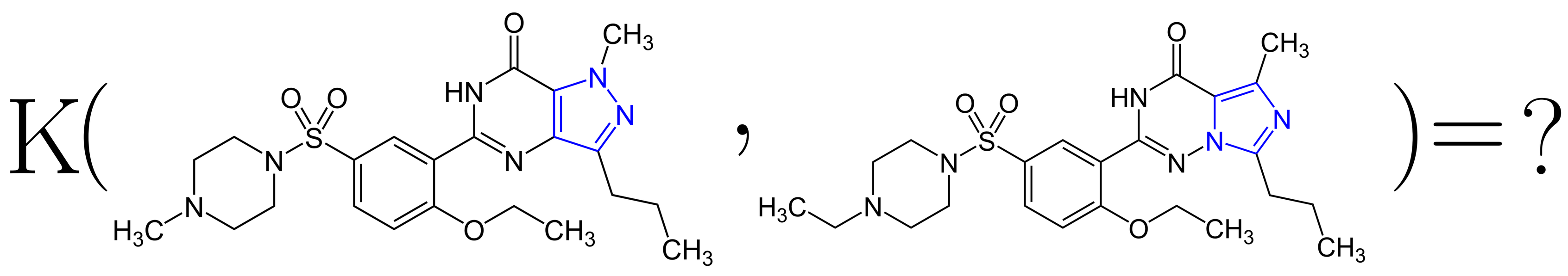
Real-world graphs . . .

- are **large**
- are **irregular**
- are **sparse**
- have a **hierarchical structure**
- have **rich information** attached to **nodes and edges**



Graph classification

How similar are two graphs?

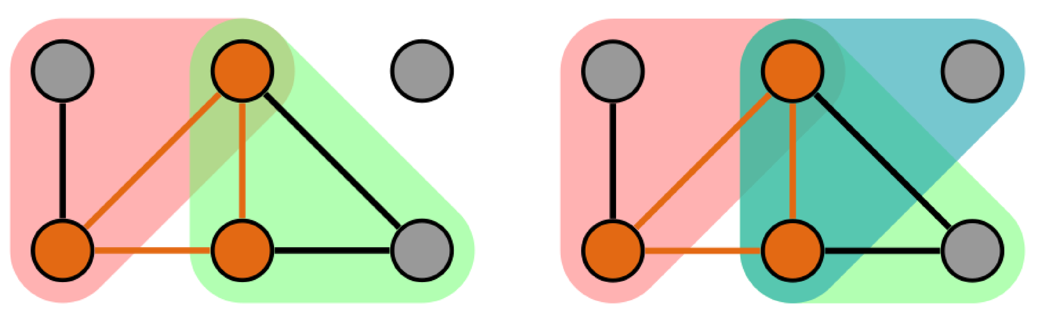


- **No obvious vector representation** for **graphs**
- **Standard** machine learning approaches **cannot be applied directly**

Methodology

Vertex refinement

Local k-dimensional Weisfeiler-Lehman kernel [ICDM 2017]

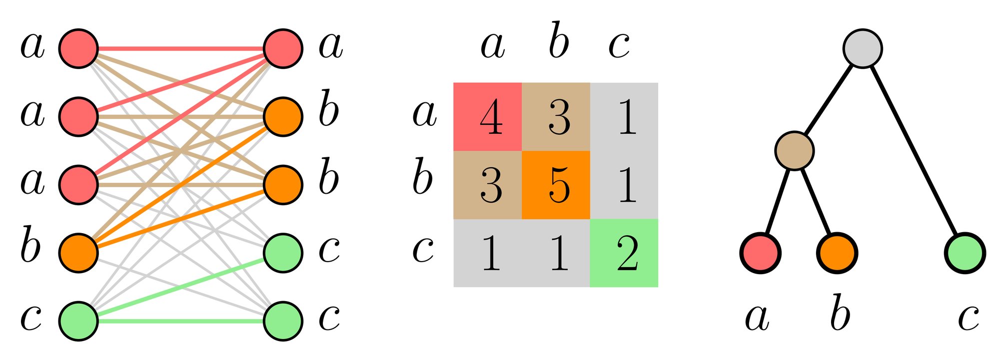


Local neighbourhood   Global neighbourhood

- **Parallel sampling algorithm** with **provable approximation guarantees**
- Considers **local** as well as **global properties**

Assignments via trees

Kernels from optimal **assignments** [NIPS 2016]

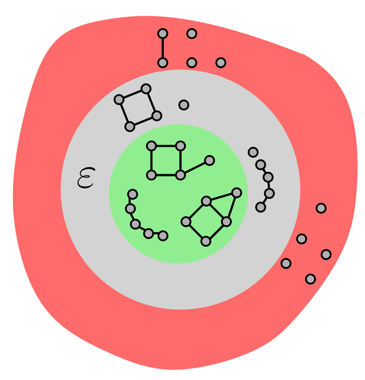


Assignment   Base-kernel   Tree

- **Base kernels** that induce PSD **assignment kernels**
- **Linear running time**
- Weisfeiler-Lehman **Optimal Assignment kernel**

Theoretical expressivity

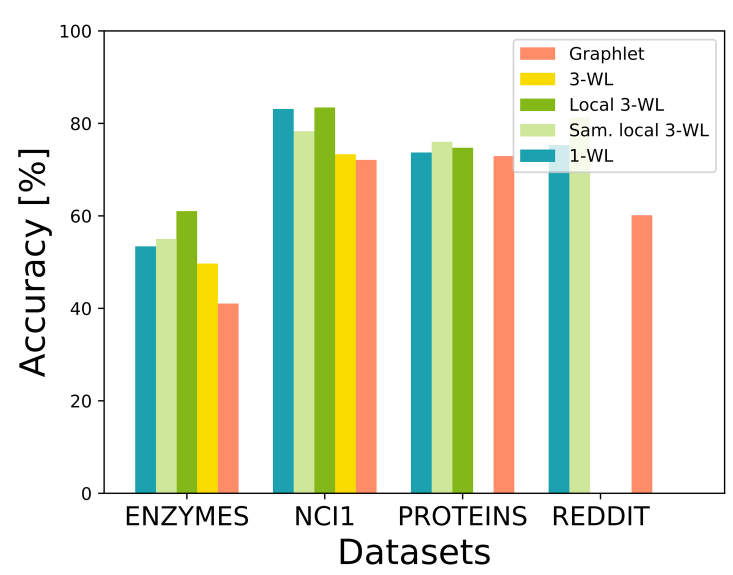
Property testing framework for **graph kernels** [IJCAI 2018]



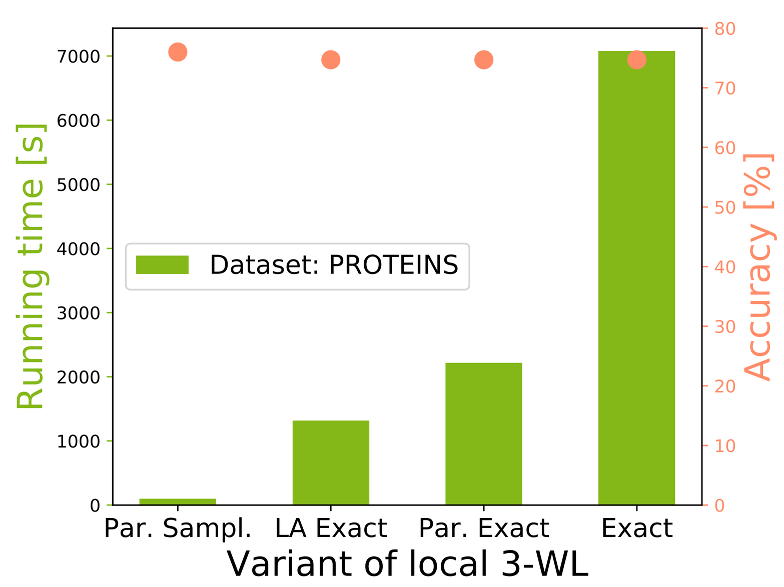
Graph property: connectivity

- Current kernels **cannot discriminate** between **simple graph properties**
- **More expressive kernel** based on **k-disks**

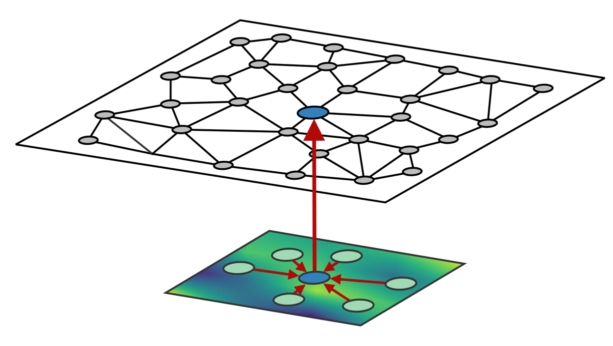
Results



Local 3-WL accuracy



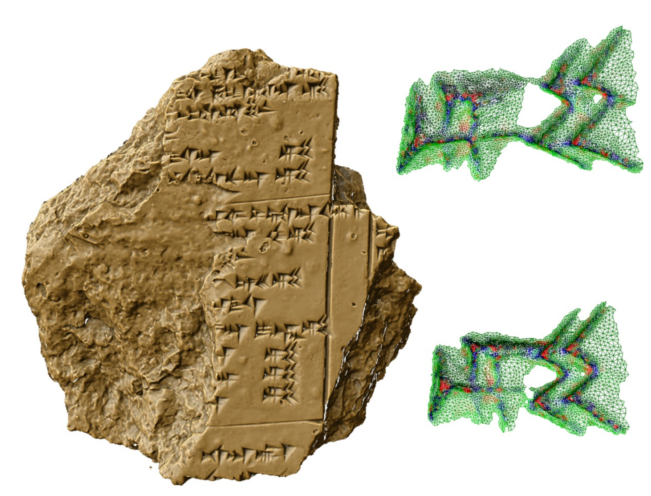
Speed-up by sampling and parallelism



SplineCNN (B2)

PyTorch geometric

Geometric deep learning framework (B2)



Cuneiform classification (B2)

Additional selected results

Scalability

- **Spectrum approximation** via random walks [KDD 2018]
- **Scalable kernels** for graphs with **continuous labels** [ICDM 2016]
- **Scalable kernels** via **propagation** [Machine Learning 2016]
- **Scaling lifted probabilistic inference** and **learning via graph databases** [SDM 2016]

Expressivity

- **Counting cycles** of large graphs [Algorithmica 2018]
- Parameterizing the **distance distribution** of undirected networks [UAI 2015]
- Architecture for **tractable multivariate Poisson distributions** [AAAI 2017] (B4)
- **Machine Learning** meets **data-driven journalism** [#Data4Good@ICML 2016] (A1)

Towards the third phase

- **Cuneiform classification** on limited data [COST@SDM 2018] (B2)
- **Hierarchical pooling** [NIPS 2018]
- **Fast deep graph learning** via **B-spline kernels** [CVPR 2018] (B2)
- **Protein complex similarity** [Preprint] (C1)
- **Group equivariant** deep learning [NIPS 2018] (B2)

Informatik 11
Algorithm Engineering

Informatik 2
Efficient Algorithms and Complexity Theory