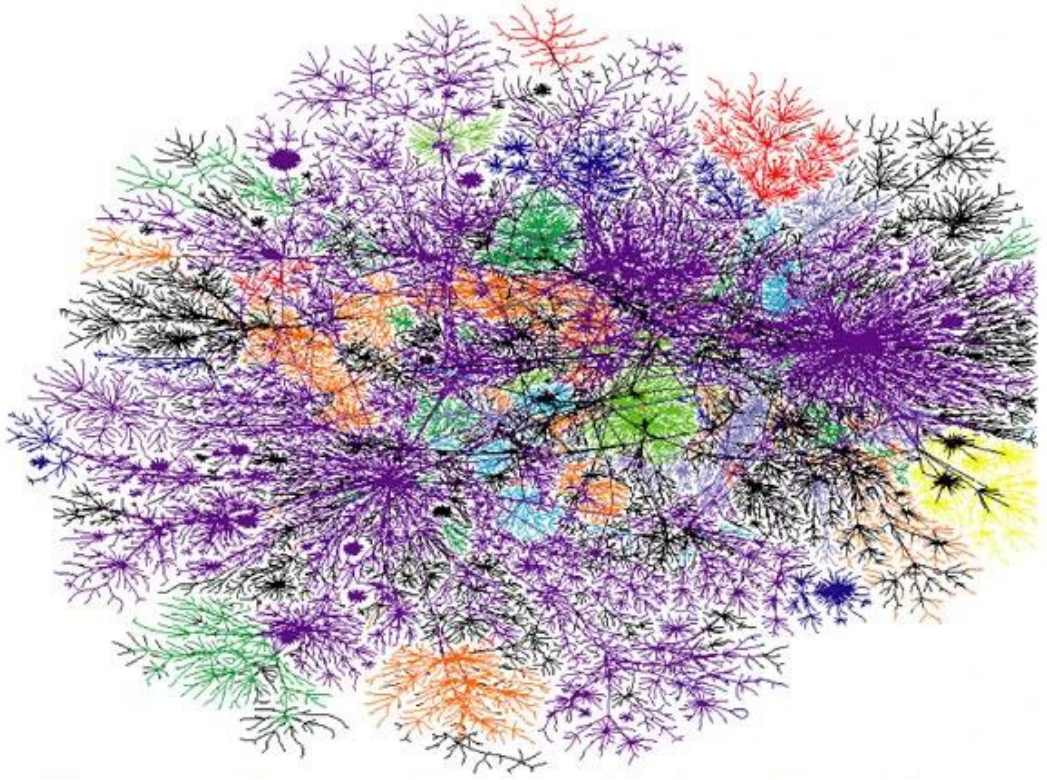
Graph Similarity Search

Team Project HWS 2025
Simon Forbat

Chair: Prof. Gemulla

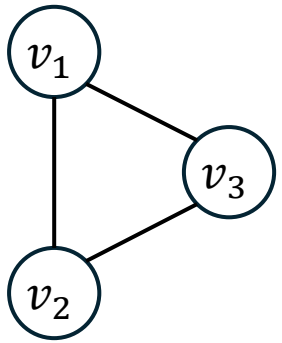
- Graphs are ubiquitous in computer science
  - Social networks, transportation networks, communication networks, molecules, recommender systems, …
- Fundamental task: **comparing (sub-)graphs**
- Highly relevant
  - Contrastive learning
  - Few-shot learning
  - Kernel methods
  - Graph mixup
  - GraphRAG
  - …



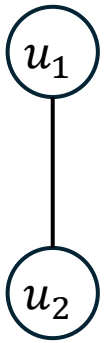
Tu, Y. How robust is the Internet?. *Nature* **406**, 353–354 (2000). https://doi.org/10.1038/35019222

Graph Edit Distance (GED)

- Minimum number of elementary edit operations
  - Insert, delete, or substitute a node or an edge
- Key technique for comparing graphs and foundation of more advanced methods
- Example
  - Delete two edges
  - Delete one node
  - $GED(G_1, G_2) = 3$



G_1



G_2

Problems

- **Efficiency:** GED computation is NP-hard
  - Many approaches proposed (combinatorical, heuristic, deep learning, …)
- **Trade-offs:** approaches have different strengths and weaknesses
  - Efficiency vs. accuracy
  - Efficiency vs. flexibility (e.g., support for weights)
  - Applicable hardware (CPU, GPU)
- **Availability:** Implementations not always available or accessible
  - E.g., may not run “out-of-the-box”
  - E.g., in C++ but required in Python / PyTorch
  - E.g., inflexible to use
- Generally not clear when to use which methods
  → **Active area of research!**

Goals of team project

- **Explore and evaluate** literature and implementations
  - Evaluation of existing methods on a set of relevant and diverse datasets
  - Find advantages and disadvantages

- **Improve** efficiency
  - Accelerate existing algorithms or implementations (on CPU/GPU)
  - Implement fast all-pairs similarity search

- Enable **generalization**
  - Support for weights per class of operations
  - Support for different weights for every individual operation
  - Why? E.g., different users in social networks are (typically) not equally important
    - Removing Taylor Swift (likely) has larger impact than removing Gringo Mayer
    - Removing Taylor Swift should incur larger cost than removing Gringo Mayer
    - Or the opposite, depending on preference

- Ideally, **publish findings** in a peer-reviewed scientific venue

Requirements

- **Prerequisites**
  - Interest/background in algorithms or algorithm engineering
  - Strong programming skills and willingness to learn and work with high-performance languages such as C / C++ or CUDA

- **TP characteristics**
  - Language: English
  - Duration: 6 months
  - No. of participants: 3-5
  - MMDS: yes
  - Online: no
  - Contact: Simon Forbat <simon.forbat@uni-mannheim.de>