Data-Intensive Systems

Bachelor project proposals 2023

Ira Assent, Cigdem Aslay, Panagiotis Karras, Davide Mottin

Our research group

Our interests

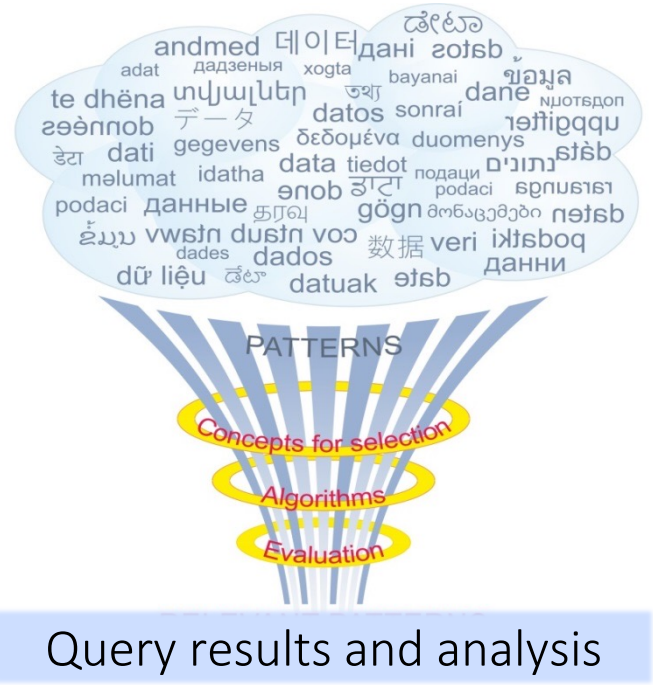
- Data management, data analysis, machine learning
- Efficiency and scalability
- Models and algorithms

Our methods

- Formalize real world problems
- Devise concepts and algorithmic solutions
- Evaluate empirically

Our tools and results

- Theoretical analysis and prototype implementations



Query results and analysis

Projects in a Nutshells

BESo: Beyond Entity Summarization (4)

SaSSo: Scalable (Spectral) Subgraph Localization (5)

Densely Connected Subgraphs in Dual Graphs (DCS-Dual) (9)

Elastic Graph Indexing (10)

Graph intelligence

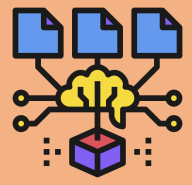
Evolutionary games and spatial upstream reciprocity (1)

Structural social balance under controversy (2)

Similarity Search with Dynamic Time Warping (3)

Using 3D-Shape Descriptors for Clustering of Molecular Dynamics Data (8)

Interaction Models



Studying the fairness of ML models (6)

Automatically labeling data for classifying rhetorical appeals (7)

ML for detecting issues

Graph intelligence

BESo: Beyond Entity Summarization

How can we describe an entity (e.g., a patient) compactly?

Densely Connected Subgraphs in Dual Graphs (DCS-Dual)

How do we find communities in a very special graph?

SaSSo: Scalable (Spectral) Subgraph Localization

How do we efficiently find a graph inside another graph?

Elastic graph indexing

Can we use the ”shape” of the graph to search fast?

Interaction models

Using 3D-Shape Descriptors for Clustering of Molecular Dynamics Data

Can we find similar proteins by using their 3D-shape?

Evolutionary games and spatial upstream reciprocity

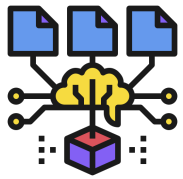
Why do we spend time, effort, and resources in cooperating with others?

Structural social balance under controversy

Can there be stability if controversial arguments are discussed?

Similarity Search with Dynamic Time Warping

How can we efficiently search time series under complex distance models?



Detecting issues

Studying the fairness of ML models

Is a ML algorithm fair? Is the underlying data fair? What are good models for fairness?

Automatically labeling data for classifying rhetorical appeals

Can we detect rhetorical appeals in texts e.g. to detect misinformation?