University of Fribourg / Faculty of Sciences and Medicine / Department of Informatics

# Unsupervised and Parameter-Free Clustering of Large Graphs for Knowledge Exploration and Recommendation

Artem Lutov

We live in an Information Age, facing a rapid increase in the amount of information that is exchanged. This permanently growing amount of data makes the ability to store, analyze, and act upon information a primary concern, raising the question: "How can one consume Big Data and transform it into actionable knowledge?". The crucial aspect in addressing the stated question is the *approach to transform the abundance of the retrieved information in a usable* (i.e., human-adapted) *and efficient way*.
In this thesis, we first overview existing research on human perception and visual analytics in Chapter 1 and Chapter 2, which leads to a set of criteria according to which the information should be transformed into knowledge to become human-readable while reflecting multiple viewpoints. Then, in Chapter 3, we discuss existing approaches to construct such a transformation and propose our novel clustering algorithm, DAOC. Our method efficiently produces a human-adapted hierarchy of clusters for any input graph without any manual tuning. To ensure that the results are accurate besides being human-readable, in Chapter 4 we analyze existing approaches of the clustering results validation. In Chapter 5, we introduce Clubmark, a new framework for benchmarking and profiling clustering algorithms on NUMA architectures. Our framework includes a dozen of state-of-the-art clustering algorithms and evaluates them on standard real-world and synthetic datasets avoiding biases. In Chapter 6, we propose an effective and efficient approach, StaTIX, to infer and complete entity types in Linked Open Data based on our DAOC clustering algorithm. In Chapter 7, we bridge a gap between graph clusters (also known as network communities) and a low-dimensional representation of the graph nodes, presenting our novel graph embedding technique, DAOR. DAOR is a highly efficient and parameter-free graph embedding technique producing metric space-robust, compact and interpretable embeddings without any manual tuning. Our approach has the ambition to greatly simplify and speed up data analysis tasks involving graph representation learning (e.g., information recommendation and search).
We conclude the thesis by suggesting the integration of our innovative clustering method, DAOC, into Visual Analytics frameworks and data search interfaces in Chapter 8. In the future, our approach might provide a higher diversity of the results without sacrificing their accuracy and, hopefully, reducing social polarization.

Jury:
Prof. Dr. Philippe Cudré-Mauroux (thesis supervisor)
Dr. Mourad Khayati (internal co-examiner)
Prof. Dr. Claudio J. Tessone (external co-examiner)
Prof. Dr. Ulrich Ultes-Nietsche (president of the jury)