EDBT-Intended Summer School Proposal

Amedeo Pachera, Lyon 1 University (France)

Research Topic

My current research focuses on the intersection of data quality and graph data management, particularly on leveraging human intelligence to support data quality tasks. Specifically, I investigate how Human-in-the-Loop paradigms can be employed to enhance data-intensive systems, with a focus on property graph models. My recent work proposes novel methods for user-centric graph repair, balancing algorithmic detection of inconsistencies with collaborative user input. This research has resulted in a paper accepted at the ACM SIGMOD 2025 research track, titled "User-centric Property Graph Repair"[1]. It introduces a framework for detecting and tracking violations in property graphs and demonstrates how multi-user collaboration can significantly improve the quality and efficiency of data repair.
The future direction of my research consists of employing deep learning algorithms (e.g. Graph Neural Network) to develop a hybrid system that combines human insights with the learning capabilities of AI models for graph data cleaning.

Motivation

My academic background includes a Master's in Computer Science and a PhD position at Lyon 1 University, where I'm co-advised by Prof. Angela Bonifati and Prof. Andrea Mauri. Throughout my studies, I developed a strong interest in the combination of human and machine intelligence, especially in the context of data management. My work spans graph databases, data cleaning, and human-computer interaction, while also drawing on methods from machine and deep learning.
What drives me is the challenge of bridging theoretical advances in data management with practical applications. My experience as a technical leader in a startup helped me develop a mindset for building scalable and production-ready systems, a skill I now apply to the development of research prototypes.
The EDBT Summer School topics this year are highly relevant to both my current research and broader academic interests. In particular, the sessions on Neuro-inspired AI for Continual Learning and AI Planning for Data Exploration align with my focus on combining human and machine intelligence to improve data-centric systems. The exploration of Interconnection of Heterogeneous Data using AI is directly tied to my work with graph data models, where integrating diverse and often noisy data sources is a key challenge. I'm also intrigued by the discussions on the mathematics of AI, which I believe will offer valuable insights into formalizing and reasoning about complex AI-driven data processes.
Attending this summer school would significantly enrich my understanding of these cutting-edge topics, allowing me to engage with fellow researchers, and inspiring new interdisciplinary approaches to my own work.

Participating in this summer school would not only enhance my technical expertise, but also offer a valuable opportunity to connect with other researchers and foster future collaborations within the data management community.

Thank you for considering my application. I look forward to engaging with peers and experts at the EDBT Summer School and contributing to meaningful discussions in the field.

[1] PACHERA, Amedeo; BONIFATI, Angela; MAURI, Andrea. User-Centric Property Graph Repairs. *Proceedings of the ACM on Management of Data*, 2025, 3.1: 1-27.

Amedeo Pachera

Academic Appointments

Amedeo Pachera
Nationality : Italian
Date of birth : 02/04/1997
Email : amedeopachera@gmail.com
GitHub : https://github.com/pake97
Linkedin : https://www.linkedin.com/in/amedeo-pachera-08427a277/
Google Scholar : https://scholar.google.com/citations?user=L84kk3AAAAAJ

University Claude Bernard Lyon 1 / PhD Student
October 2023 - now, Lyon

My research lies in the integration of artificial and human intelligence to design data-intensive applications with the goal of making them efficient, effective, and aware of people's needs and values. This includes employing Human-in-the-loop techniques to enhance existing approaches in data management applications. I mainly focus on data quality tasks on graph data models, designing algorithm and user interactions to improve existing data cleaning approaches.

Advisors : Prof. Angela Bonifati, Andrea Mauri.

Politecnico di Milano / Research fellow
June 2022 - October 2023

Research focus and responsibilities: developing Data Science and Knowledge Engineering techniques for the extraction and enrichment of textual content for the analysis of the impact of the COVID-19 pandemic. The Periscope project has received funding from the European Union's Horizon 2020 Research and Innovation programme (https://periscopeproject.eu/). I was part of the team in charge of

developing Perseus Platform (https://perseus-platform.eu/), an integrated toolkit for participatory and strategic decision-making at the EU level.

Advisor : Prof. Marco Brambilla

Experiences

Focal SRL / CTO

June 2022 - October 2023, Verona (IT)

Activities: I was the leader of the developing team of Focal, an italian startup. We developed a web platform for consultants (lawyers, accountants, project managers, …) based on a search engine that leverages semantic search with vector indices. The platform allows users to find the best consultant among different niches by prompting a specific need. We built the platform using the NextJS framework, integrating Zilliz as the vector database and AWS S3 for storage.

We also took part in several joint projects with the John Lab (https://johnbosco.it/), a laboratory that aims to help students from high schools and universities to develop their ideas and teach them entrepreneurship skills.

Education

Politecnico di Milano / Master Degree in Computer Science and Engineering

September 2019 - April 2022, Milan (IT)

Thesis: KlevR and DeepR: a benchmark for exploring deductive reasoning functionalities of variational autoencoders.

Advisor: Prof. Emanuele della Valle. Tutor: Prof. Riccardo Tommasini.

Università di Padova / Bachelor degree in in Computer Science and Engineering

September 2015 - July 2019, Padua (IT)

Thesis: Sensor fusion for Android devices

Advisor: Prof. Stefano Tommasin

Technical skills

Data structure and algorithms: definition and manipulation of data structures using c++, java and python.

Data analysis and visualization: data analysis using python libraries such as numpy and pandas and data visualization using matplotlib and seaborn.

Vector Databases and Large Language Models: using vector databases to store documents and retrieving them with similarity search. Using langChain framework to build LLMs applications and integrations.

Machine Learning, Deep Learning: defining a pipeline for training, validating and testing an AI model using TensorFlow and PyTorch, as well as using feature selection and fine-tuning framework such as Optuna.

Graph Manipulation: using cypher and Neo4j for graph querying and manipulation.

Semantic Web and Knowledge Engineering : defining OWL ontologies with different profiles (EL,QL and RL). Querying knowledge bases with SPARQL.

Web Development: development of full-stack web applications using ReactJs and NextJs framework as well as python Flask.

Publications

- Amedeo Pachera, Angela Bonifati, Andrea Mauri. **User-Centric Property Graph Repairs**. *Proceedings of the ACM on Management of Data*, 2025, 3.1: 1-27.
  Abstract : Property graphs serve as unifying abstractions for encoding, inspecting, and updating interconnected data with greater expressive power.
  They are increasingly popular across various application domains involving real users. However, graph data often contains inconsistencies that need proper transformations to address underlying constraint violations and often require specific domain knowledge.
  In this paper, we propose an interactive and user-centric approach to repair property graphs under denial constraints.
  Our approach includes a novel theoretical framework comprising a query-based inconsistency detection mechanism, a dependency graph for tracking violations, and an assignment algorithm facilitating multi-user property graph repairs by leveraging independent sets.
  We evaluate our approach through several experiments on real-world and synthetic datasets, considering different levels of user expertise and comparing against various baselines. Even with multiple non-oracle users, our approach outperforms existing

interactive and non-interactive baselines by 30% on average in terms of repair quality. Additionally, we conduct a user study to assess real user performance in property graph repairs.

- Amedeo Pachera, Angela Bonifati, Andrea Mauri. **Grafixer: Enabling User-Centric Repairs for Property Graphs.** ACM SIGMOD 2025 (to appear).
  <u>Abstract:</u> We introduce Grafixer, an innovative interactive tool that enables human-in-the-loop repair of property graphs. By leveraging user expertise, Grafixer allows collaborative identification and correction of data inconsistencies while ensuring an efficient repair process. Users can upload property graph datasets and define constraints using Cypher, a widely adopted query language. The system efficiently manages overlapping data violations, ensuring that multiple users can work simultaneously on independent repairs without conflicts. Additionally, Grafixer provides an interactive dashboard that allows administrators to monitor real-time statistics, track progress, and oversee the repair process. In our demo, we showcase the complete repair workflow from both the administrator's and users' perspectives. Administrators can configure the repair process, monitor ongoing corrections, and assess the overall status of the graph. Meanwhile, users engage in the repair process by reviewing violations, proposing corrections, and contributing to improving data quality through an intuitive and guided interface.

- Amedeo Pachera, Angela Bonifati, Andrea Mauri. **Towards User-Centric Graph Repairs**. SEA-Graph@ICDE 2024.

—

Other

I had the privilege of being a member of the ARI (Availability & Reproducibility Initiative) commission at ACM SIGMOD 2024, a PC member for the CHI workshop EmpatiCH and a reviewer for TKDE (IEEE Transactions on Knowledge and Data Engineering) journal.