KG-PRE-view: Democratizing a TVCG Knowledge Graph through Visual Explorations

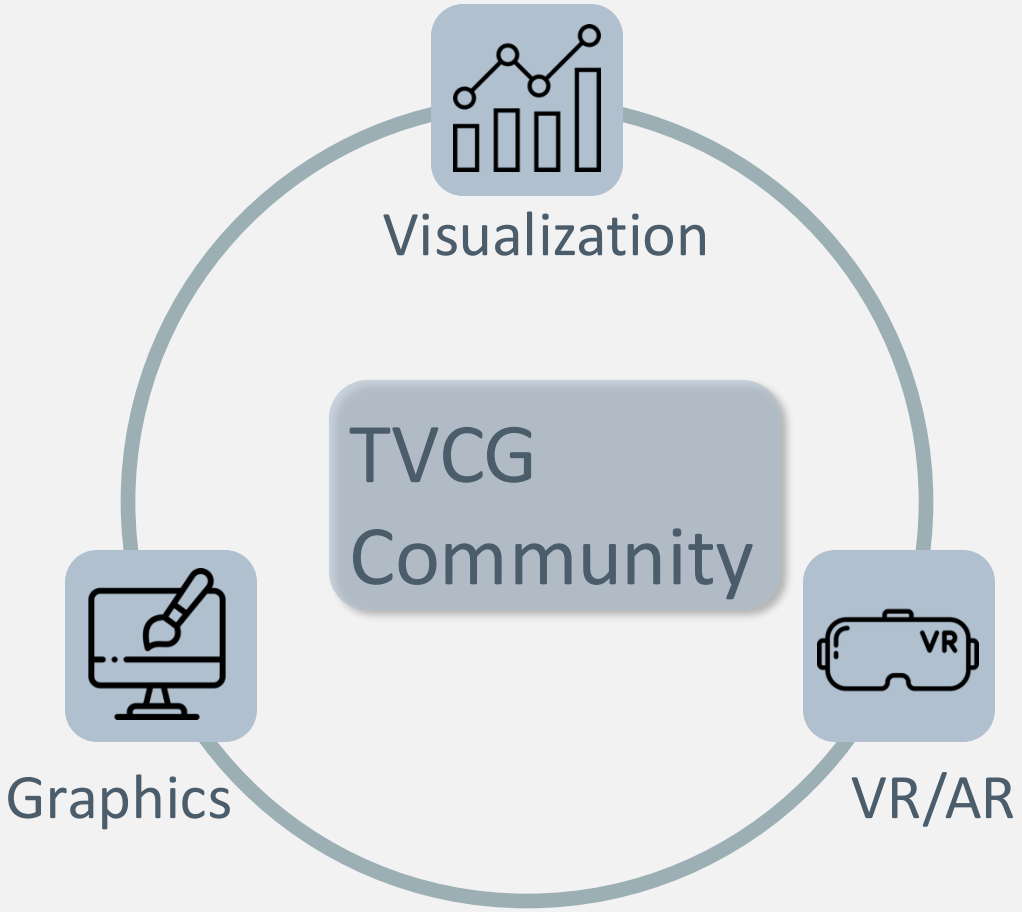
THE OHIO STATE UNIVERSITY

Author : Yamei Tu, Rui Qiu, Han-Wei Shen

Presenter: Haoyu Li

PacificVis 2024

Background



Retrieval of research ideas

Invitation of peer reviewers

selection of editorial members

Challenge

Using current online access to digital versions of papers directly in real-world scenarios poses two challenges:

- Pre-processing: many exploratory questions require **information** extracted from paper contents.
- Post-querying: it can be difficult to generate **insights** from knowledge base efficiently and interactively.

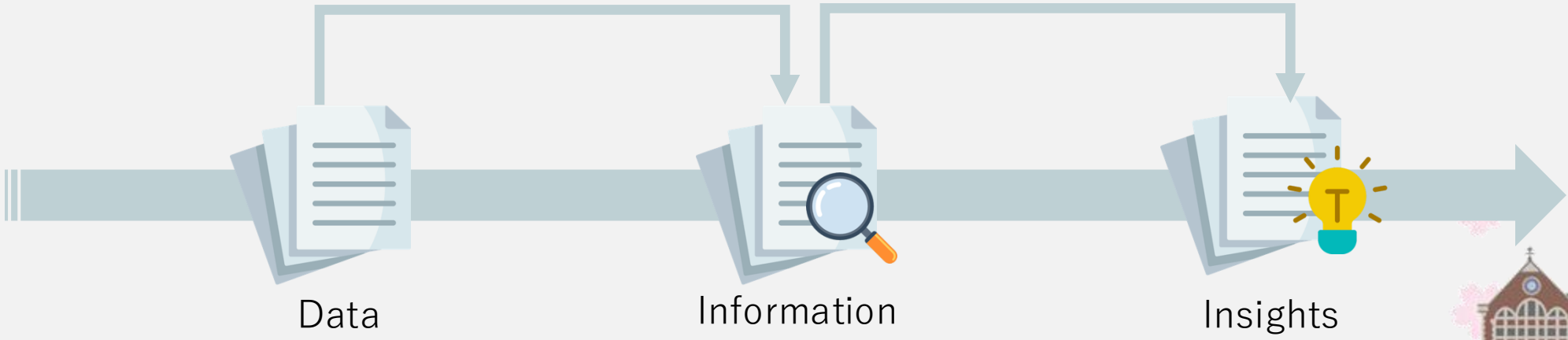
Introduction

? First, construct a knowledge graph for the TVCG community.

What + How

? Second, propose a PRE-view framework for KG visual exploration.

What + How



TVCG-KG Requirement

- KG should contain **various types of entities**, including *metadata* entity, e.g., Author, Affiliation and *semantic* entity from paper, e.g., Methods, Tasks.
- KG should contain **semantic relationships**, providing contextual information about entities.
- The format of K should offer **flexibility and expressive queries** for users to identify target information.

TVCG-KG Construction

Data Preparation

The most up-to-date TVCG dataset:

- Contain 4987 papers from 1995 to Aug, 2023
- Go through the data retrieval, cleaning, and validation.

TVCG-KG Construction

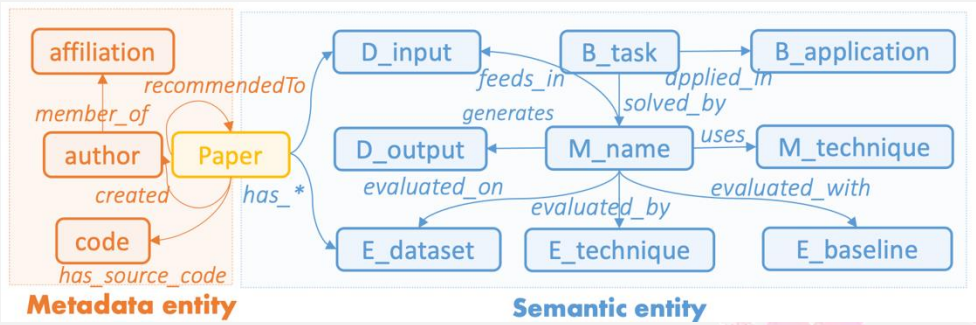
Data Preparation

Ontology Design

The most up-to-date TVCG dataset:

- Contain 4987 papers from 1995 to Aug, 2023
- Go through the data retrieval, cleaning, and validation.

- Conducted a survey of 47 survey papers from TVCG.
- identified four semantic dimensions that are widely used: background, data, method, evaluation.



TVCG-KG Construction

Data Preparation

The most up-to-date TVCG dataset:

- Contain 4987 papers from 1995 to Aug, 2023
- Go through the data retrieval, cleaning, and validation.

Ontology Design

- Conducted a survey of 47 survey papers from TVCG.
- identified four semantic dimensions that are widely used: background, data, method, evaluation.

KG Construction

- Propose an end-to-end pipeline for entity extraction and normalization.
- Experiment several design choices for prompt engineering.
- Utilize Spotlight API to normalize entities.

TVCG-KG Querying

- TVCG-KG and be imported in both RDF triplets or a Property Graph format.
- Various query languages can be utilized to query data from TVCG-KG, the basic semantics of which can be abstracted to resemble a SQL query:

SELECT {*target*}, **WHERE** {*graph pattern*}, **FILTER** {*conditions*}

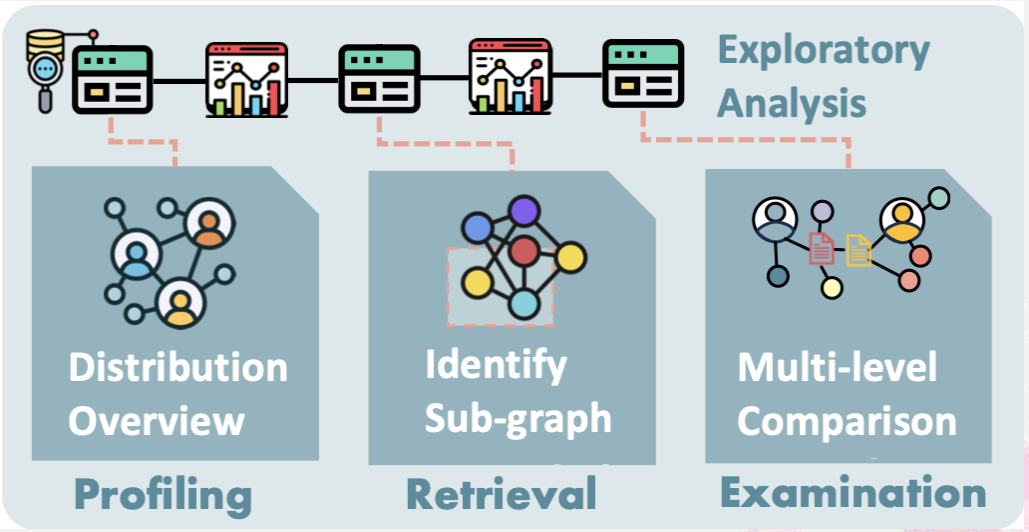
"provide a list of papers published by Mystery Rivers."

SELECT {*paper*}, **WHERE** {*paper* –created- *author*},
FILTER{*author*="*Mystery Rivers*"}

PRE-view Framework

To fill the gap between Information Metra and knowledge graph explorations, we introduce three visual exploratory tasks:

- **P**rofiling: presents the overall structure of the KGs.
- **R**etrieval: helps users retrieve information of interest from KGs.
- **E**xamination: delves into the details of target entities.



Various exploratory pipelines can be built as a sequence of tasks.

Evaluation of TVCG-KG

Structure-based Statistical Assessment

Data Quality Evaluation

Task-based TVCG-KG Evaluation

Structure-based Statistical Assessment

- Compute several metrics of ontology, data model, and graph structure.
- Compute the number of different entity and relation class.

(a)

Ontology	# of entity class	13
	# of relations	28
	# of relations per class	4.54
Data Model	# of entities	81, 033
	# of triplets	406, 291
Graph Structure	Avg. in-degree	2.42
	Avg. out-degree	5.01
	# of weakly connected components	1

(b)

technique	19, 861
author	10, 916
application	7, 257
task	5, 963
uses	50, 124
has_technique	49, 086
seeAlso	47, 742
created	42, 386

Data Quality Evaluation

- Data Consistency Evaluation:
  - Adopt a 10-fold strategy to evaluate the consistency of triplets.
  - Employ a Knowledge Graph Embedding (KGE) model trained on nine of these folds and predict the left-out fold.
  - Use Hits@K as the evaluation metric to measure the prediction performance. Higher score indicates better performance.
- Interlinking to External knowledge graphs.
  - 89.07% of author entities and 73.89% of paper entities can be mapped to the MAKG.

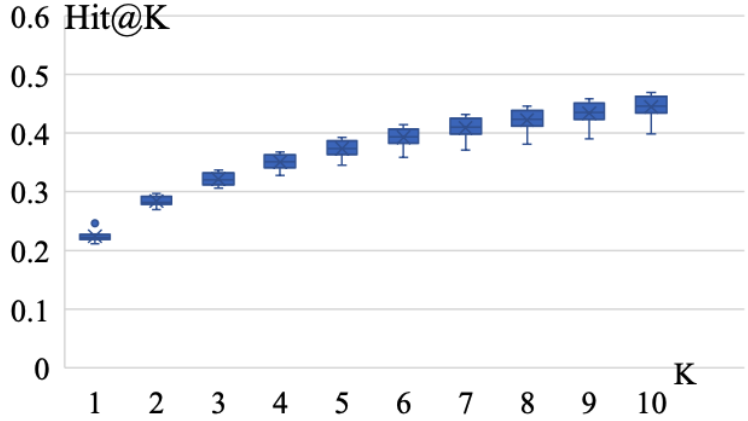
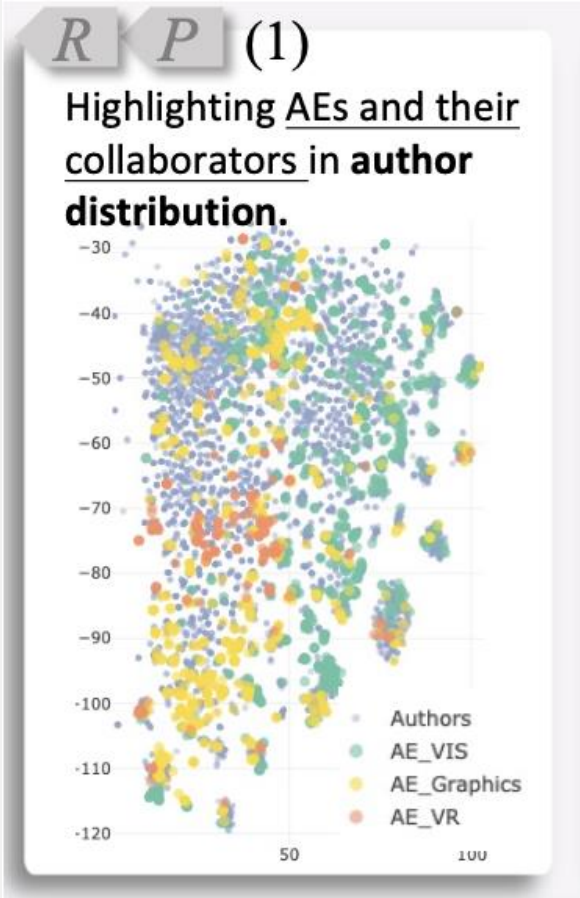


Figure 4: Statistical Distribution of 10 $Hit@K$ Scores for Each K Using a 10-fold Strategy.

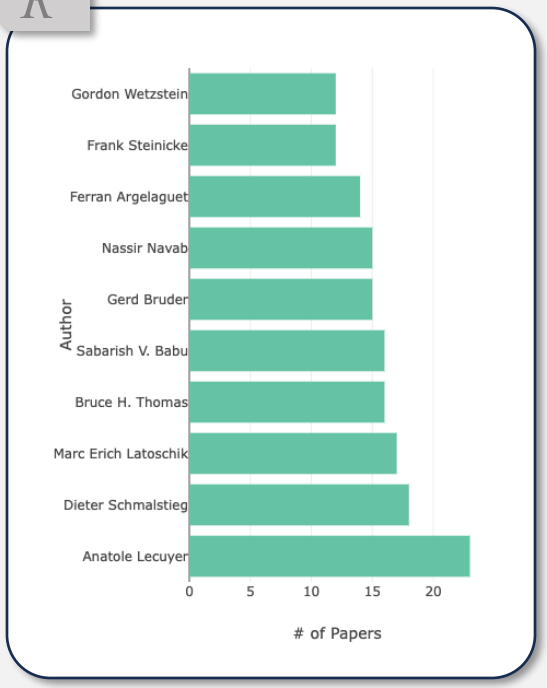
Case study – Author Profiling



- Train a TransE model on the triplets of TVCG-KG to generate embeddings and apply t-SNE to perform dimension reduction.
- To answer “Do the current Associate Editors (AEs) on the editorial board have a comprehensive coverage of the TVCG topics? “
- Color general authors as blue, and highlight AE and their collaborators by research areas (VIS, Graphics and VR).
- Indicating a good coverage of topics from distribution the AE and their collaborators across embedding space.

Case study – Author Identification & Analysis

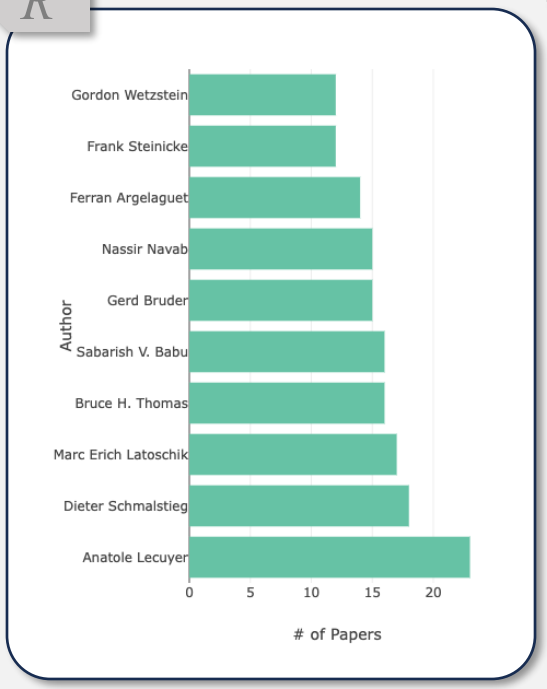
R



Retrieve **top authors** in VR/AR/XR.

Case study – Author Identification & Analysis

R



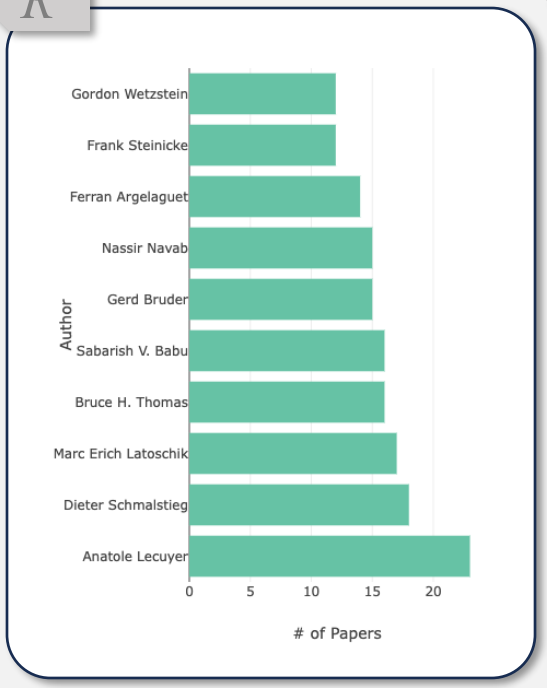
Retrieve **top authors** in VR/AR/XR.

R E

The **collaboration network** around Anatole Lecuyer.

Case study – Author Identification & Analysis

R

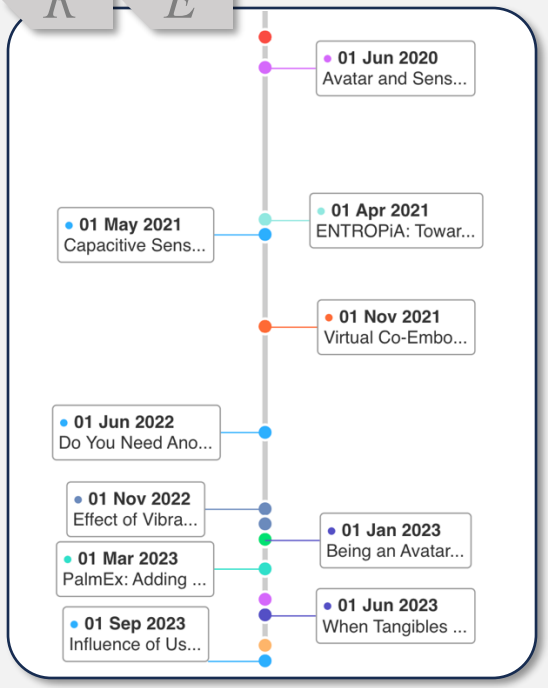


Retrieve **top authors** in VR/AR/XR.

R E

The **collaboration network** around Anatole Lecuyer.

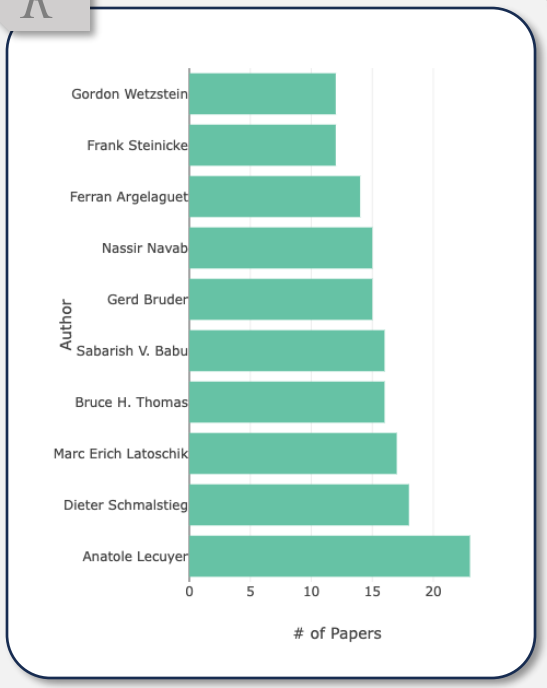
R E



Publication timelines of Anatole Lecuyer in TVCG.

Case study – Author Identification & Analysis

R

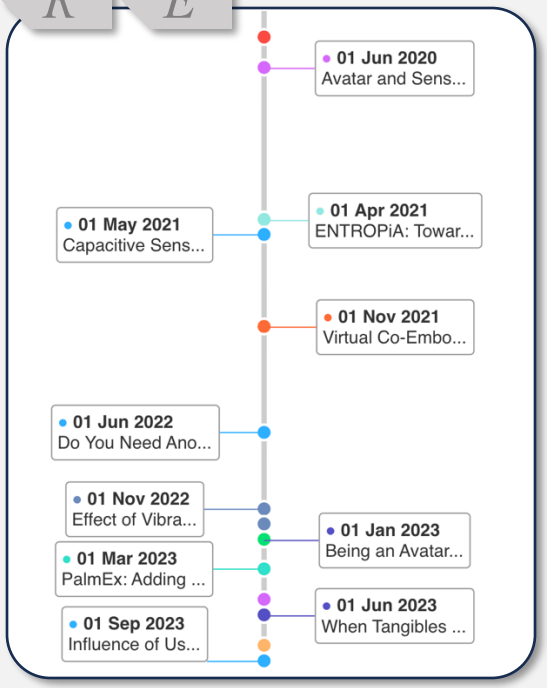


Retrieve **top authors** in VR/AR/XR.

R E

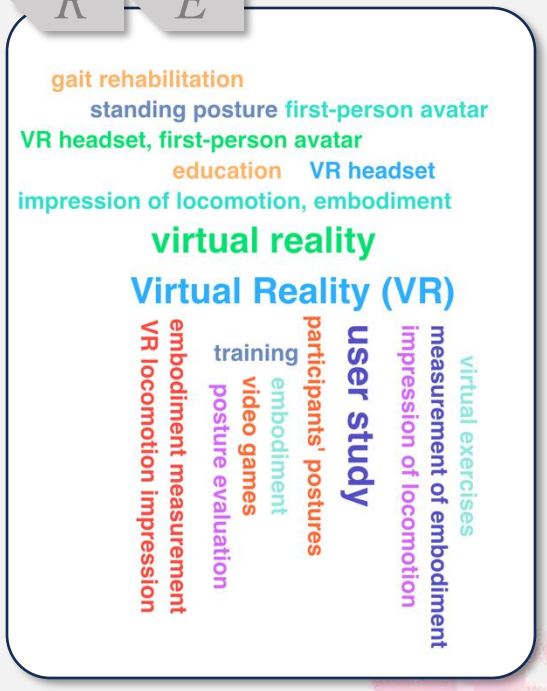
The **collaboration network** around Anatole Lecuyer.

R E



Publication timelines of Anatole Lecuyer in TVCG.

R E



Research interests of Anatole Lecuyer.

Thank you

If you have any questions, feel free to contact us

Yamei Tu (tu.253@osu.edu)