Exploring the Landscape of Recommender Systems Evaluation: Practices and Perspectives

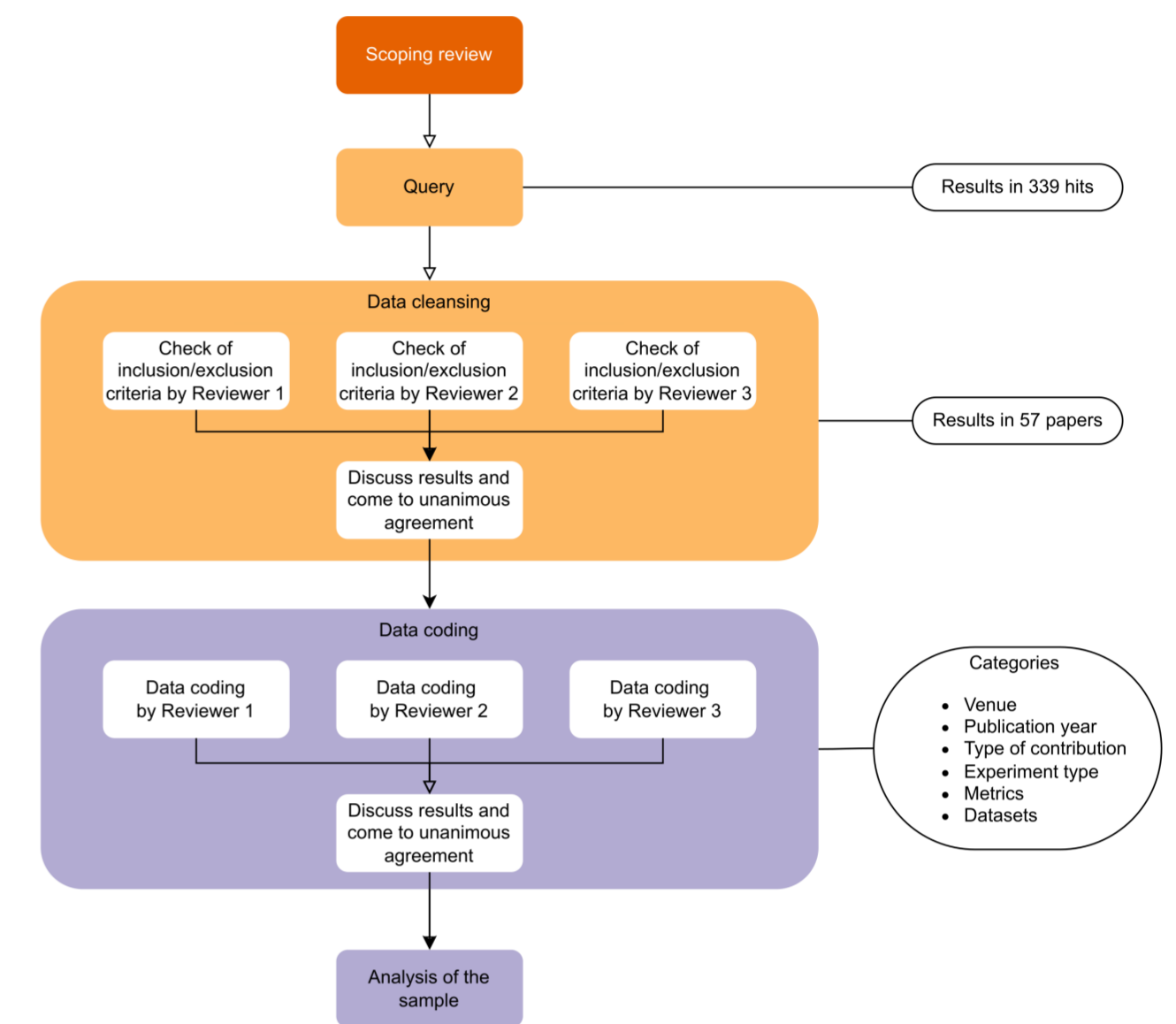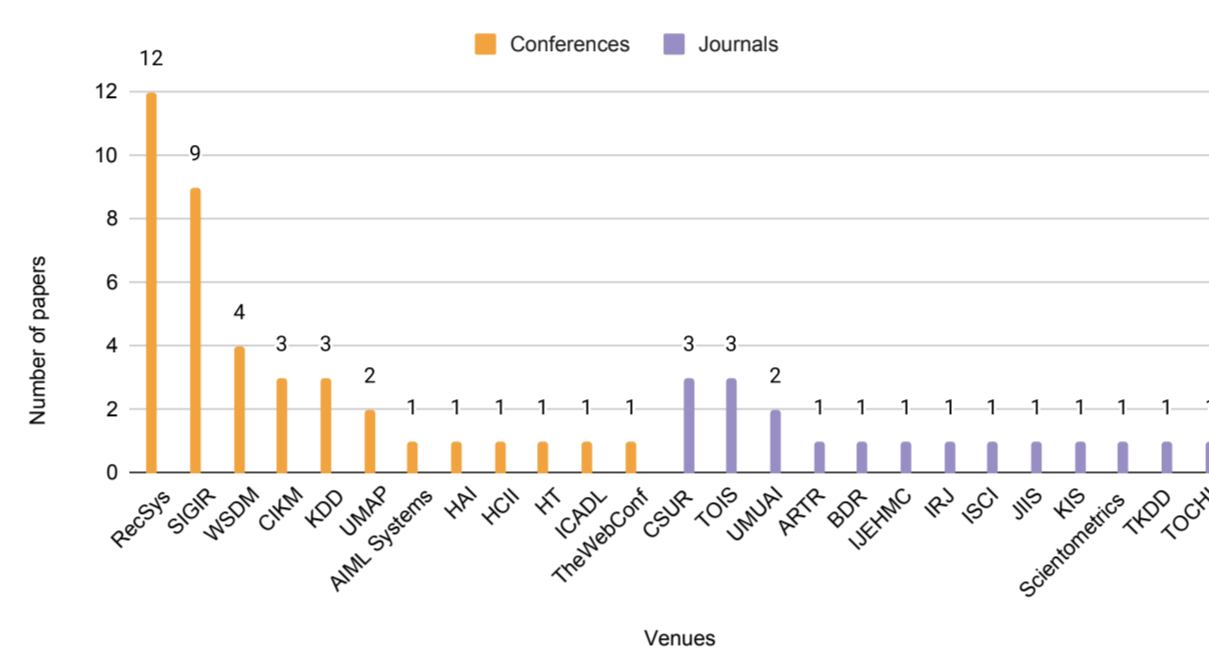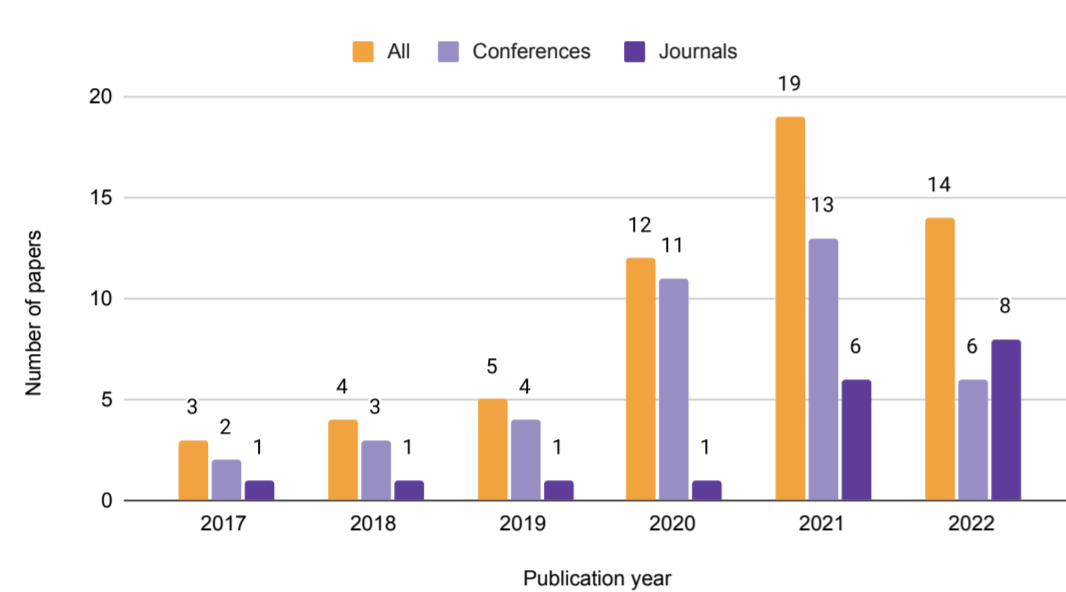Christine Bauer, Eva Zangerle, Alan Said

SCAN ME

A systematic literature review to provide a quantitative snapshot of the landscape of research on the evaluation of recommender systems over the past six years
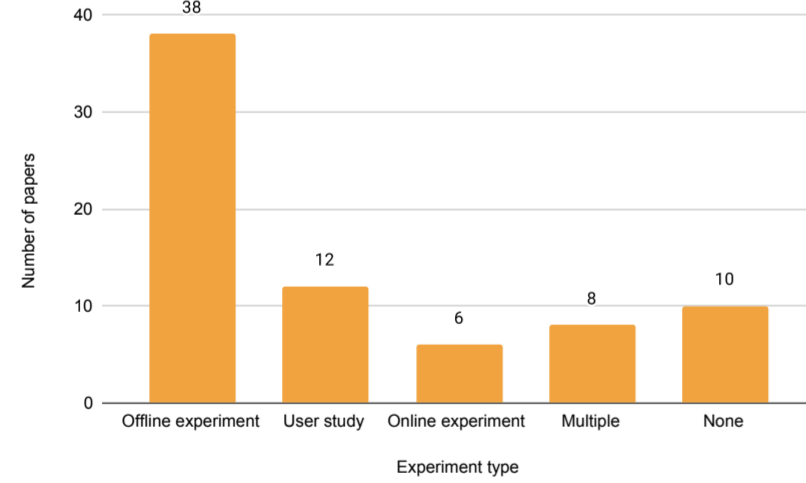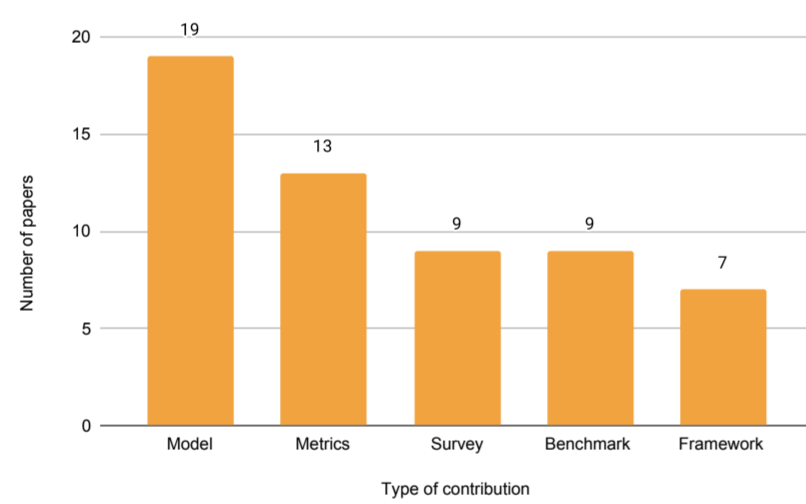
Method

ACM Digital Library as database
2017–2022
57 papers
Query: (Papers from ACM RecSys | indexed with the keyword recommend*)
 & (evalua* | reproducib* | method | methodology)
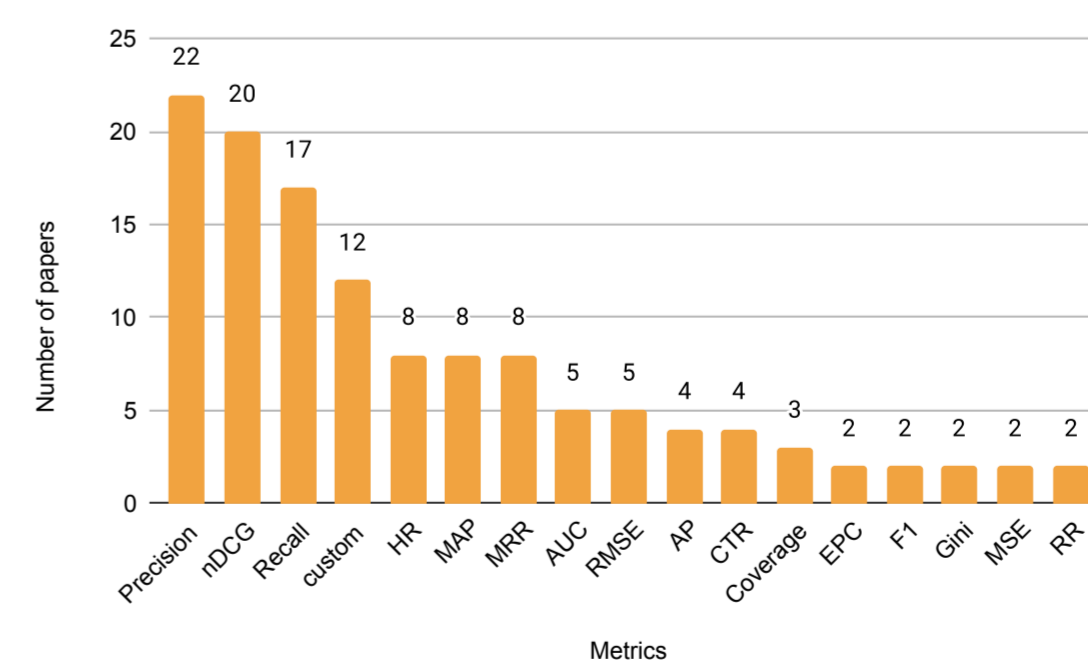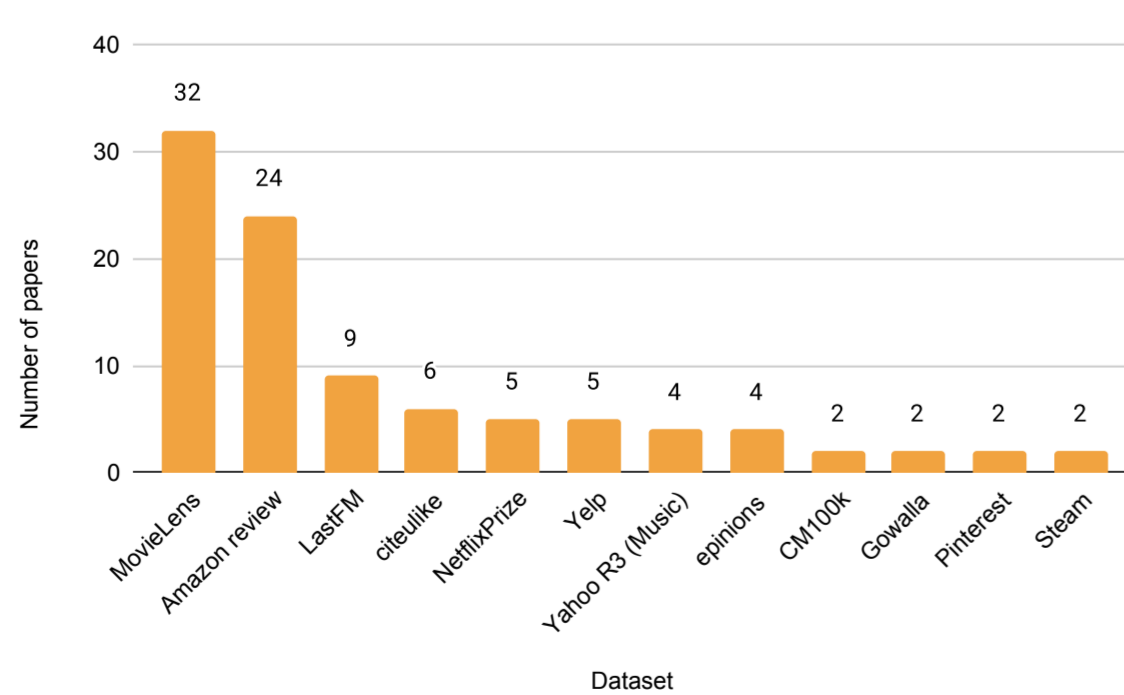

Type of Contribution and Experiment Type

Types of Contribution	Description
Benchmark	Providing an extensive critical evaluation across a (wide) set of approaches or datasets
Framework	Introducing a framework for evaluation, which may take the form of a toolkit or a conceptual framework
Metrics	Analyzing existing or introducing novel metrics of evaluation
Model	Introducing a novel recommendation or evaluation model
Survey	A literature survey

Takeaways
- Strong focus on offline experiments
- Eight papers combine evaluation methods

Datasets and Metrics

Takeaways
- Popularity gap
- Few (and old) datasets are widely used

Takeaways
- Precision, nDCG, Recall most widely used
- Many papers with custom metrics
- (Still) many error metrics

PARIS LODRON UNIVERSITY SALZBURG

universität innsbruck
Department of Computer Science

UNIVERSITY OF GOTHENBURG