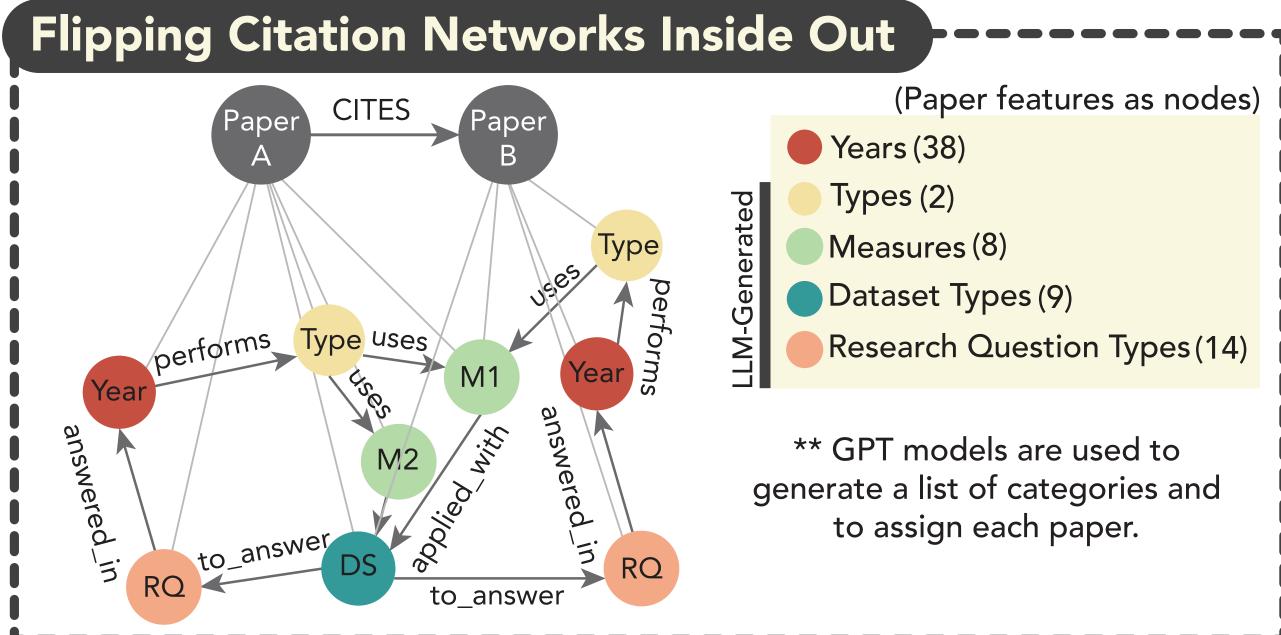
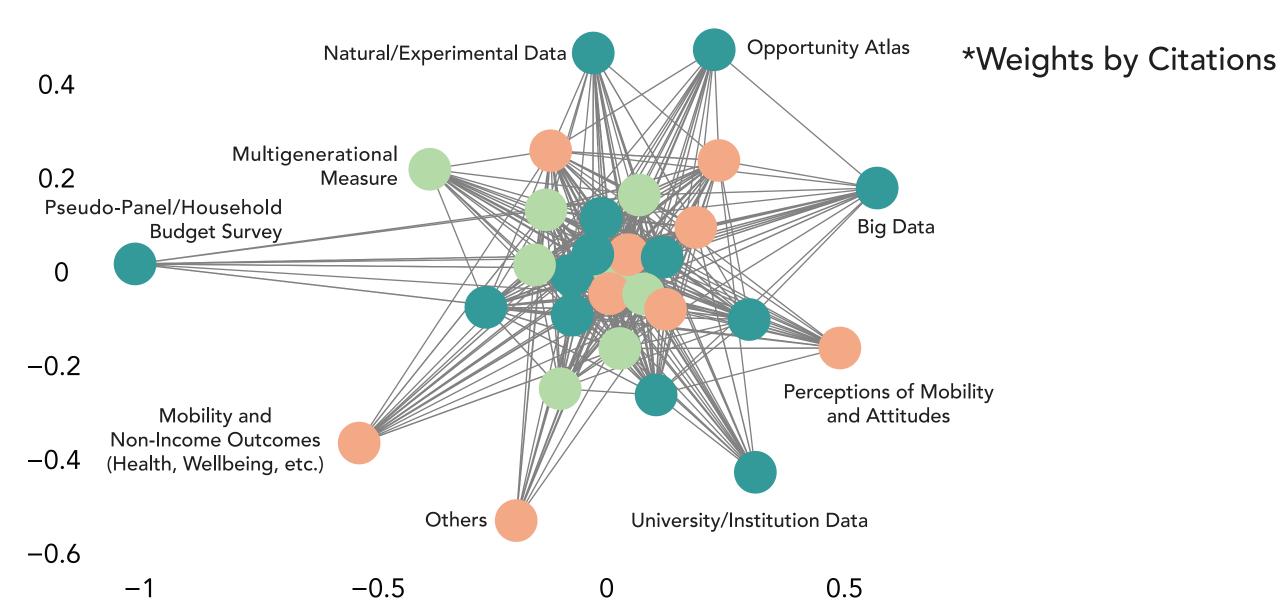
Our Data Extraction Pipelines DATA ANALYTICS **LABORATORY** was used primarily at the early stage to obtain the initial pool of the publications RESEARCH GROUP **OpenAlex** and do the simple filterations based on the metadata. generated labels for each article with prompts asking about the alignment with our LLMs research goal, and more articles were filtered out based on these labels. REVEAL filtered the abstracts that were not written in English, which were not captured by Python the metadata. Hidden Patterns {intergenerational, ''} **Topic relevance Relevance check Methodology Initial pool of** {wealth, income, earning} \ check (Detailed prompt 2) relevance check publications {mobility, transition} (General prompt) IN CITATION NETWORKS 992 16819 617 2169 Seorin Kim & Vincent Ginis 回路認同事 **Topic relevance** Journal article/English/ Filter non-English Filter with metadata BUTO, ES Faculty, VUB check Has_abstract/ > abstracts (Detailed prompt 1) 11023 888 Seorin.Kim@vub.be No duplicates 1022 Flipping Citation Networks Inside Out (Paper features as nodes)





At the core of the feature network (geometric, betweeness, top-degree):

Are there methodological (r)evolutions in

- Regression-based Measures
- Intergenerational Wealth Mobility and Inheritance
 - Empirical Estimates and Determinants

LLM-powered Labeling:

Use Case

Filtering and detecting the core characteristics of desired papers

studying {wealth/income/earning} mobility ? *Abstract only Regression-based Measure Rank-based Measure Transition Matrix/Probability Measure Absolute Mobility Measure Multigenerational Measure Decomposition / Structural Approaches Non-parametric Approaches Others 1985 1990 1995 2000 2005 2010 2015 2020 2025

Fig 1. The stream graph shows the absolute counts of each measure present in the papers. The rolling averages in the bar chart is calculated by accounting for the two previous years in order to observe a smoothed trend. Overall, regression-based measures are the dominant wealth mobility measure.

Are the measures employed in highly cited papers actually influential?

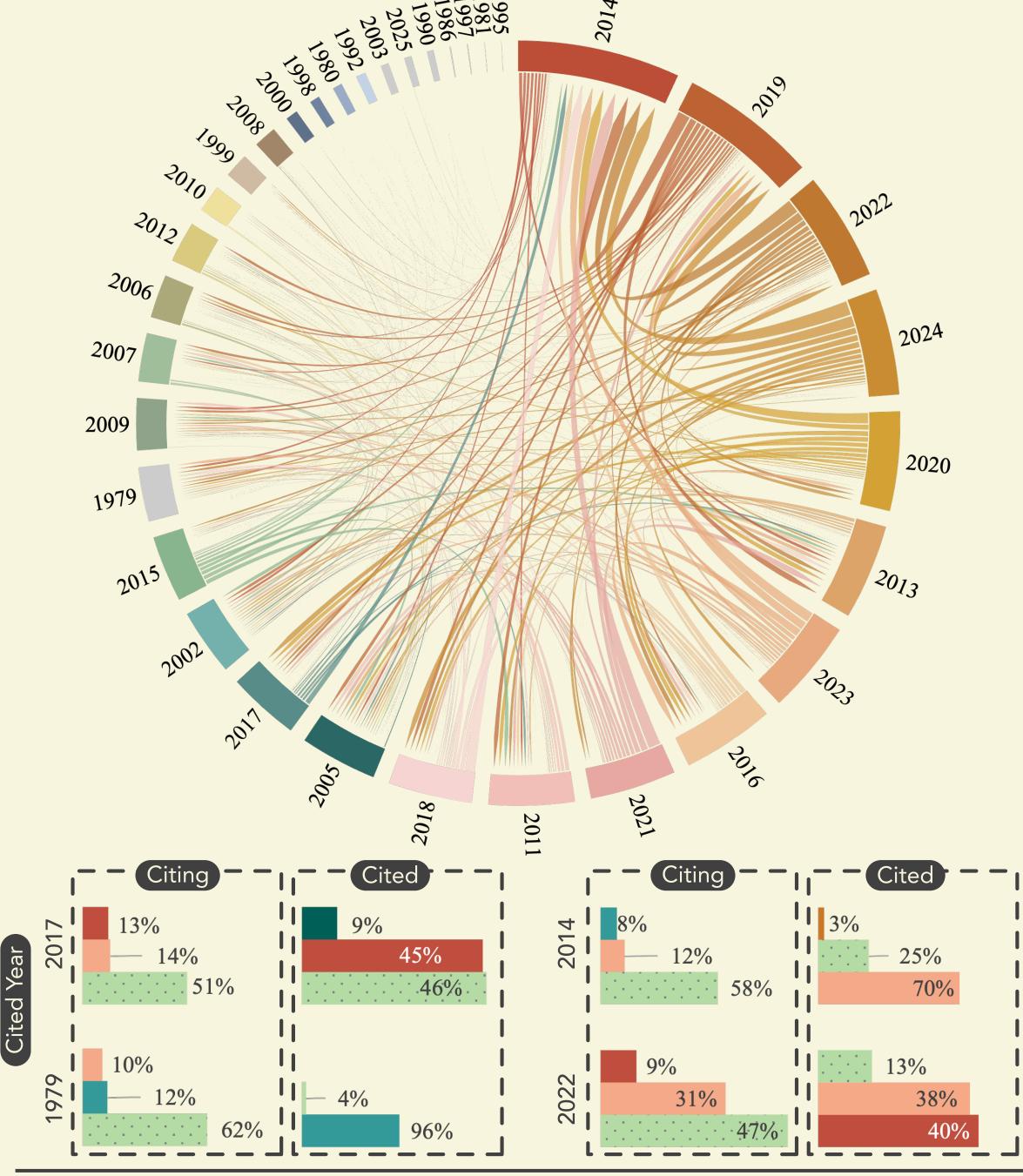


Fig 2. Citation behavior between years of the publications is shown in the chord diagram. The bar graphs compare the methods used by the cited and citing papers in the years when the cited papers were published. These four years are the only ones showing a large contrast between the citing and cited papers. The same color scheme is used as Fig1.

e.g., Assignment of categories: Cohen's Kappa 0.85 Cohen's Kappa 0.85 Rater 1 Rater 2

Takeaway Points

- LLM-powered citation networks (knowledge graphs) can help uncover hidden patterns in a research topic/field. e.g., The prevalence of measures used in the papers.
- Some categorization tasks done by LLMs—especially at higher conceptual levels like Research Question > Data > Method—can be fuzzy, leading to inconsistencies. Still, the labels are generally accurate enough to identify (ir)relevant papers for the research question at hand.
- We are currently expanding this knowledge graph and look forward to uncovering new insights.