a. Data Preparation

a1. Data Collection

Keywords — OpenAlex

{intergenerational, ‘ ’}, {wealth, income, earning}, {mobility, transition}

Social Sciences

Journal — English

Output: Corpus (N=16,819)

a2. Data Curation

GPT 4o/o3

OpenAlex — SEMANTIC SCHOLAR

- Relevance & duplication checks
- Cross-verified abstracts

Output: Curated set (N=617)

b. LLM-driven Classification

b1. Category Generation

GPT-o3

Input: all_abstracts.json
Task: Generate representative taxonomy

Output: Taxonomy JSON (K=k)
Measures (8), Data Types (14), Research Question* Types (9)

{M1: Regression-based measure,
M2: Rank-based measure,…, M8: Others}

*abbr. RQ

b2. Category Assignment

GPT-o3

Input: abstract [k]
Task: Assign paper a category

Output: Paper x Category Table

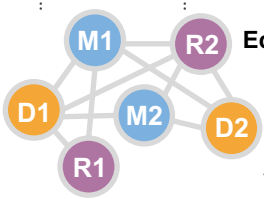
Paper	Measure	Data	RQ
Paper 1	M2	D1	R2
Paper 2	M1	D1	R1
…	…	…	…

c. Graph Synthesis

c1. KG Creation

neo4j

#1 ⇒ {M2, D1, R2} ⇒ M2, D1, R2



Edge Properties
- Year
- Papers
- Weight, w*

*$w \mathrel{+}= \mathbf{1}_{appeared}$

Output: 30 Nodes 483 Edges

c2. KG Analysis

neo4j — python

Compute: Centrality Measures, Decaying Weights*

*$\delta w_0 = \mathbf{1}_{appeared}; \delta w_{t+1} = \delta w_t e^{-\lambda} + \mathbf{1}_{reapp.}$

Output: Insights on Node, Pairs, and Triangles (Centrality, Resurgence)