

graphbrain is an open-source platform implementing semantic hypergraphs, a intrinsically recursive knowledge representation model accommodating the natural hierarchical richness of natural language – see <http://graphbrain.net>

Goals

Develop an automated pipeline to describe the socio-semantic evolution of a field

Rich text processing to extract claims validating or invalidating a target hypothesis

Stochastic block-modeling of the citation network

Focus on Environmental Kuznets Curve (EKC)

Posits an inverted U-shaped relationship between economic development and environmental indicators: first degradation, then improvement

Heated debate over the last two decades, Scopus-based corpus of 2709 papers on EKC

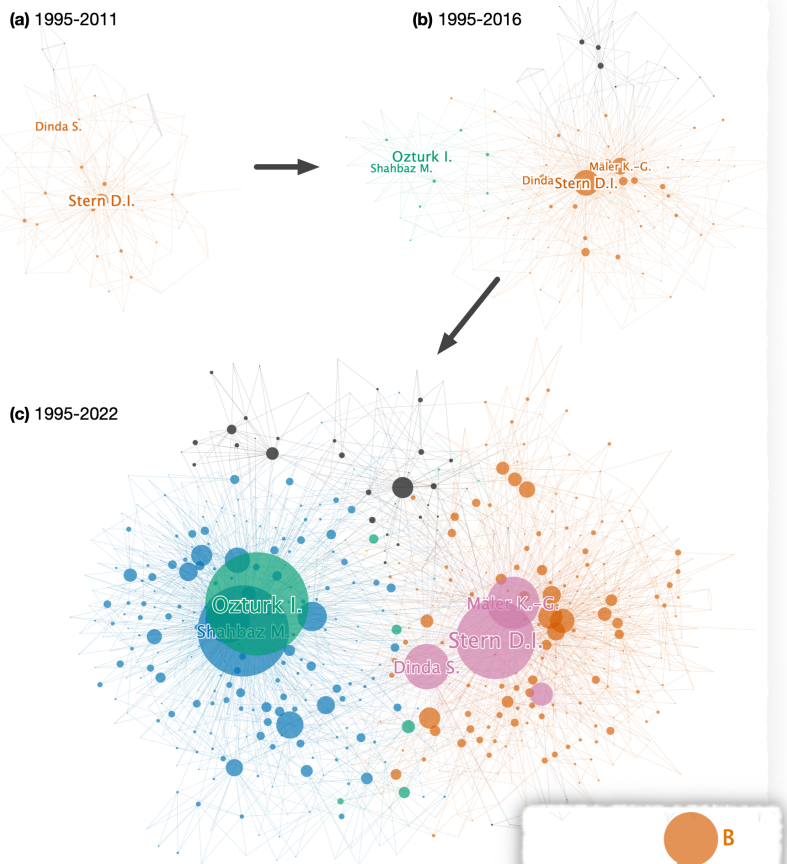
Semantic hypergraph classifier

Identify positive or negative claims

Abstracts parsed into a semantic hypergraph (SH)

Training set: 500 sentences (out of 23k i.e ~2%)

Manual iterative refinement of a handful of simple patterns/rules to extract claims (e.g., 20 predicates associated with results among the top 100), positive or negative (considering negations), or topics through relationships ("between X and Y", where X or Y is environmental or economic)



| Period | Topic | | | | | | | |
|-----------|-------------|-------------|----------------------|-------------|-------------|-------------|-------------|-------------|
| | GHG | energy | local air pollutants | water | SOx | waste | footprint | NOx |
| 1995-2011 | 1.03 | 1.29 | 0.84 | 0.81 | 1.18 | 1.10 | 0.31 | 0.91 |
| 2012-2016 | 1.05 | 1.09 | 1.15 | 1.03 | 1.00 | 0.67 | 0.42 | 0.92 |
| 2017-2021 | 1.55 | 1.39 | 1.64 | 1.01 | 1.39 | 1.01 | 1.07 | 1.50 |
| aggregate | 1.35 | 1.31 | 1.32 | 0.95 | 1.19 | 0.94 | 0.88 | 1.14 |

| Block | GHG | energy | local air pollutants | water | SOx | waste | footprint | NOx |
|-------|-------------|-------------|----------------------|-------------|-------------|-------------|-------------|-------------|
| A | .333 | .074 | .222 | .000 | .333 | .000 | .000 | .000 |
| B | .345 | .257 | .344 | .123 | .109 | .084 | .020 | .042 |
| C | .664 | .618 | .310 | .062 | .039 | .043 | .085 | .023 |
| D | .436 | .392 | .537 | .128 | .084 | .106 | .018 | .040 |

Top. Evolution of the ratio of papers with positive vs. negative results, broken down by topic (bold indicates for each topic the period of maximum ratio). Bottom. Percentage of articles mentioning each topic for each author block. Bold indicates the block where a topic has the highest presence.

Article abstracts

Hypergraph

Frequent atoms

Rule design

Rule testing

Result detection

"Our results do not support the EKC hypothesis."

(do/M (not/M support/P.so) (our/M results/C) (the/M (+/B.am ekc/C hypothesis/C)))

Frequent predicates, concepts and modifiers.

Identify predicates associated with claims
Positive: show/P, indicate/P, confirm/P, support/P, ...
Negative: reject/P, challenge/P, fail/P

Is the claim about EKC?
EKC concepts: kuznets/C, ekc/C, turning/C
Curve concepts: curve/C, shape/C, shaped/C

Parse negations
Negative modifiers: not/M, n't/M, no/M, little/M, poor/M
Result concepts: result/C, finding/C, test/C, evidence/C, ...

(do/M (not/M support/P.so) (our/M results/C) (the/M (+/B.am ekc/C hypothesis/C)))

EKC validation claim type

| | | |
|-----------------|------|------|
| positive result | .809 | .847 |
| negative result | .833 | .366 |

Stochastic block modeling

Identify author citation blocks

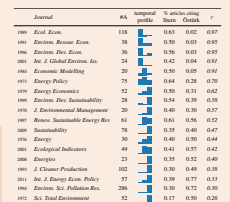
Using degree-corrected stochastic blockmodels
Reveals two main epistemic waves and blocks

An older one centered on Stern, long-lasting expert of the field, less positive on EKC, publishing less, focused on oxides of sulfur and nitrogen, more economics venues

A more recent one around Öztürk, more positive on EKC, more focused on greenhouse gases and engineering venues

| Block | #authors | pos | neg | endogamy | year | #A | #A | k |
|-------|----------|-------|-------|----------|--------|------|------|------|
| A | 3 | 0.354 | 0.084 | 0.003 | 2007.9 | 9.00 | 0.79 | 5.0 |
| B | 360 | 0.422 | 0.416 | 0.033 | 2012.1 | 3.11 | 0.76 | 4.9 |
| C | 406 | 0.952 | 0.407 | 0.102 | 2017.9 | 4.47 | 1.24 | 9.2 |
| D | 57 | 0.492 | 0.441 | 0.085 | 2016.8 | 5.96 | 0.96 | 15.7 |

Metrics per author block, including ratios of EKC claims, endogamy measured as ratio of citations received from co-authors, mean publication year, mean articles per author (#A, raw or normalized per year), and mean unique co-authors (k)



Journal publishing at least 20 articles, presented by year of foundation. All indicates the total number of articles in the corpus. # is the proportion of Stern to Stern+Öztürk, also used to rank the table. Temporal profiles are bar charts of the number of articles published over the three periods (1995-2011, 2012-2016, 2017-2021), scaled to the maximum value.