9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

The two sides of the Environmental Kuznets Curve: A socio-semantic analysis

Socio-semantic analysis; Environmental Kuznets Curve; Scientometrics; Semantic Hypergraphs; Stochastic Block Models

Extended Abstract

The relationship between environmental impacts and economic activity is a hotly debated topic in environmental economics since its inception in the 1960s. At that time, harsh criticisms against economic growth were released as economic expansion was deeemed the source of environmental degradation. The summit of this wave was the *Limits to growth* report [6], which is famous for having made the case that economic growth was not sustainable, either because of the lack of raw resources it constantly demands or because of the amount of pollution it unleashes. This call to put growth to an halt raised strong rebuttals among economists who contented that economic progress and substitution can accommodate the economy with any resource scarcity. Fears of expanding pollution were however less easily dismissed. Concerns about the need to combine development and environmental protection had led in the meantime to formulating the concept of *sustainable development* and making it a goal for the international community — calling for a new era of growth [2], even though how to achieve it without degrading further the environment was however left unaddressed. In 1992, the *World Development Report* suggested a reassuring way out of the dilemma [1]. It displayed several patterns of environmental indicators depending on country income: monotonically decreasing, monotonically increasing, or increasing then decreasing. According to the latter pattern, the state of the environment, after an initial period of degradation, will automatically improve.

The promise was clear: the economic development would somehow take care of its own drawbacks. This was soon named the environmental Kuznets curve (EKC) hypothesis, from the Kuznets curve, an inverted U-shaped relationship between inequality and economic development that Simon [5], one of the first national accountants, had observed. According to the EKC hypothesis, environmental degradation has an inverted U-shaped relationship with economic development. Many empirical contributions immediately investigated whether such a relationship exists or not, for what type of pollutants, while others pinpoint the problems in estimating its parameters. A literature subfield, both theoretical and empirical, built up quickly around "the environmental Kuznets curve".

Thirty years after the seminal contributions more than 2000 articles belong to this subfield. What motivated the initial research was to assess what the literature has finally to say on the validation or not on the EKC hypothesis. Our analysis of the EKC subfield combines two very recent computational methods in an integrated fashion: semantic and network analyses. The use of Semantic Hypergraphs (SH) [7] for the former enables us to go beyond lexicometric patterns and to extract both positive and negative claims about the validity of the EKC hypothesis from abstracts. SH enables a knowledge representation model that is intrinsically recursive and accommodates the natural hierarchical richness of natural language. One popular contemporary alternative would be to train a classifier with the help of a large language model, e.g. of the BERT family. A disadvantage of this approach is that, given its purely statistic nature, it typically require large training datsets. This disadvantage is accentuated by the fact that this NLP task is somewhat delicate, particularly in identifying if a result is positive or negative, and understanding negation appears to be a typical shortcoming of current deep learning models [3]. SH makes it possible to infer a useful classifier from a small number of cases – which is not only important because of the human effort required to annotate training datasets, but perhaps even crucial because our dataset is relatively small.

For network analysis, using degree-corrected stochastic block-modeling [4, 8] reveals the structure of the author citation network as a compact meta-graph made of a few blocks and easily-interpretable connections between them (see fig. 1). The combination of topological and semantic features, and a variety of other metrics, both temporal and structural, converges on a characterization of the field that reveals, in essence, the existence of two epistemic communities: one roughly centered around Stern, a long-lasting expert of the field, and one around Öztürk, a more recent expert that also currently dominates the field in terms of citation counts.

There also appears to be a remarkable temporal and, to a lesser extent, topical discontinuity between these communities. The first wave and epistemic community, centered around the Stern block, is on the whole less positive on EKC, publishes less often and is less endogamic and is more focused on oxides of sulfur and nitrogen (SOx and NOx pollutants). The second wave and epistemic community, centered around the Öztürk block, publishes more positive results and more results overall, is more focused on greenhouse gases and energy and is more endogamic. There is also a smaller epistemic community dominated by Chinese authors and focused on China, that appears to be a middle ground between the two waves according to the various metrics. The divergence of the two communities is also apparent in publication venues — what we call "worlds of references" that are quite distinct and whose activity closely follows the two temporal waves. Notwithstanding, we observe on the whole that the share of positive results as reported in abstracts has consistently increased over the years, yet remains of the same order of magnitude as negative results — the debate is not closed, in either of the structural communities. Beyond distinct publication and citation practices, this might more broadly suggest that the discontinuity and difference in results may be related to different understandings of the EKC, and the scientific areas, methodologies and topics that are relevant to its appraisal and validation.

References

[1] World Bank. World development report 1992. development and the environment. Technical report, World Bank, New York, 1992.

[2] Gro Harlem Brundtland, Mansour Khalid, Susanna Agnelli, Sali Al-Athel, Bernard Chidzero, Lamina Fadika, Volker Hauff, Istvan Lang, Ma Shijun, Margarita Morino de Botero, et al. *Our common future*. Oxford University Press, 1987.

[3] Allyson Ettinger. What bert is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8:34–48, 2020.

[4] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[5] Simon Kuznets. Economic Growth and Income Inequality. *The American Economic Review*, 45(1):1–28, 1955.

[6] Donella H. Meadows, Dennis L. Meadows, Jørgen Randers, and William W. Behrens. *The limits to growth: a report for the Club of Rome's project on the predicament of mankind*. Universe Books, New-York, 1972.

[7] Telmo Menezes and Camille Roth. Semantic hypergraphs. *arXiv*, 1908.10784, 2019.

[8] Tiago P Peixoto. Nonparametric weighted stochastic block models. *Physical Review E*, 97(1):012306, 2018.

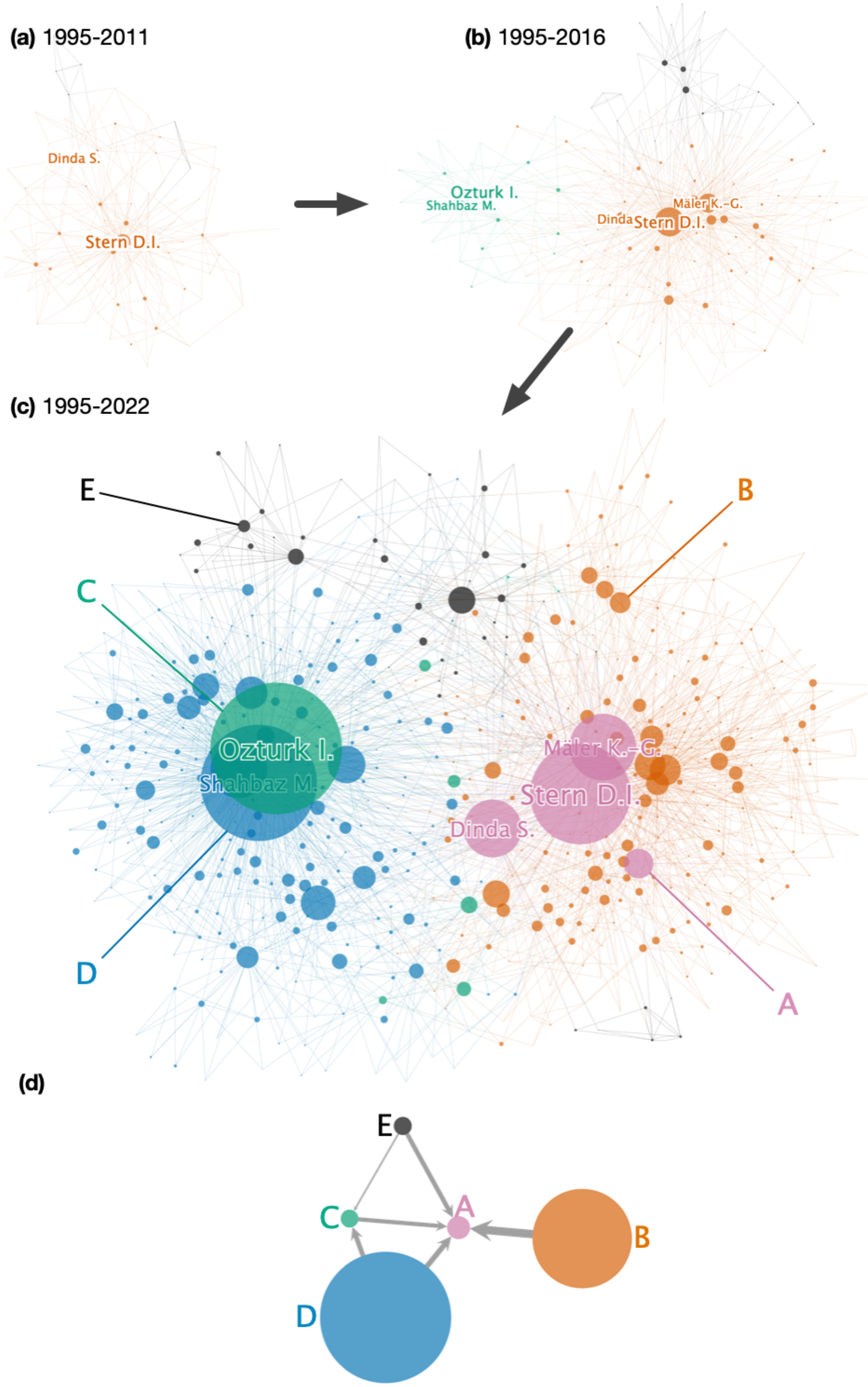

Figure 1: Graphs of citations between authors, colored by blocks determined by SBM. Node radius is proportional to in-degree. Three periods (cumulative): (a) 1995-2011 (b) 1995-2016 and (c) 1995-2022. (d) Graph of author blocks. Edge thickness is proportional to the probability of connection from authors of one block to another. Node radius is proportional to the total number of citations (for the entire corpus) received by the authors in the block.